Applied Economic Perspectives and Policy (2014) volume 36, number 3, pp. 416–437.
doi:10.1093/aepp/ppu009
Advance Access publication on 21 May 2014.

Submitted Article

Food Insufficiency and Income Volatility in US Households: The Effects of Imputed Income in the Survey of Income and Program Participation

Molly Dahl, Thomas DeLeire, and Shannon Mok*

Molly Dahl, Congressional Budget Office. Thomas DeLeire Georgetown University, IZA, and NBER. Shannon Mok, Congressional Budget Office.
*Correspondence may be sent to: shannon.mok@cbo.gov

Submitted 24 August 2012; accepted 5 March 2014.

Abstract *In this article, we explore how using imputed income data in the Survey of Income and Program Participation affects the observed relationship between household income volatility and food insufficiency. We find that measuring income volatility using imputed income data substantially understates the association between large drops in household income and food insufficiency. After excluding observations with imputed income, large drops in income are associated with a 2.1–percentage point increased probability of food insufficiency, or a 31% higher likelihood of food insufficiency.*

Key words: Data imputation, food insufficiency.

JEL codes: C18, C81, Q18.

Several studies (among others, Leete and Bania 2010; Gundersen and Gruber 2001) have used household surveys to document the prevalence of income volatility and the relationship between income volatility and food insufficiency. However, data imputation can produce misleading measures of income volatility, which, in turn, can produce misleading relationships between that volatility and food insufficiency. This poses challenges to policymakers who rely on empirical research to evaluate policies intended to reduce income volatility or understand the relationship between income volatility and other outcomes of interest.

The United States has a number of policies in place whose purpose is either explicitly or implicitly to mitigate the deleterious effects of swings in family income. Programs such as Unemployment Insurance and policies such as the progressive tax code reduce income volatility by mitigating the impacts of changes in earnings on after-tax income (Gruber 1997). Other programs, such as the Supplemental Nutrition Assistance Program (SNAP, formerly known as food stamps), don't mitigate the swings in cash income per se but rather the effects of those swings on the well-being of recipients (Blundell and Pistaferri 2003; Gundersen and Ziliak 2003). Knowledge of the extent to which programs improve outcomes, such as food sufficiency, for low-income people requires knowledge of both how programs reduce income volatility and how income volatility affects food sufficiency. In this article, we assess how data imputation may bias estimates of this latter effect.

The Survey of Income and Program Participation (SIPP), a widely used data source for analyzing income changes over time, frequently replaces missing values of income with randomly selected values from other records in the SIPP that are observationally similar and complete—a method known as hot-deck imputation. When income volatility is calculated using hot-deck imputed data, income changes are no longer based on changes within a household, but rather on differences in incomes between two households. Because income variability between households is substantially greater than income variability within households over time, using imputed data to estimate volatility leads to an overestimate of the amount of volatility. Although our analysis focuses on imputations in the SIPP, other household surveys commonly used in applied economics research contain hot-deck imputations as well.

In this article, we consider the effect of using imputed income data when examining the relationship between income volatility and an outcome of interest—in this case, the probability that a household experiences food insufficiency. We find that income volatility is a factor in food insecurity; including imputed income data when estimating that relationship tends to understate the relationship.

Background

Labor economists have approached the use of imputed data in several ways. Most studies use imputed data when they are available. However, many studies, perhaps beginning with Lillard, Smith, and Welch (1986), question the use of hot-decked data in empirical analyses. For example, in studies of the wage distribution using the Current Population Survey (CPS), DiNardo, Fortin, and Lemieux (1996) and Autor, Katz, and Kearney (2008) drop observations with imputed wage data. Other studies that drop observations with imputed earnings or income in their cross-sectional analyses include those of Hirsch and Schumacher (2004), Mellow and Sider (1983), Dooley and Gottschalk (1984), Welch (1979), and Bollinger and Hirsch (2006).

Studies that use panel data to conduct longitudinal analyses also often exclude imputed observations. For example, Bound and Krueger (1991), in their influential study comparing measurement error in survey data with administrative data from the Social Security Administration, drop observations with imputed earnings data. Their finding that "longitudinal [survey] earnings data may be more reliable than previously believed"[1] is based only on nonimputed data. Additional panel studies that drop imputed earnings data include Kim and Solon (2005), Bound and colleagues (1994), and Bollinger (1998).

