The Nature, Consequences and Geographic Variation of Misreporting of SNAP Participation

Bruce D. Meyer and Nikolas Mittag[1]

Survey data are crucial for academic research as well as to make informed policy decisions. However, the accuracy of the information obtained from surveys crucially depends on the accuracy of survey responses. Unfortunately, several indicators of survey quality, such as the non-response rate and estimates of the extent of misreporting provide evidence that the accuracy of surveys is in decline. As discussed in Meyer, Mok and Sullivan (2015a), the extent of misreporting is particularly large for questions on government transfer programs (see Bitler, Currie and Scholz 2003 for evidence on WIC under-reporting). Recent studies have linked administrative data on SNAP and other programs to administrative records from state agencies (see e.g. Marquis and Moore 1990; Meyer and Goerge, 2011; Celhay, Meyer and Mittag 2016a or Bound, Brown and Mathiowetz, 2001, for an overview). These studies reveal high rates of underreporting, showing that sometimes more than half of true SNAP recipients do not report receipt in the survey data. Even worse, they document strong correlation patterns of survey errors with individual and household characteristics. Both analytic and empirical results underline that the large extent of misreporting and its relation to key covariates will often lead to substantial and complicated biases in common analyses such as studies of program receipt and targeting as well as program effects. These analyses are important for both academics and policy makers and improving their accuracy by better understanding the nature and consequences of misreporting as well as methods to correct for it would greatly enhance the value of survey data.

We propose to use FoodAPS to make three contributions to the literature on the nature and consequences of misreporting: First, we propose to examine the extent of misreporting, its correlation with key demographic and economic characteristics as well as variables related to survey design and the interview. This would amend similar analyses based on validation data from a few states and provide the first estimates of this kind from nationally representative data. Second, we intend to examine whether (and if so, how) misreporting varies between states. Since previous studies usually only had validation data from one or two states, understanding the extent and nature of geographic variation is important to assess whether the results from these studies are likely to provide accurate descriptions of the problem in nation-wide data and whether corrections based on these geographic subsamples are likely to improve over survey reports. Third, we propose to examine the consequences for common applications such as models of program receipt by directly estimating the bias due to misreporting in such models. These biases are rarely analytically tractable and providing empirical evidence on them is important to understand when substantive conclusions are likely to be robust to survey errors.

The FoodAPS data would provide us with a unique opportunity to study these issues, since they contain validation data for a nationally representative sample. The survey focuses and hence oversamples program recipients, which makes it particularly useful to study the most common type of survey error in this context, which is failure to report program receipt. At least for subsets of the validated states, the link to administrative records is identical, so that differences between these states can only arise from differences in reporting and not from differences in the administrative data or the record linkage, making it ideal to study variation in reporting between states. The documentation suggests that the link to administrative data is of high quality, which justifies using the administrative receipt variable as a measure of truth to compare the survey responses to.[2] Throughout, we would focus

[1] Neither Meyer nor Mittag has any potential conflicts of interest to disclose.

[2] Nonetheless, a first step of the proposed research would be to analyze the reliability of the administrative measure and to address potential problems using methods we have used in previous work (e.g. Meyer and Goerge, 2011; Celhay, Meyer and Mittag 2016a). For example, we would analyze the match to the administrative records and examine whether it can be corrected for by methods such as inverse probability weighting.

on current SNAP receipt as it provides the cleanest measure from the administrative data. However, we may repeat some of the analyses using receipt in the last 12 months, since this measure is closer to the definition in other surveys.

