

ABSTRACT

Title of dissertation: **ERRORS IN HOUSING UNIT LISTING
AND THEIR EFFECTS
ON SURVEY ESTIMATES**

Stephanie Eckman, Doctor of Philosophy, 2010

Dissertation directed by: **Dr. Frauke Kreuter
Joint Program in Survey Methodology**

In the absence of a national population or housing register, field work organizations in many countries use in-field housing unit listings to create a sampling frame for in-person household surveys. Survey designers select small geographic clusters called segments and specially trained listers are sent to the segments to record the address and/or description of every housing unit. These frames are then returned to the central office where statisticians select a sample of units for the survey. The quality of these frames is critically important for the overall survey quality. A well designed and executed sample, efforts to reduce nonresponse and measurement error, and high quality data editing and analysis cannot make up for errors of undercoverage and overcoverage on the survey frame.

Previous work on housing unit frame quality has focused largely on estimating net coverage rates and identifying the types of units and segments that are vulnerable to undercoverage. This dissertation advances our understanding of the listing process, using sociological and psychological theories to derive hypotheses about lister behavior and frame coverage. Two multiple-listing datasets support

tests of these hypotheses. Chapter 1 demonstrates that two well-trained and experienced listers produce different housing unit frames in the same segments. Chapter 2 considers listing as a principal-agent interaction, but finds limited support for the ability of this perspective to explain undercoverage in traditional listing. Chapter 3 has more success explaining the mechanisms of error in dependent listing. Listers tend not to correct the errors of inclusion and exclusion on the frame they update, leading to undercoverage and overcoverage. Chapter 4 tests for bias due to the observed undercoverage, but finds little evidence that lister error would lead to substantial changes in survey estimates.

Housing unit listing is a complex task that deserves more research in the survey methods literature. This work fills in some of the gaps in our understanding of the listing process, but also raises many questions. The good news for survey researchers is that the listers' errors appear to be somewhat random with respect to the household and person characteristics, at least for the variables and datasets studied in this work.

ERRORS IN HOUSING UNIT LISTING
AND THEIR EFFECTS ON SURVEY ESTIMATES

by

Stephanie Eckman

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2010

Advisory Committee:

Dr. Frauke Kreuter, Chair/Adviser

Dr. Katharine G. Abraham

Dr. J. Michael Brick

Dr. Colm A. O'Muircheartaigh

Dr. Melissa A. Milkie

© Copyright by
Stephanie Eckman
2010

Dedication

I dedicate this dissertation to my parents, Nancy and Clif Eckman, who let me get my own library card when I was five, taught me to always sit in the front row, and put up yard signs to alert the neighbors of my academic achievements. This degree is the culmination of your years of encouragement.

Acknowledgments

I received help from many people in preparing this dissertation. I will certainly leave out a few, who I hope will accept my apologies.

Funding to support my own time while working on this research as well as for the data collection came from many sources: the Census Bureau Dissertation Fellowship, the Centers for Disease Control and Prevention Grants for Public Health Dissertations, the Maryland Population Research Center, the Charles Cannell Fund in Survey Methodology, and the Rensis Likert Fund in Research in Survey Methodology. This research would not have been possible without the support of these funders.

I was very lucky to spend the fall of 2009 at the Swiss Institute of Technology (ETH) in Zürich, in the Chair of Professor Diekmann, who welcomed me into his research community. He and his students, especially Reto Meyer, were supportive and encouraging as I struggled with address matching and Swiss German.

Thanks to Bob Groves for pestering me to explore this topic as a dissertation and making NSFG available to me to carry out the research. He put me in touch with the NSFG team, who were incredibly helpful to a stranger who wanted to get her hands on their most sensitive data. Nicole Kirgis, Shonda Kruger-Ndiaye, and Jim Lepkowski were particularly patient with me. Thanks also to Dr. Mosher at NCHS for permitting me to run my study in his survey and use the data before its public release.

Tommy Wright at the Statistical Research Division of the Census Bureau al-

lowed me to access desk space and data at the Census Bureau. He put me in touch with Jim Liu, Cliff Loudermilk, and Aliza Kwiat, who generously shared their data with me and answered my nearly endless questions.

The Joint Program in Survey Methodology is a wonderful place to work and research. I enjoyed discussions over countless lunches in the conference room and have learned from all of the students and professors. When Roger was recruiting me to join the program, he said his motto was “I love you, now get out of here” and I have benefitted from exactly that environment at JPSM.

Particular thanks to Colm O’Muircheartaigh for getting me involved in coverage research during my first few months at NORC. He showed me one could make a career out of fun survey methods research, and, over the course of many nice lunches, convinced me I could do it too.

I am especially thankful to the members of the Kreuter Research Group over the years: Michael Lemay, Carolina Cases-Cordero, and of course Frauke Kreuter herself. My best education in how to be a researcher occurred in these weekly meetings. Many thanks to each of you for creating such a special environment.

Both Frauke and Colm have been wonderful mentors and have also become good friends.

Thank you, thank you!

Table of Contents

| | |
|--|------|
| List of Tables | vii |
| List of Figures | viii |
| 1 Stochastic Coverage: Inter-lister Agreement in Repeated Housing Unit Listing | 1 |
| 1.1 Background | 2 |
| 1.2 Data & Methods | 4 |
| 1.2.1 Construction of Agreement Indicator | 7 |
| 1.2.2 Correlates of Lister Disagreement | 11 |
| 1.2.3 Models | 14 |
| 1.3 Results | 16 |
| 1.3.1 Multi-Level Models of Lister Agreement | 17 |
| 1.4 Discussion and Conclusion | 21 |
| 2 Mechanisms of Undercoverage in Traditional Housing Unit Listing | 24 |
| 2.1 Background & Hypotheses | 25 |
| 2.2 Data | 31 |
| 2.3 Analysis Methods | 35 |
| 2.4 Results & Discussion | 40 |
| 2.4.1 Comparison to Previous Results | 42 |
| 2.4.2 Tests of Hypotheses | 43 |
| 2.5 Conclusions | 49 |
| 3 Confirmation Bias in Dependent Housing Unit Listing | 51 |
| 3.1 Background & Hypotheses | 53 |
| 3.2 Data | 58 |
| 3.2.1 Manipulation of Input List | 59 |
| 3.3 Analysis Methods | 62 |
| 3.3.1 Testing for Failure-to-Add Error | 63 |
| 3.3.2 Testing for Failure-to-Delete Error | 65 |
| 3.4 Results | 66 |
| 3.4.1 Failure-to-Add | 66 |
| 3.4.2 Failure-to-Delete | 72 |
| 3.4.3 Summary and Interpretation of Results | 75 |
| 3.5 Discussion & Conclusions | 78 |
| 4 Bias Due to Undercoverage in Housing Unit Frames | 82 |
| 4.1 Introduction | 82 |
| 4.2 Data | 83 |
| 4.2.1 Survey Background | 84 |
| 4.2.2 Variable Selection | 85 |
| 4.2.3 Undercoverage in NSFG listing | 86 |

| | | |
|---------|---|-----|
| 4.3 | Methods | 87 |
| 4.3.1 | Direct Approach to Bias Estimation | 88 |
| 4.3.2 | Indirect Approach to Bias Estimation | 89 |
| 4.3.2.1 | Listing Propensity Models | 90 |
| 4.3.3 | Variance of Bias Estimates | 95 |
| 4.4 | Results of Bias Analyses | 97 |
| 4.5 | Discussion & Conclusion | 101 |
| 5 | Conclusions | 107 |
| A | Appendix: Coding of Quality of Listing Maps | 115 |
| B | Appendix: Logistic Regression Models of Traditional Listing Propensity | 118 |
| C | Appendix: Matching Addresses in NSFG Listing | 120 |
| C.1 | Matching Input List to L_3 | 122 |
| C.1.1 | Step 1: Match by ID | 122 |
| C.1.2 | Step 2: Automatic Address Match | 123 |
| C.1.3 | Step 3: Manual Address Match | 125 |
| C.1.4 | Quality Checks | 127 |
| C.2 | Matching Three Frames | 127 |
| D | Appendix: Interviewer Questionnaire | 130 |
| E | Appendix: Development of Housing Unit Level Characteristics in NSFG Dataset | 143 |
| F | Appendix: Debriefings with SRC Listers and Interviews | 145 |
| F.1 | Questions to Guide Debriefing Discussions | 145 |
| F.2 | Quotes from Transcripts of Debriefing Discussions | 149 |
| F.2.1 | Quotes from Lister A | 150 |
| F.2.2 | Quotes from Lister B | 152 |
| F.2.3 | Quotes from Lister D | 153 |
| F.2.4 | Quotes from Lister E | 156 |
| F.2.5 | Quotes from Lister F | 161 |
| G | Appendix: Census 2010 Address Canvassing Whistleblower Post | 164 |
| H | Appendix: Content of NSFG Cycle 7 Female and Male Questionnaires | 169 |
| H.1 | Female Questionnaire | 169 |
| H.2 | Male Questionnaire | 175 |

List of Tables

| | | |
|------|---|-----|
| 1.1 | Matches Identified, by Matching Step | 7 |
| 1.2 | Text Matches in Seven Passes | 9 |
| 1.3 | Summary Statistics on Variables Available in Agreement Model . . . | 15 |
| 1.4 | Agreement Rates by Housing Unit Characteristics | 17 |
| 1.5 | Model of Probability of Agreement Between Two Listers | 20 |
| | | |
| 2.1 | Number of Housing Units Listed in First and Second Listing, by Segment | 41 |
| 2.2 | Comparison of Listing Rates by Housing Unit and Segment Characteristics | 44 |
| 2.3 | Traditional Listing Linear Probability Models, Selected Cases Only . | 47 |
| | | |
| 3.1 | Number of Cases Deleted from Input to Third Listing, by Type | 60 |
| 3.2 | Number of Cases Added to Input to Third Listing, by Type | 60 |
| 3.3 | Level of Manipulation in Four Segment Sets | 62 |
| 3.4 | Percent of Unmanipulated and Deleted Units Listed in Dependent Listing | 67 |
| 3.5 | Failure-to-Add: Difference-in-Differences, in Percentage Points | 68 |
| 3.6 | Failure-to-Add: Listing Propensity Models on Unmanipulated and Deleted Cases | 70 |
| 3.7 | Percent of Added Units Listed in Dependent Listing | 72 |
| 3.8 | Failure-to-Delete: Comparison of Listing Rates between Traditional and Dependent Listing for Added Cases | 73 |
| 3.9 | Failure-to-Delete: Deletion Propensity Models on Added Cases | 74 |
| 3.10 | Comparison of Listing Rates of Manipulated Cases in Traditional and Dependent Listing, in Percentage Points | 77 |
| | | |
| 4.1 | Sample Performance in Quarter 12 of NSFG, Selected and Matched Segments | 85 |
| 4.2 | Variables Used in Bias Analysis | 86 |
| 4.3 | Percent of Cases Listed by Second and Third Listings, by Survey Stage | 88 |
| 4.4 | Listing Propensity Model | 93 |
| 4.5 | Bias Methods with 99% Confidence Intervals for All Variables and Methods | 102 |
| | | |
| B.1 | Traditional Listing Propensity Models, Selected Cases Only | 119 |
| | | |
| C.1 | Parsing and Standardizing Street Variable | 124 |
| C.2 | Automatic Matches found, by Pass | 124 |
| C.3 | Manual Matches Found | 127 |
| C.4 | Number of Housing Units Listed for Each of Three Listings, by Segment | 129 |

List of Figures

| | | |
|-----|---|-----|
| 1.1 | Distribution of Demographic Characteristics of Blocks in Census Bureau's Double-Listed Dataset | 6 |
| 1.2 | Distribution of the Distance Between Matched Pairs of Housing Units, by Matching Step | 11 |
| 1.3 | Block-level Agreement Rates. Horizontal Axis is the 215 Blocks, Sorted by Agreement Rate. | 18 |
| 2.1 | Ratio of Number of Housing Units Listed by First Lister to Number Listed by Second, by Segment | 42 |
| 4.1 | Distribution of Predicted Listing Propensities, by Listing Method | 94 |
| 4.2 | Estimates of Bias in Survey Variables Due to Undercoverage, by Listing Method and Estimation Method | 99 |
| A.1 | Example SRC Listing Map | 116 |

Chapter 1

Stochastic Coverage: Inter-lister Agreement in Repeated Housing

Unit Listing

Just as interviewers can introduce variability into survey estimates when they recruit respondents (O’Muircheartaigh and Campanelli, 1998) and administer questionnaires (Schnell and Kreuter, 2005), listers can introduce variance when they create housing unit frames. If the frames created by different listers in the same blocks are not the same over replications of the listing process, then the collected survey data will vary as well. This paper introduces the stochastic view of housing unit listing, which considers every unit in the target population to have a propensity to be covered. Some housing units will be listed on nearly every frame (i.e. have a listing coverage propensity near 100%) and others will be missing from most frames (a listing coverage propensity near 0%).¹

This paper uses a dataset collected by the Census Bureau which contains two listings of a sample of areas. Two field representatives listed each block using the same methods. Analysis reveals a good deal of inter-lister disagreement: the two listers do not create the same frame. The extent of disagreement is worrisome for all studies which use listing to create sampling frames.

¹Coverage propensity is analogous to response propensity (Dalenius, 1983; Oh and Scheuren, 1983; Bethlehem and Kersten, 1985; Groves and Couper, 1992; Lessler and Kalsbeek, 1992) with one difference: cases can be inappropriately covered (that is, overcovered), but the concept of response propensity does not allow for inappropriate response.

1.1 Background

Most studies of housing unit listing have focused on estimating net coverage rates, which is simply the number of housing units listed on the frame divided by the number that should have been listed. A net coverage rate less than 100% indicates that undercoverage exceeds overcoverage, and this is usually the case. Estimates of net coverage in listed housing unit frames range from 80% to more than 99% (Manheimer and Hyman, 1949; Kish and Hess, 1958; Hawkes, 1986; Jacobs, 1986; Joncas, 1985; Childers, 1992; Barrett et al., 2002; Pearson, 2003; Thompson and Turmelle, 2004; Turmelle et al., 2005; O’Muircheartaigh et al., 2006, 2007). Units in multi-unit buildings were found to be both undercovered and overcovered, as are renter-occupied units, vacant units and trailers (Subcommittee on Survey Coverage, 1990; Bureau of the Census, 1993; Childers, 1992; Barrett et al., 2002; Eckman and Kreuter, 2010). Several studies found low-income areas, rural areas and oddly-shaped segments to be undercovered (Subcommittee on Survey Coverage, 1990; O’Muircheartaigh et al., 2006, 2007).² The contribution of lister characteristics to coverage have rarely been studied. An exception is Pearson (2003), who found only weak support for the hypothesis that senior field staff make fewer errors of undercoverage and overcoverage than others.

A few coverage studies have made use of the stochastic conception of coverage, either explicitly or implicitly. Chang and Kott (2004) and Chhikara et al. (2007) use logistic regression to calculate the probability of a farm being undercovered in the

²Kish and Hess (1958)’s findings at the block and segment level are different, but their method of creating a gold-standard frame makes statements about coverage at these levels inappropriate, as they acknowledge.

Census of Agriculture. Several studies use logistic regression to model the likelihood that individuals will be undercovered in the Current Population Study or the decennial census (Fay, 1989; Fein, 1990; Alho et al., 1993). Much of the literature on coverage in the decennial census coverage uses the dual-system technique to estimate the number of individuals or housing units missed in both operations (for dual-system estimates of housing unit undercoverage, see Childers, 1992, 1993; Barrett et al., 2002). The notion of coverage propensity underlies these studies (Wolter, 1986; Panel on Coverage Evaluation and Correlation Bias in the 2010 Census, National Research Council, 2008). However, the focus of these studies is to estimate the number of housing units missed in both the initial and follow up efforts, not to explore the disagreement between the listers as an object of interest itself.

Using frames created by two sets of listers using identical listing methods in the same areas, this paper takes a first step towards understanding listing error by investigating inter-lister disagreement: how different are two frames created by two listers in the same segments using the same methods? I investigate agreement (and disagreement) between the listers without concern for which is accurate. Either of the two listings could have served as a survey frame, and I am interested in how and why they differ. This paper explores the degree of disagreement between two listers' housing unit frames and the correlates of this disagreement. Because this paper is the first contribution to look carefully at inter-lister disagreement, the analyses presented below are largely exploratory and descriptive. The next chapters take a more theoretical approach to exposing the mechanisms of lister error.

1.2 Data & Methods

In 2007, as part of its ongoing effort to evaluate the coverage of the Master Address File (MAF), the Demographic Statistical Methods Division of the Census Bureau listed 5700 census blocks. This listing effort was called the Frame Assessment for Current Household Surveys (FACHS). This listing used the standard Census Bureau listing methods. When the MAF contained any addresses for the selected block, the listers were given these addresses in the listing software and updated this list in the field (this method of listing is called *dependent listing*). When the MAF contained no addresses, listers traveled around the block and created a list of all housing units (this method is called *traditional listing*). The in-field listing (traditional or dependent) was assumed to be the gold standard frame, and the coverage of the un-updated MAF was evaluated against this standard.³

As a check on the assumption that the field listing was in fact the gold standard frame, the Census Bureau repeated the in-field listing in a subsample of 301 of the 5700 blocks. A second lister, different than the first, was sent to each of these selected blocks to list it again. If the first lister used traditional listing, so did the second. If the first lister used dependent listing, the second lister did so as well, and the input listing to the second listing was identical to the input to the first. (That is, the second listing was not dependent on the first.) The second listing was completed quickly, always within five months of the first (personal communication with Clifford Loudermilk, Demographic Statistical Methods Division, Census Bu-

³For more background on the Master Address File, the FACHS evaluations and other assessments of the MAF's coverage properties, see Kennel (2007); Liu (2008, 2009); Loudermilk and Li (2009). This paper does not address the important question of MAF coverage.

reau). All listers who participated in the listings were trained Census Bureau field representatives. This paper uses the data from this subset of the FACHS study to explore what factors lead two listers doing the same task to create different housing unit frames.

The subsample for this double listing exercise is not nationally representative. Blocks in which the United States Postal Service's Delivery Sequence File, a major component of additions to the MAF, showed no growth were excluded from selection. (See Kwiat (2009) for details on the selection of the 301 blocks.) Four of the blocks selected for double listing were not listed a second time due to staffing constraints. Of those fielded twice, not all contained housing units. For these reasons, 215 of the 301 selected blocks are available for my analyses. 37 blocks used traditional listing because there were no cases on the Master Address File to serve as an input frame, and 178 blocks used dependent listing.

These 215 blocks are in 43 states. The number of selected blocks per state ranges from one to 30; in 14 states there is just one block. The completed blocks are in 156 block groups, meaning the sample is not very clustered. 124 of the blocks are individual selections that do not border any other selected blocks. All other block groups that contain a selected block contain just two or three, except one block group where 22 blocks were selected. Figure 1.1 gives the distribution of selected housing unit and demographic characteristics for the blocks.

Both listings used the Census Bureau's listing software which provides listers with a map of the blocks they are to list and displays the addresses on the input listing (if any). Listers can add units that are not on the input listing, delete units,

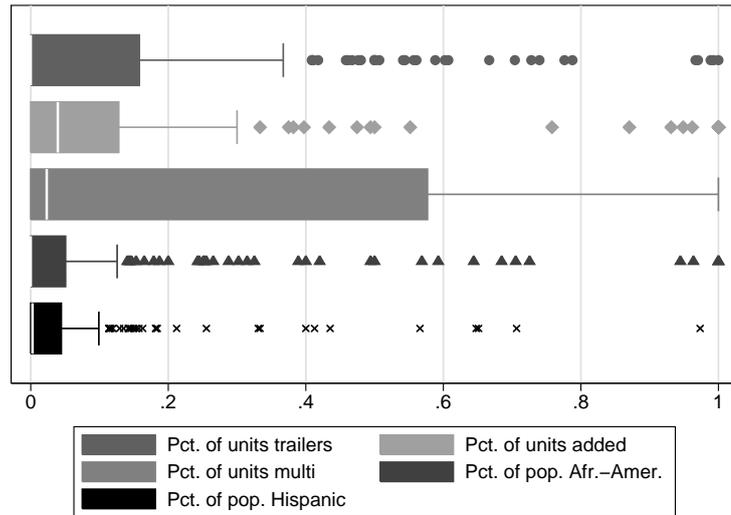


Figure 1.1: Distribution of Demographic Characteristics of Blocks in Census Bureau's Double-Listed Dataset

move units from one block to another, or simply indicate that the unit is correct on the input listing. Listers can also move housing units by creating new mapspots. As they confirm and add units, they record whether the unit is a trailer and whether it looks to have been constructed after 2000. When adding units, listers parse addresses into several fields which greatly simplifies the task of matching the two listings to identify which housing units were listed by both listers.

For each of the 215 completed blocks, I have two housing unit frames, created by two listers using the same methodology. I use these frames to identify the housing unit and block characteristics that correlate with agreement and disagreement between the listers.

Table 1.1: Matches Identified, by Matching Step

| Step | Matches | Percent |
|--|---------|---------|
| Step 1: Units from MAF | 44,753 | 84.3% |
| Step 2: Address matches, in SAS | 8,249 | 15.5% |
| Step 3: Address matches, manual review | 93 | 0.2% |
| Total | 52,995 | |

1.2.1 Construction of Agreement Indicator

The first step in preparing the data for analysis was matching the two listings of each block to identify agreement and disagreement between the two listers. There were 59,363 housing units listed by the first lister and 60,943 units listed by the second lister.⁴ These listed cases represent 67,205 unique housing units. To construct the agreement indicator, I matched these two frames against each other in a multi-step process. This sort of matching work always requires judgments. In this section, I describe my matching protocol in detail. Other researchers might make different judgments and would thus create a slightly different agreement indicator, which would impact the results given below. However, I feel that all of the matching decisions I made are justifiable and defensible.

I matched only housing units that at least one lister included on the frame. Housing units on the MAF that were deleted by both listers as nonexistent or out-of-block were not matched, even though these two deletions are a kind of agreement. All matching was performed within blocks, i.e. if two listers included units with the same address but in different blocks, I do not consider these units a match.⁵ I

⁴The division into first and second listers is based on the times at which the listings were performed: the first listing is the one that happened earlier.

⁵Aliza Kwiat of the Demographic Statistical Methods Division of the Census Bureau, working

permitted only one to one matches.

The guiding principle in the matching work was whether an interviewer would go to the same housing unit if the two addresses were selected. When I believed that s/he would, I considered the two listed cases a match. For example, unit A and unit 1 at the same address most likely refer to the same unit, and selecting either one would lead to the same unit being approached for an interview.

Step 1: ID Matching of Units on the MAF

In 178 of the blocks, the two listers used dependent listing: they started with a list of units on the Master Address File (MAF) in the assigned block and confirmed or deleted these in the field. Because every housing unit on the MAF has a unique ID, in the first matching step I simply matched units in the two frames by this ID. This step identified housing units on the MAF that both listers agreed were on the housing unit frame. As shown in table 1.1, these are the majority of the matches.

Step 2: SAS Matching

The next step compared the addresses of the remaining unmatched units, both those on the input list and those added in the field. Here the parsing of the addressees in the listing software was a great help. The first pass required that all of the address fields match exactly, which identified 7,676 matches. Subsequent passes dropped fields from the matching routine. For example, the second pass did not require a match on the direction prefix field, so that 932 E Elm St would match with the same repeated listing dataset, takes a slightly different approach to the matching process and thus her results do not match my results in this paper (Kwiat, 2009). Her approach is *lister-centric*: she looks at agreement between the two listers about the appropriate action for each housing unit in the field and on the input list. My approach is *frame-centric*: I am interested in whether the two frames created by the listers are the same, or not.

Table 1.2: Text Matches in Seven Passes

| Address Field | Pass 1 | Pass 2 | Pass 3 | Pass 4 | Pass 5 | Pass 6 | Pass 7 |
|---------------------|--------|--------|--------|--------|--------|--------|--------|
| Block | X | X | X | X | X | X | X |
| House number | X | X | X | X | X | X | X |
| House number suffix | X | X | X | X | | | |
| Direction Prefix | X | | | | | | |
| Street Type Prefix | X | X | X | X | X | | |
| Street Name | X | X | X | X | X | X | X |
| Street Type | X | X | X | X | X | X | |
| Direction | X | X | X | | | | |
| Extension | X | X | | | | | |
| Apartment | X | X | X | X | X | X | X |
| Matches Found | 7676 | 12 | 0 | 6 | 30 | 0 | 531 |

to 932 Elm St. This pass identified 12 additional matches. These passes identify more accurate matches first, to ensure that a low quality match could not crowd out a better match. After each pass I reviewed the matches before accepting them and discarded six that seemed inappropriate. All passes required that the house number, street name and apartment designator match exactly. See Table 1.2 for the match criteria at each pass and the number of matches found.

At every pass, I insisted that the two addresses agree on the block number. Although pass seven would allow 115 1st Ave to match to 115 1st St, it also required that the two listers place these units in the same block. My justification in matching these two addresses was that interviewers assigned to interview at 115 1st Ave and 115 1st St in the same block would likely end up at the same housing unit. Thus I am comfortable matching these two addresses if they were still unmatched in the seventh pass of the second matching step.

Step 3: Manual Matching

Units still unmatched were then output to spreadsheets and matched manually. This step caught many spelling and parsing errors that were not detectable by the programs above⁶ as well as different apartment designators (A, B, C versus 1, 2, 3) or (101, 102, 103 versus 1, 2, 3). When one lister included two units at an address and the other only one, I matched the single unit to the first unit and left the second unit unmatched, because interviewers are trained to interview at the first unit if the single family home selected for the survey turns out to be a multi-unit home. This work identified 93 additional matches.

Use of Mapspots to Review Matching

As they list, Census Bureau listers capture the coordinates (latitude and longitude) of each unit they confirm or add. These data help direct interviewers back to housing units selected for interview. Figure 1.2 shows the distribution of the distance between the mapspots of matched housing units, broken down by matching step. The average distance between matched units is 0.06 kilometers and the median is 0.03. The maximum distance is 3.3 kilometers. There are some extreme outliers among the units matched in the first two steps. However, because these two steps are the most precise of all the matches, these outliers point to the inherent variability in the capturing of the coordinates (Sando et al., 2005), not of improper matching routines. Thus although there is some variability in the distances between the mapspots for units matching in the last two steps, this is likely due to

⁶In several cases, I fixed these spelling and parsing errors in the datasets and reran the step 2 matching routines. The match counts in Tables 1.1 and 1.2 reflect the results after these cleaning steps were applied.

variability in the GPS readings and in listers' mapspotting procedures, rather than to inappropriate matches. I manually reviewed all matches where the distance between the housing units was greater than one kilometer and found no evidence in the address fields or listing notes that these units were not true matches.

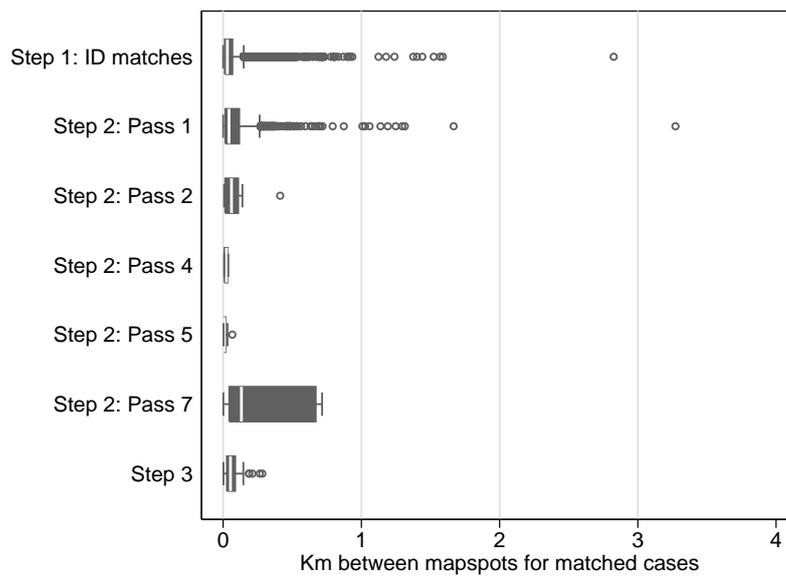


Figure 1.2: Distribution of the Distance Between Matched Pairs of Housing Units, by Matching Step

1.2.2 Correlates of Lister Disagreement

To this dataset of listed housing units I added variables that I expect will explain some of the disagreement between the two listers who listed these blocks.

Block Characteristics

Earlier work on housing unit listing coverage has found low-income, rural and oddly shaped segments to be undercovered (Subcommittee on Survey Coverage, 1990; O’Muircheartaigh et al., 2006, 2007). I will expand upon this work by exploring the

other characteristics that correlate with agreement (and disagreement) between the two listers.

In dangerous neighborhoods, listers may be less likely to walk down alleys and gangways, enter multi-unit buildings and talk to residents, and listing quality can suffer. If high crime rates suppress the listing propensity of some units, we should see more disagreement between listers in these areas. The FBI's Uniform Crime Reports (UCR) dataset provides violent crime rates at the county level (United States Department of Justice. Federal Bureau of Investigation., 2009).⁷ To separate the effects of crime itself from those demographic characteristics which affect the perception of crime and disorder, I separately control for race, ethnicity, housing unit density, and percent of households with low income (Sampson and Raudenbush, 2004).

Each lister had a map of the area they were to list. However, these maps are not easy to use and may even be out of date. Comparing the listing maps for each block with the same area on Google Maps (both map and satellite views). I coded several features of each block map: whether the block had a non-visible boundary, a water boundary or was simple and rectangular. (See Appendix A for details on the coding of these variables.) Of the 215 completed blocks, 59 (27%) had non-visible boundaries, 63 (29%) had a water boundary; 15 blocks had both of these characteristics. Thirty-three blocks (15%) were rectangular-shaped with no irregularities.

⁷Violent crime is defined as murder, rape, robbery and aggravated assault.

Housing Unit Characteristics

Because the dataset could contain two observations of each unit, coding housing unit level variables was not straightforward. The three available housing unit variables are binary indicators of trailer, multi-unit and add (whether a unit was initially on the MAF or was added by the listers).

Each lister is supposed to indicate every unit that is a trailer in the listing software. Unless a lister takes this action, a unit is assumed not to be a trailer. I coded a unit as a trailer if either lister indicated it was, because false negatives are more likely than false positives given the default behavior of the software. 4.4% of the listed housing units were trailers.

Units with any text in the Apartment field of the address (except those flagged as trailers) are designated to be in multi-unit buildings. In those rare cases where the two lister disagreed about whether a unit should have an apt designator, the unit *was* marked as a multi-unit. This situation occurred in only 166 cases and nearly all of these were matched during the first matching step by MAF ID. 58% of the housing units in the dataset are flagged as in multi-unit buildings.

Units that were not on the MAF were added by the listers in the field. Because listers sometimes delete a unit from the MAF and later add that same unit back in to the frame, there are a few cases of units on the two frames that match, but one was added and one was not. In those cases, the unit is *not* marked as an add in my dataset. All cases in traditionally listed blocks are marked as adds. 9.6% of all units were added by both listers.

Listers are certain to differ in their skills related to listing, such as map reading, spatial ability and comfort with the laptop and listing software. Furthermore listers received different kinds of training and have different work histories: some do more listing than others, some are more used to urban listing, etc. Ideally I would control for all of these lister-level characteristics in my models. Unfortunately I am not able to do so, due to Census Bureau restrictions on use of data about employees. The Census Bureau also was not able to identify which listers worked in which blocks. Thus my models do not contain any lister data or any clustering by listers.

Table 1.3 gives the means, ranges and standard deviations of the variables in my analyses.