Several previous studies have examined the link between income volatility and food insufficiency.[2] Using the 1991 and 1992 panels of the SIPP, Gundersen and Gruber (2001) showed that food-insufficient households had higher income variability, were more likely to have experienced income

[1] *Bound and Krueger (1991), p. 1.*

[2] *Food insufficiency is generally defined as sometimes or often not having enough to eat.*

shocks (such as loss of earnings or food stamps), and were less likely to have savings. Others have examined the effects of resources that mitigate the effects of income volatility. Using the Panel Study of Income Dynamics, Blundell and Pistaferri (2003) found that food assistance programs reduced but did not eliminate the effect of permanent income shocks on food expenditures at home between 1978 and 1992. Ribar and Hamrick (2003) found that the ownership of assets (which may be negatively related to income volatility) is negatively related to food insufficiency.

More recently, Leete and Bania (2010) found that households with relatively volatile incomes have an increased likelihood of experiencing food insufficiency. Their measure of income volatility is the level of difference between monthly income and average income over the year. They point out that, if income volatility is due to measurement error, then coefficient estimates will be biased toward zero and thus will show no effect of income volatility on food insufficiency. In sensitivity tests, Leete and Bania (2010) show that excluding observations with imputed income results in larger estimates of the association between income drops and food insufficiency.

Our analysis extends previous work by more explicitly considering how income imputations in the SIPP affect the relationship between income volatility and food insufficiency. First, we document that large income changes are more likely to be observed among households that have imputed income even though these households are observationally similar to households without imputations. Second, our measures of household income and income volatility differ from those used in previous studies to focus attention on income volatility. We measure income volatility using the arc percentage change in household income over two four-month periods. Measurement of volatility in percentage terms allows households at different income levels to have different responses to the same difference in income amounts. Also, people moving from zero to positive income and those moving from positive to zero income are included symmetrically. In addition, our measure of household income does not include the value of food assistance—for instance, food stamps or Special Supplemental Nutrition Program for Women, Infant, and Children (WIC) benefits—whereas others' measure of income does often include the value of food assistance. Researchers have found that households tend to receive food assistance when they are most food insecure[3] and higher rates of food insecurity among food assistance recipients can be attributed to self-selection among food assistance recipients (Nord and Golla 2009; Mykerezi and Mills 2010; Ratcliffe, McKernan, and Zhang 2011) or reporting errors in measures of food insecurity and food assistance (Gundersen and Kreider 2008). The inclusion of the value of food assistance in household income does not allow distinction between the association of income volatility with food insufficiency and the association of food assistance receipt with food insufficiency.

Data and Methods

Data Source

The SIPP comprises a set of panel surveys that were conducted annually from 1984 to 1988, from 1990 to 1993, and then again in 1996, 2001, 2004, and

[3]Food security measures access by all people at all times to enough food for an active, healthy life and is related to food insufficiency.

2008. In each panel, interviews are conducted at four-month intervals (known as a wave). Some information (including income, program participation, and household composition) is collected in each interview; other information (including measures of well-being) is collected less frequently. The SIPP was significantly redesigned in 1996, and subsequent panels use the same imputation procedures. In this study, we use the 1996, 2001, 2004, and 2008 panels. The sample sizes for these panels range from 40,000 to 52,000 households, and the panels range in duration from nine to twelve interviews (about three years to four years).

At each interview, the survey collects information on the labor market earnings for each household member and all other sources of cash income for the household in each of the previous four months, as well as a comprehensive set of demographic information. Nonlabor income includes income from a wide array of possible sources, including unemployment insurance, welfare payments, retirement income (which includes Social Security, railroad retirement, and pension income), disability and Supplemental Security Income payments, and interest and dividends. We construct a wave-based measure of total (pretax) household income, which is the sum of earnings and nonlabor income for each household member in the four-month period referenced in the interview. (For more details, see the appendix.) We do not include the value of food assistance benefits such as food stamps and WIC.

We examine volatility in several measures of income because it is not clear, a priori, how changes in income translate into changes in economic well-being. We first construct the arc percentage change (APC) in total household income (Y) between two waves for each household[4]:

$$\text{APC in income} = 100 \times (Y_t - Y_{t-1})/((Y_t + Y_{t-1})/2) \tag{1}$$

The APC is symmetric with respect to the measures of income in the two waves and is defined even when income is zero in either period (but is not defined when income is zero in both periods, and we drop observations with zero income in both waves). For the remainder of the article for expositional purposes, we will often refer to the APC as the percentage change.