1. The Extent and Nature of Misreporting

As a first step, we would examine the extent and nature of misreporting in the FoodAPS data, which provides a unique opportunity to do so in a nationally representative data set. We propose to conduct analyses similar to those in Meyer and Goerge (2011) and Celhay, Meyer and Mittag (2016a). In particular, we would first compute rates of false negatives (the fraction of true recipient households that do not report receipt) and false positives (true non-recipients reporting receipt) in the overall sample. These rates provide a good first summary of the accuracy of the data. Previous studies have found high rates of false negatives and low rates of false positives using validation data from New York (Celhay, Meyer and Mittag, 2016a) as well as Illinois and Maryland (Meyer and Goerge, 2011). Comparing the FoodAPS rates for the entire US to these state-level studies can only provide limited evidence on the representativeness of these state-level studies since surveys and definitions of receipt differ, so we propose to examine this by other means below. Nonetheless, comparing them is of interest from a survey design perspective as the differences may provide further evidence on the theories of misreporting discussed in, among others, Bound, Brown and Mathiowetz (2001) and Celhay, Meyer and Mittag (2016b). For example, common conjectures are that the difficulty of information retrieval or that the stigma of reporting receipt prevent accurate reporting. The low false negative rate given in the FoodAPS User Guide suggests that the focus of the survey or the motivation of interviewers may be more important.

Second, we propose to examine the nature of misreporting, in particular its relation to variables that are commonly used to study program receipt and its effects. To do so, we would estimate Probit models for both types of error. That is, to examine the correlates of false negatives, we would estimate a Probit model among true recipient households where the dependent variable is whether SNAP receipt was reported or not. To examine false positives, we would estimate a Probit model with the same dependent variable among true non-recipients. We would use the detailed information on demographic (e.g. household composition, age, race, education) and economic (e.g. income, employment status, receipt of other programs) characteristics of the household (or the household head) that FoodAPS collected as covariates. In addition, the FoodAPS data contain an unusual amount of interview related variables, such as the language of the interview, several variables recording respondent difficulties (e.g. due to memory issues or reading difficulties), assistance by other household members and even an interviewer identifier. Including these variables in our misreporting models would provide further evidence on the survey-design related determinants of misreporting examined among others in Celhay, Meyer and Mittag (2016b).

A better understanding of which variables are correlated with reporting errors is crucial to assess likely biases in similar data and to address them. For example, Meyer and Mittag (2014) show that corrections of binary choice models with misclassification in the dependent variable work well when estimates of these models of misreporting are available. Furthermore, it would provide evidence on previous finding, such as that program reporting decreases with income and whether these patterns found in household surveys extends to surveys with a different focus. This is important to assess whether these relations are an artefact of a specific survey design or more general. Obtaining further information on how robust these relationships are is important for researchers to assess whether the correlations (and consequently the biases) are likely to be similar in data sets that have not been validated previously.

2. Assessing Geographic Variation in Misreporting

An important limitation of previous validation studies that examined misreporting of government transfers is that they are usually only able to link survey data to administrative data for one or two states. This raises the important questions of whether the findings in these studies would be similar if they had validated data for all states and whether one can use data from one or a few states to correct estimates in data on the entire US. FoodAPS contains a validated measure for the same question for multiple states, which would allow us to test both issues.[3] Both the survey and the question asked differ from the data used in previous validation studies, so the results from these studies for a specific state and the results obtained from FoodAPS for the same state are likely to differ for many reasons. Nonetheless, testing whether reporting in the FoodAPS data differs by state would provide novel and important evidence on the extent of geographic variation in misreporting. While misreporting may well differ between surveys, finding substantial geographic variation of misreporting in FoodAPS would cast doubt on the external validity of previous studies and our ability to correct estimates based on geographic subsamples.

To examine the extent of geographic variation, we would first estimate false positive and false negative rates by state and analyze the amount of variation in them. We would conduct analyses similar to Harris (2014), who analyzes county-level misreporting rates, to provide a descriptive summary of the variation in unconditional misreporting rates. We would particularly focus on comparing states that have previously been validated (e.g. Colorado, Illinois, New York, Maryland, Texas) to assess whether the extent of misreporting found in these states is likely to over- or understate the problem at the national level. Harris (2014) finds substantial variation in misreporting rates at the county level, suggesting that unconditional misreporting rates are unlikely to be constant across states. However, most analyses and corrections only assume misreporting to be constant conditional on the covariates used rather than unconditionally. To examine this issue further, we would re-estimate the Probit models for false negatives and false positives using state fixed and random effects models in addition to the covariates discussed above. We would test whether the level of misreporting differs between states after conditioning on these covariates by testing for the joint significance of these state-specific intercepts. In addition, we can examine the substantive importance of this residual variation in misreporting levels by examining the share of the variation in misreporting rates explained by the random effects. Depending on the sample size, we may also interact key covariates such as age and income with state dummies to assess whether the relation of misreporting to important covariates varies between states. While heterogeneity in levels can potentially be addressed by incorporating aggregate information as discussed in Mittag (2013), heterogeneity in these slopes would be particularly problematic for corrections based on some states.