1.2.3 Models

The final dataset is at the housing unit level and contains 67,205 cases. For each observation I have a binary variable indicating whether the two listers agreed that this unit should be included on the frame or whether only one lister thought so. Agreement between the two listers is the dependent variable I attempt to explain in the models below.

I run multi-level models that can account for both the clustering of units into blocks and the inclusion of block-level characteristics as explanatory variables of interest (Snijders and Bosker, 1999; Raudenbush and Bryk, 2002; Gelman and Hill, 2007). The models do not include the selection weights for the block sample due to the unusual universe from which these blocks were selected: blocks where the

Table 1.3: Summary Statistics on Variables Available in Agreement Model

| Variable | Mean | Std. Dev | Min | Max | N |
|--|-------|----------|--------|-------|-------|
| Unit added by listers | 0.096 | 0.295 | 0 | 1 | 67205 |
| Unit is in multi-unit building | 0.583 | 0.493 | 0 | 1 | 67205 |
| Unit is a trailer | 0.044 | 0.206 | 0 | 1 | 67205 |
| Proportion of population Hispanic | 0.095 | 0.127 | 0 | 0.973 | 67205 |
| Percent of HHs with income less than 45k | 0.472 | 0.186 | 0.071 | 0.879 | 67205 |
| Proportion of population African-American (only) | 0.16 | 0.251 | 0 | 1 | 67205 |
| HUs per land square mile, block level (standardized) | 0 | 1 | -0.542 | 3.156 | 67205 |
| Block is rural | 0.191 | 0.393 | 0 | 1 | 67205 |
| Map, simple shape, no interior streets | 0.048 | 0.213 | 0 | 1 | 67205 |
| Map, block has water boundary | 0.258 | 0.438 | 0 | 1 | 67205 |
| Map, block has invisible boundary | 0.233 | 0.423 | 0 | 1 | 67205 |
| Per capita crime rate, violent crimes (standardized) | 0 | 1 | -0.968 | 5.128 | 67205 |

Postal Service Delivery Sequence File showed positive growth.

Because the dependent variable is dichotomous, a logistic model is an obvious choice. However, I am interested in several interaction effects in these models and interpretation of interaction effects in nonlinear models is complex (Ai and Norton, 2003). In addition, I cannot include lister effects in my models due to data limitations, and coefficient estimates in nonlinear models are particularly prone to unobserved variables bias.⁸ For these reasons, I use a linear probability model (as suggested by Wooldridge (2009, pp. 454–457) and Mood (2010)). I fit all models with the multilevel linear regression command `xtreg` command in Stata 11 (StataCorp LP, 2009).

1.3 Results

Across all of the 215 blocks, the listers agreed about the inclusion of 79.0% of the housing units. Two listers using exactly the same methods create frames that are quite different. Table 1.4 breaks these agreement rates down by the housing unit characteristics available on the dataset. The agreement rate is higher among units on the Master Address File, the input to the dependent listing process, than among those that were added. There is a lower agreement rate among units in multi-family structures than single-family, and among trailers than non-trailers.

The block-level agreement rates are shown in Figure 1.3. In the upper right corner of this graph are 18 blocks where the two frames are in complete agreement.

⁸The unobserved variables effect in logistic regression is due to the fixed error term, see (Mood, 2010).

Table 1.4: Agreement Rates by Housing Unit Characteristics

| | n | Agreement Rate | <i>F</i> stat |
|-------------|--------|----------------|---------------|
| Overall | 67,205 | 79.0% | |
| On MAF | 60,746 | 80.5% | 4.26* |
| Added | 6,459 | 64.9% | |
| Multi-Unit | 39,191 | 76.7% | 1.30 |
| Single Unit | 28,014 | 82.2% | |
| Trailer | 2,982 | 65.8% | 5.24* |
| Non-trailer | 64,223 | 79.6% | |

F statistic tests significance of difference within each pair

F statistics reflects clustering of HUs by block

* Significant at 5% level

These are not only small blocks: two-thirds of these blocks have only five or fewer listed housing units, but three have more than 30. In the lower left corner there are 28 blocks where the two frames do not agree at all. In 22 of these blocks, one lister listed zero units and the other included from one to 105 units. (As discussed above, I exclude blocks where the two listers agreed that there were no housing units.) There is a good deal of diversity in these block level agreement rates. The goal of this paper is to explain this diversity, using characteristics at the housing unit and block levels.

1.3.1 Multi-Level Models of Lister Agreement

Table 1.5 presents the estimates from the multi-level model described above. The dependent variable is whether the two listers agreed about the inclusion of a housing unit (1) or not (0). Positive coefficients indicate characteristics that make agreement more likely.

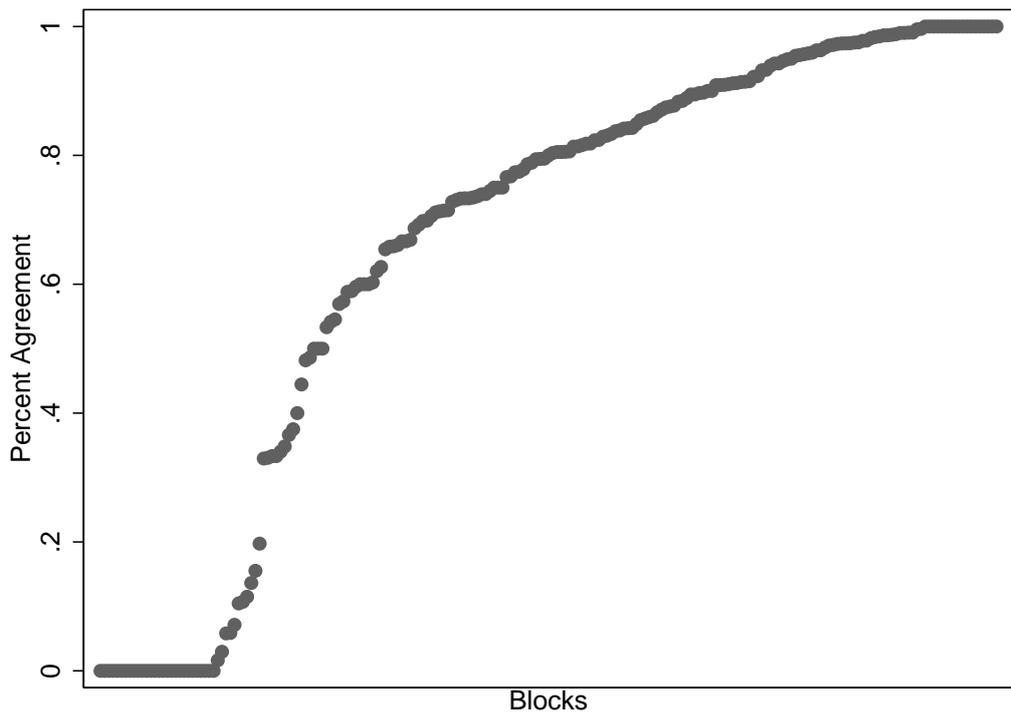


Figure 1.3: Block-level Agreement Rates. Horizontal Axis is the 215 Blocks, Sorted by Agreement Rate.

In the first row, the agreement probability for single family units that listers added is 22 percentage points lower than the agreement probability for single-family units they did not add, and this result is strongly significant ($\hat{\beta} = -0.220, z = -31.27$).⁹ The agreement probability for units in multi-unit buildings that were not added (and in segments with average crime rates) is seven percentage points lower than for single family units that were not added ($\hat{\beta} = -0.0773, z = -17.11$). There is also a strong and significant interaction effect between these two characteristics in the opposite direction, such that units that are both added and multi-unit (and in segments with an average crime rate) are 12 percentage points less likely to be listed by both listers than those which are neither added nor multi-unit. Trailers are also associated with inter-lister disagreement, though this coefficient is not significant.

The second set of independent variables in Table 1.5 refers to segment characteristics. Listers are more likely to agree about the inclusion of units in blocks with Hispanic residents: the agreement probability increases by 45 percentage points when the Hispanic population of a block increases from 0% to 100%, holding all other characteristics constant ($\hat{\beta} = 0.452, z = 2.61$). The larger the share of households earning less than \$45,000 per year in a segment,¹⁰ the less likely is agreement between the listers ($\hat{\beta} = -0.359, z = -2.77$). The proportion of the population that is African-American does not have a significant association with agreement.

Housing unit density and rural blocks are also not significantly associated with the

⁹This finding prefigures the discussion in Chapter 3 on confirmation bias.

¹⁰\$45,000 is approximately 200% of the federal poverty level for a family of four (DeNavas-Walt et al., 2009, pg. 43).

Table 1.5: Model of Probability of Agreement Between Two Listers

| | $\hat{\beta}$ | z |
|--|---------------|----------|
| Unit Added (1), on MAF (0) | -0.220*** | (-31.27) |
| Multi-Unit (1), Single Family (0) | -0.0773*** | (-17.11) |
| Multi-Unit * Added | 0.177*** | (18.99) |
| Unit is Trailer (1) or Not (0) | -0.0115 | (-1.31) |
| Proportion Pop. Hispanic | 0.452** | (2.61) |
| Proportion of HHs with income less than 45k | -0.359** | (-2.77) |
| Proportion Pop. African-American (only) | 0.0653 | (0.57) |
| HUs per land square mile, block level (standardized) | 0.0169 | (0.34) |
| Block rural (1) or not (0) | 0.0637 | (1.27) |
| Map, simple shape (1) or not (0) | -0.0722 | (-1.03) |
| Map, block has water boundary (1) or not (0) | 0.0236 | (0.46) |
| Map, block has non-visible boundary (1) or not (0) | 0.0520 | (1.01) |
| Per capita violent crime rate (standardized) | -0.0811* | (-2.53) |
| Violent crime rate * Multi-unit | 0.0197*** | (3.72) |
| Constant | 0.853*** | (11.23) |
| StdDev(Blocks) | 0.304 | |
| StdDev(Residual) | 0.303 | |
| rho | 0.502 | |
| Observations | 67205 | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

agreement probability. None of the three codes of the quality of the listing maps are significant predictors of agreement. In fact these map quality variables are all in the unexpected direction: simple blocks are associated with less agreement, and blocks with water or nonvisible boundaries are associated with more agreement.

The last two independent variables concern the effect of county-level per capita violent crimes on agreement between the listers. A one standard deviation increase in the crime rate leads to an eight percent decrease in the probability that two listers will agree about the inclusion of a single-family housing unit ($\hat{\beta} = -0.0811, z = -2.53$). This variable also interacts significantly with the indicator for units in multi-unit buildings. The effect of crime on the agreement probability is dampened by about two percentage points for multi-units ($\hat{\beta} = 0.0197, z = 3.72$).

1.4 Discussion and Conclusion

Segments with many multi-unit buildings, in high-crime areas and with low income households are those where listers are most likely to disagree. Surveys that focus on the poor and those in high crime neighborhoods may wish to use a multiple-lister design to capture housing units with low listing propensities and avoid the undercoverage that would likely result from the use of a single listing. Multiple listings capture housing units with low listing propensities, though they place burden on central office staff who must deduplicate the frames to control the probabilities of selection. Additional work is needed on how to raise the listing propensities of cases at risk of undercoverage.

Although no previous studies have looked at inter-lister agreement, many of the results above are consistent with previous research on lister error. These studies have shown that listers have trouble correctly covering trailers and multi-unit homes, and this finding is consistent with my results that listers tend to disagree about these units as well.

The failure of the map quality variables to account for inter-lister disagreement suggests that listers do not vary in how they interpret the low-quality listing maps. The coefficients on the three map quality variables are all in the opposite direction from what I expected, and have low t statistics. More research into how listers use and interpret these maps is clearly needed. (The dissertation by Rusch (2008) on the spatial abilities of listers and the design of listing software is a step in this direction.)

To reduce data collection costs, some surveys are moving away from housing unit listing to commercially-available address databases, or are considering such a move.¹¹ However, in-field listing is still thought to be the gold-standard. In fact, listed frames are often used as a benchmark against which the databases are compared (O'Muircheartaigh et al., 2003; Thompson and Turmelle, 2004; Turmelle et al., 2005; Dohrmann et al., 2006; O'Muircheartaigh et al., 2006; Dohrmann et al., 2007; O'Muircheartaigh et al., 2007). The central finding of this paper is that there is a good deal of variability in the frames different listers create, suggesting that these frames have limitations as gold standards and raising concerns about the

¹¹Those I know of include the General Social Survey, the National Study of Drug Use and Health, the National Health Interview Study, the National Children's Study, and the Survey of Consumer Finance.

estimates of the coverage rates of address databases.

Few studies of housing unit listing have explicitly modeled coverage propensities. But the findings of substantial inter-lister disagreement above demonstrate that the stochastic model of the listing process deserves a more central role in our thinking and research on listing error.

While this double listing dataset is unique and allows for interesting analyses, it does have several drawbacks. Most important, there is no gold standard. I have no grounds to assert that one lister's frame is more accurate than the other's. However, both of these listings did pass the Census Bureau's quality control procedures and thus each could serve as a frame for the many important household surveys they carry out.

The second drawback is that I do not have access to data about the listers. The most important differences between the first and second listings in this study are at the lister level. Information about the listers, their experience levels, training, education, etc., would make for a richer analysis. The dataset used in the remaining chapters does contain lister level data and allows for experimental manipulation to expose the mechanisms of lister error. Driven by the findings in this chapter, the next chapters make use of the stochastic conception of housing unit coverage.

Chapter 2

Mechanisms of Undercoverage in Traditional Housing Unit Listing

Chapter 1 used a large double-listing dataset to explore the correlates of lister disagreement and showed that inter-lister agreement rates vary quite a bit across segments and housing units. While the first chapter explored the correlates of this variation, the analyses were constrained by several limitations of the dataset. I was not able to manipulate the listing method to explore the different mechanisms of error in traditional and dependent listing. I also had no data about the listers who participated in the study.

In this chapter I provide a theoretical basis for errors in traditional listing, derive hypotheses, and test them using a smaller but more appropriate dataset collected for this purpose. I pay particular attention to the incentives listers face and how these can lead to frame error. Just as interviewers' incentives can lead to errors at other stages of the survey process (nonresponse bias (Manheimer and Hyman, 1949; Kennickell, 2000, 2003), sampling error (Boyd and Westfall, 1955, 1965, 1970; Alt, 1991; Eyerman et al., 2001) and measurement error (Matschinger et al., 2005)), the incentives built into the listing task can encourage listers to make inadvertent or even purposeful mistakes while listing. However, with this dataset, I find limited support for the hypotheses derived from this perspective.

2.1 Background & Hypotheses

There are two methods of in-field housing unit listing. In *traditional listing* (also called scratch listing), listers travel around each selected block in the segment and record the address or description of every housing unit (Kish, 1965; Survey Research Center, 1969, 1976). In *dependent listing* (also called update or enhanced listing), listers are provided with a list of addresses and travel around the segment, correcting the list to match what they see in the field. Estimates of net coverage¹ of listed housing unit frames range from 80% to more than 99% (Manheimer and Hyman, 1949; Kish and Hess, 1958; Hawkes, 1986; Jacobs, 1986; Joncas, 1985; Childers, 1992, 1993; Barrett et al., 2002; Pearson, 2003; Thompson and Turmelle, 2004; Turmelle et al., 2005; O’Muircheartaigh et al., 2006, 2007). At the segment level, low-income areas tend to be undercovered, as do rural areas and oddly-shaped segments (Subcommittee on Survey Coverage, 1990; O’Muircheartaigh et al., 2006, 2007).² At the housing unit level, units in multi-unit buildings are both undercovered and overcovered, as are renter-occupied units and vacant units (Subcommittee on Survey Coverage, 1990; Childers, 1992; Barrett et al., 2002). Lister characteristics have not been found to be statistically-significant predictors of coverage (Pearson, 2003). None of these coverage studies move beyond the correlates of coverage

¹The net coverage rate is the number of housing units listed on the frame divided by the number that should have been listed. A net coverage rate less than 100% means that undercoverage exceeds overcoverage, and this is usually the case. However, this metric obscures important differences between frames: a frame that contains a large amount of overcoverage offset by an equally large amount of undercoverage will appear to be just as accurate as a frame that contains no undercoverage or overcoverage. Unfortunately, most coverage studies do not collect the data that will allow them to separately calculate undercoverage and overcoverage rates.

²Kish and Hess (1958)’s findings at the block and segment level are different, but their method of creating a gold-standard frame make statements about coverage at these levels inappropriate, as they acknowledge.

rates to test theoretically-derived hypotheses about the mechanisms of lister error.

My hypotheses are motivated by the theories from economic sociology and behavioral economics, which in broad terms hold that individuals act in their own interests and respond to incentives, but are also subject to social norms that constrain self-interested behavior (Coleman, 1994). In this context it is important to understand the conditions in which listers work. I conducted 30 to 60 minute debriefings with seven listers and interviewers from the Survey Research Center at the University of Michigan to gain insight into their work situations. Extended quotations from these debriefings are given in Appendix F.

We send our listers and interviewers to neighborhoods they might never visit otherwise. These include very wealthy neighborhoods with gated homes, very poor neighborhoods with public housing, and neighborhoods where few people speak English. In my one-on-one debriefings with listers, they mention feeling unsafe or uncomfortable in some neighborhoods. One had a gun pulled on her in a remote part of Alaska. Others are nervous in secluded wooded areas. Listers have been pulled over by the cops for driving slowly and stopping often while listing. One lister mentioned the importance of knowing what colors the local gangs wore and avoiding them (lister debriefings). But even when people feel physically safe in strange neighborhoods, the novel environment can disturb behaviors and perceptions (Taylor et al., 1984; De Young, 1999).

In these unfamiliar settings, listers usually work entirely on their own, the only member of the project staff within hundreds of miles. When a lister encounters a difficult or unclear listing situation while in the neighborhood, she should

call the central office for guidance, and listers report doing this (lister debriefings). However, calling in is not always feasible. Listers are not provided with cell phones and some segments are out of range. Central office staff may not be available at the time the lister calls. Faced with obstacles to reaching project staff, the lister may decide to make her own choices about the appropriate listing behavior, especially if she is far from home or feels awkward in the neighborhood. But listers are not sampling statisticians or principal investigators and do not know the larger goals of the survey. For these reasons the judgments they make may not be in the best interests of data quality.³ If listers do behave in the ways suggested by these theories, then they likely weigh what they know about the survey and the goals of the listing task as well as their own interests in getting home in time to pick up children or prevent another drive out to a remote area.

The last important piece to understanding lister working conditions is that their frames cannot be thoroughly reviewed. Traveling in another lister to check the work is prohibitively expensive.⁴ Particularly in the NSFG design (discussed in more detail below) where the same person lists a segment and later interviews there, no one else associated with the project may ever visit the segment. Listers may well be aware that their work cannot be checked in the field.

The listing task is one where listers work on their own, in unfamiliar neigh-

³To be fair, the effects of housing unit coverage on overall survey quality are not known to sampling statisticians or principal investigators, either. This dissertation is an effort to fill in this knowledge.

⁴Only the Census Bureau performs in-field relisting of segments as a check on the accuracy of their listed frames, and they check at most two blocks per interviewer per year (personal communication with Rodrick Marquette, Mathematical Statistician, Decennial Statistical Studies Division, Census Bureau.)

neighborhoods, knowing that no one will discover any errors they make. These are the hallmarks of the principal-agent model. A principal-agent problem arises whenever a principal hires an agent to perform a task, and the agent has information about the work that principal does not have. The classic principal-agent problem involves a landlord (principal) and a sharecropper (agent). The landlord wants the sharecropper to put forth a high level of effort, but he cannot observe her effort level. He can observe only the amount of final output, which is a function of both the agent's effort level and environmental conditions (soil quality, pests, rainfall, etc.) that are known only to the agent. Knowing that the landlord cannot tell how hard she is working, the sharecropper will not work as much as she could, or as much as the landlord would like. This model has also been applied to insurance markets, industrial regulation, executive compensation, and financial markets (Sappington, 1991; Stiglitz, 2008).

Thinking about listing as a principal-agent problem reveals that a careful consideration of lister incentives is important to understanding error in housing unit frames. In this chapter I develop and test hypotheses inspired by the principal-agent model about the mechanisms of undercoverage in traditional listing.

Crime

Interviewers pursue different contact strategies in neighborhoods where they do not feel safe (Martin, 1981; Groves and Couper, 1992). It is likely that listers behave differently in these areas as well. High quality listing involves walking down alleys and gangways, entering multi-unit buildings and often talking to residents.

In a dangerous neighborhood, listers may be less likely to do each of these. Thus I suspect that listers will create frames of lower quality in neighborhoods where they do not feel safe.⁵ The impact of crime on undercoverage is likely stronger with units in multi-unit dwellings, which often require more investigation to list accurately.

Invisible Boundaries

Some blocks have boundaries that are not obvious on the ground; they may correspond to power lines, underground streams, political boundaries, or natural features that no longer exist. When a lister is assigned such a block, she often has trouble determining which units near the boundary should be included on the frame. Listers report sometimes speaking with local residents and officials or using a GPS device (which is not provided) to find the boundary (lister debriefings), all of which mean extra effort for the lister. Errors in determining the boundaries can lead to undercoverage of units near the boundary.

Language Spoken in Segment

Accurate listing often requires reading signs ("For Rent" or "Coming Soon, Luxury Apartments") and asking questions of segment residents. If the residents and the lister do not speak a common language, however, listers cannot ask how many units are in a building or whether anyone lives above the strip of commercial stores. Thus segments where the primary language is not one the lister speaks are likely to be undercovered and this undercoverage should show up most strongly in

⁵Listers' perceptions of crime are likely related not only to actual crime rates in the neighborhood but also to the physical state of the area and the demographic makeup (Sampson and Raudenbush, 1999, 2004). I will use listers' own reports of the safety of the segment to test this hypothesis, not arrest or victimization data, to isolate the effect of listers' perceptions of crime.

dwellings in multi-unit buildings.

Driving

Driving while listing certainly makes the task more comfortable: a lister can stay warm (or cool), listen to the radio, and drink a cup of coffee. Many interviewers prefer to drive around their segments while listing, even when instructed not to (personal communication with NSFG Cycle 7 Staff, 2008). Listers drive when they are uncomfortable on the street with a laptop or when the sun makes the screen of the laptop difficult to see (lister debriefings). But driving can lead to undercoverage of units not visible from the street: a lister in a car cannot check around the back of a structure for a second entrance or multiple gas meters. Here again I expect more undercoverage of units in multi-unit buildings with this behavior.

Lister Motivation

Interviewers of course differ in their interests and motivations. Some listers likely take on the interests of the principal investigator as their own and others likely see the work as a job like any other. I expect that those listers who are more motivated by the pay of the interviewing task will commit more errors of undercoverage.

Hypothesis: Listers who find a segment unsafe commit more errors of undercoverage than those who do not find the segment unsafe.

Hypothesis: Units in multi-unit buildings are more likely than single-family units to be undercovered in high crime segments.

Hypothesis: Housing units on blocks with invisible block boundaries are undercovered.

Hypothesis: Listers undercover housing units in segments where the primary language is not one they themselves speak.

Hypothesis: Units in multi-unit buildings are more likely than single-family units to be undercovered in segments where the lister does not speak the language.

Hypothesis: Driving while listing is associated with undercoverage of multi-units.

Hypothesis: Units in multi-unit buildings are more likely than single-family units to be undercovered while driving.

Hypothesis: Listers more strongly motivated to take the interviewing job by financial concerns will undercover housing units.

2.2 Data

To test these hypotheses I had listers from the National Survey of Family Growth (NSFG) relist a nationally-representative sample of segments used in the survey. The NSFG is a national area probability study conducted by the Division of Vital Statistics at the National Center for Health Statistics to study fertility behavior (Groves et al., 2005). Data collection for NSFG Cycle 7 is carried out by the Survey Research Center at the Institute for Social Research, University of Michigan. In the current design, all interviewers are also listers: in every quarter they interview cases in their active segments and list housing units in segments that

will be active the next quarter. (For this reason in this paper I use the terms interviewer and lister interchangeably.) NSFG listers record housing units into a tablet computer while in the field. The software ensures that listers parse addresses into fields (street number, street name, apartment designator) and provides a drop down menu of known street names to minimize spelling errors and standardize abbreviations. Listers also record segment level observations on the computer: method of travel around the segment, languages spoken in the segment, safety and accessibility concerns, and type of housing units.

Segment Selection

I randomly selected 13 primary sampling units (PSUs) containing 49 segments from the NSFG quarter 12 sample. The three goals of the selection process were to overrepresent segments that are likely to be more difficult to list according to previous findings in the existing literature, ensure a diverse representation on the variables involved in my hypotheses, and select a nationally representative sample. I split the quarter 12 PSUs into two strata. The first stratum contained all segments that were particularly helpful in meeting the first two goals above, as well as segments in the same PSUs as these segments. The second stratum contained the 64 segments in the remaining PSUs. I selected all 10 PSUs (40 segments) from the first strata and three PSUs (nine segments) from the second. Because each quarter of data collection for NSFG is nationally representative, my random sample from quarter 12 could also represent the entire country.⁶

⁶As discussed below, however, I do not have segment selection weights for the quarter 12 segments, thus all analyses are unweighted. Without weights, the sample of segments cannot be said to be nationally representative.

The resulting sample is diverse with respect to per capita violent crime rates and percent of multi-family units and also with respect to income and percent African-American, as desired.⁷ According to the observations collected by the first lister, 17 of the selected segments contain Spanish speakers, 13 contain gated buildings, three contain at least one trailer, and in 16 of the segments the lister reported safety concerns.

The first listing of each segment was conducted by the interviewer assigned by the project. That lister used dependent listing in 38 of the 49 selected segments, and traditional listing in the other 11. SRC central office staff performed several quality checks of the first listing as part of their usual protocol. They checked the order of house numbers, checked the ratio of listed counts to the Census housing unit counts and reviewed the blocks with online mapping software (personal communication with NSFG Cycle 7 Staff, 2009).

The segment was then listed a second time using traditional listing in every segment. The second lister was also a trained NSFG lister and interviewer who used the listing software and made independent segment observations. The second listing was never done by the same lister who did the first and was not subject to SRC quality checks.

⁷The crime data is at the county-level and comes from the Federal Bureau of Investigation's Uniform Crime Reports. I calculated violent crimes per capita at the county level and then flagged those segments in the quarter 12 sample which were above the median. Violent crimes are defined in the UCR as murder, rape, robbery and aggravated assault (United States Department of Justice, Federal Bureau of Investigation., 2009). The multi-family indicator is derived from the first listing: for each segment I calculated the percent of all addresses that were in multi-family buildings (based on a non-empty apartment field) and then flagged those segments above the median. The income and race variables were similarly dichotomized. The data for these comes from the Census 2000 SF3 files (U.S. Census Bureau, 2002a).

Listers

NSFG administers a questionnaire to all of their interviewers (see Appendix D). All interviewers who participated in my study completed this questionnaire, though one interviewer chose not to respond to several of the questions. The questions cover interviewer experience, motivation and attitudes, as well as race, ethnicity and language measures.

Eleven interviewers performed the second listing of the segments. The number of segments listed by each interviewer ranges from one to nine. More than half the listers have a bachelor's or master's degree and only one has no college at all. All listers have at least one year of interviewing experience. Five report holding another job while working for NSFG, three of these have another interviewing job. Three interviewers are African-American (the others are white) and three speak Spanish; one is both African-American and speaks Spanish.

Several items on the interviewer questionnaire attempt to capture the interviewers' motivations in taking the job with NSFG. The questionnaire asks interviewers to use a ten point scale (where 1 is low and 10 is high) to report the attractiveness of four aspects of the job: flexible working hours, importance of survey research, pay and interacting with a variety of people (see Q25 in Appendix D). Only on the pay variable did the interviewers use the entire scale (on the other three motivation measures, all selected eight or higher). However, this pay variable is not appropriate for my motivation analyses. First, it has missing data (one lister declined to answer this question). Second, the use of the work "attractiveness" in the question text is unfortunate: interviewers may indicate that the pay is not

attractive because they feel the rate is too low, not because they do not need the income. Thus I don't think this variable truly captures interviewers' motivations in the sense I mean. A fifth variable, which has no missing data, captures whether the lister has another job in addition to working as an NSFG interviewer. Because the NSFG job requires 30 hours per week, those listers who hold another job are likely motivated more by financial concerns than those who do not.

Matching

Before testing my hypotheses I matched the frames created by the first and second listing of each segment. I used several techniques of increasing permissiveness to match units that would result in the same household being contacted for screening and interview. The quality of this matching work will impact the quality of all my analyses. I provide details on the procedures used in matching all three listings of these segments (the third is not involved in this paper) in Appendix C.

Three of the 49 segments posed particular challenges to matching. These are among the most rural segments, where listers used descriptions to identify housing units because there are no addresses. Because the matching in these three segments is not yet complete, most results below refer only to the 46 segments, rather than to all 49.

2.3 Analysis Methods

To test the above hypotheses about the effect of interviewer incentives on frame quality, I fit multilevel listing propensity models. A binary variable indi-

cating whether each unit was listed by the second lister or not is the dependent variable in the listing propensity models.

Explanatory Variables

The independent variables in the models are at three levels: listers, segments, and housing units. See Table 2.3 for a summary of the variables, and their means and ranges. The first three sets of variables given in this table are those that are used in the multivariate models discussed below. Those in the fourth set are not used in the models.

Lister At the lister level are the responses to the NSFG interviewer questionnaire.

The content of the questionnaire is given in Appendix D.

Segment At the segment level are Census demographic data on linguistic isolation, urban versus rural, income and percent of the population that is African-American from the 2000 Census (U.S. Census Bureau, 2001a,b, 2002a,b).⁸ I also have the lister's segment observation and her method of travel around the segment. There are several measures of map quality at this level as well (see Appendix A on the creation of the map quality measures.) Also at this level are interactions of the segment and lister level characteristics, such as whether the interviewer speaks the language of the segment residents.

Housing Unit At the housing unit level are indicators for a unit in a multi-unit

⁸The first three of these variables (linguistic isolation, urban versus rural, and income) are available at the block group level. Most segments include blocks from only one block group. For those segments that cross block group boundaries, I averaged the data across the relevant block groups to calculate segment level statistics. Data on the percent of the population that is African-American is available at the block level and thus this variable refers precisely to the segment.

building, for a trailer, and for units with no house number. Details on the creation of housing unit level variables are given in Appendix E.

Case Base

I run these models on the housing units listed by the first lister to find the characteristics at the different levels that are associated with higher and lower listing propensities by the second lister. I do not include those units listed by the second listers and not the first (there were 1029 of these across the 46 segments) because I have no basis for distinguishing between overcoverage by the second lister and undercoverage by the first. For example, if the second lister included many multi-unit buildings on Jackson Ave that the first lister did not include, is it because those units were missed by the first lister? Or because the units should not have been listed? Perhaps the units on Jackson Ave are professional offices and not residential, or perhaps that block of Jackson is outside of the segment. Ideally I would run one model for the units that all listers should have covered, to find the characteristics that make those units more or less likely to be correctly listed, and another model for the units that no lister should have included. Unfortunately, I do not have a gold standard frame for these segments.