An alternative method of measuring income volatility accounts for the level of income change instead of the percentage change.[5] Two households with different levels of income can have the same percentage change in income but differ in their likelihood of experiencing food insufficiency. We measure income change relative to the federal poverty guideline (FPL) as

$$100 \times (Y_t - Y_{t-1})/((FPL_t + FPL_{t-1})/2). \tag{2}$$

We impose several sample restrictions. We exclude households headed by individuals younger than 25 years or older than 55 years at the time of the survey because their incomes may be especially volatile as they enter or leave the labor market. We also exclude households with incomes in the top 1% or bottom 1% in any year to reduce the effect of outliers. In the analysis of the association between income volatility and food insufficiency, we restrict the sample to households with incomes less than 200% of the FPL

[4]*See Allen and Lerner (1934), pp. 226–9, or Hensher, Rose, and Greene (2005), p. 392. This measure is identical to a growth rate used in the literature measuring patterns in employment growth (e.g., Davis, Haltiwanger, and Schuh 1996; Tornqvist, Vartia, and Vartia 1985).*

[5]*We are grateful to the editor and to an anonymous reviewer for this suggestion.*

because it is unlikely that income changes are associated with food insufficiency in higher-income households.[6]

Our main outcome variable of interest is food insufficiency, which is measured consistently across SIPP panels. Starting in 1991, the SIPP includes questions about food insufficiency once in each panel. We base our food insufficiency measure on respondents' answers to the following question: "Getting enough food can also be a problem for some people. Which of these statements best describes the food eaten in your household in the last four months?" We code households as being food insufficient if they respond that there is "sometimes not enough to eat" or "often not enough to eat."[7] There is another question regarding food insufficiency in the previous month. In this article, we present results using both measures of food insufficiency.

Because we are interested in the association between income change and an outcome, we examine income changes occurring immediately before the measurement of food insufficiency. In each panel, we calculate the percentage change in the measures of income between the wave before food insufficiency is measured and the wave in which food insufficiency is measured. This enables the length of time over which we observe income changes to be the same for all observations.

Imputed Data

The Census uses a variety of methods to impute missing data (US Census Bureau 2001, chapters 4 and 6). The most common method is a hot-deck imputation method. The hot-deck imputation replaces missing values with randomly selected values from complete records that are observationally similar based on a small number of variables in the same dataset. (The exact set of variables used to create the hot-deck cells that define similar records varies depending on what is being imputed.) Other widely used datasets, including the CPS and the American Community Survey, also use this method to impute missing values.

Before 1996, the SIPP data files released by wave relied mainly on cross-sectional imputation; only current wave characteristics were used to match complete records (known as donors) to records with missing values (known as recipients). Beginning with the 1996 panel, information from the previous wave for donors and recipients was used to define the hot-deck cells. The values used to replace missing values, however, only come from complete respondents in the current wave. For new sample members who do not have a previous wave interview, the hot-deck would rely solely on characteristics reported in the current wave.

The SIPP details which imputation method was used to replace missing values of a variable for an observation. These flags indicate whether the imputation was hot-decked and whether the data used to match recipients to donors used current wave or previous wave data. Both hot-deck methods,

[6]Researchers have also found food insecurity among higher-income households (Coleman-Jensen et al. 2011). We estimate similar effects (but with more precision) when households of all income levels are included.

[7]In our sample, approximately 8% of responses to the food insufficiency question were imputed. Observations with imputed food insufficiency are kept in the main analysis, and restricting the analysis to observations without food insufficiency and without imputed income results in similar estimates. Observations with imputed income are slightly more likely to have imputed food insufficiency too.

however, only replace missing values with values from the current wave. In a small number of cases, missing values were not imputed using a hot-deck but rather derived using other information from the respondent's own interview. In those cases, it is likely that income volatility calculated from imputed data is underestimated.

We classify households as having imputed income if any household member has any source of income imputed in any of the eight months included in calculating percentage change.[8] Although any missing income amount results in the observation being labeled as imputed, among households with imputed income, an average of 60% of their income was imputed. Most of the income imputations involve the use of previous wave values.

In most cases, imputation of missing data can result in improved estimates of the cross-sectional means and variances (see Rubin 1987). However, the use of hot-deck imputed data can be problematic when constructing measures based on the change in the potentially imputed variable over multiple periods.[9] Observed changes are not "real" in that they are not calculated from differences in reported values over time for a given observation, but rather are calculated from differences in values across observations. For example, consider an observation in which the respondent provided income data, Y_1, in wave 1 but not in wave 2 (thus Y_2 was missing and was imputed using a hot-deck imputation). The measure of APC for that observation, $100 \times (Y_2 - Y_1)/((Y_2 + Y_1)/2)$, is based on the difference between the observation's actual income in wave 1 and some other observation's income in wave 2. This measure using imputed data is closely related to an observation's percentage deviation from the sample average—a measure of cross-sectional variability. Because it is likely that cross-sectional income variability (the variability of income between households) is substantially greater than the variability in income within households over time, using hot-deck imputed data to estimate the percentage change in income likely leads to an overestimate of the amount of variability.