Besides providing information on the extent of geographic variation and thereby evidence on the external validity of previous validation studies, the analyses above test the validity of assumptions that are commonly made when correcting for underreporting based on one state. In addition, we would also use the FoodAPS data to examine whether these corrections work well directly. Even if the assumptions are violated, a crucial question is whether such corrections still improve over the survey data. To do so, we would implement imputation methods based on parametric models such as those proposed in Davern et al. (2009), Mittag (2013) and Scholz et al. (2009) and compare the resulting receipt rates and the FoodAPS reports to the administrative rates we consider “truth”. Ideally, we would use the data to assess whether one can extrapolate from one state that has been validated (e.g. New York) to the

[3] The validated data is likely to be more reliable for some states as discussed in section 2.3.4 of the codebook. Depending on the available sample sizes, it may be preferable to conduct these tests within groups of states for which the same administrative data was available and the record linkage was the same.

remainder of the US, since that is the problem usually encountered in practice. However, this is unlikely to be feasible given the sample size of FoodAPS. Two potential alternatives would be to extrapolate from the set of states that have recently been validated or to split the sample by geography.

3. Consequences of Misreporting

Several studies document that misclassification of binary variables can lead to complicated biases in multivariate models (e.g. Carroll et al. 2006; Meyer and Mittag 2014) and that these biases can severely affect substantive conclusions (e.g. Gundersen and Kreider 2008; Haider, Jacknowitz and Schoeni 2003; Meyer and Mittag 2015). While these biases are only tractable in specific cases, the FoodAPS data would allow us to examine how misreporting of government programs affects common analyses, such as models of program receipt, directly. This would shed light on the substantive importance of misclassification in survey data. In addition, there are few formal results for bias due to measurement error, particularly in non-linear models and when measurement error is not classical. This makes it difficult to assess whether substantive conclusions are likely to be robust to measurement error, since one can only appeal to stylized facts from these special cases. For example, it is frequently conjectured that estimates are likely to be attenuated due to measurement error, even though this result only applies to special cases (e.g. univariate linear models with classical measurement error, but not binary choice models). FoodAPS would allow us to assess whether the actual results conform to these patterns even when the conditions under which they are known to hold do not apply.

In particular, we propose to examine these biases by estimating these models twice, once using the administrative SNAP receipt variable and once using the reported variable. The difference between the estimated coefficients (or marginal effects) provides an estimate of the bias due to misreporting. We could conduct tests of parameter equality as in Meyer and Goerge (2011) and compute measures of the overall bias as in Meyer and Mittag (2014). We would primarily focus on the bias in models of SNAP receipt as in Meyer and Goerge (2011), where the dependent variable in a Probit model is an indicator for current receipt. Such analyses are crucial for program targeting and for evaluating efforts to increase program participation. If possible, we would like to extend these analyses to models where SNAP receipt is an independent variable, such as estimating the effect of SNAP receipt on food security or expenditures in certain food categories. However, SNAP receipt is endogenous in these analyses (see Bitler 2014), so whether they yield meaningful results depend on the validity and strength of potential instrumental variables. Variation in state policies has been used as an instrument for SNAP receipt in such analyses (see e.g. Ratcliffe, McKernan and Zhang, 2011; Schmidt, Shore-Sheppard and Watson, 2013). However, while there is still substantial variation in asset limits, state policies are far more homogeneous in the time period covered by FoodAPS, so it is unclear whether these instruments would be powerful enough to yield meaningful results, particularly since we are interested in the difference between two sets of estimates (from models using reported and administrative SNAP receipt). Consequently, it would be interesting to examine these models in addition to Probits for program receipt, but whether it is feasible depends on factors we can only assess with access to the data.