However, for the cases that were selected for the NSFG interview, I have disposition data that indicates whether they were proper listings or not. NSFG selected 2,114 cases in these segments. (Only cases listed by the first lister were eligible for selection due to cost and timing constraints.) For these cases, I have disposition data from the NSFG screening effort about whether each case was eligible

| Variable | Mean | Std. Dev. | Min. | Max. | N |
|--|------|-----------|------|------|------|
| Multi-Unit | 0.19 | 0.4 | 0 | 1 | 1970 |
| Vacant | 0.1 | 0.3 | 0 | 1 | 1970 |
| Trailer | 0.01 | 0.07 | 0 | 1 | 1970 |
| Proportion Pop. Spanish language | 0.05 | 0.1 | 0 | 0.39 | 1970 |
| Proportion Pop. African-American | 0.17 | 0.25 | 0 | 1 | 1970 |
| Map, invisible boundary | 0.4 | 0.49 | 0 | 1 | 1970 |
| Lister feels unsafe | 0.11 | 0.32 | 0 | 1 | 1970 |
| Lister drove herself while listing | 0.67 | 0.47 | 0 | 1 | 1970 |
| Lister and segment language match | 0.8 | 0.4 | 0 | 1 | 1970 |
| Proportion HUs rural | 0.06 | 0.19 | 0 | 1 | 1970 |
| Proportion HHs with income $\leq 50,000$ | 0.56 | 0.21 | 0.12 | 0.89 | 1970 |
| Map, blocks have simple shape | 0.2 | 0.4 | 0 | 1 | 1970 |
| Map, segment has external water boundary | 0.13 | 0.33 | 0 | 1 | 1970 |
| Interviewer Hispanic | 0.17 | 0.38 | 0 | 1 | 1970 |
| Interviewer African-American | 0.26 | 0.44 | 0 | 1 | 1970 |

for the screener or not, which allows me to distinguish somewhat between overcoverage by the first lister and undercoverage by the second listers. The fact that the first listing was more thoroughly checked by the SRC central office staff than the second listing also supports my use of the cases selected from the first listing as the universe for my analyses.

Of the 2,114 cases selected from the first listing, just over one percent (25) were dispositioned by the interviewers as improperly listed lines (non-residential or out-of-segment).⁹ These lines should not have been listed by any of the listers. In the 46 segments where all of the three-way matching is complete, 1,994 lines were selected and 24 of these were improperly listed.

Using the properly-listed lines among the selected cases in the segments where the matching is complete, I can run the models described above on the observations from the second listing of each unit. (Including the first units would skew the results as all the selected cases were listed by the first lister.)

Models

Because the dependent variable in my models is binary, logistic regression models are an obvious choice. However, given the complexities in interpreting interaction effects in logistic and other nonlinear models (Allison, 1999; Ai and Norton, 2003; Mood, 2010) I present and interpret linear probability models, as suggested by Wooldridge (2009) and Mood (2010). I did run the same models as logistic regressions and they do not change the substantive conclusions. The logistic models

⁹I cannot tell for which reason the interviewers coded these housing units as inappropriate listings.

are given in Appendix B.

Each segment was relisted using traditional listing by only one lister, and each housing unit is in only one segment. Thus my multilevel models could have three nested levels: listers, segments and housing unit. Alternatively, the segments are nested within 13 PSUs and the models could be specified in that way. However, the highest level of clustering (whether lister or PSU) does not have any effect in the models I run below (the intraclass correlation coefficient for listers or PSUs is always very near 0), so the models I run below include random effects at only the segment level. I include fixed effects for the 11 listers to remove their idiosyncratic influence on listing propensity and isolate the underlying mechanisms (Snijders and Bosker, 1999; Kohler and Kreuter, 2005).

2.4 Results & Discussion

Table 2.1 shows the number of units listed in the first and second listing. The first pair of columns in this table gives the number of cases listed by each lister in each segment and does not rely on the matching work. Figure 2.1 shows the ratio of the size of the two listings graphically (number listed by the second lister divided by the number listed by the first). Points on the left, below the reference line, correspond to segments where the second lister listed fewer units than the first and those in the upper right where the second lister included many more. (The labels in the figure refer to the segment numbers in Table 2.1.) Already we can see quite a bit of diversity both between the two listings and across the segments.

Table 2.1: Number of Housing Units Listed in First and Second Listing, by Segment

| Segment | All Cases | | Selected Cases | |
|---------|-----------|-----------|----------------|-----------|
| | Listing 1 | Listing 2 | Listing 1 | Listing 2 |
| 1 | 55 | 48 | 19 | 7 |
| 2 | 106 | 110 | 19 | 17 |
| 3 | 155 | 154 | 19 | 17 |
| 4 | 164 | 182 | 19 | 17 |
| 5 | 89 | 79 | 39 | 34 |
| 6 | 87 | 87 | 34 | 34 |
| 7 | 130 | 162 | 34 | 34 |
| 8 | 210 | 204 | 29 | 27 |
| 9 | 108 | 101 | 27 | 26 |
| 10 | 93 | 106 | 27 | 25 |
| 11 | 122 | 119 | 54 | 54 |
| 12 | 98 | 94 | 62 | 52 |
| 13 | 149 | 124 | 62 | 48 |
| 14 | 96 | 101 | 15 | 9 |
| 15 | 159 | 194 | 55 | 28 |
| 16 | 109 | 111 | 54 | 50 |
| 17 | 96 | 93 | 43 | 42 |
| 18 | 108 | 89 | 43 | 33 |
| 19 | 83 | 83 | 43 | 42 |
| 20 | 74 | 81 | 34 | 34 |
| 21 | 80 | 80 | 38 | 38 |
| 22 | 84 | 84 | 38 | 38 |
| 23 | 122 | 139 | 38 | 35 |
| 24 | 584 | 580 | 38 | 38 |
| 25 | 95 | 96 | 38 | 34 |
| 26 | 88 | 84 | 71 | 67 |
| 27 | 626 | 626 | 63 | 61 |
| 28 | 103 | 100 | 63 | 62 |
| 29 | 165 | 198 | 89 | 81 |
| 30 | 271 | 312 | 86 | 64 |
| 31 | 152 | 226 | 86 | 86 |
| 32 | 95 | 97 | 42 | 40 |
| 33 | 233 | 131 | 42 | 17 |
| 34 | 2,337 | 2,146 | 43 | 36 |
| 35 | 99 | 95 | 55 | 53 |
| 36 | 82 | 82 | 54 | 54 |
| 37 | 95 | 95 | 55 | 55 |
| 38 | 162 | 162 | 55 | 55 |
| 39 | 417 | 403 | 50 | 46 |
| 40 | 236 | 239 | 50 | 50 |
| 41* | 94 | 95 | 46* | * |
| 42 | 110 | 110 | 47 | 26 |
| 43* | 110 | 104 | 46* | * |
| 44 | 118 | 122 | 7 | 2 |
| 45 | 82 | 86 | 8 | 7 |
| 46 | 144 | 141 | 8 | 6 |
| 47 | 78 | 119 | 28 | 22 |
| 48* | 160 | 158 | 28* | * |
| 49 | 110 | 113 | 71 | 63 |
| Total | 9,423 | 9,345 | 1,994 | 1,766 |

* Matching not complete, segment dropped in models

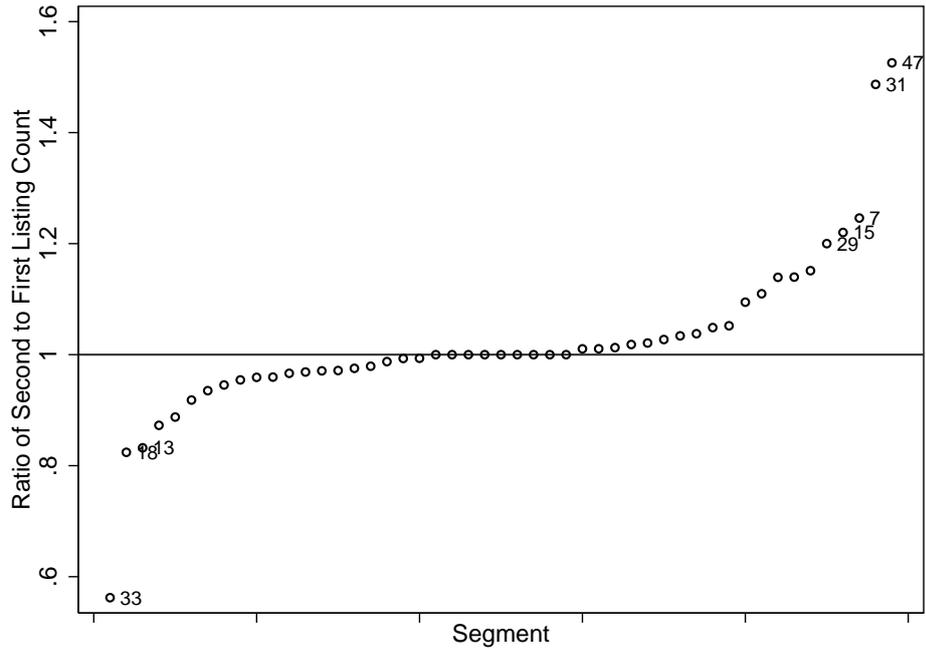


Figure 2.1: Ratio of Number of Housing Units Listed by First Lister to Number Listed by Second, by Segment

Looking at only the 1,994 selected lines in the segments where matching is complete, 88.5% were listed by the second listers. (Note that all selected lines were listed by the first lister, because cases were selected for the NSFG survey from the frame created by the first lister.) Among the 1,970 selected cases that the interviewers found to be residential units inside the segments, the cases which are the universe for my models below, 89.2% were listed by the second lister.

2.4.1 Comparison to Previous Results

As discussed above, previous work has shown that trailers, vacant units and those in multi-unit buildings, as well as housing units in low-income, rural and oddly-shaped blocks, are prone to undercoverage. To ground my results in previous

work, the upper half of Table 2.2 gives the percent of the selected units listed by the second lister, broken down by these characteristics.

Housing units in multi-unit buildings, vacant units and trailers are all more likely to be undercovered than single family buildings, occupied units and non-trailers. Units in complex blocks, rural blocks and low-income blocks are also undercovered. All of these results are in the expected direction. The F statistics in the last column account for the clustering of housing units into 13 PSUs. Only two pairs, trailers versus non-trailers and rural versus non-rural, show significantly different coverage rates. However, power analyses revealed that due to the high degree of clustering in my data, I do not have enough PSUs to detect significant differences on the segment and lister characteristics. Thus the lack of significant findings on these variables is not surprising.

The fact that my dataset confirms previous findings in listing research provides reassurance that the matching work is accurate and that there is nothing unusual about the segments in my study. However, the results of these earlier studies do not connect with a larger theoretical framework. I now turn to testing the hypotheses motivated by consideration of lister incentives in section 2.1.

2.4.2 Tests of Hypotheses

I hypothesized that high crime segments, listing while driving, unclear segment boundaries, and mismatches between lister and segment languages contribute to undercoverage in traditional listing. The lower half of Table 2.2 compares the

Table 2.2: Comparison of Listing Rates by Housing Unit and Segment Characteristics

| | n | Listing % | <i>F</i> stat |
|----------------------|------|-----------|---------------|
| Multi Unit | 382 | 83.0% | 3.33 |
| Single Family | 1588 | 90.7% | |
| Vacant | 200 | 83.0% | 3.64 |
| Occupied | 1770 | 89.9% | |
| Trailer | 10 | 60.0% | 14.44* |
| Non-Trailer | 1960 | 89.4% | |
| Simple shape | 386 | 93.0% | 1.50 |
| Complex shape | 1584 | 88.3% | |
| Rural | 315 | 75.9% | 5.49* |
| Not Rural | 1655 | 91.8% | |
| Low Income | 918 | 88.1% | 0.16 |
| High Income | 1052 | 90.2% | |
| Invisible Boundary | 797 | 85.3% | 2.36 |
| Visible Boundaries | 1173 | 91.9% | |
| Lang. Matched | 1575 | 89.5% | 0.06 |
| No match | 395 | 88.4% | |
| Safety concerns | 225 | 88.9% | 0.004 |
| No concerns | 1745 | 89.3% | |
| Drove alone | 1318 | 90.8% | 1.30 |
| Walked or was driven | 652 | 86.0% | |
| Other job | 970 | 90.7% | 0.54 |
| No other job | 1000 | 87.8% | |
| HH Screened | 1602 | 90.1% | 3.84 |
| Not Screened | 368 | 85.3% | |

* Difference significant at 5%

listing rates by these characteristics and provides an initial test of the hypotheses. Units in segments with invisible boundaries are less likely to be covered than those in segments without invisible boundaries. This result is as expected. Segments where the lister speaks the language of the residents are covered somewhat better than those where the lister does not. Units in segments where the traditional lister had safety concerns are covered at a lower rate than those in segments where the lister did not have concerns, but this difference is very small. Driving while listing is associated with higher listing propensities, in contrast to my hypothesized effect.¹⁰ Listers who hold other jobs also do better, which contradicts my hypothesis. None of the differences in the lower half of the table are significant due in part to the lack of power in my highly clustered dataset.

Given the mixed support for my hypotheses in the bivariate analyses of Table 2.2, I turn to the multi-level models described above, which simultaneously control for possibly confounding segment and housing unit characteristics. The models in Table 2.3 show the impact of lister, segment and housing-unit level variables on the listing propensity of the selected cases. In each model the dependent variable is a binary indicator of whether the second lister listed the selected housing unit (1) or did not (0). Positive coefficients indicate characteristics that are associated with a greater likelihood of being listed by the traditional lister. Accepting the premise that all of these housing units should have been listed by all listers, positive coefficients suggest that those characteristics make a unit *less likely* to be undercovered,

¹⁰Note that driving was not randomly assigned in this dataset. Each lister could decide for herself whether to drive or walk while listing. Thus segment characteristics that lead a lister to choose to drive may affect the estimate of this difference.

and negative coefficients suggest those that make a unit *more likely* to be undercovered.

The first model in the table simply provides a baseline for comparison. The intra-segment correlation coefficient (ρ) is 24.4%, meaning that almost one-quarter of the variability in listing propensities is within segments and three-quarters is between segments.

The second model includes explanatory variables that previous studies have found to be correlated with listing propensity. Units in multi-unit buildings are 17 percentage points less likely to be covered in traditional listing than single family units ($\hat{\beta} = -0.173, z = -6.96$). Vacant units and trailers also have lower listing propensities, by three and one percentage points respectively, but these effects are not significant. The more rural a segment is, the more likely the units in that segment are to be undercovered. The positive coefficient on the poverty measure, proportion of households in the block group with incomes less than \$50,000, suggests that the larger the share of households below this income threshold, the higher the listing probability of all units in the segment. This finding is the only one that is not in the expected direction. A likelihood ratio test of this model against the first is statistically significant ($p < 0.001$). The share of variance within segments (ρ) drops to 17.5% controlling for these five variables.

The third model adds segment and lister characteristics that are unique to this listing study. Because the study involves 46 segments and only 11 listers, only a few variables could be added without overwhelming the model. This model adds two demographic variables, percent Spanish speakers and percent African-

Table 2.3: Traditional Listing Linear Probability Models, Selected Cases Only

| | (1) | (2) | (3) | (4) |
|--------------------------------------|---------------|---------------|---------------|---------------|
| | $\hat{\beta}$ | $\hat{\beta}$ | $\hat{\beta}$ | $\hat{\beta}$ |
| | z | z | z | z |
| Multi-Unit | | -0.173*** | -0.171*** | -0.173*** |
| | | (-6.96) | (-6.80) | (-3.44) |
| Vacant | | -0.032 | -0.033 | -0.032 |
| | | (-1.51) | (-1.52) | (-1.50) |
| Trailer | | -0.009 | -0.006 | -0.006 |
| | | (-0.10) | (-0.06) | (-0.06) |
| Proportion HUs rural | | -0.236* | -0.243* | -0.238 |
| | | (-2.00) | (-1.99) | (-1.95) |
| Proportion HHs with income <= 50,000 | | 0.091 | 0.046 | 0.045 |
| | | (0.83) | (0.32) | (0.31) |
| Proportion Pop Spanish language | | | 0.001 | 0.018 |
| | | | (0.00) | (0.04) |
| Proportion Pop. Afr.-Amer. | | | 0.175 | 0.193 |
| | | | (0.84) | (0.93) |
| Map, invisible boundary | | | -0.060 | -0.058 |
| | | | (-1.03) | (-0.99) |
| Lister feels unsafe | | | -0.165 | -0.149 |
| | | | (-1.30) | (-1.16) |
| Lister drove herself while listing | | | 0.011 | 0.017 |
| | | | (0.17) | (0.27) |
| Lister and segment language match | | | -0.038 | -0.041 |
| | | | (-0.51) | (-0.55) |
| Multi * Lister feels unsafe | | | | -0.053 |
| | | | | (-0.80) |
| Multi * Lister drove | | | | -0.015 |
| | | | | (-0.27) |
| Multi * Language match | | | | 0.024 |
| | | | | (0.35) |
| Constant | 0.877*** | (36.20) | 0.950*** | (9.68) |
| | | | 1.006*** | (7.36) |
| | | | | 1.001*** |
| | | | | (7.31) |
| StdDev(segment) | 0.157 | 0.125 | 0.119 | 0.119 |
| StdDev(residual) | 0.276 | 0.272 | 0.272 | 0.272 |
| rho | 0.244 | 0.175 | 0.161 | 0.161 |
| Log Likelihood | -316.426 | -279.600 | -277.991 | -277.667 |
| Pseudo R2 | | 0.116 | 0.121 | 0.122 |
| Observations | 1970 | 1970 | 1970 | 1970 |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

American, and those variables needed to test the hypotheses developed above. The coefficients on the controls from the first model are largely unchanged and maintain their signs and significance patterns. As expected, units in segments with invisible boundaries ($\hat{\beta} = -0.060$) and those where the lister feels unsafe ($\hat{\beta} = -0.165$) have lower listing propensities. After controlling for segment characteristics that might capture listers' decisions to drive while listing, such as percent rural, driving still has an unexpected, positive effect on coverage by traditional listers ($\hat{\beta} = 0.011$). When the lister speaks the language of the segment residents, the listing propensity drops by almost 4 percentage points ($\hat{\beta} = -0.038$). However, none of the variables added in this model are statistically significant. A likelihood ratio test of this model against the second fails to reject, indicating that this model does not explain listing propensity better than the previous.

The fourth model adds interactions of three variables with the multi-unit indicator. Coefficients on the other variables are largely unchanged between models 2 and 3, though the coefficient on percent rural is now just below the threshold for significance at the 5% level. As discussed in the hypothesis section, the effects of safety concerns, driving, and speaking the language of segment residents on undercoverage are expected to be stronger for units in multi-unit buildings. When a lister feels unsafe, the listing propensity of single family units is reduced by 15 percentage points and multi-units by 20 percentage points. Both of these results are in the expected direction. The effect of driving on listing probabilities is 1.7 percentage points (in the positive direction) for single family units, but only 0.2 percentage points for multi-units. A lister who speaks the language of the segment reduces the

listing propensity of single family units by four percentage points and multi-units by 1.7 percentage points. These effects are in the unexpected direction.

The main effects in model 3 and the interaction effects in model 4 test the hypotheses developed above. None of the coefficients on these variables is significant, but several are in the expected direction. Units in blocks with invisible boundaries, and in segments where the lister feels unsafe, have reduced listing propensities, and safety concerns reduce the propensities of units in multi-unit buildings even more than single family. However, driving increases listing propensities and speaking the language of the segment decreases them; these effects are not as expected, though both effects do move towards the expected sign when interacted with multi-unit status.

2.5 Conclusions

This paper has used theories from economic sociology, specifically the principal-agent model, to develop hypotheses about how listers' incentives contribute to error in housing unit frames. For this purpose I collected a multiple-listing dataset of 49 segments throughout the United States. While the data replicate earlier findings about the correlates of listing propensity, both bivariate and multivariate tests of my hypotheses lead to limited support for the hypotheses. Given the variables and data available, it does not appear that listers' incentives are the main drivers that lead them to make errors of undercoverage in traditional listing. However, the listing data studied here was quite clustered, involving only 11 listers, which makes

tests of the effects of lister characteristics on coverage propensity quite underpowered. It is possible that this lack of power in part explains the poor support for the hypothesized explanations.

One clear finding from this work is that there is a good deal of undercoverage in traditional listing, and we should continue to look for theories that will explain the phenomenon. I am excited by work in environmental psychology about the effects of built environments on individual's perception, spatial understandings and cognitive skills. Future work may find these theoretical approaches more valuable than those tested here.

I note that one interesting finding in this dataset is that housing units that completed the NSFG screener were four percentage points more likely to be listed by the second lister ($p < 0.08$, Table 2.2), suggesting a connection between undercoverage and nonresponse that warrants additional exploration.

Chapter 3

Confirmation Bias in Dependent Housing Unit Listing

Chapter 2 explored the mechanisms of error in traditional listing. While the findings in that paper corroborated previous work in the literature, I found only limited support for my hypotheses concerning the effects of listers' incentives on the quality of their listing work. This chapter focuses on dependent listing, where the error mechanisms are likely to be different.

In the summer of 2009, the Census Bureau conducted a very large listing operation to prepare for the 2010 census. More than 100,000 Address Canvassers updated the Master Address File to ensure that enumeration forms would go out to every household. They essentially listed the entire country.¹ One canvasser wrote a whistle-blowing post for the blog My2Census.com, a decennial census watchdog website, detailing the problems of the operation in New York City. The post, included as Appendix G, points out the difficulties of urban listing.

There are small tenement buildings in Chinatown and Harlem brownstones; where there are illegal subdivisions. It is very difficult to gain entry or make contact even if you speak the language.

The list of addresses on the master Address File was loaded onto the listers' handheld computers (HHC) and they updated this list in the field. The whistle-blower

¹The Address Canvassing operation did not include remote parts of Alaska and Maine (personal communication with Robin Pennington, Census Bureau, November 6, 2008.)

reports that s/he received pressure to simply confirm the prior listing.

We were told that if we couldn't gain access to a building after two visits we had to accept what was in the HHC as correct. Many of us were tempted to falsify work and accept what was in the HHC. . . One of the other listers found an entire building with over 200 single illegally divided rooms. The HHC had less than 10 units listed in it. If they accepted what was in the HHC as true they would have missed over 200 housing units.

Incentives to finish quickly also led listers and supervisors to confirm the existing list rather than carefully check each building for missed, or inappropriate, units.

It was alleged that some of the crew leaders and field operations supervisors told their listers since there was no regard to quality that they could skip making contact even going as far as not conducting field work and enter the units at home. There is no way that listers who were reassigned work magically gained access to buildings people couldn't access for weeks unless they accepted what was in the HHC as true. The crew leaders and field supervisors who finished first were rewarded with additional work. Those who finished last were sometimes "written up" as unproductive and the office terminated their employment.

The type of listing described by this lister is *dependent* listing. Listers are sent into the field to update an existing list of addresses called the *input listing*. They delete

non-existent or nonresidential units and add units that are missing. Dependent listing is used to create housing unit frames for many surveys as well. Often the input list is a previous traditional listing of the area or a geocoded address database. This paper provides strong evidence that listers using dependent listing do indeed tend to confirm errors of exclusion and inclusion on the input list, leading to errors of undercoverage and overcoverage on their frames.

3.1 Background & Hypotheses

In housing unit listing, whether for the decennial census or for household surveys, we worry about two types of error: undercoverage and overcoverage.² Undercoverage occurs when housing units inside the selected area are not listed. Chapter 2 focused only on undercoverage, which can raise concerns about bias in survey estimates, if the undercovered units are different than the correctly covered units (Wright and Tsao, 1983; Groves, 1989; Lessler and Kalsbeek, 1992). Overcoverage is the inclusion of elements on the frame that should not have been listed and it can be further classified into two types. *Out-of-scope overcoverage* is the inclusion of non-residential or non-existent units. While this kind of overcoverage increases data collection costs slightly, it does not affect survey estimates.³ *Multiple-probability overcoverage* is the inclusion of elements that are in the target popula-

²Kish (1965) lists four kinds of frame error: undercoverage, overcoverage, duplicates and clustered units. Lessler (1980) adds two others: inaccurate auxiliary data (stratification and size variables) and insufficient locating data. In this paper I am only concerned with undercoverage and overcoverage.

³Out-of-scope overcoverage does introduce variability into the sample size and can therefore increase the variance of estimates slightly, but I will not investigate this effect here.

tion but whose probability of selection should come from elsewhere. For example, a lister might misread her segment maps and list an additional block along 16th St. While these housing units are part of the survey's target population, their chance to be selected comes from their own segment; including them twice inappropriately and unknowingly inflates their probability of selection. If selected, these units may well be interviewed and brought into the final sample and thus can affect estimates (Wright and Tsao, 1983; Lessler and Kalsbeek, 1992).

In dependent listing, undercoverage can occur in two ways: either (1) the units were on the list and the lister inappropriately removed them, or (2) the units were not on the input list and the lister failed to add them. The two types of overcoverage can occur because either (3) the lister added inappropriate units, or (4) the units were on the input list and the lister failed to delete them. Situations 1 and 3 are similar—in each case the input list is correct and the lister introduces an error. In situations (2) and (4), it is the list that is incorrect, and the lister fails to correct it. This paper concentrates on these two situations where the lister introduces undercoverage and overcoverage by failing to correct problems with the input list. I call this phenomenon confirmation bias.

Eckman and Kreuter (2010) provided the first study of confirmation bias in dependent listing. In a small listing in Ann Arbor and Ypsilanti, Michigan, they introduced errors of undercoverage and overcoverage into the input list. They found that listers tend to confirm that the input listing is correct and thus to transfer those errors to the housing unit frame. When the input list includes an incorrect unit, the lister has a tendency not to delete it, a *failure-to-delete* error. Seventeen percent of

the units added to the input list were confirmed by the listers. Conversely, when a lister sees units inside the segment that are not on the list, she has a tendency not to add them, a *failure-to-add* error. Suppressing a housing unit from the input listing decreased its listing propensity by 13.4 percentage points. Eckman and Kreuter find some support for the hypothesis that both types of confirmation bias are more likely with units in multi-unit buildings. This paper extends these findings by investigating the confirmation bias phenomenon in more depth. First, this work uses a larger geographic sample to explore whether the phenomenon exists on a larger scale. Second, it digs deeper into the mechanisms at work in confirmation bias.

Confirmation bias is not unique to housing unit listing. In the social psychology literature, the term confirmation bias has several meanings (Klayman, 1995). The one closest to the use in the survey field is the tendency to look for corroborating evidence (Hitchcock, 1995, pg 324) or for “the presence of what you expect” (Klayman, 1995, pg 386). In dependent interviewing, when a second interviewer can see the result recorded by the first, s/he is more likely to collect the same response (O’Muircheartaigh, 2004; Lynn and Sala, 2006). We see the same result in dependent coding when the second coder sees the code assigned by the first (Biemer and Lyberg, 2003). In both cases the work of the second person is affected by expectations set by prior information. Dependent listing is similar: the second lister has prior information that may create expectations about what she will find in the field.⁴

⁴The psychological confirmation bias literature has also influenced another thread of research in survey methodology on the effect of interviewers’ expectations, about incentives, response rates and item sensitivity, on their production (Singer and Kohnke-Aguirre, 1979; Singer et al., 1983, 2000).

My use of the term confirmation bias is closer in spirit to its use in the coding literature. I suspect that the underlying mechanism in confirmation bias in housing unit listing is an *appeal to authority*, where the input list serves as the authority. Several of the listers I spoke to talked about unclear or difficult listing situations, much like those described in the blog post quoted above, where they looked to the list for guidance about the existence and designation of the housing units in the area. They mentioned locked buildings where they could get no information about the number of units, and segments with invisible boundaries where they could not tell what was inside and what outside the segment. In these situations a few listers said all they could do was trust the list. Another lister talked about a more general hesitancy to contradict the list in any situation. (To be fair, two listers also mentioned the importance of not relying too much on the list, suggesting some inter-lister variability in the confirmation bias phenomenon.) See Appendix F for details from the interviewer debriefings.

These discussions with listers, as well as the post by the address canvasser above, suggest that confirmation bias will be more likely in difficult listing situations. We know a bit about the kinds of units and segments that are difficult for listers from Chapters 1 and 2. In these chapters I found that listers tend to disagree about the inclusion of units in rural segments, those with complex shapes, those with invisible boundaries, and those with more high income households. They also disagree about the inclusion of units in multi-unit buildings.⁵ If the confirmation

⁵While Chapter 1 finds strong disagreement about trailers, I do not have enough trailers in this dataset I use in this chapter to analyze them separately.

bias phenomenon found by Eckman and Kreuter stems from the input list serving as an authority in difficult listing situations, we should see more confirmation bias in these situations.

I also suspect that the overall level of error in the list affects the listers' tendency to confirm errors in the list. If the input listing is very inaccurate, confirmation bias may be less likely, as the authority of the list is undermined. The lister may essentially turn to traditional listing if the input list contains too many errors. I suspect that confirmation bias of both types (failure-to-add and failure-to-delete) is most likely when the input list is largely accurate. (Of course, confirmation bias cannot occur when the input list is entirely accurate.)

Hypothesis: Listers tend to fail to add units not on the input list.

Hypothesis: Listers tend to fail to delete units that do not exist from the input list.

Hypothesis: Both types of confirmation bias are more likely in rural segments.

Hypothesis: Both types of confirmation bias are less likely in segments where the blocks all have a simple square shape.

Hypothesis: Both types of confirmation bias are more likely in segments with an invisible external boundary.

Hypothesis: Both types of confirmation bias are more likely in segments with a high percentage of high income households.

Hypothesis: Both types of confirmation bias are more likely in multi-unit buildings.

Hypothesis: Both types of confirmation bias are less likely in segments where the input list contains many errors.

3.2 Data

To test these hypotheses I again use the NSFG repeated listing dataset described in Chapter 2. In the previous chapter, I discussed only two listings in this dataset. The first listing was conducted by the project to support normal sampling and interviewing procedures. The second listing used traditional listing in every segment. Each of the 49 selected segments was also listed a third time, using dependent listing. Like the second listing, the third was not subject to SRC's quality review.

The listers who performed the third (dependent) listing of the segments were drawn from the same pool as those who did the other two listings—trained and experienced NSFG lister and interviewers. However, those who participated in the third listing happen to have somewhat different qualities as reported in the interviewer questionnaire (the questionnaire is given in Appendix D). There were 11 listers involved in the third listing and each lister listed between three and nine segments. Only one is African-American, and six speak Spanish. All have completed at least two years of college. These listers are quite experienced with between four and fourteen years of interviewing work before they began working on NSFG Cycle 7. The dependent listers are as a whole better educated and more experienced than the traditional listers in my study.⁶

⁶These differences in lister attributes are not by design. I was not able to select the listers who

3.2.1 Manipulation of Input List

To test the above hypotheses about confirmation bias, I experimentally manipulated the input to the third (dependent) listing. The foundation of the input was the frame created by the first listers, prior to the quality checks performed by SRC. I added addresses not listed by the first lister and deleted addresses that were listed by the first lister. If confirmation bias exists, the third lister should show a tendency not to correct these errors.⁷

Units Deleted from Input Listing

I deleted 556 units from the input list, 5.9% of those listed by the first listers. The lowest deletion rate by segment was 1.5% and the highest was 18.3%. The deleted housing units are of four types as shown in Table 3.1.

These deletions were performed quasi-randomly. Every housing unit in the first listing was assigned a probability of being deleted depending on whether the case was selected⁸ and the manipulation group of the segment (as discussed below). For each housing unit I generated a number between zero and one from a uniform distribution and if the number was less than the probability assigned to the case I flagged the unit for deletion. The allocation of the flagged cases to deletion type was not entirely random, but based on the overall deletion rate in the segment and participated in my study nor assign them to segments or methods.

⁷The 1980 Census used a similar technique to review the work on precavass enumerators (Fan et al., 1984), but I have not seen any results from this check.

⁸In doing the deletions, I gave preference to deleting units which had been selected for the NSFG screener and interview. My thinking at the time was that deleting selected cases would allow me to control for cases which the dependent lister failed to add because they were not real units. However, so few of the selected cases were found to be improper listings (less than 1%, as discussed in Chapter 2) that this safeguard was unnecessary and is not utilized in the analyses below.

plausibility constraints.