References

Bitler, M. The Health and Nutrition Effects of SNAP: Selection Into the Program and a Review of the Literature on Its Effects. University of Kentucky Center for Poverty Research Discussion Paper Series, DP2014-02.

Bitler, M.P., Currie, J., & Scholz, J.K.. (2003). WIC eligibility and participation. *Journal of Human Resources.* 38(S): 1139-1179.

Bound, J., C. Brown, and N. Mathiowetz. 2001. “Measurement error in survey data." In *Handbook of Econometrics*. Vol. 5, , ed. J. J. Heckman and E. Leamer, Cp. 59, 3705-3843. Amsterdam:Elsevier.

Carroll, Raymond J., David Ruppert, Leonard A. Stefanski, and Ciprian M. Crainiceanu. 2006. *Measurement Error in Nonlinear Models: A Modern Perspective. 2nd ed.,* Boca Raton: Chapman & Hall.

Celhay, P., B. Meyer, and N. Mittag. 2016a. "Measurement Error in Program Participation". Unpublished Manuscript.

Celhay, P., B. Meyer, and N. Mittag. 2016b. "What Leads to Errors in Surveys? Evidence from Multiple Government Programs". Unpublished Manuscript.

Davern. M., J. Klerman, J. Ziegenfuss, V. Lynch, and G. Greenberg. 2009. "A Partially Corrected Estimate of Medicaid Enrollment and Uninsurance: Results from an Imputational Model Developed off Linked Survey and Administrative Data." *Journal of Economic and Social Measurement.* 34(4):219-240.

Gundersen, C., and B. Kreider. 2008. "Food Stamps and Food Insecurity What Can Be Learned in the Presence of Nonclassical Measurement Error?" Journal of Human Resources, 43(2): 352-382.

Haider, S. J., A. Jacknowitz, and R. F. Schoeni. 2003. "Food Stamps and the Elderly: Why Is Participation so Low?" *The Journal of Human Resources*, 38: 1080-1111.

Harris, B. C. 2014. "Within and Across County Variation in SNAP Misreporting: Evidence from Linked ACS and Administrative Records". *CARRA Working Paper #2014-05.* U.S. Census Bureau.

Marquis, K.H, and J.C Moore. 1990. "Measurement Errors in SIPP Program Reports." U.S. Census Bureau.

Meyer, B. D. and Goerge, R. 2011. Errors in survey reporting and imputation and their effects on estimates of food stamp program participation. *US Census Bureau CES-WP-11-14*.

Meyer, B.D. and Mittag, N. 2014. "Misclassification in Binary Choice Models". *NBER Working Paper 20509*.

Meyer, B. D. and N. Mittag. 2015. "Using Linked Survey and Administrative Data to Better Measure Income: Implications for Poverty, Program Effectiveness and Holes in the Safety Net," NBER Working Paper 21676.

Meyer, B.D., Mok, W.K.C. and Sullivan, J.X. 2015a. Household Surveys in Crisis. *Journal of Economic Perspectives*, 29(4), 199-226.

Meyer, B.D., Mok, W.K.C. and Sullivan, J.X. 2015b. "The Under-Reporting of Transfers in Household Surveys: Its Nature and Consequences". *Harris School of Public Policy Studies Working Paper.*

Mittag, N.. 2013. "A Method of Correcting for Misreporting Applied to the Food Stamp Program." Harris School of Public Policy, University of Chicago. Chicago IL.

Ratcliffe, C., McKernan, S. M., & Zhang, S. 2011. „How much does the Supplemental Nutrition Assistance Program reduce food insecurity?" *American Journal of Agricultural Economics*, March, 1-16.

Schmidt, L., Shore-Sheppard, L., and Watson, T. 2013. "The effect of safety net programs on food insecurity". NBER Working Paper 19558.

Scholz, J.K., R. Moffitt and B. Cowan. 2009. "Trends in Income Support." In Changing Poverty, Changing Policies, eds. M. Cancian and S. Danziger. New York: Russell Sage Foundation.