Table 3.1: Number of Cases Deleted from Input to Third Listing, by Type

| Type of Deleted Unit | Cases | Percent |
|-------------------------------------|-------|---------|
| Entire multi-unit building | 87 | 15.7% |
| Unit in multi-unit building | 50 | 9.0% |
| Single-family | 263 | 47.3% |
| All housing units on street segment | 156 | 28.1% |
| Total | 556 | |

Units Added to Input Listing

To test for failure-to-delete error, I added 421 housing units to the input listing. The lowest addition rate by segment was 0.64% and the highest was 14.2%.

The added units are of five types, as given in Table 3.2.

Table 3.2: Number of Cases Added to Input to Third Listing, by Type

| Type of Added Unit | Cases | Percent |
|---|-------|---------|
| Units on new street segment | 45 | 10.7% |
| Unit in multi-unit building | 59 | 14.0% |
| Building in midst of others | 167 | 39.7% |
| Turn single-family into make multi-unit | 88 | 20.9% |
| Units outside of segment | 62 | 14.7% |
| Total | 421 | |

These manipulations were also randomized using a method similar to that described above for the deletions. However, it was not always possible to add a unit at the point specified by the randomization, for example, between units five and six in an eight unit building. Thus I gave myself some leeway to deviate from the

randomly selected spots when adding units. I tried to add units that could seem plausible to the listers. For example, in a building with three units numbered 1,2 and 3, I might add a unit 4, or a basement unit. In a street with house numbers increasing by four (504, 508, 512) I might add a two unit building at 510. When adding units across the street, I used online satellite images and real estate websites to find addresses of housing units that very likely were across the street from the segment.

Manipulation Groups

To test whether the overall level of error also affects confirmation bias, I varied the level of manipulation at the segment level. I randomly split the 49 selected segments into four sets and varied the degree of manipulation, as shown in table 3.3. The addition and deletion rates given in the table are at the unit level, not at the manipulation level: one manipulation could have added or deleted many units.

These manipulation groups allow me to test the hypothesis that when the input list is quite inaccurate, listers commit fewer errors of confirmation bias. I compare the fourth group (low deletion and addition rates) to the other three groups because this group will likely appear to the lister as obviously different than the other three. This fourth group is the one where the input list is of highest quality; in each of the other three groups, the sum of the addition and deletion rates is greater than 10%.

Matching Frames

To prepare the dataset for analysis, I again had to match the frames together,

Table 3.3: Level of Manipulation in Four Segment Sets

| Deletions | Additions | Segments | Deletion rate | Addition rate |
|-----------|-----------|----------|---------------|---------------|
| High | High | 12 | 6.3% | 5.3% |
| High | Low | 12 | 8.6% | 2.4% |
| Low | High | 12 | 4.6% | 7.6% |
| Low | Low | 13 | 3.4% | 2.6% |
| Overall | | 49 | 5.9% | 4.7% |

but the matching steps were more complicated than those described in Chapters 1 and 2. This matching task involved two separate matching operations, each using multiple steps.

I first matched the input list given to the third lister (with the suppressed lines added back in) to the frame created by these listers, to identify which lines were confirmed, deleted and added by the listers. Then I matched the three frames to each other. The quality of this matching work will impact the quality of all my analyses. Details on the procedures used in both rounds of matching are in Appendix C. As discussed in Chapter 2, the matching could not be completed in three very rural segments. The incomplete matching affects some of my analyses in this chapter, but not all.

3.3 Analysis Methods

I use a variety of techniques to test the hypotheses developed above. Testing for failure-to-add and failure-to-delete errors requires slightly different techniques for reasons described below. Most tests use the traditional listers as a control group not subject to the input list manipulation. These analyses can use only the cases in

the 46 segments where the matching work is complete.

3.3.1 Testing for Failure-to-Add Error

I use three techniques to test my hypotheses about failure-to-add confirmation bias. First, I compare listing rates for the unmanipulated and deleted units. Because the deletions to the input list are nearly random, these rates provide evidence for confirmation bias. If the dependent listers show a tendency towards confirmation bias, the deleted units should be listed at a lower rate than the unmanipulated units.

Second, I calculate a difference-in-differences estimate of the effect of deleting cases from the input listing. This technique uses the traditional listers, who were not subject to the manipulation, as a control group.⁹ Let L_{unm}^{dep} be the fraction of unmanipulated cases on the input list that were listed by the dependent lister. L_{del}^{trad} is the fraction of cases deleted from the input list that were listed by the traditional lister. L_{del}^{dep} and L_{unm}^{trad} are defined similarly. Then $L_{del}^{dep} - L_{del}^{trad}$ captures the difference in the listing rates for the deleted units between those listers who were and were not subject to the manipulation. This difference is part of the effect I am interested in, but it does not take advantage of the experiment by comparing the manipulated to the unmanipulated units. Conversely, $L_{unm}^{dep} - L_{del}^{dep}$ captures the difference in the listing rates between the unmanipulated and manipulated cases.

⁹The difference-in-differences technique is commonly used with panel data to derive treatment effects from non-randomized designs: each case serves as its own control and the difference in the change from period 1 to period 2 between those who did and did not receive the treatment is the average treatment effect (Angrist and Pischke, 2009, pp. 221–247).

(This difference is the same as the first analysis method discussed above.) The shortcoming of this approach is that any systematic variation in the deleted and unmanipulated units could bias the estimate. The more appropriate estimator of the treatment effect is the difference-in-differences estimate, which adjusts for anything that is unique about the deleted cases and takes advantage of the experiment:

$$D - in - D = (L_{unm}^{trad} - L_{del}^{trad}) - (L_{unm}^{dep} - L_{del}^{dep}) \quad (3.1)$$

This estimate depends on the accuracy of the match between the traditional and dependent listers. It can use only the 46 segments where the matching between the second and third frames is complete.

The third analysis technique expands upon these difference-in-differences results by simultaneously controlling for housing unit and segment level characteristics in estimating the size of the failure-to-add effect. Just as in Chapter 2, I use linear probability models with fixed effects for the eleven listers and random effects for the segments. The dataset used in the model is at the housing unit and listing level: each of the unmanipulated and deleted housing units appears in the dataset twice, once for the traditional and once for the dependent listings. The binary dependent variable is whether a given lister listed a housing unit (1) or did not (0). The independent variables of interest are a dummy variable indicating method (traditional listing is the reference category), a dummy variable indicating whether the housing unit was deleted from the input listing, and the interaction of these two, which captures the effects of the deletion on the dependent listers. Including

the indicator of deletion at the housing unit level controls for any unobserved attributes that make the suppressed units harder or easier to list. The models also control for many of the other housing unit and segment characteristics used in the models in the previous chapters. Just as in Chapter 2, the models do not account for the selection probabilities of the segments. While the area sample in each quarter of NSFG data collection is nationally representative, the weights to make inference from a single quarter do not exist.¹⁰

3.3.2 Testing for Failure-to-Delete Error

All of the analysis techniques discussed in the previous section compared the deleted units to the unmanipulated. For the cases added to the input list, a different logic applies. Here relying on authority would mean failing to delete the added units, giving them a positive listing propensity. No comparison to the unmanipulated units is necessary.

I again use three techniques to explore failure-to-delete errors in dependent listing. The first technique compares the listing rates of the added cases to the null hypotheses that none of these units should have been listed. The second analysis compares the listing rates for the added housing units across the two listing methods. If the traditional listers also included some of these added units, that suggests the units do in fact exist and the estimate of confirmation bias should be reduced. In the notation introduced above, the second analysis calculates $L_{add}^{dep} - L_{add}^{trad}$. The third analysis, using linear probability models, expands on the comparison between

¹⁰I plan to soon develop weights for the 49 segments in my sample and rerun the models.

the traditional and dependent listing of the added units by controlling for housing unit and segment characteristics.

3.4 Results

Together the results of these analyses give a clear picture of both the failure-to-add and the failure-to-delete effects in dependent listing. The manipulation of the input list provides evidence that listers commit errors of both kinds in dependent listing.

3.4.1 Failure-to-Add

As discussed above, if dependent listers are susceptible to failure-to-add confirmation bias, then they should be less likely to list housing units deleted from the input list than those not deleted. Table 3.4 shows that only 63.8% of the deleted units were included in the dependent-listed frame, versus 95.7% of the units not deleted from the input listing. That is, removing cases from the input listing reduced the likelihood that these units would be listed by almost 32 percentage points. The difference in listing rates is even larger for housing units in multi-unit buildings, and less pronounced in single family units. All differences in the listing rates between the deleted and unmanipulated lines in this table are highly statistically significant.

To test the sensitivity of the difference in overall listing rates to individual lister behavior, I re-estimated this overall result, dropping the segments listed by

each of the eleven listers in turn. The estimates of the difference in the listing rates between the unmanipulated and deleted units ranged from 25.0% to 34.0%, suggesting that the estimate shown in Table 3.4 is not due to unusual behavior by any one lister. Every lister failed to add back one or more deleted units in her segments.

Table 3.4: Percent of Unmanipulated and Deleted Units Listed in Dependent Listing

| | | n | Percent Listed | F Statistic |
|---------------|-------------------|------|----------------|-------------|
| Overall | Unmanipulated | 8862 | 95.7% | |
| | Deleted | 556 | 63.8% | |
| | <i>Difference</i> | | 31.8% | 31.94* |
| Multi-Unit | Unmanipulated | 2808 | 98.7% | |
| | Deleted | 152 | 57.2% | |
| | <i>Difference</i> | | 41.5% | 21.51* |
| Single Family | Unmanipulated | 6054 | 94.3% | |
| | Deleted | 404 | 66.3% | |
| | <i>Difference</i> | | 28.0% | 22.56* |

* Significant at the 5% level

The difference-in-differences estimate is given in Table 3.5. In the second column are the listing rates for the unmanipulated and deleted lines in the traditional (second) listing. (While these listers were not subject to the manipulation, by matching the frames together I can determine the share of these lines listed by the traditional listers.) The listing rates for the deleted units are consistently smaller than those for the unmanipulated lines, even among the traditional listers. This suggests that there is something different about the deleted lines that makes them harder to list, as expected due to the way in which lines were selected for deletion. For this reason, the results in Table 3.4 may overstate the failure-to-add effect by

not controlling for this fact. Table 3.5 compares the listing rates for the unmanipulated and deleted units across the two listing methods. In every case the dependent listers included a larger share of the unmanipulated units and a smaller share of the deleted units. In the last column are the difference-in-differences estimates of the effect of deleting units from the input to the dependent listing, as defined in Equation 3.1. In these 46 segments, removing cases from the input listing reduces the likelihood that those cases will be included on the frame by 26.5 percentage points. The results are again larger in magnitude for units in multi-family buildings.

Table 3.5: Failure-to-Add: Difference-in-Differences, in Percentage Points

| | | Housing Units | Traditional Pct. Listed | Dependent Pct. Listed | D-in-D Estimate |
|---------------|---------------|------------------|----------------------------|--------------------------|--------------------|
| Overall | Unmanipulated | 8519 | 86.67% | 96.11% | -26.46 |
| | Deleted | 535 | 81.68% | 64.67% | |
| Multi-unit | Unmanipulated | 2803 | 84.94% | 98.90% | -34.31 |
| | Deleted | 152 | 77.63% | 57.23% | |
| Single family | Unmanipulated | 5716 | 87.51% | 94.77% | -22.93 |
| | Deleted | 383 | 83.29% | 67.62% | |

Three segments dropped because matching not complete, data will not match Table 3.4

The results of the multi-level regression models are given in Table 3.6. The first set of variables are controls which have been found to be correlated with listing propensity in bivariate analysis in previous work (though they are not significant here). In the first model, the first variable in the second set shows that dependent listers are nine percentage points more likely to list the unmanipulated units in high manipulation segments than are traditional listers ($\hat{\beta} = 0.094, z = 22.92$).

The next row in the table shows that the deleted units are in four percentage points less likely to be listed by the traditional listers than those that were not manipulated ($\hat{\beta} = -0.041, z = -3.40$), which we also saw in Table 3.5. Unmanipulated units in multi-unit buildings are less likely to be listed by traditional listers ($\hat{\beta} = -0.085, z = -12.68$). In the next row, unmanipulated housing units in segments selected to be part of the low manipulation group do not have significantly different listing probabilities in traditional listing, as expected, because assignment to manipulation groups was random.

The third set of variables tests the effect of the deletion of cases from the input list on the dependent listers. These variables are all interaction effects. The first row in this set is the two-way interaction of the deletion of units from the input list and the dependent listing method. The manipulation of the input list has a strong and negative effect on the listing propensity of single family units in high manipulation segments in dependent listing and this effect is strongly significant ($\hat{\beta} = -0.252, z = -13.32$). Deleting these cases from the input list reduces their propensity to be listed by dependent listers by 20 percentage points ($0.094 - 0.041 - 0.252 = -0.199$) relative to the propensity of unmanipulated single-family cases in high manipulation segments in traditional listing.

In the next row of this section of independent variables, the tendency of dependent listers to add back deleted single family units is stronger in segments in the low manipulation group (that is, where the input list is of high quality) than in segments where the manipulation rate is higher ($\hat{\beta} = 0.131, z = 3.62$). Said the other way around, dependent listers are thirteen percentage points more likely to

Table 3.6: Failure-to-Add: Listing Propensity Models on Unmanipulated and Deleted Cases

| | (1) | | (2) | |
|--|---------------|----------|------------------|----------|
| | $\hat{\beta}$ | z | $\hat{\beta}$ | z |
| Map, simple | 0.018 | (0.29) | 0.018 | (0.27) |
| Map, invisible boundary | -0.060 | (-1.11) | -0.059 | (-1.05) |
| Pct. HUs rural | -0.072 | (-0.62) | -0.073 | (-0.60) |
| Pct. HHs with income <= 50,000 | 0.044 | (0.36) | 0.044 | (0.35) |
| Pct. Pop. Afr.-Amer. | 0.032 | (0.27) | 0.030 | (0.24) |
| Pct. Pop Spanish language | 0.039 | (0.12) | 0.046 | (0.13) |
| Dependent method (1), Traditional (0) | 0.094*** | (22.92) | 0.094*** | (22.97) |
| Unit Deleted | -0.041*** | (-3.40) | -0.041*** | (-3.40) |
| Multi-Unit | -0.085*** | (-12.68) | -0.086*** | (-12.85) |
| Low rate of manipulation in segment | 0.051 | (1.13) | 0.052 | (1.10) |
| Unit deleted * Dependent method | -0.252*** | (-13.32) | -0.444*** | (-13.98) |
| Unit deleted * Dependent * Low rate of manipulation | 0.131*** | (3.62) | 0.097*** | (2.66) |
| Unit deleted * Dependent * Multi-unit | -0.100*** | (-3.79) | | |
| Unit deleted * Dependent * Entire multi-unit building deleted | | | <i>reference</i> | |
| Unit deleted * Dependent * Unit in multi-unit building deleted | | | 0.219*** | (4.55) |
| Unit deleted * Dependent * Single family unit deleted | | | 0.269*** | (7.92) |
| Unit deleted * Dependent * All units on street deleted | | | 0.084* | (2.29) |
| Constant | 0.837*** | (9.05) | 0.835*** | (8.65) |
| StdDev(segment) | 0.116 | | 0.121 | |
| StdDev(residual) | 0.269 | | 0.269 | |
| rho | 0.156 | | 0.169 | |
| Observations | 18108 | | 18108 | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

commit failure-to-add confirmation bias in segments where the list is of low quality. This result contradicts the hypothesis that when the list contains more errors, listers notice the problem and do a better job of fixing the input list, making fewer confirmation errors.

In the last row of this section of independent variables, the interaction of the deletion of units and dependent listing with multi-units is negative and significant ($\hat{\beta} = -0.100, z = -3.79$). That is, when units in multi-unit buildings are deleted from the input list, they are ten percentage points less likely to be added back by dependent listers than are deleted single family units. The increase in failure-to-add error for multi-units is as expected and as found in the other analyses above.

The second model in Table 3.6 compares the different types of deleted units. Here the deletion of an entire multi-unit building is the reference category. Each of the other types of manipulations has a positive coefficient, meaning that the reference category is the one which experiences the most confirmation bias. When the deleted unit is a single unit in a multi-unit building, it is 22 percentage points more likely to be added back to the frame by dependent listers than are the units in the buildings which were deleted entirely ($\hat{\beta} = 0.219, z = 4.55$). Deleted single family units are 27 percentage points more likely to be added back ($\hat{\beta} = 0.269, z = 7.92$). When all units on a street segment were deleted (such as all units on the even side of the 400 block of Baltimore Ave) they are eight percentage points more likely to be added back than are the units in an entirely-suppressed multi-unit building ($\hat{\beta} = 0.084, z = 2.29$).

3.4.2 Failure-to-Delete

Turning attention to the 421 units added to the input list, if dependent listers are susceptible to failure-to-delete confirmation bias, then they should show a tendency not to delete these units. Table 3.7 shows that 24.9% of the added units were confirmed by the lister using dependent listing. A sensitivity analysis showed that this effect is not due to just one lister; dropping each lister in turn yielded a range of estimates from 14.4% to 27.7%. Two listers did not confirm any of the added units in their segments, though they did have among the fewest added cases (20 and 7). A larger share of the added units in multi-unit buildings were confirmed than those in single family homes (27.5% versus 23.5%), though the difference between these two is not significant.

Table 3.7: Percent of Added Units Listed in Dependent Listing

| | n | Percent Listed |
|---------------|-----|----------------|
| Overall | 421 | 24.9% |
| Multi-Unit | 153 | 27.5% |
| Single Family | 268 | 23.5% |

The second failure-to-delete analysis, shown in Table 3.8, compares the listing rates for the added cases in each of the two listing methods. In the first row we see that 6.6% of the housing units that were added to the input list were also listed by the traditional lister. This result suggests that in a few cases I fabricated units that really did exist. (Eckman and Kreuter also found that one of the units they added existed in Ann Arbor. Of course, another possibility I must acknowledge for this finding is matching error.) The dependent listers confirmed 25.1% of the

added cases. Adding units to the input list raises their listing propensity by 18.5 percentage points. Again the effect is stronger in units in multi-family buildings.

Table 3.8: Failure-to-Delete: Comparison of Listing Rates between Traditional and Dependent Listing for Added Cases

| | Housing Units | Traditional Pct. Listed | Dependent Pct. Listed | Difference (Pct. Points) |
|---------------|------------------|----------------------------|--------------------------|-----------------------------|
| Overall | 410 | 6.6% | 25.1% | 18.5 |
| Multi-Unit | 149 | 0.67% | 26.8% | 26.2 |
| Single family | 261 | 10.0% | 24.1% | 14.2 |

The regression models in Table 3.9 expand upon these results. This linear probability models is run on both the traditional and dependent listings of the 410 added housing units in the 46 segments where the matching is complete. The dependent variable is a binary indicator of whether the lister *deleted* the added units (1) or not (0) (Note that the dependent variable in this model, in contrast to other regression models in Chapters 2 and 3, is not listing but deletion.) Negative coefficient estimates again signify characteristics associated with confirmation bias just as in Table 3.6. The first set of independent variables in the table are again controls, and again none are significant. The second set of independent variables show no strong effects for units in the low manipulation segments, as expected. Units in multi-unit buildings have a deletion propensity nine percentage points higher than single family units ($\hat{\beta} = 0.092, z = 2.74$), meaning that listers of both methods were more likely *not to list* the 149 added units that were in multi-unit buildings.

The third set of independent variables tests hypotheses about failure-to-delete confirmation bias. The first variable in the third set estimates the difference in the

Table 3.9: Failure-to-Delete: Deletion Propensity Models on Added Cases

| | (1) | | (2) | |
|--|---------------|---------|------------------|---------|
| | $\hat{\beta}$ | z | $\hat{\beta}$ | z |
| Map, simple shape | -0.013 | (-0.23) | -0.003 | (-0.06) |
| Map, invisible boundary | 0.023 | (0.45) | 0.026 | (0.48) |
| Pct. HUs rural | -0.005 | (-0.04) | 0.015 | (0.11) |
| Pct. HHs with income <= 50,000 | 0.073 | (0.64) | 0.133 | (1.07) |
| Pct. Pop. Afr.-Amer. | -0.193 | (-1.56) | -0.186 | (-1.40) |
| Pct. Pop Spanish language | 0.143 | (0.44) | 0.183 | (0.52) |
| Low rate of manipulation in segment | 0.035 | (0.63) | 0.036 | (0.63) |
| Multi-Unit | 0.092** | (2.74) | 0.037 | (1.23) |
| Dependent method (1), Traditional (0) | -0.145*** | (-5.03) | -0.018 | (-0.35) |
| Dependent method * Low Manip. Rate | 0.025 | (0.37) | 0.015 | (0.23) |
| Dependent method * Multi-Unit | -0.121** | (-2.62) | | |
| Dependent * Units added in new street | | | <i>reference</i> | |
| Dependent * Unit added in multi-unit building | | | -0.208** | (-2.99) |
| Dependent * Building between others | | | -0.224*** | (-3.97) |
| Dependent * Single Family turned into multi-unit | | | -0.177** | (-2.70) |
| Dependent * Added unit outside segment | | | -0.094 | (-1.41) |
| Constant | 0.908*** | (11.44) | 0.878*** | (10.02) |
| StdDev(segments) | 0.053 | | 0.069 | |
| StdDev(residual) | 0.315 | | 0.313 | |
| rho | 0.028 | | 0.047 | |
| Observations | 820 | | 820 | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

listing rates between the two methods for the added lines, the simple failure-to-delete effect. Listers using the dependent method are 15 percentage points less likely to delete the added single-family cases in high manipulation segments than the traditional listers ($\hat{\beta} = -0.145, z = -5.03$). The failure-to-delete effect does not interact significantly with segments in the low manipulation group. The effect is stronger for units in multi-unit buildings ($\hat{\beta} = -0.121, z = -2.62$): dependent listers are 12 percentage points less likely to delete these units than they are single family units. The stronger effect for multi-units is as expected and as found in the other failure-to-delete (and failure-to-add) analyses.

The second model adds in a test of the strength of the effects among the different kinds of added units. All of the estimated coefficients are negative and all but the last are significant, meaning that the reference category (units added in a new street) is the type of manipulation that listers were *least* likely to confirm.

3.4.3 Summary and Interpretation of Results

The results above are clear evidence that both failure-to-delete and failure-to-add confirmation bias exist in dependent listing. Housing units inside the segment which are not on the input list are at risk of undercoverage. Those on the input list inappropriately are at risk of overcoverage. These results replicate previous findings on confirmation bias and demonstrate that the phenomenon is not localized to student listers or to southeast Michigan, as in the Eckman and Kreuter study. These analyses used a national dataset and quite experienced listers. I find strong

support for my first two hypotheses about the existence of the types of confirmation bias.

This paper expands on previous work not only geographically but also by exploring the housing unit and segment characteristics associated with confirmation bias. I hypothesized that confirmation bias, of both types, is more common in segments where traditional listers also have difficulties with undercoverage: rural segments, complex-shaped blocks, segments with invisible boundaries and those with a high percent of high income households.

Table 3.10 summarizes the difference-in-differences results for failure-to-add and the single difference results for failure-to-delete by important segment and housing unit characteristics. These bivariate results provide tests of my hypotheses about the effect of segment characteristics on confirmation bias. (While I could have added tests of these hypotheses as additional manipulation effects in the multivariate models, I believe this table is easier to interpret). Nonrural segments are more susceptible to both kinds of confirmation bias, in contrast to my hypothesis. I suspect this contradictory finding is due to the concentration of multi-unit buildings in non-rural segments. Segments made up of complex-shaped blocks, those with nonvisible boundaries, and those with more high income households, are more likely to experience confirmation bias, as hypothesized. These findings support my supposition that dependent listings commit confirmation error in situations that also give traditional listers trouble, suggesting that the input list serves as an authority in difficult listing situations.

I do not find strong support for my hypothesis that when the input list con-

Table 3.10: Comparison of Listing Rates of Manipulated Cases in Traditional and Dependent Listing, in Percentage Points

| | Failure-to-Add (Diff-in-Diff) | Failure-to-Delete (Difference) |
|------------------------------|----------------------------------|-----------------------------------|
| Overall | -26.46 | 18.54 |
| Rural ¹ | -19.23 | 15.00 |
| Not Rural | -27.45 | 18.92 |
| Map, simple shape | -14.8 | 13.69 |
| Map, complex shape | -28.03 | 19.58 |
| Invisible boundary | -26.68 | 22.84 |
| No invisible boundary | -25.68 | 12.92 |
| Low Income ² | -15.34 | 7.75 |
| High Income | -34.26 | 23.49 |
| Low Manip. Rate ³ | -18.82 | 17.31 |
| High Manip. Rate | -27.19 | 23.68 |

Shading indicates the member of each pair which is greater in absolute value

¹ Segments above median, percent of housing units in rural blocks

² Segments above median, percent of households with income < \$50,000

³ Low-low group compared to high-high group, see Table 3.3

tains a good deal of error, listers realize it is no longer an authority and commit fewer confirmation errors. If that were the case, then a high error rate in the input list should lead listers to question the authority of the list and reduce instances of confirmation bias. Instead I find that the more error in the input list, the more likely is confirmation bias. The bivariate differences in Table 3.10 show less confirmation bias in segments where the list has fewer errors (is of high quality), not more. The multivariate results also show that listers commit fewer errors, of both types of confirmation bias, when the manipulation rates are low (though the coefficient on this term is significant only in the failure-to-add model). These results suggest that appeal to authority is not the right framework for understanding the confirmation bias phenomena. Another possible explanation is that the error rates in this study (see Table 3.3) were not large enough. Or perhaps listers perceive error rates differently than I measure them here. Additional work is needed to uncover the mechanisms underlying confirmation bias.

3.5 Discussion & Conclusions

These results have important implications for surveys which use dependent listing in whole or in part to create housing unit frames, which includes many Census Bureau household surveys (U.S. Census Bureau, 2006), the National Survey of Family Growth (NSFG Cycle 7 Staff, 2008), the Residential Energy Consumption Survey (personal communication with Krishna Winfrey, National Opinion Research Center, January 14, 2010), and of course the decennial census. Surveys which rely

on dependent listing are much more reliant on the quality of their input listing than has been thought. Errors of inclusion and exclusion on the input listing are likely to be transmitted to the final frame due to confirmation bias. If the kinds of units undercovered and overcovered by the input listing are different than those which are properly covered, then confirmation error can introduce coverage bias into survey data.

For this reason, survey organizations that use dependent listing should have a good understanding of the determinants of the quality of their input lists. Unfortunately, our understanding of the error in the listings and databases that serve as input lists are often based on studies which use dependent listing to identify errors (studies which use this technique include O’Muirheartaigh et al., 2003; Thompson and Turmelle, 2004; Turmelle et al., 2005; O’Muirheartaigh et al., 2006, 2007). The results in this paper suggest that those estimates of input frame quality are too high as they themselves suffer from confirmation bias.

Indeed the confirmation bias findings in this paper call into question the findings of all listing studies which use dependent listing to create a gold standard frame. The Census Bureau routinely checks listers’ work by sending a senior field representative to list the segment, using the original frame as an input (personal communication with Rodrick Marquette, Mathematical Statistician, Decennial Statistical Studies Division, Census Bureau.) Other studies which use this technique include Hansen and Steinberg (1956) and Pearson (2003). The assertion in these studies is that experienced listers using dependent listing produce a gold standard frame. All dependent listers in this study, however, had at least four years of ex-

perience and all committed confirmation bias. I discuss the implications of these three chapters for gold standard frame construction in Chapter 5.

This paper has largely treated failure-to-add and failure-to-delete confirmation bias as similar phenomena—both can lead to errors on housing unit frames and to bias in survey estimates. However, the two are very different in the minds of many survey practitioners. The listers (and SRC trainers) I spoke to were much more concerned with undercoverage than overcoverage (lister debriefings, personal communication with NSFG staff). I have heard trainers encourage listers to err on the side of inclusion, rather than exclusion: “When in doubt, list it.” This attitude comes from a belief that overcoverage is less of a concern than undercoverage, and that a reasonable amount out-of-scope overcoverage is not a problem for surveys. While this instruction has good intentions, it may contribute to the failure-to-delete tendency observed in this analysis. During the debriefings conducted as part of this study, listers indicated that they were sometimes hesitant to remove units from the input listing (see Appendix F). Thus it seems that lister training may encourage failure-to-delete confirmation bias, but I have not seen any statements by lister trainers that would encourage failure-to-add confirmation bias.

It is true that it does not take much interviewer time to identify nonresidential or nonexistent units after they are selected. I am concerned, however, with multiple-probability overcoverage and the ways that the two types of overcoverage can blur together. For example, consider a single family home that has been incorrectly listed as two units, and the second unit has been selected. The appropriate behavior for the interviewer is to disposition the case as nonexistent (out-of-scope

overcoverage). However, it is conceivable that some interviewers would think that since the two units had been combined into a single-family that the selected case now points to that single unit. If the interviewer proceeds to interview, then that single-family unit has two chances of selection and is overcovered in the sense of multiple-probability overcoverage defined above. More research is needed to understand how interviewers really do handle these sorts of situations. Yet the viewpoint by listers and managers that overcoverage is not a problem likely explains some of the tendency towards failure-to-delete bias I find in this study.

These three chapters together paint a picture of the difficulties listers face in creating housing unit frames. Listing is a complex task and no method seems clearly superior. When the input list is very good, dependent listers can create high quality frames, though their improvements may not be worth the cost of time and travel. When the input list is not good, dependent listers have difficulties in the same areas that traditional listers do.

Clearly frame quality deserves more attention in the literature on household surveys. However, the most important issue for survey data quality is the contribution of undercoverage and overcoverage to bias in survey estimates. If the errors traditional and dependent listers make are not related to variables of interest to survey researchers, then we need not devote resources to quantifying, understanding and reducing them. The next chapter uses NSFG data to estimate the impact of housing unit undercoverage on estimates in that survey.

Chapter 4

Bias Due to Undercoverage in Housing Unit Frames

4.1 Introduction

The previous chapters have made important contributions to our understanding of errors in listed housing unit frames. However, questions still remain about the mechanisms of error in housing unit listing. Future studies of lister error should use larger samples with more segments and listers to test hypotheses derived from alternative theories. But before pursuing additional research into the mechanisms of error in housing unit listing, it is wise to check first if these errors impact survey estimates.

Coverage bias is a risk whenever a frame contains undercoverage or multiple-probability overcoverage. Undercoverage occurs when listers fail to include one or more housing units that lie inside the segment on their frame. Multiple-probability overcoverage occurs when listers include units from outside the segment. In a survey of residents of households, such as the Current Population Survey, if the people who live in undercovered or overcovered housing units are different than those who live in the correctly covered units, survey data can be biased. In a study of housing conditions, such as the Residential Energy Consumption Survey, it would be enough that the undercovered or overcovered units themselves were different, regardless of the characteristics of the inhabitants.

The NSFG dataset used in Chapters 2 and 3 supports estimation of bias due to undercoverage in housing unit frames. Using the completed interviews in the segments selected for relisting, I can estimate the bias in important variables if the cases vulnerable to undercoverage in the second and third listings were not covered. The dataset does not permit estimates of bias due to multiple-probability overcoverage because it does not contain indicators of which units, if any, are outside of the segment or otherwise had more than one chance to be selected.

There are few estimates of coverage bias in the literature on area probability surveys. Coverage bias is difficult to estimate: one must know not only which units are undercovered and overcovered, but must also have data about the undercovered cases. There is indirect evidence that implausible estimates of relative school enrollment and labor force participation rates from the Current Population Survey by race and sex (Clogg et al., 1989) and victimization rates from the National Crime Study (Martin, 1981; Cook, 1985) are due to coverage bias. However, these studies do not look specifically at bias due to housing unit listing. To my knowledge this chapter offers the first estimates of bias due to errors in housing unit frames.

4.2 Data

This chapter again uses the multiple listing dataset discussed in Chapters 2 and 3. This dataset contains three listings of a sample of 49 segments from quarter 12 of Cycle 7 of the National Survey of Family Growth (NSFG). See Appendix C for a discussion of how the three frames were matched together. Using the response data

collected by NSFG, I can estimate undercoverage bias in means of NSFG variables.

4.2.1 Survey Background

NSFG data produce important estimates on fertility and family formation behavior that are used by demographers to model the size and composition of the population in the future. These data are also a valuable resource to researchers studying marriage, divorce, fertility, adoption, sexually transmitted infections, and more. The survey is conducted for the Vital Statistics Division at the National Center for Health Statistics (NCHS).

Data collection is underway for Cycle 7 of NSFG. Each quarter, interviewers receive an assignment of selected cases in their segments and approach each household to participate in the survey. They first attempt to screen the household to determine if any residents are eligible (15-44 years old), then select an eligible member and continue to the interview. After eight weeks, the remaining nonrespondent cases are subsampled to concentrate the interviewer efforts on fewer cases and thus improve the representativeness of the respondent sample. More details on all stages of the training, sample selection, contact, screening, subsampling and interview procedures are available in Groves et al. (2009).

From the segments in the listing study, 1,994 cases were selected for the NSFG study.¹ 24 of these cases were found to be improper listings, nonresidential or outside of segment, leaving 1,970 cases approached for the screener. 81.3% of the

¹Although 49 segments were selected and listed for this study, the counts given here and in the rest of the paper refer only cases in the 46 segments where matching is complete. See Appendix C.

proper listings completed the screener. 56.1% of the screened households contained one or more eligible persons, and 75.5% of the selected respondents completed the interview (see Table 4.1).

Table 4.1: Sample Performance in Quarter 12 of NSFG, Selected and Matched Segments

| | Selected | Good HUs | Screened | Eligible | Interviewed |
|-------------------------|----------|----------|----------|----------|-------------|
| Cases | 1994 | 1970 | 1602 | 898 | 678 |
| Pct. of Selected | | 98.8% | 80.3% | 45.0% | 34.0% |
| Pct. of Previous Column | | 98.8% | 81.3% | 56.1% | 75.5% |

Percent unweighted for selection probabilities

Just over half of the completed interviews (56%) were with female respondents. Due to the topic of the survey, the questionnaires are quite different for male and female respondents. See Appendix H for outlines of the questionnaires.

4.2.2 Variable Selection

For the bias analyses I chose variables that are in both the male and female questionnaires to keep the sample size large. The data from Cycle 7 will not be released or reported on until 2011. However, the Vital Statistics Division has allowed me to access response data in advance of their release. These data are confidential and sensitive. As part of my agreement with NCHS, I cannot present any data that could be used to make forecasts about the Cycle 7 NSFG data. All of the variables for which I estimate coverage bias below are disguised. I use uninformative names and center the data. The continuous variables are also standardized (divided by the standard deviation of the variable). These manipulations ensure that readers

of this dissertation cannot glean advance information about the Cycle 7 data still being collected. Table 4.2 gives some information about the variables selected for the bias analyses.

Table 4.2: Variables Used in Bias Analysis

| Variable | Topic | Type | n |
|----------|-------------|------------|-----|
| M1 | Health | Proportion | 678 |
| M20 | Health | Proportion | 678 |
| M2 | Sexual | Proportion | 665 |
| M7 | Sexual | Proportion | 677 |
| M19 | Sexual | Proportion | 667 |
| M22 | Sexual | Proportion | 668 |
| M31 | Sexual | Proportion | 658 |
| M28 | Demographic | Proportion | 678 |
| M17 | Demographic | Proportion | 678 |
| M27 | Financial | Proportion | 656 |
| M4 | Demographic | Count | 678 |
| M6 | Demographic | Count | 678 |
| M24 | Demographic | Count | 678 |
| M15 | Financial | Continuous | 678 |
| M32 | Financial | Continuous | 669 |

4.2.3 Undercoverage in NSFG listing

Despite these limitations on the identification of the variables in my dataset, the multiple listing conducted in conjunction with NSFG offers a unique resource for the estimation of undercoverage bias in housing unit frames. The second and third listings of the segments contained a good deal of undercoverage and response data is available for many of the undercovered cases. The first row of Table 4.3 gives the number of cases in my segments at each stage of the interview process. The next two rows show the percent of cases at each stage that were listed by the second and

third listers. 88.6% of the selected cases were on the traditionally listed frame, and slightly more, 93.0%, were covered by the dependent-listed frame. The second column refers to those cases which were found by the interviewer to be appropriate listings—residential and inside the segment. Moving from left to right in the table progresses through the stages of the survey: screening, eligibility and interview. In the last column, the traditional listers undercovered almost ten percent of the cases that completed the interview, and the dependent listers just over five percent. The listing rates in both frames increase across the columns, meaning that the cases that continued through later stages of the interviewing process were easier to list than those that did not progress, an interesting finding I will return to in the discussion.

The last two rows of Table 4.3 separate the cases on the dependent frame into those that were not manipulated and those that were deleted from the input list. 338 of the 1,994 selected cases were deleted from the input listing. (None of the selected cases correspond to housing units added to the input listing.) As found in Chapter 3, dependent listers were much less likely to list the cases that were deleted than those that were not deleted. This finding is reinforced by the coverage rates in the last two rows of Table 4.3.

4.3 Methods

These coverage rates indicate that there is quite a bit of undercoverage in the second and third listings. However, just as nonresponse rates are not neces-

Table 4.3: Percent of Cases Listed by Second and Third Listings, by Survey Stage

| | Selected | Good HUs | Screened | Eligible | Interviewed |
|---------------------|----------|----------|----------|----------|-------------|
| Total | 1994 | 1970 | 1602 | 898 | 678 |
| Traditional Listing | 88.6% | 89.2% | 90.1% | 90.2% | 90.6% |
| Dependent Listing | 93.0% | 93.5% | 93.6% | 94.1% | 94.8% |
| Unmanipulated | 96.9% | 97.3% | 97.3% | 97.6% | 97.5% |
| Deleted | 74.0% | 74.8% | 75.0% | 75.9% | 81.1% |

Refers to only 46 segments where the matching is complete.

sarily good predictors of nonresponse bias (Groves, 2006; Groves and Peytcheva, 2008), coverage rates are unlikely to be good predictors of coverage error. This paper uses two methods to estimate hypothetical coverage bias due to undercoverage of housing units by the traditional and dependent listings in NSFG variables. Each approach has strengths and weaknesses, and together they can provide a sense of the risk of bias due to undercoverage.

4.3.1 Direct Approach to Bias Estimation

The first method, which I call the direct method, is simply the difference between the mean calculated on the covered cases and the mean on all the cases. Let \bar{Y} be the estimate of the mean or proportion of a given variable on the 678 responding cases in my 46 segments.² This mean will be 0 for all variables, due to the centering as discussed above. Let \bar{Y}_{trad} be the same mean calculated on only those selected and completed cases which were also included in traditional (second) listing. There are 614 such cases. Let \bar{Y}_{dep} be the mean calculated on only those cases

²Due to some missing data among the completed cases, as shown in Table 4.2, the exact number of cases may be smaller for some variables.

covered by the dependent listing, 643 cases. Then the direct estimates of bias are:

$$\begin{aligned}\text{bias}_{\text{trad}}^{\text{direct}}(\bar{Y}) &= \bar{Y}_{\text{trad}} - \bar{Y} \\ &= \bar{Y}_{\text{trad}} \\ \text{bias}_{\text{dep}}^{\text{direct}}(\bar{Y}) &= \bar{Y}_{\text{dep}} - \bar{Y} \\ &= \bar{Y}_{\text{dep}}\end{aligned}$$

(because $\bar{Y} = 0$ for all variables Y).

While this estimation method has intuitive appeal, the bias estimates it produces for the dependent listing method necessarily reflect the manipulation of the input list. Dependent listers were much less likely to include the deleted units, and the manipulation may affect the bias estimates if the deleted cases are different than the unmanipulated on the survey variables. Using the direct method, it is not possible to remove the effect of the manipulation of the input list from the bias estimate.

4.3.2 Indirect Approach to Bias Estimation

The second approach to estimating bias, which I call the indirect approach, takes advantage of the listing propensity models used in previous chapters. The indirect estimate of bias is:

$$\text{bias}^{\text{indirect}}(\bar{Y}) \approx \frac{\sigma_{Y\lambda}}{\bar{\lambda}} \tag{4.1}$$

where λ_i is the coverage propensity of housing unit i , $\bar{\lambda}$ is the average propensity among the covered and undercovered cases, and $\sigma_{Y\lambda}$ is the covariance between the propensity and the variable of interest (adapted from Bethlehem, 2002). Bias will be large when the listing propensity is highly correlated with the survey variable, or the average listing propensity is low. If the listing propensity model is correctly specified, the indirect method should give the same estimates as the direct method.

The indirect method can produce estimates of bias due to undercoverage by traditional and dependent listing, just as the direct method does. I fit two logistic models to estimate the propensity of each completed case to be covered by each of the two listing methods, λ_{trad} and λ_{dep} . Plugging each of the propensities into the numerator and denominator of equation 4.1 leads to indirect estimates of bias due to the traditional and dependent listing methods.

A strength of the indirect method is that I can use the listing propensity models to simulate bias under different conditions. Specifically the indirect method allows for the estimation of bias in dependent listing if the input list had not been manipulated. The weakness of this method is that it is quite model dependent. Different models of listing propensity may lead to different estimates of bias.

4.3.2.1 Listing Propensity Models

The models used to predict listing propensity for the indirect bias estimation differ in several ways from the models presented in earlier chapters. Here the goal of the models is not to study the characteristics that make listing more or less likely,

but to estimate the probability that each unit will be listed, given its values on the relevant characteristics. I fit separate models for traditional and dependent listing.

The dependent variable in the models is again a binary indicator of whether the lister included the housing unit (1) or did not (0). Both models use logistic regression to ensure that all predicted propensities are within the range (0, 1). Only the 678 completed cases are included in the models. This reduction in sample size necessitates rethinking the structure of the models as there are fewer cases within each segment and lister.³

The models in this chapter do not contain any random or fixed effects for segments or listers. In the earlier versions of the models, fixed effects for listers removed the idiosyncratic effects of the particular lister used in each segment to isolate the effect of lister characteristics in the regression coefficients. However, coefficient estimation is not the goal of the models in this chapter and thus this precaution is not necessary. The simpler models used here do account for the clustering of housing units into segments in calculating standard errors, but without explicitly modeling the segments with random effects. The benefit to running unclustered models without fixed or random effects is that more post-estimation and diagnostic tools are available.

The estimated odds ratios for the explanatory variables and fit statistics are given in Table 4.4 for the traditional and dependent listing propensity models, though I do not interpret them here.⁴ I tested several versions of the models with

³Two of the 46 segments contain zero completed interviews. Only one respondent was selected in these segments, due to very low eligibility rates, and s/he did not complete the interview.

⁴For interpretation of listing propensity in the two listing methods, see Chapters 2 and 3.

different covariates until I found those that had high AUC values (area under the ROC curve), indicating a strong ability to discriminate between the listed and unlisted cases, and low dbeta statistics, indicating the model fits all the data points rather well (Hosmer and Lemeshow, 2000; Long and Freese, 2005).

Figures 4.1(a) and 4.1(b) show the kernel densities of the predicted propensities from the traditional and dependent listing models. The distributions have very similar shapes and ranges. The majority of the housing units have very high propensities, near 100%, due to the high coverage rates in each method (89.2% in the traditional listing and 93.5% in the dependent listing).

The listing propensities from the dependent listing model incorporate flags for the effect of the cases deleted from the input listing. 111 of the 678 completed cases were deleted from the input list. This manipulation has a strong effect on listing propensity, the failure-to-add effect, as shown in the odds ratio on the last variable in Table 4.4. These deleted cases are disproportionately missing from the frames created by the dependent listers.⁵ The low listing propensities of these cases could affect the estimates of bias from the direct method. If these cases are different than the unmanipulated cases on any variables chosen for the bias analysis, then the estimates of bias will be affected by the manipulation.

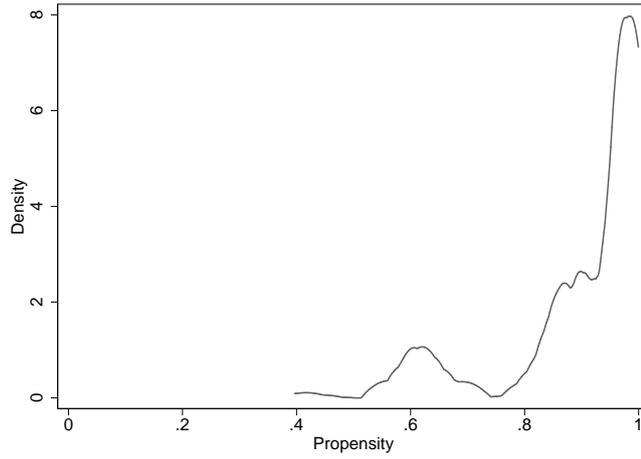
The indirect method presents an opportunity to remove the effect of the manipulation of the input list from the bias calculation. Recoding the relevant variable

⁵Note that the other type of input list manipulation, the addition of units, does not affect the propensities of the selected cases because none of these cases was on the first listing and thus none was eligible for selection for NSFG screening and interviewing. For this reason, the only manipulation I need to worry about when computing propensities for the selected cases is the deletion of units listed by the first lister.

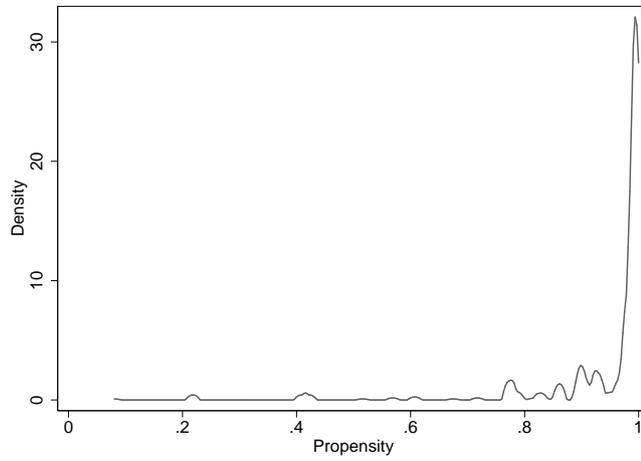
Table 4.4: Listing Propensity Model

| | Traditional Listing | | Dependent Listing | |
|---|---------------------|----------|-------------------|----------|
| | Odds Ratio | <i>z</i> | Odds Ratio | <i>z</i> |
| Multi-Unit | 0.130** | (-2.93) | 1.910 | (0.77) |
| Map, invisible boundary | 0.040* | (-2.53) | 0.893 | (-0.13) |
| Map, simple shape | 1.624 | (0.42) | | |
| Segment has external water boundary | | | 1.020 | (0.02) |
| Pct. HUs rural | 0.599 | (-0.37) | 0.010* | (-2.09) |
| Pct. HHs with income <= 50,000 | 0.564 | (-0.45) | 10.009 | (1.06) |
| Pct. Pop. Afr.-Amer. | 1.802 | (0.32) | 0.509 | (-0.28) |
| Pct. Pop Spanish language | 0.035 | (-0.63) | 0.036 | (-0.39) |
| Gated communities in segment | 19.566* | (2.15) | 4.598 | (0.55) |
| Lister drove herself while listing | 0.549 | (-0.60) | 5.567 | (1.08) |
| Lister feels unsafe | 0.082* | (-2.28) | 0.180 | (-0.71) |
| Lister and segment language match | 5.423 | (1.64) | 13.290 | (1.65) |
| Years of interviewer experience | 1.095 | (0.56) | 1.403 | (1.41) |
| Lister African-American | 1.003 | (0.00) | 0.026 | (-1.36) |
| Lister speaks Spanish | 0.061* | (-2.32) | 0.799 | (-0.10) |
| Lister reports Spanish speakers | 2.199 | (0.56) | 3.933 | (0.56) |
| Units in segment predominately trailers | 0.017** | (-2.94) | 0.045*** | (-3.37) |
| HU deleted from input listing to L3 | | | 0.082** | (-3.20) |
| Constant | 269.261*** | (3.70) | 0.399 | |
| AUC | 0.853 | | 0.927 | |
| Pseudo- R^2 | 0.252 | | 0.374 | |
| Observations | 678 | | 678 | |

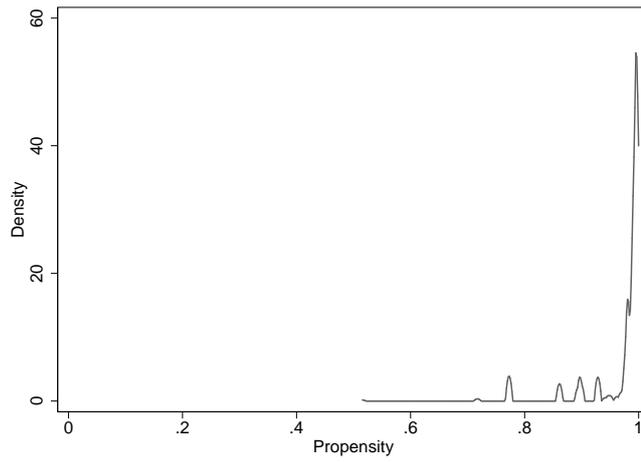
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$



(a) Traditional Listing



(b) Dependent Listing, With Manipulations



(c) Dependent Listing, Without Manipulations

Figure 4.1: Distribution of Predicted Listing Propensities, by Listing Method

on the manipulated units to reflect no manipulation and re-estimating the dependent listing propensity of these cases produces a third predicted listing propensity for every case, $\lambda_{\text{dep, no manip}}$.⁶ Using this propensity, the indirect method of bias estimation can estimate the bias in the survey variables had the manipulation of the input list not been done.

Figure 4.1(c) shows the distribution of the predicted listing propensities for this counterfactual situation, had the input to the dependent listing not been manipulated. The predicted propensities are all larger than 50%, higher than the predicted propensities from the other models.

The outcome of these models are three predicted listing propensities for each of the interviewed cases. The indirect method of bias estimation can approximate bias under each of these listing scenarios.

4.3.3 Variance of Bias Estimates

These bias estimates are useful only when accompanied by a measure of their precision. The Jackknife procedure can put a confidence interval around each of the estimates (Wolter, 2007, Chapter 4). The Jackknife is useful in estimating variances in many complex designs involving clustering and weighting (Lohr, 1999, pp. 304–306).

The form of the Jackknife used here, the delete-1 procedure, involves dividing the sample into R groups, and repeating the relevant estimation procedures R times, dropping each of the R groups one at a time. The R groups were the 13 PSUs

⁶ $\lambda_{\text{dep, no manip}} = \lambda_{\text{dep}}$ for all unmanipulated cases.

selected into the listing project. (See Chapter 2 for more details on the selection of PSUs and segments for this project.)

Consider the direct estimates of bias in the mean of variable Y due to under-coverage in the traditional listing, $\text{bias}_{\text{trad}}^{\text{direct}}(\bar{Y})$. Let $\text{bias}_{\text{trad}}^{\text{direct}}(\bar{Y}_{(r)})$ be the estimate of bias when the r^{th} PSU is dropped. This calculation involves re-estimating both the mean on all the cases in the undropped PSUs ($\bar{Y}_{(r)}$) as well as the mean on the cases covered by the second lister ($\bar{Y}_{\text{trad},(r)}$), and then taking the difference. This process yields R different direct estimates of the bias.

The variance of these R estimates around the bias estimate on all PSUs⁷ gives an estimate of the variance of the original estimate. The form of the Jackknife estimator of the variance of the estimate of bias in variable Y , due to traditional listing, is:

$$\hat{V}[\text{bias}_{\text{trad}}^{\text{direct}}(\bar{Y})] = \frac{R-1}{R} \sum_{r=1}^R [\text{bias}_{\text{trad}}^{\text{direct}}(\bar{Y}_{(r)}) - \text{bias}_{\text{trad}}^{\text{direct}}(\bar{Y})]^2 \quad (4.2)$$

The square root of this variance is the standard error of the bias estimate.

The Jackknife variance calculation for the indirect estimate is very similar. For each survey variable and each of the three listing propensities (λ_{trad} , λ_{dep} , and $\lambda_{\text{dep, no manip}}$), drop each PSU in turn, recalculating both the numerator and the denominator of the bias estimator. This procedure leads to R estimates. The sum of the squared differences between each estimate and the full sample estimate, times

⁷As originally developed, the squared deviations in the Jackknife formula were taken around the average across the R estimates, not around the full sample mean. However, it is now common to use the full sample mean in this calculation (Wolter, 2007, p. 153).

an adjustment factor, estimates the variance in the full sample estimate.

$$\hat{V}[\text{bias}_{\text{dep}}^{\text{indirect}}(\bar{Y})] = \frac{R-1}{R} \sum_{r=1}^R [\text{bias}_{\text{dep}}^{\text{indirect}}(\bar{Y}_{(r)}) - \text{bias}_{\text{dep}}^{\text{indirect}}(\bar{Y})]^2 \quad (4.3)$$

These estimated variances and standard errors reflect the impact of the experimental design only, not any sampling variance due to the selection of segments or cases.⁸

4.4 Results of Bias Analyses

The two methods of bias estimation produce five estimates of bias for each variable: one for each of the two listing methods, from both the direct and indirect methods, plus one additional estimate from the indirect method using the propensity that removes the effect of the listing manipulations. These estimates are presented below in tabular and graphical form to permit comparisons among methods and variables.

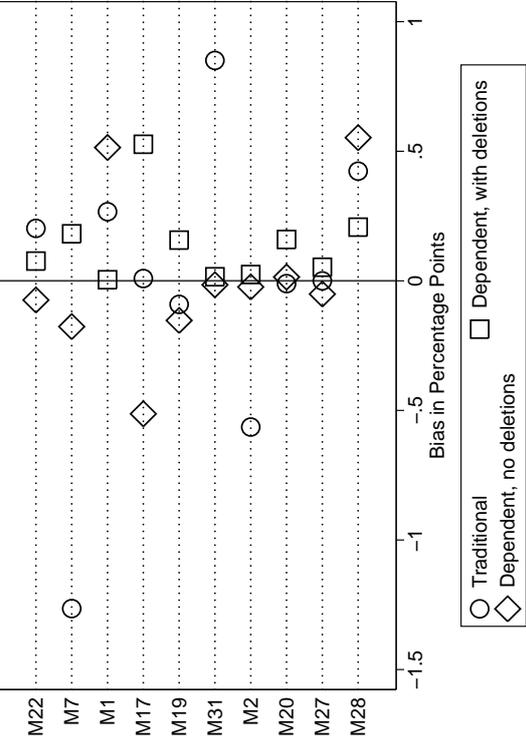
Direct estimates of bias in NSFG variables due to the traditional and dependent listings of each segment are shown on the left side of Figure 4.2. The top left panel of Figure 4.2 shows bias in proportions and the bottom panel shows bias in continuous variables. These are estimates not of the bias in official Cycle 7 data, but of the bias that would be introduced had these alternative frames been used instead of the initial listing.

⁸The variance estimation procedure also does not account for the uncertainty in the predicted listing propensities.

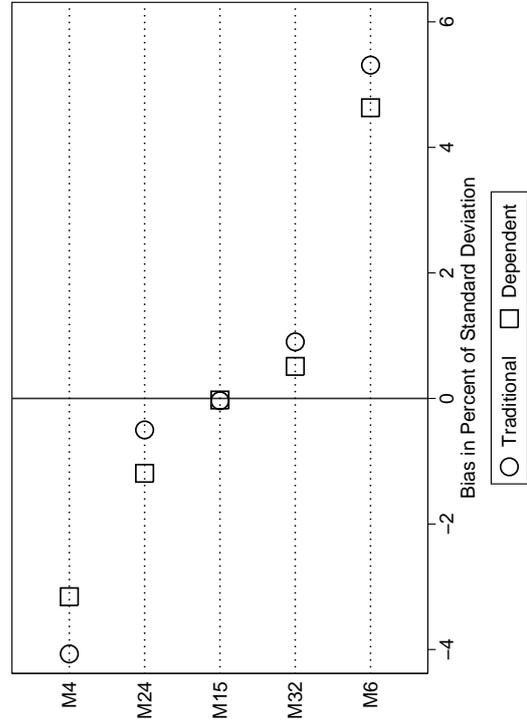
Each row in the top left panel corresponds to one of the ten proportions selected for bias estimation. Along the horizontal axis is the bias scale in percentage points, with 0 representing no bias. For each variable there are two points. The circle shows the bias in the proportion if the completed cases in housing units that were not listed by the traditional lister were dropped from the calculation. The square shows the bias if the cases not covered by the dependent lister were dropped. The rows are sorted by the size of the bias due to traditional listing.

In the first row of Figure 4.2(a), undercoverage in the traditional listing would have led to bias of about -0.35 percentage points in proportion M22. If the proportion calculated on the full sample for this variable was 50%, then the calculation on only the cases covered by the traditional listers would be 49.65%. The bias due to undercoverage in dependent listing is smaller on this variable: calculating the proportion on only the cases listed by the dependent lister would yield an estimate with a very small positive bias (0.05 percentage points), indicated by the square just to the right of the reference line in the first row. (While these are small effects, they could be large in relation to the mean. A change of 0.35 percentage points if the mean of this variable were 2% would be a large relative effect. Because the variables are centered, we cannot see the relative effect here.) The bias due to dependent listing in the second row is the largest in absolute value, just under one percentage point in the negative direction. No proportion would be biased by more than one percentage point in either direction.

The lower left panel of Figure 4.2 shows the bias estimates for the continuous variables. The horizontal axis in this graph is in percent of standard deviation

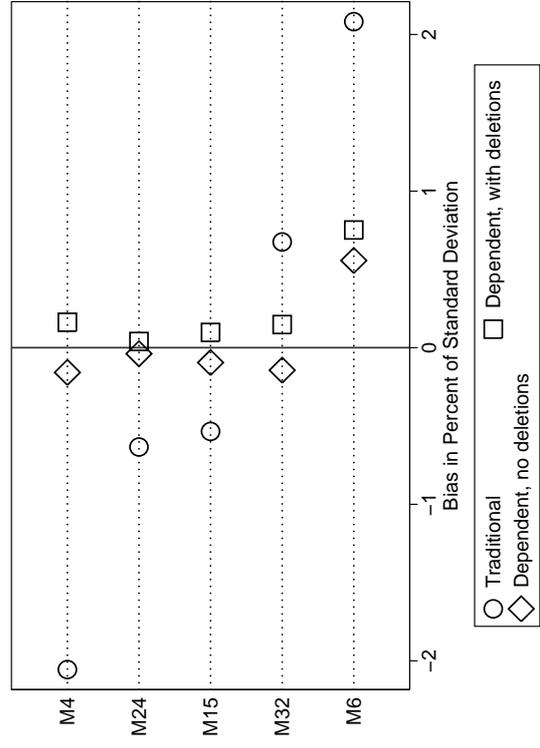


(a) Direct Estimates of Bias, Proportions



(c) Direct Estimates of Bias, Continuous

(b) Indirect Estimates of Bias, Proportions



(d) Indirect Estimates of Bias, Continuous

Figure 4.2: Estimates of Bias in Survey Variables Due to Undercoverage, by Listing Method and Estimation Method

units. For example, in the first row of Figure 4.2(c), the bias in variable M4 due to undercoverage in traditional listing would be four percent of one standard deviation in the negative direction. The bias in the same variable due to undercoverage in dependent listing would be slightly smaller in absolute value but also in the negative direction. Listers using both methods tended to undercover cases where the value on this variable was greater than the full sample mean. No estimate would move more than six percent of one standard deviation in either direction due to undercoverage.

There is no bias due to either method for M15, a financial variable. M32 is also financial and shows little bias. Results in earlier chapters suggested a correlation between the percent of households in a segment living below the poverty line and the coverage rate in the segment. We might then expect to find bias in these financial variables due to undercoverage, but we do not.

The two graphs on the right side of Figure 4.2 display the indirect bias estimates for the same variables (in the same order). The horizontal axes in each panel on the right are in the same units as the corresponding panels on the left. Figures 4.2(b) and 4.2(d) each show three estimates corresponding to the bias due to undercoverage in traditional listing, dependent listing with the manipulations of the input list, and dependent listing without the manipulations in the input list.

In the first row of Figure 4.2(b), traditional and dependent listing would each lead to slight positive bias in variable in M22, and dependent listing without the manipulations would lead to a small negative bias. In the second row, traditional listing leads to negative bias of 1.2 percentage points. The two dependent methods

would lead to biases that are approximately equal in magnitude but of different signs. Most of the bias estimates in the upper right panel are close to zero, though the overall range of estimates is slightly larger than the direct estimates for these variables, given in Figure 4.2(a).

The graph in the lower panel shows the three bias estimates for each of the five continuous variables. For each variable, the bias associated with both kinds of dependent listing is smaller (in absolute value) than that from traditional listing. The range here is narrower than the direct estimates of bias for the same variables in the lower left panel. The indirect estimates of variance for these variables are in general smaller than the direct estimates. Here we do not see smaller bias effects for the two continuous financial variables, M15 and M32.

Table 4.5 presents the estimates behind these graphs, as well as confidence intervals for each estimate. When assessing the significance of as many estimates as are given in Table 4.5 (30 direct + 45 indirect), it is wise to use a higher probability cutoff than 95%, which would find on average three or four significant values in any set of 75 estimates. For this reason, the table shows 99% confidence intervals rather than the standard 95%. None of the bias estimates are significantly different than zero at the one percent level.

4.5 Discussion & Conclusion

The bias estimates developed and presented in this chapter reflect the hypothetical risk of bias in several variables collected by NSFG. These estimates are

Table 4.5: Bias Methods with 99% Confidence Intervals for All Variables and Methods

| Variable | Direct Bias Estimates | | | | Indirect Bias Estimates | | | | | |
|--|-----------------------|------------------|--------|------------------|-------------------------|-----------------|--------------|-------------------|---------------|------------------|
| | Trad | CI | Dep | CI | Trad | CI | Dep w/ manip | CI | Dep, no manip | CI |
| Proportions, units: percentage points | | | | | | | | | | |
| M22 | -0.342 | (-1.068, 0.384) | 0.050 | (-0.376, 0.475) | 0.202 | (-0.865, 1.27) | -0.074 | (-0.574, 0.426) | 0.076 | (-1.495, 1.647) |
| M7 | -0.338 | (-1.335, 0.659) | -0.968 | (-1.948, 0.011) | -1.265 | (-2.871, 0.341) | -0.177 | (-3.696, 3.343) | 0.181 | (-3.319, 3.681) |
| M1 | -0.312 | (-1.092, 0.468) | 0.240 | (-0.366, 0.847) | 0.267 | (-0.738, 1.272) | 0.514 | (-0.333, 1.361) | 0.004 | (-0.346, 0.354) |
| M17 | -0.309 | (-2.523, 1.905) | 0.214 | (-2.39, 2.818) | 0.009 | (-6.288, 6.306) | -0.513 | (-19.077, 18.051) | 0.527 | (-12.94, 13.994) |
| M19 | -0.135 | (-1.045, 0.775) | -0.070 | (-1.227, 1.087) | -0.091 | (-0.656, 0.474) | -0.153 | (-1.558, 1.252) | 0.157 | (-3.12, 3.434) |
| M31 | -0.030 | (-0.996, 0.936) | 0.392 | (-0.838, 1.621) | 0.850 | (-1.082, 2.783) | -0.015 | (-1.788, 1.758) | 0.016 | (-0.416, 0.448) |
| M2 | 0.033 | (-0.971, 1.038) | 0.528 | (-0.575, 1.631) | -0.564 | (-1.741, 0.613) | -0.024 | (-2.4, 2.353) | 0.024 | (-0.501, 0.55) |
| M20 | 0.128 | (-0.033, 0.289) | -0.080 | (-0.556, 0.396) | -0.011 | (-0.306, 0.284) | 0.015 | (-0.174, 0.204) | 0.160 | (-0.162, 0.481) |
| M27 | 0.379 | (-0.852, 1.611) | 0.166 | (-1.631, 1.962) | 0.000 | (-0.685, 0.684) | -0.051 | (-2.132, 2.029) | 0.052 | (-1.426, 1.531) |
| M28 | 0.460 | (-0.981, 1.901) | 0.602 | (-0.888, 2.091) | 0.423 | (-0.966, 1.813) | 0.552 | (-0.821, 1.926) | 0.207 | (-0.68, 1.094) |
| Continuous, units: percent of standard deviation | | | | | | | | | | |
| M4 | -4.070 | (-11.774, 3.634) | -3.158 | (-12.107, 5.79) | -2.055 | (-5.668, 1.557) | -0.158 | (-11.882, 11.566) | 0.162 | (-2.828, 3.153) |
| M24 | -0.501 | (-2.934, 1.932) | -1.193 | (-5.003, 2.618) | -0.634 | (-3.193, 1.925) | -0.039 | (-1.852, 1.775) | 0.040 | (-1.215, 1.294) |
| M15 | -0.036 | (-0.12, 0.049) | -0.025 | (-0.12, 0.07) | -0.535 | (-4.096, 3.026) | -0.095 | (-6.386, 6.196) | 0.098 | (-2.807, 3.002) |
| M32 | 0.902 | (-0.2, 2.004) | 0.511 | (-0.881, 1.903) | 0.675 | (-1.1, 2.45) | -0.144 | (-1.525, 1.237) | 0.148 | (-2.746, 3.041) |
| M6 | 5.310 | (-2.657, 13.277) | 4.631 | (-3.354, 12.616) | 2.082 | (-2.599, 6.762) | 0.556 | (-2.526, 3.639) | 0.752 | (-2.663, 4.166) |

No bias estimates are significant at the 1% level

of course survey-specific and variable-specific (and statistic-specific: I have looked only at means, not other statistics that could be calculated from the same variables such as totals, medians, etc.). None of the 75 estimates of bias are significant at the one percent level. A larger sample of cases or segments would perhaps detect significant differences in these estimates, but the current study is too small and clustered to do so. To my knowledge, these are the first estimates of coverage bias due specifically to housing unit coverage in the survey literature.

Coverage error is rarely studied because of the difficulties involved in both identifying the undercovered cases and collecting data about them. The unique design of this listing study made these estimates possible. The multiple listing approach revealed the cases at risk of undercoverage in each listing method, while the survey collected data on these cases.

Despite the unique contribution of this study, several shortcomings should be noted. This paper can calculate bias due only to undercoverage by the second and third listers. Although the first listing received more quality checks than the others and is likely of higher quality, it too may contain some undercoverage. There were quite a few housing units listed by the second and third listers yet missed by the first. Some of these cases may be good listings, but they were not eligible for selection and no data exists for them. A richer dataset would have data about these cases and support estimates of bias due to undercoverage in the first listing as well.

The adjusted dependent listing propensity (Figure 4.1(c)) represents a listing situation that is not much more realistic than the one which does reflect the manipulations (Figure 4.1(b)). With the manipulations modeled away, the input listing

already contains all of the cases to be covered. But the frames that are used as input to dependent listing do contain errors of inclusion and exclusion (O’Muircheartaigh et al., 2006, 2007; Montaquila et al., 2010). A more realistic approximation of dependent listing as it is actually used would involve modeling errors in the input list that mimic the errors these lists really contain. However, this approach is outside the scope of this chapter.

The bias and standard error calculations in this chapter are unweighted for the PSU, segment, housing unit, respondent and subsampling selection probabilities. This is unfortunate and may very well affect my estimates of bias and variance. For example, Chapter 2 showed that housing units in rural areas are less likely to be covered by traditional listers. The rural segments in this study have low probabilities of selection, due to their low housing unit counts, and thus cases selected from these segments should have larger weights than the cases in other segments, where coverage rates are also higher. If the coverage bias in rural segments has a different pattern than in the non-rural segments, including the selection weights will change the overall estimates of bias in favor of the pattern in the heavily-weighted rural segments. I will rerun all bias and variance calculations later this year when the weights are available from the Survey Research Center, the NSFG data collection contractor. At that time I will also incorporate adjustments to the weights for both the selection of quarter 12 from the 16 quarters of Cycle 7 data collection and for the selection of segments for my study.

The bias estimates found in this chapter are rather small in magnitude. Looking back at Equation 4.1, for a given set of listing propensities, the size of the bias is

due to the relationship between the propensities and the survey variables. Many of the variables collected in the NSFG survey are related to marriage, reproduction, and sexual histories (see Appendix H). The theories and models tested in Chapters 2 and 3 do not suggest strong correlations between these variables and listing propensities, which should lead to low bias. For the bias analysis, I chose some variables on fertility and sexuality topics and others on more general topics, to speak to bias in other surveys as well. Bias estimates for all the items explored were small. However, bias that is small in absolute terms may be large in relative terms.

One finding in this chapter and also mentioned briefly in Chapter 2 is the relationship between listing propensity and response propensity. The coverage rates on all cases are lower than those on the screened cases, which are lower than coverage rates on the interviewed cases (see Table 4.3). The cases that progress through the survey, those that are most cooperative, are also more likely to be covered. This interesting relationship warrants additional research.

While undercoverage does exist in housing unit frames, it appears in this study to be only weakly related to variables on the survey questionnaire. The findings of low bias due to undercoverage above do not apply directly to the official NSFG frame, the first of the three listings, from which the interviewed cases were selected: data on cases undercovered by that frame were not available. However because that official frame received more scrutiny and quality checks than the other two frames, we can presume that undercoverage, and bias due to undercoverage, is lower on the official frame than on the two frames examined closely. Overall the findings here of small bias due to undercoverage are good news for NSFG and a

positive sign for all household surveys that use listing to create household frames.

Chapter 5

Conclusions

This dissertation has used two datasets to explore errors in housing unit frames and how these errors can lead to bias in survey data. I find that even well-trained and experienced listers do produce different frames, making errors of undercoverage and overcoverage. I show for the first time that experienced listers tend not to fix the errors in the listing they are given. I call this phenomenon confirmation bias, after a similar finding in research on dependent coding and interviewing. However, none of the errors of overcoverage and undercoverage in this study lead to substantial absolute bias in means derived from survey estimates. These findings are good news for surveys which depend on listing to create sampling frames, but also suggest directions for future research.

Chapter 1 used a repeated listing dataset from the Census Bureau to estimate the degree of disagreement between two experienced listers using the same methodology. The two listers do produce different frames which would lead to different samples and different final datasets. The overall agreement rate was only 79%, and this varied quite a bit among the blocks. While I could not separate undercoverage by one lister from overcoverage by another with this dataset, the agreement rate indicates that listers do make errors. These findings motivate the need for additional work to uncover the mechanisms of lister error.

Unfortunately analyses of these data were constrained by a lack of information about interviewers. In the two listings of each segment, most of the characteristics that varied were those due to listers, such as experience, training, etc., yet the dataset did not permit analysis of these characteristics. I was also not able to manipulate the listing process.

To address these shortcomings, I worked with the Survey Research Center at the University of Michigan to collect data specifically for my dissertation. Experienced listers from the National Survey of Family Growth conducted three listings of a nationally-representative sample of 49 segments. The dataset contains lister observations of the segments, lister demographics and background information as reported in the interviewer questionnaire, as well as response data for a sample of housing units. I experimentally manipulated the listing task to test hypotheses about lister error.

The analyses of both datasets rest on the quality of the matching work, which is always imperfect. Undoubtedly another researcher would match the Census Bureau and NSFG datasets slightly differently. For this reason, Chapter 1 and Appendix C report the details of the procedures used in matching. I gave the listing task the care and attention it deserved as the foundation of all my analyses. Nevertheless, I was not able to complete the matching work in three very rural NSFG segments where few units had house numbers. These segments were dropped from most analyses in Chapters 2, 3 and 4.

The second chapter used the NSFG dataset to test hypotheses about the mechanisms of error in traditional listing. These hypotheses were motivated by an un-

derstanding of the listing task as a principal-agent problem in which monitoring is costly and the agent (lister) has more information than the principal (survey researchers). The overall coverage rate for the traditional listers was 89%. Breaking this overall rate down by housing unit and segment characteristics replicated findings of earlier work: multi-unit and vacant units were undercovered, as well as units in poor and rural segments. The hypotheses derived from the principal-agent model found limited support, suggesting that we should look to alternative theoretical approaches for future work on the mechanisms of error in traditional listing.

I suspect this limited support is due in part to the small size of the NSFG listing dataset. While large in terms of housing units, it contained 49 segments and only 11 listers within each method. Testing the principal-agent involved interacting lister attributes with segment characteristics and also with housing unit characteristics. The dataset was not powerful enough to detect significant contributions at these higher levels. I have some suggestions below on how future studies can avoid these problems.

The third chapter tested hypotheses about the mechanisms of error in dependent listing related to confirmation bias. Analyses revealed that listers do show a tendency to confirm the list that they are given and not to add missing units or remove inappropriate ones. Units in multi-unit buildings are particularly vulnerable. Results are quite strong in both bivariate and multivariate analyses.

These findings indicate that confirmation bias should be a concern in dependent listing. But the results lack external validity—the introduced errors are not the same as the errors listers are likely to encounter in the input lists actually used

in listing. Future research into confirmation bias should look more carefully at the kinds of errors that are typical in input lists and whether these are the types that listers tend not to correct.

The implications of more realistic confirmation bias for coverage bias in survey data should also be explored. The errors introduced here were random. Perhaps the errors in the commercial address databases are not random and are related to survey variables, raising the risk of coverage bias due to confirmation bias. For example, new construction is often missing from the commercial address lists as it takes a while for these units to be picked up by the postal service and then make their way into survey frames. The families in these units may be younger or less wealthy than those in the older units; undercovering them due to failure-to-add error could lead to bias.

The last substantive chapter looks at bias due to undercoverage in both traditional and dependent listings. Using two methods of bias calculations and 15 NSFG variables, I found only small bias would result if the alternative frames had been used.

Together these chapters break new ground in listing research and lead me to several suggestions for surveys which use listing. The findings on confirmation bias are the strongest results from my research. The quality of the frame produced via dependent listing is in part a function of the quality of the list provided to the listers. When using dependent listing, I recommend that the threshold for the size of the input list be set quite high, particularly in areas with many multi-unit buildings. For example, dependent listing could be used only when the size of the input list is

greater than or equal to the housing unit count from the most recent Census. This requirement would help reduce the chances of failure-to-add error.

Furthermore I suggest that lister training emphasize the point that input listings do contain errors of both omission and inclusion. Some training practices encourage failure-to-delete error by emphasizing a preference to err on the side of overcoverage. These instructions are based on a belief that overcoverage does not lead to bias and can be cleaned up at low cost during data collection. I have more concerns about overcoverage and believe we need to understand better how interviewers handle instances of multiple probability overcoverage in their assignments. Failure-to-delete error, however, is a larger threat to bias and not the intention of any lister training. Training should include a discussion of confirmation error to warn listers against it. Perhaps manipulations like those in Chapter 3 could be used to periodically check up on how carefully listers examine the input list for errors.

Another implication concerns the coverage of units in multi-unit buildings. That listers undercover and overcover these sorts of units is a robust finding in previous research and in my dissertation. For those surveys where full coverage is critical or where multi-units status is believed to correlate strongly with the survey variables, I propose a procedure similar to the missed housing unit procedure. When a selected case is in a multi-unit building, the interviewer could be asked to do additional work to determine the number of units in the building. Interviewers often gain access to buildings and speak to residents and are thus in a better position than listers to get an accurate unit count. If the number of units found by

the interviewer is greater than the number on the frame, appropriate adjustments, including selection of the new units, could be made. This procedure would increase coverage in buildings with multiple units.

Of course, the findings in this dissertation raise more questions than they answer and I have several ideas for ways to expand up on this research in the future. In retrospect, drawing a national sample of segments was not necessary. (In fact the final matched dataset was not nationally representative due to the absence of weights and the difficulties matching three rural segments.) If I were to redesign this study, I would focus on just three or four purposefully-chosen areas with several listing segments each. One would be very urban, with many multi-unit buildings and another should be very rural (more on rural listing below).

Most importantly, more listings should be done of each segment, by more listers. The repeated listings should use both methods but also repetitions within method as in the Census Bureau dataset. Such a design would permit stronger analyses of inter-lister variation in frame quality and improve the ability to separate the effects of lister characteristics from those of segments. In my study, only 11 listers participated in each method. The models in Chapters 2 and 3 did not contain enough variability at the lister level to test many interesting interactions of lister and segment characteristics, such as race of lister and race of segment residents. Future research should aim to do so. Of course, additional listings would complicate the matching task.

The technique of partnering with a survey already in the field worked out quite well and I would do so again. However, the dataset contained no responses

from cases that were undercovered by the official NSFG listing. In the future I would hope to gather data about cases undercovered by each listing.

Looking to the future of housing unit listing, I believe we will see a move towards the use of commercial address lists without in-field updating. NORC used this method of frame construction in its most recent National Frame (Harter et al., 2010). However, there are still many parts of the country, particularly rural areas, where the lists' coverage is quite low. I suspect that in five years or so, these rural areas will be where most listing work is carried out. Yet we know little about how to address the particular challenges posed by rural listing. In these areas, listers often drive because the distances are too great to walk. Housing units often do not have numbers, and streets may not have names. I recommend additional research into rural listing to identify the difficulties listers in these areas face and develop procedures to address them.

Finally I want to return to the larger research interests that prompted this dissertation. While I am quite interested in coverage research and have been working in this area for many years now, I am also fascinated by the role of interviewers in survey work. A dissertation on listing error nicely combined these two interests, exploring how interviewers affect coverage.

Interviewers can contribute to nearly every component of total survey error. We most often think of interviewers' behaviors leading to measurement error in the ways they administer questions, or to nonresponse error by the kinds of respondents they recruit, or fail to recruit. Interviewers can bias samples when allowed to select their own cases (Manheimer and Hyman, 1949; Eyerman et al., 2001). Inter-

viewers as coders can introduce bias and variance (Campanelli et al., 1997; Biemer and Lyberg, 2003). This dissertation has focused quite narrowly on the role of interviewers (listers) in housing unit coverage, but in my future career I hope to study the role of interviewers in other error sources as well. Interviewers are the agents of data collection. Most of the responsibility for the quality of the final data product rests with them. The field needs a better understanding of the influences on their behavior and their decisions, and how these affect survey data.

Chapter A

Appendix: Coding of Quality of Listing Maps

The maps provided to the listers by both the Census Bureau and the Survey Research Center (SRC) are often out of date and can even be misleading. Nearly all of the listers I spoke to expressed frustration with the quality of the listing maps. Most reported that they purchased commercial maps or downloaded online maps for all their segments. Both the SRC and Census maps are derived from TIGER (Topologically Integrated Geographic Encoding and Referencing) data released by the Geography Division of the Census Bureau. See Figure A.1 for an example SRC listing map.

The most important function of the maps is to let the lister know which block or blocks are selected for listing. Any errors in the maps, such as incorrect street names, unclear boundaries, missing streets, etc. can lead to errors in listing. (See Roberts (2010) for a striking example of how map errors led to substantial under-coverage in the decennial Address Canvassing effort.) For both my listing datasets, I compared the TIGER-based maps with Google maps (both map and satellite view) of the same area and coded discrepancies that I noticed.

Map_Simple Block is simple and rectangular without interior streets, in both TIGER and Google maps.

Map_Interior Google map indicates additional interior streets not shown on the

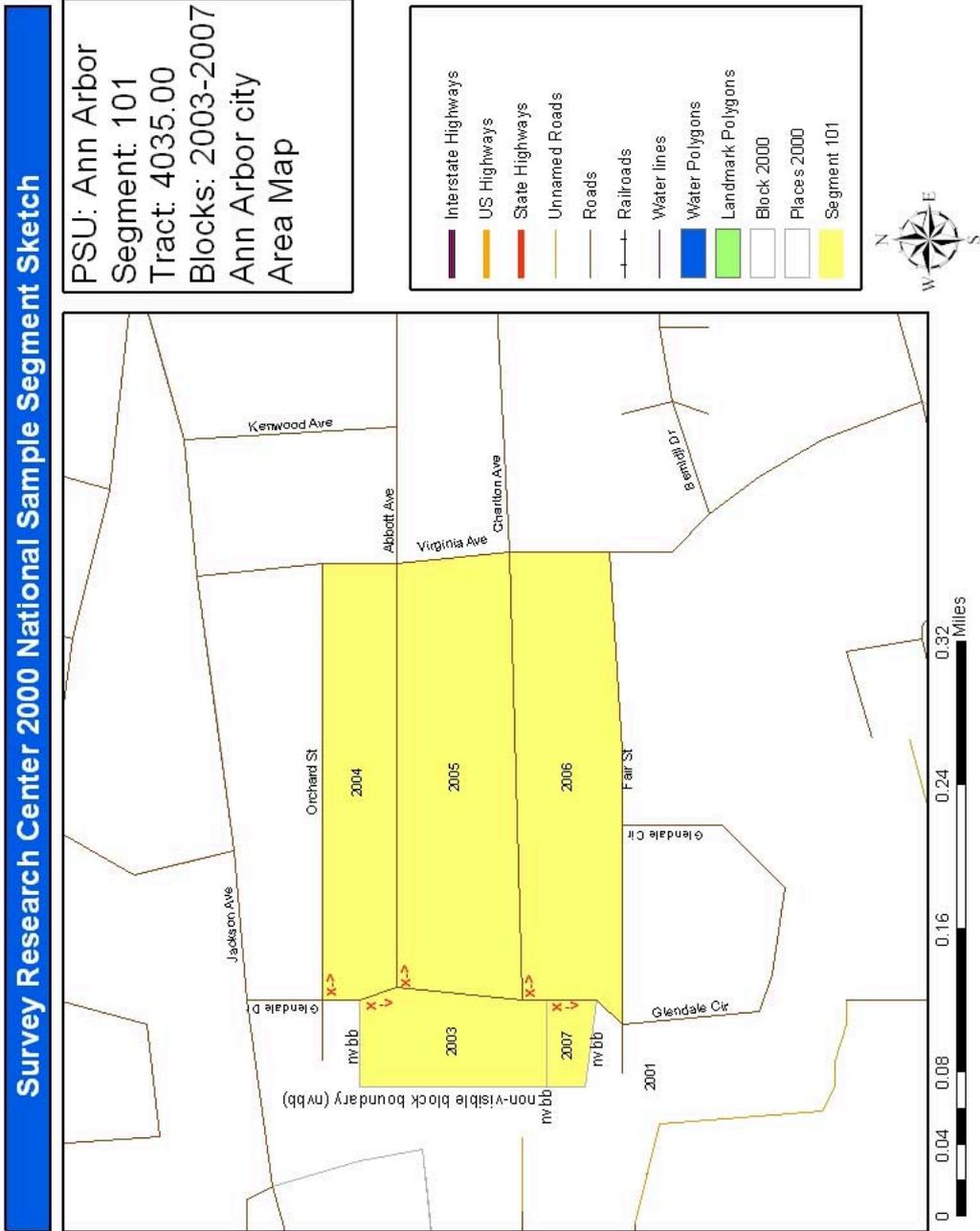


Figure A.1: Example SRC Listing Map

TIGER map. Listers specifically mentioned these anomalies as making the listing task difficult (lister debriefings).

Map_NVBB Block appears to have a nonvisible boundary, i.e. at least one of the block boundaries is not obviously a street or a water feature. Nonvisible boundaries are often political boundaries (town, county, etc.) but may also be overhead power lines or the previous path of a stream. One lister said she discovered a nonvisible boundary to correspond to an underground cable. Listers reported difficulties understanding where to start and stop listing when their segments have nonvisible boundaries (lister debriefings).

Chapter B

Appendix: Logistic Regression Models of Traditional Listing

Propensity

Table B.1: Traditional Listing Propensity Models, Selected Cases Only

| | (1) | | (2) | | (3) | | (4) | |
|------------------------------------|-----------|--------|----------|---------|----------|---------|----------|---------|
| | OR | z | OR | z | OR | z | OR | z |
| Multi-Unit | | | 0.208*** | (-5.49) | 0.217*** | (-5.28) | 0.350* | (-2.09) |
| Vacant | | | 0.683 | (-1.52) | 0.675 | (-1.56) | 0.684 | (-1.51) |
| Trailer | | | 0.978 | (-0.03) | 1.005 | (0.01) | 1.001 | (0.00) |
| Pct. HUs rural | | | 0.170 | (-1.54) | 0.179 | (-1.44) | 0.174 | (-1.46) |
| Pct. HHs with income <= 50,000 | | | 1.369 | (0.24) | 0.556 | (-0.33) | 0.570 | (-0.32) |
| Pct. Pop Spanish language | | | | | 0.289 | (-0.24) | 0.164 | (-0.34) |
| Pct. Pop. Afr.-Amer. | | | | | 2.109 | (0.32) | 1.839 | (0.26) |
| Map, invisible boundary | | | | | 0.308 | (-1.72) | 0.293 | (-1.78) |
| Lister has safety concerns | | | | | 0.342 | (-0.83) | 0.384 | (-0.71) |
| Lister drove herself while listing | | | | | 1.557 | (0.57) | 1.565 | (0.55) |
| Lister has safety concerns | | | | | 0.342 | (-0.83) | 0.384 | (-0.71) |
| Lister and segment language match | | | | | 2.229 | (0.86) | 2.741 | (1.05) |
| Multi * Lister has safety concerns | | | | | | | 0.990 | (-0.01) |
| Multi * Lister drove | | | | | | | 1.073 | (0.10) |
| Multi * Language match | | | | | | | 0.489 | (-0.92) |
| Constant | 17.721*** | (9.42) | 1.3e+09 | (0.01) | 2.4e+10 | (0.00) | 7.5e+09 | (0.01) |
| StdDev(segments) | 1.753 | | 1.234 | | 1.145 | | 1.141 | |
| rho | 0.483 | | 0.316 | | 0.285 | | 0.284 | |
| Log Likelihood | -549.148 | | -517.053 | | -514.690 | | -514.054 | |
| Observations | 1970 | | 1970 | | 1970 | | 1970 | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Chapter C

Appendix: Matching Addresses in NSFG Listing

The analyses presented in this paper involve two rounds of address matching, comparing listed addresses to identify those which refer to same housing unit. This sort of matching work always requires judgments. Other researchers might make different judgments and would thus create a slightly different agreement indicator, which would impact the results. However, I feel that all of the matching decisions I made are justifiable and defensible. The quality of this matching work will greatly affect the quality of the results; both false matches and false non-matches would cause errors in my analysis. In this appendix I explain the two rounds of matching in detail.

The first round of matching involved comparing the input listing given to the dependent listers (the listers who performed the third listing of each segment) to the frame created by those listers. In the second round I matched the three frames to each other to determine which units were listed by only one lister, which by two listers and which by all three. Each round involved several matching steps using both computerized and manual matching and several quality checks. Although the procedures used in each of the two rounds were quite similar, in this appendix I discuss all steps in both rounds separately.

In all of this matching work, my goal was to match addresses that would

lead to the same housing unit being selected. Using this principle, I did allow matches of what at first glance seem to be different addresses, e.g. 1146 Juniper Ln and 1164 Juniper Ln. After discussion with listers and Survey Research Center central office staff, it became clear that interviewers notice and correct these sorts of mistakes in the field. If 1146 Juniper Ln is selected and does not exist, but 1164 Juniper Ln does, the interviewer will often make the judgment about which address was intended herself or she will call in to the central office for confirmation. I also did not allow any many-to-one matches: each address on a frame could have one and only one match on another frame. All matching was done only within segment.

Although there have been great advancements recently in probabilistic matching algorithms (Herzog et al., 2007; Schnell et al., 2009), these sophisticated techniques are not needed here. The segments in my study are quite small: the average number of housing units per segment is less than 200 (range from 50 to 2300, see table 2.1). Thus manual review of all lines within each segment was possible. Additionally, the probabilistic matching routines are very good at resolving spelling errors in street names, but due to the drop-down menu of street names in the listing software, spelling differences are uncommon in these listings.

To protect respondent confidentiality, none of the addresses shown in the examples below are true addresses from the quarter 12 NSFG listing.

C.1 Matching Input List to L_3

The third listing of each segment used the dependent method of listing. I derived the input to this dependent listing from the frame created by the first lister (L_1), with manipulations (additions and suppressions) as discussed in the text. The manipulations allowed me to test for failure-to-add and failure-to-delete error, but only after matching the two lists together to find which added and suppressed lines were included on the frame after listing. The input listing contained 9283 lines plus the 561 housing units I deleted, and L_3 contained 10445 lines.¹ Matching these two lists involved three steps.

C.1.1 Step 1: Match by ID

In the first matching step, I took advantage of the unique key that tracks addresses from the input list to the final frame. However, because listers have the ability to edit addresses on the input list as well as delete them and add them back, not all matches by address ID are true matches.² In the first matching step, I matched only cases that had the same ID as well as the same address (house number, street, and apartment number). Only cases that matched on all of these attributes and were confirmed by the lister were considered matches. These criteria led to 7670 matched pairs. All cases where the address changed in any way between

¹The frame created by the third listers contains 9597 units. The software, however, retains the units removed by the dependent listers and thus the database of units listed and deleted by the third listers is larger, 10445. To capture failure-to-delete bias I had to match to this larger list.

²Listers tell me that they edit addresses on the input list and delete them and add them back when the order of the input list is not correct. They find these techniques easier than reordering the input list using the software's reordering tool.

the input list and the final frame, or where the address was not confirmed, were sent on for further matching.

C.1.2 Step 2: Automatic Address Match

There were 2174 remaining unmatched lines from the input list (including the deleted lines) and 2775 remaining unmatched lines from L_3 . The next step in the matching process involved parsing the addresses into standardized pieces and matching these pieces to identify addresses that referred to the same housing unit.

The Survey Research Center listing software collects addresses in three parts: house number, street and apartment number. I used a SAS macro to parse the street variable on both the input frame and L_3 into four fields:

- Pre-direction: Street direction, when it precedes the street name (e.g. N, E, NW)
- Street Name: Street name (e.g. Main, 37th, Martin Luther King)
- Street Type: Type of street (e.g. Ave, St, Dr, Circle)
- Post-direction: Street direction, when it follows the street name (e.g. N, E, NW)

The parser also standardizes the address parts to improve matching. Table C.1 gives examples of both the parsing and the standardization.

With the full address parsed into 6 pieces (the four shown in Table C.1, plus the house and apartment numbers), the cases ran through the matching programs

Table C.1: Parsing and Standardizing Street Variable

| Full street | Parsed Fields | | | |
|----------------------|---------------|-------------|-------------|----------------|
| | Pre-direction | Street Name | Street Type | Post-direction |
| Brooklyn Ave | | Brooklyn | Ave | |
| North 49th Av | N | 49th | Ave | |
| NW Cherry Hill Drive | NW | Cherry Hill | Dr | |
| Dry Creek Road S | | Dry Creek | Rd | S |

Table C.2: Automatic Matches found, by Pass

| Field | Pass 1 | Pass 2 | Pass 3 | Pass 4 |
|----------------|--------|--------|--------|--------|
| Segment | X | X | X | X |
| House number | X | X | X | X |
| Pre-direction | X | | | |
| Street Name | X | X | X | X |
| Street Type | X | X | X | |
| Post-direction | X | X | | |
| Apartment | X | X | X | X |
| Matches Found | 987 | 0 | 0 | 28 |

(SAS macros). The first pass required matches on all 6 fields, and identified 987 matches. Subsequent passes relaxed the matching criteria. For example, the second pass would match 1495 Beard Ave to 1495 S Beard Ave. The final automatic matching pass would match 659 Wayne Dr to 659 Wayne Rd. Table C.2 shows the matching criteria at each pass as well as the number of matches found. All passes required matching house numbers, street names and apartment numbers.³

Note that only cases which did not match in step 1 or in pass 1 went on to the later, more permissive, passes. That is, only if both 659 Wayne Rd and 659 Wayne Dr had no better partners would they be matched together. I carefully reviewed all matches from the later passes to ensure that they seemed reasonable.

³Cases on either list where the house number was missing or was some variant of “No #” were excluded from all matching in step 2 to avoid false matches. These cases with missing house numbers were matched in step 3.

C.1.3 Step 3: Manual Address Match

To find all remaining matches, in the third step I reviewed all unmatched addresses. I created spreadsheets of all addresses in each segment, both matched and unmatched. These spreadsheets contained not only the full address of each case on the input list and on L_3 but also the description provided by the lister, if any.⁴ The addresses on each list were sorted in street, house number and apartment order. I reviewed these spreadsheets carefully to identify matches that were not picked up in the previous steps. I identified five kinds of matches in this review.

Inconsistent Apartment Numbers Listers used different designators to refer to apartments in a building: for example A, B, C; 1, 2, 3; Front, Rear; 1st floor, 2nd floor, 3rd floor; 101, 102, 103. Because steps 1 and 2 required perfect matches on apartment numbers, these units were not matched. When the two lists used different numbering schemes but agreed on the number of units in a building, I matched all the units (in the order implied by the numbering). In those cases where one list contains more units at an address than the other list, I tried to deduce which designators referred to the same units.

This set of manual matches also includes cases where one list thought a structure was a single family home and the other saw more than one unit at the address. In fact, my manipulation of the input list made this quite likely as I both turned two unit buildings into single-family buildings and vice versa. In-

⁴Listers are instructed to provide descriptions of the housing units whenever the address does not include a house number or when it might be unclear which unit is meant.

interviewers are trained to approach the first unit when a selected single-family case turns out to be a multi-unit building, so I matched the single family unit to the first unit and left the other units unmatched.

Typo in Street Name The automatic matching routines in step 2 always required a perfect match on street name. But in a few situations, the third lister edited the street name to correct misspellings. I matched these addresses during my manual review.

Typo in House Number When two units appeared to be the same except for small differences in the house number, I matched these units. These matches were made only when there were no or very few other unmatched units with the same street name, because in these situations I felt that the interviewer would probably come to the same conclusion in the field if the case were selected. If an interviewer cannot find 12195 Willow Rd, but does see 12159 Willow Rd, she is likely to interview at the second address.

No House Number When a house number is not available, listers are taught to write “No #” in the house number field and include a description that uniquely identifies the housing unit. These kinds of units appear on both the input to the third listing and the output to the third listing. I was able to match these lines when they were the only unmatched cases on a street or when the descriptions made it clear they referred to the same unit.

This process led to 251 additional matches, more than half due to inconsistent apartment numbers (see Table C.3).

Table C.3: Manual Matches Found

| Type | Matches Found |
|--------------------------------|---------------|
| Inconsistent Apartment Numbers | 136 |
| Typo in Street Name | 56 |
| Typo in House Number | 11 |
| No House Number | 45 |
| More Dubious | 3 |
| Total | 251 |

C.1.4 Quality Checks

After matching I performed several quality assurance steps. I carefully reviewed all computer and manual matches and the match types.

C.2 Matching Three Frames

The process of matching the three final frames was similar to that used in matching the input to the third listing to the output. I first matched addresses unit strict matching criteria on all address parts and then relaxed these criteria, allowing divergent street directions and typos. I then performed manual matching that could account for spelling errors and other small differences. The principles used in matching were the same in this round as in the previous round. All matches were within segment and were one-to-one. Once again, my goal in matching addresses was to identify those listed cases that would lead to the same housing unit being approached for an interview.

There are three ways in which this matching round was different from that described in section C.1 of this Appendix. First there is no ID across the three

frames so no ID matching was possible. Second, I matched only listed addresses on the three frames. Any addresses not verified, or added and then deleted, are not part of what they think is the housing unit frame for the segment and were not matched. Third, this round of matching involves three address sources rather than just two, which complicates the process as discussed below.

The first listing contained 9423 listed lines, the second 9345 and the third 9597 (see Table C.4). After parsing and standardizing the addresses as described in section C.1, I used SAS to identify perfect matches across all of the address pieces in all three frames. This process found 6169 triples. I then used the matching macros, the same ones described above, to find identical pairs of addresses across listings that matched to a unit in the third listing at decreasing levels of strictness. This process identified another 419 matched triples. Next I created spreadsheets of all units in each listing (both matched and unmatched) for manual matching, in the same way described above, and found another 1165 three-way matches. I reviewed these spreadsheets several times, ordering the units within segments in listing order and also in street name and numbers order.

The total number of three-way matches is 7751.⁵ There are also 1322 pair matches without a third.

I recognize that there will likely be some false matches and false nonmatches that I cannot eliminate. If these errors are correlated with any of the independent variables in my models, my results will be biased (Carroll et al., 2006, pg. 345-352).

⁵During manual review, I dissolved some of the matches found in the less-strict SAS matching steps. For this reason, the number of final three-way matches does not quite equal the sum of the numbers given above.

Table C.4: Number of Housing Units Listed for Each of Three Listings, by Segment

| Segment | L1 Lines | L2 Lines | L3 Lines |
|---------|----------|----------|----------|
| 1 | 55 | 48 | 72 |
| 2 | 106 | 110 | 106 |
| 3 | 155 | 154 | 141 |
| 4 | 164 | 182 | 184 |
| 5 | 89 | 79 | 83 |
| 6 | 87 | 87 | 87 |
| 7 | 130 | 162 | 128 |
| 8 | 210 | 204 | 211 |
| 9 | 108 | 101 | 108 |
| 10 | 93 | 106 | 94 |
| 11 | 122 | 119 | 123 |
| 12 | 98 | 94 | 98 |
| 13 | 149 | 124 | 148 |
| 14 | 96 | 101 | 155 |
| 15 | 159 | 194 | 239 |
| 16 | 109 | 111 | 106 |
| 17 | 96 | 93 | 95 |
| 18 | 108 | 89 | 104 |
| 19 | 83 | 83 | 81 |
| 20 | 74 | 81 | 76 |
| 21 | 80 | 80 | 79 |
| 22 | 84 | 84 | 84 |
| 23 | 122 | 139 | 115 |
| 24 | 584 | 580 | 553 |
| 25 | 95 | 96 | 96 |
| 26 | 88 | 84 | 87 |
| 27 | 626 | 626 | 642 |
| 28 | 103 | 100 | 104 |
| 29 | 165 | 198 | 165 |
| 30 | 271 | 312 | 284 |
| 31 | 152 | 226 | 154 |
| 32 | 95 | 97 | 95 |
| 33 | 233 | 131 | 202 |
| 34 | 2337 | 2146 | 2342 |
| 35 | 99 | 95 | 98 |
| 36 | 82 | 82 | 79 |
| 37 | 95 | 95 | 94 |
| 38 | 162 | 162 | 163 |
| 39 | 417 | 403 | 411 |
| 40 | 236 | 239 | 234 |
| 41 | 94 | 95 | 94 |
| 42 | 110 | 110 | 102 |
| 43 | 110 | 104 | 105 |
| 44 | 118 | 122 | 130 |
| 45 | 82 | 86 | 77 |
| 46 | 144 | 141 | 145 |
| 47 | 78 | 119 | 158 |
| 48 | 160 | 158 | 155 |
| 49 | 110 | 113 | 111 |
| Total | 9423 | 9345 | 9597 |

Chapter D

Appendix: Interviewer Questionnaire

This interview is voluntary and confidential.

Your answers will not be identified with your name, and they will not be shared with your supervisors or human resources department. Your answers to these questions will help researchers do basic research on the survey process and better understand the data in this survey.

Q1. In addition to the University of Michigan, have you ever worked as an interviewer at any other survey or market research organization?

1. Yes
5. No

Q1a. In addition to the National Survey of Family Growth (NSFG) Cycle 7, have you worked as an interviewer on any other University of Michigan survey projects?

1. Yes; have worked on other UM survey projects
5. No; NSFG Cycle 7 is my first UM survey project

[IF Q1=NO AND Q1A=NO, GO TO Q3]

Q1b. Which previous NSFG cycles have you worked on? Please check all that apply.

- Cycle 1 (1973)
- Cycle 2 (1976)
- Cycle 3 (1982)
- Cycle 4 (1988)
- Cycle 5 (1995)
- Cycle 6 (2002)
- I did not work on any previous cycles of NSFG

Q2. Was your previous interviewing experience doing in-person, telephone interviews, or both?

1. In-Person
2. Telephone
3. Both

Q3. Including working on NSFG Cycle 7, how many months or years have you been an in-person field interviewer? 0 for less than one month, Enter months and years

Q4. Including all types of interviewing, about how many survey projects have you ever interviewed on?

0. NSFG Cycle 7 is my FIRST survey project [GO TO Q7]

----- # of survey projects

Q5. Have you ever worked on a survey with the following content? (SELECT ALL THAT APPLY)

1. Sexual activity
2. Drug use
3. Criminal activity
4. None of the above

Q6. On how many survey projects have you used a computer to do interviewing?

----- # of survey projects

Q7. Before coming to training had you ever used a “stylus” (the electronic pen used on the tablet) on a computing device (including on a PDA or a Palm Pilot)?

1. Yes: have used stylus
5. No: never used stylus before NSFG training

Q8. Besides being a Field Researcher on the NSFG, do you currently have any other paying jobs?

1. Yes

5. No [GO TO Q9]

Q8a. How many hours per week, on average, do you work on your other job(s)?

----- # of hours per week

Q8b. Is/Are your other paying job(s) as an interviewer or something else?

1. Interviewer

2. Something Else

3. Both

Q9. What is the highest level of school you have completed?

1. High school graduate or GED

2. Some college but no degree

3. 2-year college degree (e.g., Associate's degree)

4. 4-year college graduate (e.g., BA, BS)

5. Graduate or professional school

Q10. Are you currently attending school or college to get a degree or certificate?

1. Yes

5. No

Q11. What is the month and year of your birth?

----- month ----- year

Q12. Are you Hispanic or Latina, or of Spanish origin?

1. Yes

5. No [GO TO Q13]

Q12a. Are you ...

1. Puerto Rican

2. Cuban

3. Mexican

4. Central or South American

5. Some other Hispanic or Latina group (specify)

Q13. Which describes your racial background? Please select one or more groups.

1. American Indian or Alaska Native

2. Asian

3. Native Hawaiian or Other Pacific Islander
4. Black or African American
5. White

[IF ONLY ONE RACIAL GROUP SELECTED, GO TO Q14]

Q13a. Which of these groups would you say best describes your racial background?

1. American Indian or Alaska Native
2. Asian
3. Native Hawaiian or Other Pacific Islander
4. Black or African American
5. White

Q14. Do you speak any languages other than English?

1. Yes
5. No [GO TO Q15]

Q14a. What language(s) do you speak?

- 1 Spanish

7 Other (specify) [GO TO Q15]

Q14b. On a scale from 1 to 5, where 1 means 'barely conversational' and 5 means 'native Spanish speaker,' how proficient do you think you are in Spanish?

Q15. What religion are you now, if any?

1. None
2. Catholic
3. Jewish
4. Baptist or Southern Baptist
5. Methodist, Lutheran, Presbyterian, or Episcopal
6. Other Protestant or Christian religion
7. Hindu, Buddhist, or Muslim
8. Other Religion [specify]

Q16. Currently, how important is religion in your daily life? Would you say...

1. Very
2. Somewhat important
3. Not important

Q17. What is your current marital status?

1. Married
2. Not married but living with a partner
3. Widowed
4. Divorced
5. Separated
6. Never been married

Q18. Which category represents the total yearly income for your household during the past 12 months?

1. Under \$25,000
2. \$25,000-\$34,999
3. \$35,000-\$49,999
4. \$50,000-\$74,999
5. \$75,000 or more

Q19. How many babies, if any, have you ever given birth to?

----- # of babies

[IF Q19=0, GO TO Q20]

Q19a. How old are your children? Please enter all that apply.

- 0. Under 5 years old
- 1. 5-12 years old
- 2. 13-18 years old
- 3. 19 years old or older

Q20. In what city and state do you live?

City/Town_____

State_____

Q21. What county do you live in?

_____ COUNTY

Q21a. How many years have you lived in that county?

_____ # years

Now we have some questions about your opinions and attitudes about the interviewing process. This information will be used for statistical purposes only and has no effect on your employment with the University of Michigan.

Q22. How strongly do you agree or disagree with the following three statements?

Q22a. Most of the time I can/I will be able to figure out what a respondent's real objections are to a survey.

- 1. Strongly Agree

2. Agree
3. Disagree
4. Strongly Disagree

Q22b. I can/I will be able to persuade people to agree to interviews better than most other interviewers.

1. Strongly Agree
2. Agree
3. Disagree
4. Strongly Disagree

Q22c. No matter what I do, there are/there will be some respondents who will never agree to participate.

1. Strongly Agree
2. Agree
3. Disagree
4. Strongly Disagree

Q23. Interviewers sometimes have difficult decisions to make while doing their jobs. Which one of the following statements comes closest to how you feel as an interviewer?

1. IT'S BETTER TO PERSUADE A RELUCTANT RESPONDENT TO PARTICIPATE THAN TO ACCEPT A REFUSAL, EVEN WHEN YOU FEEL THEY WON'T GIVE VERY ACCURATE ANSWERS.
2. IT'S BETTER TO ACCEPT A REFUSAL FROM A RELUCTANT RESPONDENT THAN TO PERSUADE THEM TO PARTICIPATE WHEN YOU FEEL THEY WON'T GIVE VERY ACCURATE ANSWERS.

Q24. This question uses a scale from 1 to 10, where 1 means 'dislike very much' and 10 means 'like very much'. You can use both of these numbers, plus all of the numbers in between. Please answer with a number from 1 to 10 to describe how much you like or dislike each of the following tasks:

- a. Approaching the household/doorstep introduction
- b. Gaining cooperation
- c. Conducting the interview
- d. Working with a supervisor (field operations coordinator/production manager)
- e. Working on a team
- f. Completing paperwork associated with an interview
- g. Converting reluctant informants and respondents

Q25. This question uses an attractiveness scale where 1 means 'very unattractive' and 10 means 'very attractive'. You can use both of these numbers, plus all of

the numbers in between. Using this attractiveness scale, please signify the attractiveness of each of the following tasks in your job as an interviewer:

- a. Flexible work hours
- b. Relevance/importance of survey research
- c. Pay
- d. Interacting with a variety of people

Now we have some questions about your opinions and attitudes on a few topics, including some covered by the NSFG survey. This information will be used for statistical purposes only and will have no effect on your employment.

Q26a. Sexual relations between two adults of the same sex are all right.

1. Strongly Agree
2. Agree
3. Disagree
4. Strongly Disagree

Q26b. Any sexual act between two consenting adults is all right.

1. Strongly Agree
2. Agree

3. Disagree

4. Strongly Disagree

Q27. A young couple should not live together unless they are married.

1. Strongly Agree

2. Agree

3. Disagree

4. Strongly Disagree

Chapter E

Appendix: Development of Housing Unit Level Characteristics in NSFG Dataset

The entire NSFG listing dataset has three observations of every housing unit. Thus deriving housing unit level characteristics is not straightforward. The variables at the housing unit level in my models are multi-unit, multi-unit large, trailer, no number and several variables related to the interviewer dispositions. below I provide details on how I created these variables.

Trailer If any lister indicated that a unit was a trailer, in the apartment or description field, then the housing unit is a trailer on all observations of the unit.

Multi-unit Units in multi-unit structures can be designated with apartment number or in the street number field (for example: 1950A, 49 1/2, 1175-1). If the housing unit was listed by only one lister, and the lister included any of these designations, then the unit is flagged as in a multi-unit building in my dataset. If the housing unit was listed by two or three listers and more than half of the them designated it as multi-unit, then it is flagged as multi-unit at the housing unit level. If exactly half of the listers indicated the unit was in a multi-unit structure (one of the two listers who included the unit) then I flagged the unit as a multi-unit (unless I had manipulated the multi-unit

status of the unit in the third listing). Any units coded by this process as both trailers and multi-unit were recoded as single family.

Multi-unit in small building Housing units flagged as multi-unit were categorized as in small buildings by counting the number of units in the same building within each lister (usually those with the same house number and street name). If the minimum number of units in the building across the three listers was fewer than 19, the unit was coded as in a small building.

No Number If any lister recorded an address as not having a street number, I flagged the unit as a no number case.

Disposition data For each case selected for interview, the assigned interviewer, who was also the first lister, assigned a code capturing the outcome of the screening and interviewing process. These codes are, broadly: no contact, refusal, screener complete not eligible, screener complete eligible, out of scope (vacant) and improper listing (nonresidential or out of segment). The codes were assigned to each selected unit by only one lister and thus no decisions were needed to create housing unit level variables.

Chapter F

Appendix: Debriefings with SRC Listers and Interviews

In the fall of 2009 I conducted individual debriefings with trained listers and interviewers from the Survey Research Center. These were interviewers who had completed some listing work for my project. The questions I prepared for discussion are given in Section F.1 but the meetings were informal and I encouraged the subjects to talk with me about anything they thought was relevant. Section F.2 gives quotes from the transcripts themselves that support the discussions in Chapters 2 and 3. I promised confidentiality to the listers and thus cannot reveal their names or print the entire transcripts.

I am very thankful to the listers both from their listing work for this project and for their participation in these discussions. The assistance provided by these listers greatly improves my research.

F.1 Questions to Guide Debriefing Discussions

INTRODUCTION:

Thanks for your time

Little bit about who I am:

- Student at Maryland
- Dissertation research on listing

- Working with NSFG folks, but not employee of Michigan
- happy to answer questions about research, at the end

Calling only as a researcher

- won't report individual discussions back to Sharon, Nicole or any ISR
- YOU ARE THE EXPERT

May I record so I don't have to take a lot of notes?

What interviewing experience do you have?

What listing experience?

NSFG only or other studies too?

Other companies?

Questions for experienced interviewers:

Some studies do a missed housing unit check, looking for housing units that were not on the frame. Does your study do that?

- Do you understand why we do it?
- Do you think most FIs do this correctly?

Interviewers sometimes notice errors in the listing. Have you noticed listing errors while interviewing?

- If yes, please describe most common error
- What causes these errors?

Another issue I'm concerned with is overcoverage-when HUs outside of the selected segment are listed and selected.

- Do your maps show you where the segment boundaries are?
- Have you ever been asked to check if the selected housing units are actually inside the segment?
- Does this work well?
- Do you think listers list outside of the segment?

Are there characteristics of a housing unit that make you think "this one is not going to participate"?

Questions for experienced listers:

DEFINE SCRATCH / UPDATE LISTING. Have you done traditional or dependent listing?

Have you done listing on paper or just on the computer?

What differences have you noticed in listing procedures if you work for different projects and/or companies?

Follow up method:

- Do you feel that the list you're updating sometimes gets in the way, and it would be easier to list from scratch?
- Or do you think update listing is usually easier?

Follow up paper/computer:

- Do you prefer paper or computer listings?
- Are there some things that are easier on paper?
- Are there some things that are easier on the computer?

I know that the central office reviews all the NSFG listings and sometimes asks questions or has a lister redo part of a listing.

- Did ask you to redo any of your listings?
- What did they think was wrong?
- Was it?

What kinds of units do you think we might miss in listing?

I'm sure it's sometimes hard to figure out how many units are in a multi-unit building. What do you do in these situations?

Listers have said that some kinds of segments are more difficult to list-those with invisible boundaries or really large rural segments. What do you think?

- PROMPT: What do you think are the kinds of segments that have more listing errors?
- What about urban areas might make them difficult?

Have you listed a segment with a nvbb (invisible boundary)?

- Did you figure out what the boundary was? STREAM, POWER LINE, POLITICAL BOUNDARY

- How did you decide what to list?

Are there other kinds of segments that you think we list poorly?

Do you talk to neighborhood residents when listing? As informants, or just chit-chat?

Do you speak to other informants, such as post offices, fire stations etc.?

From what I've heard, listers have been questioned by residents or even police when listing. Has something like this happened to you? How did you handle the situation?

Do you drive while listing? Have someone drive for you? Walk?

Do you enjoy listing work? More or less than interviewing work?

Have you heard of listers ever making deliberate listing errors? Why might a lister do that?

ASK ABOUT EXPERIENCES IN LISTING FOR MY PROJECT (WHICH THEY KNOW AS THE ST-TEST PROJECT)

F.2 Quotes from Transcripts of Debriefing Discussions

I conducted seven debriefing sessions and have recordings of six of them. (Due to technical difficulties, one was not recorded, though I do have notes I took during the debriefing.) I promised each respondent confidentiality, thus I cannot print the transcripts in full and cannot reveal the names of the interviewees I spoke to. Below are selected quotes (with some identifying details removed) that support the claims made out the listers' behaviors and concerns in Chapters 2 and 3. Note that

the listers use the terms scratch and update to refer to traditional and dependent listing.

F.2.1 Quotes from Lister A

LISTER A: The hardest ones to determine when you're listing are one of the non-visible barrier.

MS. ECKMAN: Right, I've heard about those. What have you seen?

LISTER A: Well, I've only really had those, I think – did I have one in a scratch listing? I try to look for a big ditch –

MS. ECKMAN: Mm-hmm.

LISTER A: – which has very often been a barrier, because in many tracts of land, what looks like a simple ditch is some kind of a creek run on a flat map.

MS. ECKMAN: Oh, I see.

LISTER A: Sometimes I'm not real sure that I got the right spot. But in – there's usually some kind of gap between houses if you're in a subdivision.

MS. ECKMAN: Mm-hmm.

LISTER A: Because it would be wetter than somebody would really want to build on anyway at certain portions of the time probably (inaudible). But I had one that was a county – not a county line, a township line, and I called the township office and asked them, can you please tell me what the last address would be –

MS. ECKMAN: Oh.

LISTER A: – in this township? Because I figured out that it was a township

line and I called and it was. They verified that and what the last address would be. And this house was their township; that house was not.

...

LISTER A: Well, I'm driving up this path which is rough blacktop, not very wide with the brushy trees not greened out yet on both sides. And I haven't made it up around – and I'm just starting to go around the bend and I – it just feels awful.

MS. ECKMAN: Hmm.

LISTER A: And I don't want to go any farther and I worked very hard at backing out of the driveway. Well, I told my team leader, I said, it just felt wrong.

MS. ECKMAN: Hmm.

LISTER A: And it was about 3:30 on an afternoon on a weekday. It was bright, sunshiny, spring, whatever, and I said, there was no reason that it should have felt bad, but it did. And when they had a traveling interviewer come and one of the places they sent her was there and I ran into her in the hotel to trade some info, she said, Martha, there was nothing wrong back there. Well, it was a home that was up for sale. There was nobody living in it. It was very secluded back there.

MS. ECKMAN: Uh-huh.

LISTER A: So, who knows what kids could have been doing what at 3:30 after school got out if they know there's a place where nobody is right now.

MS. ECKMAN: Right, right.

LISTER A: And who knows what was going on the day that I was driving up that driveway.

MS. ECKMAN: Right.

LISTER A: Well, for some reason, I felt awful and I left.

...

LISTER A: But you're driving down streets very slowly, you have flashers on and someone had come out and – I made a second pass down the road to check on something, and somebody had come out and even like almost knocked on my car, this older lady. And I'm saying older my age now, like 65. She said, what are you doing? I said, okay. And by George, within 15 minutes, there was a policeman on my tail.

MS. ECKMAN: Oh, wow.

LISTER A: And he was saying, what are you doing? And I showed him all my paperwork because I am legit. But it was – it was one of the newer subdivisions and it was a subdivision that's not in town or near town, it's a bit of country and a bit of subdivision, and those are the hardest, those are the hardest.

F.2.2 Quotes from Lister B

LISTER B: Yeah, I would say the rural and big apartment complexes are difficult because – yeah, in <LOCATION>, they have that invisible boundary line.

MS. ECKMAN: Oh, did they?

LISTER B: Yeah. I could not – I'd finally think I'd got it, but I don't know for sure. I don't have anything – I wish I had something to definitely say that was it. I remember Judy, my FOC, saying, somebody there could tell you if that's a – you

know, that that's the invisible boundary line. But I didn't have anybody to ask.

MS. ECKMAN: Oh, I see.

LISTER B: (Inaudible) that myself before I get out there, you know, instead of asking. But, yeah, that's difficult, invisible boundary lines

...

LISTER B: Yeah, I prefer to drive because you don't – you're not as suspicious-looking for the most part, like in an urban area, where there's apartment buildings, you can just drive and rest, keep going, not draw too much attention to yourself. If you're out listing, it does kind of – it can rouse a little suspicion, a little more suspicious. So, that's why I would prefer to –

MS. ECKMAN: Yeah.

LISTER B: You know, just a little more incognito by driving.

F.2.3 Quotes from Lister D

LISTER D: Well, in the updating, you have the list and sometimes you think, well, you know, something's there and maybe I'm just not seeing it, you know. So, you really kind of are tempted sometimes to just, you know, confirm it.

MS. ECKMAN: Mm-hmm.

LISTER D: And if we can't get into a locked area, of course, then that's what we do, we just confirm them. Sometimes there's a gated and guarded community and we can't get in. So, we just confirm all the ones that were listed there. So, it could be right, it could not be. We don't know.

...

LISTER D: Yeah. Well, we – the only way we can do that is list like the doorbells or whatever and go up there and see what mailboxes are there. Usually it's inside the door or – it's usually inside the front door there and look in and see what the numbers are or by the mailboxes. And if we find another gas meter or something like that, we can just put down no number, you know, and then when – if that one is selected for our screen, then we'll be more – then we're able to go into the building, knock on the doors and find out, well, there's not really another one, so then we can just close that one out as improper listing. But we'd rather have more listed there, you know, as a possibility than less. We don't want to miss any.

...

MS. ECKMAN: And has anyone – the residents, have they kind of challenged you while you're listing?

LISTER D: Oh, oftentimes, yeah. I had a guy with a gun on his hip, you know (inaudible) in Alaska. What are you doing on this street? This is private property. And I said, well, jeez, there's several houses here, I thought it was just a public street, you know. He says, no, and I want you off here. So, you know, you kind of telling them you're doing – you're just updating maps and getting information. They usually accept that. You know, you show them your identification from the University of Michigan and you have a little sign in the window that says you're on official business.

...

MS. ECKMAN: Right. Now, the one thing I worry about specifically with rural areas is these non-visible boundaries.

LISTER D: Yeah. I mean, yes.

MS. ECKMAN: So, I guess – you're giggling. I guess that means that they can be a problem for you.

LISTER D: Oh, yes.

MS. ECKMAN: So, just tell me about those. Do you get out there and then you can figure out what it is or sometimes you can't figure it out or –

LISTER D: Sometimes there's just no indication what it is and the only way you figure it out is by looking at the map and seeing where the curve in the road is and how far it should be – you know, there's a mileage gauge on the map, and so, you can figure it out that way. And sometimes you – it might be just a county boundary and there's no sign saying, okay, you're going into another county.

MS. ECKMAN: Mm-hmm.

LISTER D: It's just, you know, invisible.

MS. ECKMAN: Yep.

LISTER D: So, that's the way you figure it out. You just figure out the curves in the road and how far it is from the road you've just come from, a crossroad or something like that.

...

LISTER D: And you can just go down there and check and easily just do it

from your car. And that's – there's no need to walk. Of course, busy road, of course, you have to walk because you can't be slow-moving traffic there. And rural, you cannot walk, of course, because it's way, way, way too far between places.

MS. ECKMAN: Mm-hmm.

LISTER D: If you're going to have an apartment, you need to walk. You need to find out the apartment and also the office to get some information there.

MS. ECKMAN: So, you're saying you prefer to walk areas where there's apartments that –

LISTER D: Oh, yes, definitely, mm-hmm.

MS. ECKMAN: Mm-hmm.

LISTER D: But I prefer to drive if I can. But if it's going to – if it's too difficult, then, of course, I'll walk. But, you know, in most subdivisions that are pretty evenly spaced, it's easy to just drive along and find the – if I find that it's difficult finding the house numbers, of course, then I'll get out and walk. But I think in most of the areas, I was able to just drive.

F.2.4 Quotes from Lister E

LISTER E: So, if it looks like there are two apartments, it's better to list them. Then you might go later and they say, oh, no, we've been combined into one now.

...

LISTER E: So, you just go through it and click confirm, you know, pretty much, maybe have to add or delete something here or there, but it can be really

helpful. So, yeah, you're clicking. Now, there sometimes you might have a house that has – you know, it will say apartment one, apartment two, and you don't see apartment two at all.

MS. ECKMAN: Right.

LISTER E: But from what I'm recalling in the training – and I usually do this – we're encouraged to leave it in and then later try to see if that apartment two is, in fact, there. Unless I'm just really positive there cannot possibly be an apartment two, then I'll delete it. But sometimes I get into a little decision making there. I think, wow, I just don't see it at all.

MS. ECKMAN: Mm-hmm.

LISTER E: So, it can also – you know, it can be easier and it can help include things that aren't visible because sometimes there is an apartment and the entrance is in the back. There's no way I would have seen that.

MS. ECKMAN: Right.

LISTER E: And that's – you know, it's more likely that you will include that. I mean, that's important with a scratch listing not to get kind of just in automatic mode, that you're just going down the street, confirm, confirm, because sometimes there is something that isn't included in that original update list and you have to make sure you get it. There can be a whole block that was missed or a whole little cul-de-sac.

MS. ECKMAN: Really?

LISTER E: So, you have to be really – not lazy about it, you know, really alert. Sometimes – I've had it in the past where I've had update listings where the

information provided to me was not good.

...

LISTER E: Or an unsafe area where I feel like I shouldn't necessarily get out of my car a whole lot. I should, you know, do it mostly by car. So, those are hard to – or if it's a long drive and a no trespassing sign, well, I kind of venture up that drive. But, again, you're thinking about these safety issues and maybe you're holding back a little more than you would otherwise.

...

MS. ECKMAN: Right, right, I can see that. So, my next set of questions here is about segment characteristics. You've already mentioned segments that are dangerous being difficult to list.

LISTER E: Mm-hmm.

MS. ECKMAN: And let me just confirm, I think what you're saying there is that a dangerous neighborhood, you're more likely to drive while listing and, therefore, maybe miss some things. Is that what you were saying?

LISTER E: It's possible. And, again, it depends on the kind of street. If it's not a busy street, if it's just a side street and you can drive along the side of the street and stop and park your car and then maybe get out and look if you have to, that's not a problem. But the problem is if it's not a real good (inaudible) and it's a busy street.

MS. ECKMAN: I see.

LISTER E: And you can't park at all, you can't stop along the street. Well, if need be, then I'll have to park somewhere and get out and walk and then go back to my car. I mean, there's some you just can't look from your car if it's dangerous with traffic.

MS. ECKMAN: I see. So, you're not talking so much about fear of crime, more about fear of traffic.

LISTER E: Well, it's more traffic. Although if there's fear of crime, I will try to stay in my car more.

MS. ECKMAN: Mm-hmm.

LISTER E: I mean, obviously, when I work the segment, I'll have to get out and walk around. But, you know, I don't walk unnecessarily. I will stay in my car where it's feasible. But I'll also try to go at a time of day when it's probably okay, you know, in the morning, Sunday morning maybe.

...

MS. ECKMAN: Okay, yeah. I've seen in some of the training materials these non-visible boundaries, NVBB.

LISTER E: Oh, they can be difficult.

MS. ECKMAN: Yeah.

LISTER E: Can be. It can be easy. It can be really easy once you get, you know, kind of in the habit of it. You know, it could be a – like maybe it's indicated on the map that it's water. Well, then there might not be any water, but you see a little dip or a little vegetation line where you think, yeah, some times of the year,

there might be a little trickle of water in that. It's not always, you know, a flowing river. Or it can be a power line or you might notice suddenly the road isn't paved or the numbering system on the mailboxes can change.

...

MS. ECKMAN: We talked about driving and walking. It sounds like driving can be more convenient.

LISTER E: It can, uh-huh. You can – you know, you can get more of an overview maybe and cover an area more quickly certainly. But if you can't pull over, I mean, then it doesn't work. You have to be able to stop your car there and enter. I mean, I could pull ahead three houses and enter those three, or if need be, just quick write the addresses on a piece of paper if I can't stay there very long on the side.

MS. ECKMAN: I see, uh-huh.

LISTER E: Then enter them into the computer. Or sometimes I – if it's an update, I'm hardly able to stop, you know, but I can go maybe slowly. So, I can't be looking down at my computer and then up at the house, so I write down – for that block, I write down all the addresses on a pad of paper and I hold that pad up near my steering wheel, and then I take a pen and I kind of check off these addresses as I pass them. Then I might have to go around that block three or four times to do that because of the traffic where you can't stop. So, that can be – and I guess you could get out and walk in an instance like that. It's kind of a toss up.

F.2.5 Quotes from Lister F

LISTER F: But I think I'd almost like to do one by scratch because the updating you kind of – I don't know, it just – I guess it may be hard to see the mistakes sometimes and then you realize, wait a minute, this is really on the opposite side of the street and it throws you off. So, you got to really think that you really have to do it as if you were doing it on your own and not relying on what's there, because some houses have been missed that have added and some of them you wonder why they're not included. So, I was just listing in Rhode Island and it was a group house, but there was no way of knowing that that was a group house –

...

LISTER F: Yeah, just, you know, if there's two families living in one unit and that you may not come – or sometimes going around to the back end of an apartment, you know, if it's an old housing, like a tenement.

MS. ECKMAN: Mm-hmm.

LISTER F: You know, if you go around to the back, you may miss something in the back of the households.

MS. ECKMAN: Right.

LISTER F: I remember doing that in Michigan and Ann Arbor seeing where there may be a household right in the back of the building.

MS. ECKMAN: Right.

LISTER F: That you may not even know from the street. So, you really – yeah, this really can be on foot when there – it's less rural, more buildings closer

together.

MS. ECKMAN: Mm-hmm.

LISTER F: It really (inaudible) me to – to be really thorough, you need to get – do it on foot and really check in the back, too.

...

MS. ECKMAN: Mm-hmm. Well, now, we've talked a lot about the kinds of housing units that are difficult. What about segment level characteristics? Are there segments that are harder to list and easier to list?

LISTER F: Yes, the ones that have an invisible boundary.

MS. ECKMAN: Mm-hmm.

LISTER F: The town lines. I think the ones – I think most interviewers stress this over and over again. You know, the quality of the maps that we receive when especially we're going into an area we haven't been and there's an invisible boundary. We're on the lookout for it, but it's not always visible when we're there if it's a town line.

...

MS. ECKMAN: Mm-hmm. So, you have, yourself, encountered some of these non-visible boundaries, huh?

LISTER F: Yes, mm-hmm.

MS. ECKMAN: Yeah, everyone mentions those as being particularly troublesome, but I'm not – I can't get my mind around them. I'm not exactly sure what

that's about. I think I should go out and –

LISTER F: Well, there could be – there could be – I guess mostly it's town boundaries that you don't see or any markers. Sometimes they – if it's wintertime and they have a stream that you want to start at or something like that –

MS. ECKMAN: Oh, mm-hmm.

LISTER F: – you cannot see it, or if it's in the fall when it's dry, you can't tell, and overgrown. So, you cannot tell if one house is included or not, where your starting point is.

MS. ECKMAN: Right.

LISTER F: That can be unclear.

Chapter G

Appendix: Census 2010 Address Canvassing Whistleblower Post

Post from My Two Census blog, 10/5/2009

<http://www.mytwocensus.com/2009/10/05/>

[feature-real-stories-from-the-census-bureau/](http://www.mytwocensus.com/2009/10/05/feature-real-stories-from-the-census-bureau/)

I worked in the New York City area as a lister during address canvassing and was disappointed with how the operation was conducted. One of my colleagues pointed me to this website some time ago and I felt compelled to share my story. We had alot [sic] of the technology glitches in the hand held computers [HHC] that are widely know by now which included:

- software issues such the [sic] program freezes
- transmission problems such as the Sprint cellular network being down and missing assignments and map spots
- hardware issues such as the fingerprint swipe not working

But New York City has its own problems and is a completely different beast in itself. New York City is the most densely populated city in the United States and each neighborhood has its own unique character. The Census Bureau tries to monitor productivity but the very nature of the city makes it very hard to monitor. Since all the units of multi unit apartment buildings are listed separately a

lister has to key in every entry. Comparing someone who has an assignment with high rise apartment buildings versus someone who has single family homes is like comparing apples with oranges.

During address canvassing we were instructed to find someone who was knowledgeable about where people live or could live. But locating a knowledgeable respondent was easier said than done. There are small tenement buildings in Chinatown and Harlem brownstones; where there are illegal subdivisions. It is very difficult to gain entry or make contact even if you speak the language. There are also a lot of abandoned construction sites where developers tried to take advantage of the real estate boom after September 11th but found themselves out of money in the current recession.

Luckily for the Census Bureau, the current recession produced a talented pool of very intelligent and highly educated workers. My crew leader was knowledgeable and a great leader. From the very beginning he was committed to doing things right. He said that he was continuously told a proper address canvassing operation would be the cornerstone of a successful enumeration. He was thorough and all the work was quality checked by one of the other listers or his assistant. When we couldn't gain access to a building, he encouraged us to try again and gave us additional work to keep us productive. In the end we had all these partially complete assignments where we had one or buildings we either couldn't get into or make contact with anyone. However the office was less than empathetic to our thoroughness. Our crew leader told us that Assistant Manager of Field Operations, field operations supervisors (FOS) and crew leaders in other districts would belittle those who were

behind. They would constantly say things like “John’s district is 40% complete why aren’t you 40% complete?” We were told that if we couldn’t gain access to a building after two visits we had to accept what was in the HHC as correct. Many of us were tempted to falsify work and accept what was in the HHC as correct but my crew leader and FOS were adamant about not doing that. One of the other listers found an entire building with over 200 single illegally divided rooms. The HHC had less than 10 units listed in it. If they accepted what was in the HHC as true they would have missed over 200 housing units.

At the beginning of the fourth [sic] week, my crew leader and several others were written up for being unproductive because they weren’t working fast enough to complete their assignments. They asked the Field Operations Supervisor to approve the writeups. One of the Field Operations Supervisors refused to sign the writeups and they wrote him up also for being insubordinate.

During address canvassing we were to document any additions, or deletions to the address list on an INFO-COMM which is a carbon copy paper. They said that they were hiring clerks to reconcile INFO-COMMs between the production and quality control. The sheer volume of having to go through 2000 pieces of paper is mind boggling. Originally, the plan was to use the INFO-COMMs to help the quality control listers, but they wanted to keep the operation independent so quality control wrote an additional INFO-COMM. All told we wrote out over 2000 INFO-COMMs.

The handheld computer also had glitches. They switched crew leaders in districts that weren’t working fast enough and sometimes just reassigned work. When listers saw their timesheets weren’t approved they submitted additional timesheets

electronically. The new crew leader approved it and then they accused these listers of intentionally trying to milk the government clock. They accused half of an entire crew of listers of clocking overtime.

Nonetheless with all the problems most of the listers worked quickly and breezed through their assignments. By the end of the first week we were about 25% done but they decided to train another 100 listers, by the end of the second week we were halfway done and some crews were almost done but they trained another group of listers. Some of these listers were trained and received no field work because there was none. All told we trained over 100 listers who received less days of work than the four and half days worth of training they received.

The thing to realize is that this was a poorly planned operation from the very beginning. The Census Bureau will waste money for government contracts on hand held computers that are shoddy and unreliable and training staff for which there is no work. But they will try to cut corners when it comes to their mission of counting each person accurately. In order to try to save money and finish ahead of other regions they used intimidation and the threatening of employees. I'm glad that Field Operations Supervisor stood up to the higher ups because like my crew leader said to me... they're just of [sic] bullies.

When the address canvassing operation finished up it was alleged that some of the crew leaders and field operations supervisors told their listers since there was no regard to quality that they could skip making contact even going as far as not conducting field work and enter the units at home. There is no way that listers who were reassigned work magically gained access to buildings people couldn't access

for weeks unless they accepted what was in the HHC as true. The crew leaders and field supervisors who finished first were rewarded with additional work. Those who finished last were sometimes “written up” as unproductive and the office terminated their employment.

Luckily this story has a happy ending. My crew leader didn’t fire any of us for clocking overtime. What they found was that the payroll system was mistakenly rewarding people overtime if they worked over eight hours during a work day even though they were below forty hours in a week. Someone was able to view the timesheet submissions in the office and prove all these listers weren’t clocking overtime. It was rumored that someone who discovered this was the same FOS who refused to sign the writeups.

As for thousands of INFO-COMMs they are sitting in the office file cabinets gathering dust maybe someday someone will go through them. I highly doubt it given the sheer magnitude. I think my crew leader was incredible. And from what I heard from some of the listers that met him their Field Operations Supervisor was even better. I never got the chance to see him but I am honored to have worked with someone who is willing to jeopardize his job for what was morally right. I am surprised I received a phone call the other day to work in the next operation Group Quarters Validation. But I’m pretty sure that my crew leader or FOS won’t be returning anytime soon.

Chapter H

Appendix: Content of NSFG Cycle 7 Female and Male

Questionnaires

H.1 Female Questionnaire

Adapted from NSFG Cycle 7 Staff (2008, Figure 6, pp. 13–14).

Section A

- Respondent demographic characteristics (age; DOB; marital/cohabitation status; race and Hispanic origin)
- Household roster (age; sex; relationship to respondent)
- Introduction to Life History Calendar
- Education (degrees; highest grade completed; date last attended)
- Childhood family background and parents

Section B

- Onset of menstruation (menarche)
- Current pregnancy status
- Number of pregnancies
- Detailed pregnancy history (more details if in last 5 years)

- Confirmation of pregnancy history
- Care of nonbiological children (women 18D44)
- Relinquishment of biological children for adoption
- Adoption plans and preferences (women 18D44)

Section C

- Marital history and characteristics of each husband
- Details on current cohabiting partner, if there is one
- Cohabitation history; characteristics of former cohabiting partners
- Whether the respondent has had biological children with each of her husbands and cohabiting partners
- Ever had sexual intercourse (asked if never married, never pregnant and never cohabited): Age and date of first intercourse; Characteristics of first sexual partner (if not already discussed); Date and age of first intercourse after menarche
- Sex education (if 15D24); timing relative to first sex
- Number of sexual partners (in lifetime; in past 12 months; before first marriage)
- Recent (last 12 months) partner history, up to 3 partners (or last partner ever, if none in the past 12 months); more details on current partners

Section D

- Sterilizing operations (respondent and husband/cohabiting partner)
- Desire for sterilization reversal (tubal ligations and vasectomies only)
- Nonsurgical sterility and fertility problems (respondent and husband/cohabiting partner)

Section E

- Ever-use of contraceptive methods, how emergency contraception was obtained, discontinuation of use, and reasons for dissatisfaction with selected methods
- Details on first method ever used (even if before first intercourse)
- Method use at first sexual intercourse
- Months during which she had intercourse for past 3D4 years or since first intercourse (if within the last 3D4 years)
- Contraceptive method history by month, for past 3D4 years or since first method used
- Method used at first and last sex, for up to three partners in last 12 months
- Wantedness of each pregnancy (by respondent and by father of pregnancy)
- Happiness to be pregnant scale
- Further details on circumstances surrounding pregnancies in last 3 years (including wantedness with that partner)

- Current method use, reasons for current nonuse
- Recent pill use (reasons; brand and type, consulting the Pill Chart)
- Consistency of condom use in last 4 weeks
- Frequency of sex in past 4 weeks

Section F

- Use of medical services related to birth control and reproduction in the last 12 months (services include receipt of: birth control method; checkup or medical test related to using birth control; counseling about birth control; sterilization; counseling about getting sterilized; emergency contraception; counseling about emergency contraception; pregnancy test; abortion; Pap smear; pelvic exam; prenatal care; post-pregnancy care; testing or treatment for sexually transmitted disease (STD))
- Provider and payment information for each visit for these services in last 12 months (more detail if specific clinic is cited)
- Activation of clinic lookup if service was received at a clinic
- First service ever received is asked of women 15-24 years of age, including when received, and type of provider
- If clinic is regular source of medical care
- Ever visited a clinic

Section G

- Do you want a/another baby (respondent and partner)
- Intentions to have a/another baby
- Number of additional children (respondent/respondent and husband) intend to have

Section H

- Infertility services (help to get pregnant; help to prevent miscarriage)
- Infertility diagnoses received, if ever pursued medical help to get pregnant
- Vaginal douching
- Health problems related to childbearing (pelvic inflammatory disease; diabetes (gestational & nongestational); ovarian cysts; uterine fibroids; endometriosis; problems with ovulation or menstruation)
- Physical disabilities/limitations
- HIV testing experience (some items limited to last 12 months)
- Where HIV test was received, if within the last 12 months
- HPV vaccine-related knowledge and experience

Section I

- Health insurance coverage in last 12 months

- Current residence and residence as of April 1, 2000
- Place of birth (date came to the United States, if born outside of the United States)
- Religion and attendance at religious services, at age 14 and currently
- Work in past 12 months and current work status (respondent and husband/cohabiting partner)
- Child care arrangements used (if any) in past 4 weeks for children under 13
- Attitudes: relationships, sex, condom use, gender roles, parenthood

Section J (ACASI)

- General health, including height and weight
- Pregnancy history (numbers ending in live birth, abortion, or other outcomes)
- School suspension/expulsion (for respondents 15–24 years old)
- Substance use (cigarettes; alcohol; marijuana; cocaine; crack; crystal meth, IV drugs)
- Sexual experience with males (vaginal intercourse, oral sex, and anal sex; condom use at last occurrence of each type of sex; timing of oral sex relative to vaginal intercourse (if age 15–24 and have had both types of sex); nonvoluntary intercourse with males (asked only for age 18 or older); numbers of male partners in lifetime; numbers of male partners in last 12 months (including numbers by specific type of sex); and other HIV/STD risk behaviors)

- Sex with females, including number of female partners
- Sexual orientation and attraction
- STD experience (some items limited to last 12 months)
- Individual earnings, family income, and public assistance during the previous year

H.2 Male Questionnaire

Adapted from NSFG Cycle 7 Staff (2008, Figure 7, p. 15).

Section A

- Respondent demographic characteristics (age; DOB; marital/cohabitation status; race and Hispanic origin)
- Household roster (age; sex; relationship of each member to respondent)
- Education (degrees; highest grade completed; date last attended)
- Basic information about his childhood family background and parents
- Numbers of marriages and cohabitations

Section B

- Ever had sexual intercourse
- Sex education received (Rs aged 15–24 only)
- Sterilizing operations

- Ever had biological child(ren); how many
- Enumeration of (up to) three most recent female sexual partners, or last partner ever

Section C

- Marital and cohabitation dates for current wife/partner
- Surgical sterilization and infertility (wife/partner)
- Biological children with current wife/partner (more details if born in last 5 years)
- Other children his current wife/partner had from previous relationships (more details if he lived with the child)
- Other nonbiological children he or his current wife/partner ever cared for
- First and last sex: dates and contraceptive use
- Contraceptive use in last 12 months

Section D

- Characteristics of (up to) three sexual partners in the past 12 months or last partner ever, contraceptive use at first and most recent sex, and date of first sex in the last 12 months
- Information on children with these partners (collected as above in C)

- First intercourse ever (if not already discussed): date, contraceptive method use, and characteristics of partner

Section E

- Characteristics of former wives and first cohabiting partner

Section F

- Other biological children (information collected as above in C)
- Other nonbiological children ever raised
- Pregnancies fathered in his lifetime that did not result in live birth (total number and numbers by outcome)
- Exact number of female partners lifetime and in last 12 months

Section G

- Activities with the children living in his household
- Activities with his biological and adopted children living elsewhere
- Financial support of his biological and adopted children living elsewhere

Section H

- Desire for (wanting) a/another baby (respondent & wife/cohabiting partner)
- Intentions to have a/another baby, asked individually or jointly, as appropriate

Section I

- Usual source of health care
- Health insurance coverage in last 12 months
- Health services received in last 12 months (more details if under age 25)
- Infertility services received
- HIV testing experience

Section J

- Current residence and residence as of April 1, 2000
- Place of birth (date came to United States, if born outside of the United States)
- Religion and attendance at religious services, at age 14 and currently
- Military service
- Work status (respondent and wife/cohabiting partner)
- Attitudes: relationships, sex, condom use, gender roles, parenthood

Section K (ACASI)

- General health questions
- Significant life events
- School suspension/expulsion (for respondents 15-24 years old)

- Pregnancies fathered
- Substance use (alcohol; marijuana; cocaine; crack; crystal meth; IV drugs)
- Sexual experience with females (vaginal intercourse, oral sex, and anal sex; condom use at last occurrence of each type of sex; timing of oral sex relative to vaginal intercourse (for respondents 15-24 who have had both types of sex); nonvoluntary intercourse with females (asked only for respondents 18 or older); numbers of female partners in lifetime; numbers of female partners in last 12 months (including numbers by specific type of sex); and other HIV/STD risk behaviors)
- Sexual experience with other males; condom use at last occurrence of anal or oral sex; nonvoluntary sex with males (asked only for respondents 18 or older); HIV/STD risk behaviors, including number of male partners)
- Sexual orientation and attraction
- STD experience (some items limited to last 12 months)
- Individual earnings, family income, and public assistance during the previous year

Bibliography

- Ai, C. and E. C. Norton (2003). Interaction Terms in Logit and Probit Models. *Economics Letters* 80(1), 123 – 129.
- Alho, J. M., M. H. Mulry, K. Wurdeman, and J. Kim (1993). Estimating Heterogeneity in the Probabilities of Enumeration for Dual-System Estimation. *Journal of the American Statistical Association* 88(423), 1130–1136.
- Allison, P. D. (1999). Comparing Logit and Probit Coefficients Across Groups. *Sociological Methods & Research* 28(2), 186–208.
- Alt, C. (1991). Stichprobe und Repräsentativität. In H. Bertram (Ed.), *Die Familie in Westdeutschland. Stabilität und Wandel familialer Lebensformen*, pp. 497–531. Opladen: Leske & Budrich.
- Angrist, J. and J. Pischke (2009). *Mostly Harmless Econometrics: an Empiricist's Companion*. Princeton University Press.
- Barrett, D. F., M. Beaghen, D. Smith, and J. Burcham (2002). Census 2000 Housing Unit Coverage Study. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 146–151.
- Bethlehem, J. (2002). Weighting Nonresponse Adjustments Based on Auxilliary Information. In R. M. Groves, D. A. Dillman, J. L. Eltinge, and R. J. A. Little (Eds.), *Survey Nonresponse*, Chapter 18. Wiley-Interscience.
- Bethlehem, J. and H. Kersten (1985). On the Treatment of Nonresponse in Sample Surveys. *Journal of Official Statistics* 1(3), 287–300.
- Biemer, P. P. and L. E. Lyberg (2003). *Introduction to Survey Quality*. Wiley-Interscience.
- Boyd, H. W. and R. Westfall (1955). Interviewers as a Source of Error in Surveys. *Journal of Marketing* 19(4), 311–324.
- Boyd, H. W. and R. Westfall (1965). Interviewer Bias Revisited. *Journal of Marketing Research* 2(1), 58–63.
- Boyd, H. W. and R. Westfall (1970). Interviewer Bias Once More Revisited. *Journal of Marketing Research* 7(2), 249–253.
- Bureau of the Census (1993). Programs to Improve Coverage in the 1990 Census. Technical report. 1990 CPH-E-3.
- Campanelli, P. C., K. Thomson, N. Moon, and T. Staples (1997). The Quality of Occupational Coding in the UK. In L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, and C. Dippo (Eds.), *Survey Measurement and Process Quality*, pp. 437–457. Wiley-Interscience.

- Carroll, R. J., D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu (2006). *Measurement Error in Nonlinear Models: A Modern Perspective* (Second ed.), Volume 105 of *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC.
- Chang, T. and P. S. Kott (2004). Modeling NML Using the Area Frame Survey. In *Technical Manuscript, National Agricultural Statistics Service, United States Department of Agriculture*.
- Chhikara, R. S., F. M. Spears, C. R. Perry, and P. S. Kott (2007). Supplemental Samples for the 2007 Area Frame: A Design for Estimating Numbers of NML Farms for the 2007 Census of Agriculture. In *Research and Development Division, National Agricultural Statistics Service, United States Department of Agriculture, Research Report No. RDD-07-01*.
- Childers, D. R. (1992). The 1990 Housing Unit Coverage Study. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 506–511.
- Childers, D. R. (1993). Coverage of Housing in the 1990 Decennial Census. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 635–640.
- Clogg, C., M. Massagli, and S. Eliason (1989). Population Undercount and Social Science Research. *Social Indicators Research* 21(6), 559–598.
- Coleman, J. S. (1994). A Rational Choice Perspective on Economic Sociology. In N. J. Smelser and R. Swedberg (Eds.), *The Handbook of Economic Sociology*, pp. 166–180. Russel Sage Foundation.
- Cook, P. (1985). The Case of the Missing Victims: Gunshot Woundings in the National Crime Survey. *Journal of Quantitative Criminology* 1(1), 91–102.
- Dalenius, T. (1983). Some Reflections on the Problem of Missing Data. In W. G. Madow and I. Olkin (Eds.), *Incomplete Data in Sample Surveys*, pp. 411–413. Academic Press.
- De Young, R. (1999). Environmental Psychology. In D. E. Alexander and R. W. Fairbridge (Eds.), *Encyclopedia of Environmental Science*, pp. 223–224. Kluwer Academic Publishers.
- DeNavas-Walt, B. D. P. Carmen, and J. C. Smith (2009). Current Population Reports, P60-236, Income, Poverty, and Health Insurance Coverage in the United States: 2008. Technical report, U.S. Government Printing Office, Washington, DC.
- Dohrmann, S., D. Han, and L. Mohadjer (2006). Residential Address Lists vs. Traditional Listing: Enumerating Households and Group Quarters. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 2959–2964.

- Dohrmann, S., D. Han, and L. Mohadjer (2007). Improving Coverage of Residential Address Lists in Multistage Area Samples. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*.
- Eckman, S. and F. Kreuter (2010). Confirmation Bias in Housing Unit Listing. Under Review.
- Eyerman, J., D. Odom, and J. Chromy (2001). Impact of Computerized Screening on Selection Probabilities and Response Rates in the 1999 NHSDA. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*.
- Fan, M. C., M. L. Sutt, and J. H. Thompson (1984). Evaluation of the 1980 Census Precanvass Coverage Improvement Operations. In *Proceedings of the Section on Government Statistics, American Statistical Association*.
- Fay, R. E. (1989). An Analysis of Within Household Undercoverage in the Current Population Survey. In *Proceedings of the Bureau of the Census Annual Research Conference*, pp. 156–175.
- Fein, D. J. (1990). Racial and Ethnic Differences in U.S. Census Omission Rates. *Demography* 27(2), 285–302.
- Gelman, A. and J. Hill (2007). *Data Analysis Using Regression and Multi-level/Hierarchical Models*. Oxford University Press.
- Groves, Robert M., M. W. D., J. Lepkowski, and N. G. Kirgis (2009). Planning and Development of the Continuous National Survey of Family Growth. *Vital Health Statistics* 1(48).
- Groves, R. M. (1989). *Survey Errors and Survey Costs*. John Wiley and Sons.
- Groves, R. M. (2006). Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly* 70(5), 646–675.
- Groves, R. M., G. Benson, and W. D. Mosher (2005). Plan and Operation of Cycle 6 of the National Survey of Family Growth. *Vital Health Statistics* 1(42).
- Groves, R. M. and M. Couper (1992). *Nonresponse in Household Surveys*. John Wiley and Sons.
- Groves, R. M. and E. Peytcheva (2008). The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis. *Public Opinion Quarterly* 72(2), 167–189.
- Hansen, M. H. and J. Steinberg (1956). Control of Errors in Surveys. *Biometrics* 12(4), 462–474.
- Harter, R., S. Eckman, N. English, and C. O’Muircheartaigh (2010). Applied Sampling for Large-Scale Multi-Stage Area Probability Designs. In P. Marsden and J. Wright (Eds.), *Handbook of Survey Research* (Second ed.). Emerald.

- Hawkes, W. (1986). Census Data Quality: A User's View. *Journal of Official Statistics* 2(4), 531–544.
- Herzog, T. N., F. J. Scheuren, and W. E. Winkler (2007). *Data Quality and Record Linkage Techniques*. New York: Springer.
- Hitchcock, D. (1995). Do the fallacies have a place in the teaching of reasoning skills of critical thinking? In H. V. Hansen and R. C. Pinto (Eds.), *Fallacies: Classical and Contemporary Readings*, pp. 319–327. The Pennsylvania State University Press.
- Hosmer and Lemeshow (2000). *Applied Logistic Regression* (2nd ed.). New York: Wiley-Interscience.
- Jacobs, C. (1986). Interim Evaluation of Listing Process Audit. Unpublished memorandum to Housing Working Group, U.S. Bureau of Labor Statistics. [Cited in (Subcommittee on Survey Coverage, 1990)].
- Joncas, M. (1985). Cluster Listing Check Program for the Redesigned LFS Sample. Unpublished report, Ottawa: Statistics Canada. [Cited in (Subcommittee on Survey Coverage, 1990)].
- Kennel, T. (2007). A Coverage Profile of Area Frame Blocks on the United States Census Bureau's Master Address File. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*.
- Kennickell, A. B. (2000). Asymmetric Information, Interviewer Behavior, and Unit Nonresponse. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*.
- Kennickell, A. B. (2003). Reordering the Darkness: Application of Effort and Unit Nonresponse in the Survey of Consumer Finances. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*.
- Kish, L. (1965). *Survey Sampling*. John Wiley and Sons.
- Kish, L. and I. Hess (1958). On Noncoverage of Sample Dwellings. *Journal of the American Statistical Association* 53(282), 509–524.
- Klayman, J. (1995). Varieties of Confirmation Bias. *Psychology of Learning and Motivation* 32, 385–418.
- Kohler, U. and F. Kreuter (2005). *Data Analysis Using Stata*. Stata Press.
- Kwiat, A. (2009). Examining Blocks with Lister Error in Area Listing. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*.
- Lessler, J. T. (1980). Errors Associated with the Frame. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*.

- Lessler, J. T. and W. D. Kalsbeek (1992). *Nonsampling Error in Surveys*. John Wiley and Sons.
- Liu, X. (2008). Using A MAF-Based Frame For Demographic Household Surveys. In *Proceedings of the Section on Government Statistics, American Statistical Association*.
- Liu, X. (2009). Impact of MAF-Based Frame Coverage on Survey Estimates. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*.
- Lohr, S. L. (1999). *Sampling: Design and Analysis*. Duxbury Press.
- Long, J. S. and J. Freese (2005). *Regression Models for Categorical Dependent Variables Using Stata* (2nd ed.). Stata Press.
- Loudermilk, C. L. and M. Li (2009). A National Evaluation of Coverage for a Sampling Frame Based on the Master Address File (MAF). In *Proceedings of the Section on Survey Research Methods, American Statistical Association*.
- Lynn, P. and E. Sala (2006). Measuring Change in Employment Characteristics: The Effects of Dependent Interviewing. *International Journal of Public Opinion Research* 18(4), 500–509.
- Manheimer, D. and H. Hyman (1949). Interviewer Performance in Area Sampling. *Public Opinion Quarterly* 13(1), 83–92.
- Martin, E. (1981). A Twist on the Heisenberg Principle: Or, How Crime Affects its Measurement. *Social Indicators Research* 9(2), 197–223.
- Matschinger, H., S. Bernert, and M. C. Angermeyer (2005). An Analysis of Interviewer Effects on Screening Questions in a Computer Assisted Personal mental Health Interview. *Journal of Official Statistics* 21(4), 657–674.
- Montaquila, J. M., V. Hsu, and J. M. Brick (2010). Using a Match Rate Model to Predict Areas Where USPS-Based Address Lists May Be Used in Place of Traditional Listing. Under Review.
- Mood, C. (2010). Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It. *European Sociological Review* 26(1), 67–82.
- NSFG Cycle 7 Staff (2008). Interviewer Training for NSFG Cycle 7, June 16 - June 20, 2008. Survey Research Center, University of Michigan.
- Oh, H. and F. Scheuren (1983). Weighting Adjustment for Unit Nonresponse. *Incomplete Data in Sample Surveys* 2, 143–184.
- O’Muircheartaigh, C. A. (2004). Simple Response Variance: Estimation and Determinants. In P. Biemer, R. M. Groves, L. Lyberg, N. Mathiowetz, and S. Sudman (Eds.), *Measurement Errors in Surveys*. Wiley-Interscience.

- O'Muircheartaigh, C. A., S. A. Eckman, and C. Weiss (2003). Traditional and Enhanced Field Listing for Probability Sampling. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 2563–2567.
- O'Muircheartaigh, C. A., E. M. English, and S. A. Eckman (2007). Predicting the Relative Quality of Alternative Sampling Frames. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 551–574.
- O'Muircheartaigh, C. A., E. M. English, S. A. Eckman, H. Upchurch, E. Garcia, and J. Lepkowski (2006). Validating a Sampling Revolution: Benchmarking Address Lists against Traditional Listing. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 4189–4196.
- O'Muircheartaigh, C. and P. Campanelli (1998). The Relative Impact of Interviewer Effects and Sample Design Effects on Survey Precision. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 162(3), 63–77.
- Panel on Coverage Evaluation and Correlation Bias in the 2010 Census, National Research Council (2008). *Coverage Measurement in the 2010 Census*. National Academies Press.
- Pearson, J. M. (2003). Quality and Coverage of Listings for Area Sampling. In *Proceedings of the Section on Government Statistics, American Statistical Association*.
- Raudenbush, S. and A. Bryk (2002). *Hierarchical Linear Models* (Second ed.). Sage.
- Roberts, S. (2010). New York's Nooks are a Challenge to Census Takers. *New York Times February 23*.
- Rusch, M. L. (2008). *Relationships Between User Performance and Spatial Ability in Using Map-Based Software on Pen-Based Devices*. Ph. D. thesis, Iowa State University.
- Sampson, R. J. and S. W. Raudenbush (1999). Systematic social observation of public spaces: A new look at disorder in urban neighborhoods. *American Journal of Sociology* 105(3), 603–651.
- Sampson, R. J. and S. W. Raudenbush (2004). Seeing Disorder: Neighborhood Stigma and the Social Construction of "Broken Windows". *Social Psychology Quarterly* 67(4), 319–342.
- Sando, T., R. Mussa, J. Sobanjo, and L. Spainhour (2005). Quantification of the Accuracy of Low Priced GPS Receivers for Crash Location. *Journal of the Transportation Research Forum* 44(2), 19–32.
- Sappington, D. (1991). Incentives in Principal-Agent Relationships. *Journal of Economic Perspectives* 5(2), 45–66.

- Schnell, R., T. Bachteler, and J. Reiher (2009). Privacy-preserving Record Linkage using Bloom Filters. *BMC Medical Informatics and Decision Making* 9(1), 41.
- Schnell, R. and F. Kreuter (2005). Separating Interviewer and Sampling-Point Effects. *Journal of Official Statistics* 21(3), 389.
- Singer, E., M. R. Frankel, and M. B. Glassman (1983). The Effect of Interviewer Characteristics and Expectations on Response. *Public Opinion Quarterly* 47(1), 68–83.
- Singer, E. and L. Kohnke-Aguirre (1979). Interviewer Expectation Effects: A Replication and Extension. *Public Opinion Quarterly* 43(2), 245–260.
- Singer, E., J. van Hoewyk, and M. P. Maher (2000). Experiments with Incentives in Telephone Surveys. *Public Opinion Quarterly* 64(2), 171–188.
- Snijders, T. A. B. and R. J. Bosker (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modelling*. Sage.
- StataCorp LP (2009). *Stata statistical software: Release 11*. College Station, TX: StataCorp.
- Stiglitz, J. E. (2008). Principal and agent (i). In S. N. Durlauf and L. E. Blume (Eds.), *The New Palgrave Dictionary of Economics* (Second ed.). Palgrave Macmillan.
- Subcommittee on Survey Coverage (1990). Survey Coverage. Technical report, Federal Committee on Statistical Methodology.
- Survey Research Center (1969). *Interviewer's Manual*. Institute for Social Research, The University of Michigan.
- Survey Research Center (1976). *Interviewer's Manual* (Revised ed.). Institute for Social Research, The University of Michigan.
- Taylor, R. B., S. D. Gottfredson, and S. Brower (1984). Neighborhood Naming as an Index of Attachment to Place. 7(2), 103–125.
- Thompson, G. and C. Turmelle (2004). Classification of Address Register Coverage Rates: A Field Study. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 4477–4484.
- Turmelle, C., J.-F. Rodrigue, and G. Thompson (2005). Using the Canadian Address Register in the Labour Force Survey Implementation, Results and Lessons Learned. In *Proceedings of the Conference of the Federal Committee on Statistical Methodology*.
- United States Department of Justice. Federal Bureau of Investigation. (2009, 6). Uniform crime reporting program data [united states]: Arrests by age, sex, and race, 2007 [computer file]. icpsr25108-v1.

- U.S. Census Bureau (2001a). Census 2000 Summary File 1. Technical report.
- U.S. Census Bureau (2001b). Census 2000 Summary File 3 Technical Documentation. Technical report.
- U.S. Census Bureau (2002a). Census 2000 Summary File 3. Technical report.
- U.S. Census Bureau (2002b). Census 2000 Summary File 3 Technical Documentation. Technical report.
- U.S. Census Bureau (2006). Technical Paper 66: Design and Methodology, Current Population Survey. Technical report.
- Wolter, K. M. (1986). Some Coverage Error Models for Census Data. *Journal of the American Statistical Association* 81, 338–346.
- Wolter, K. M. (2007). *Introduction to Variance Estimation* (2nd ed.). Springer-Verlag.
- Wooldridge, J. M. (2009). *Introductory Econometrics: A Modern Approach* (fourth ed.). South-Western.
- Wright, T. and H. J. Tsao (1983). A Frame on Frames: An Annotated Bibliography. In T. Wright (Ed.), *Statistical Methods and the Improvement of Data Quality*. Academic Press.