Methods

First, we document the increased likelihood of observing a large income change among all households, including those with imputed income. We estimate a logistic model relating the probability that we observe a large change in income for a household, which we define as an increase or decrease in income of 25% or more (V), to demographic and employment characteristics. Specifically, we estimate:

$$\ln\left(\frac{V_{i,t+1}}{1 - V_{i,t+1}}\right) = \alpha Impute_i + \beta Z_{i,t} + \varepsilon_{i,t} \tag{3}$$

where $Impute_i$ is an indicator for imputed income in either wave and where Z_i is a set of household characteristics, including age of household reference person, number of children, education, race, and household composition, all measured at the end of the first wave. We also include the average of household income relative to FPL over three waves as a control.

[8] *For this article, we do not consider missing values that arise when no interviews are collected from a household. In general, these missing interviews are accounted for by using weights.*

[9] *For example, see the concerns raised over the use of imputed data in the SIPP in Williams (1992).*

Next, we see how imputed income affects the observed relationship tween income volatility and food insufficiency in households with incomes less than 200% of FPL. Controlling for income and other household characteristics, we estimate logistic regressions relating the probability that a household experiences food insufficiency to the household's income volatility. In particular, we estimate the following relationship:

$$ln\left(\frac{I_{i,t+1}}{1-I_{i,t+1}}\right) = BX_{i,t} + \gamma \cdot income\ volatility + e_{i,t+1} \qquad (4)$$

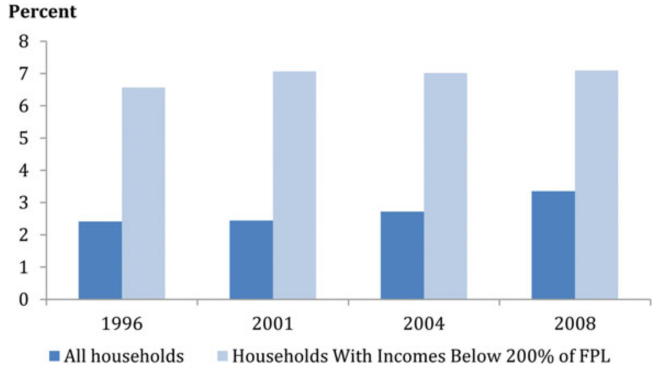
where I_i is the indicator for household food insufficiency, X_i includes dummies for the education level of the household head (less than high school, high school, some college, and college or more [excluded]), marital status, the number of children in the household, the panel, and household income relative to FPL. We take the average of household income relative to FPL over three waves (the wave food insufficiency is measured in and the two preceding waves). All specifications are estimated using SIPP person-level sample weights for the reference month so that estimates are nationally representative.[10]

Income volatility is specified in several ways. We measure the APC in income as specified in equation (1). In our preferred specifications, we include indicators for a drop in income exceeding 25% or a rise in income exceeding 25%. Large increases and large decreases in income are included separately in the regression because large income drops (but not large income rises) are likely to be associated with food insufficiency. We check whether the results depend on our choice of 25% as a threshold in several ways. First, we include the percentage change directly. Second, we define the change in income relative to FPL, as specified in equation (2), and we define a large drop or rise as a change in income relative to FPL exceeding 10%. Finally, income volatility in either direction—rises or drops—could be associated with food insufficiency because it indicates that income is uncertain. In our final specification, we include the variance in income over the three waves ending with the measurement of food insufficiency. Both large drops and rises in income would result in high variance.

We determine whether this relationship changes when we drop observations with imputed income. Because the effects of imputed income may depend on the imputation method used (for example, income volatility is more likely to be overestimated when missing income values are hot-decked than logically derived), we drop observations with income imputed using any method and then separately by each imputation method (hot-deck using current wave data only, logically derived imputation, and hot-deck using previous wave data). On the one hand, dropping observations with imputed income reduces the bias in our estimates of the relationship between food insecurity and income volatility. On the other hand, dropping observations from the data calls into question whether the remaining sample is nationally representative.

[10]Specifically, we use the "logit" command in STATA 13.0. All results are quite similar if we either estimate unweighted results or use SIPP replicate weights to account for the complex survey design of the SIPP. We prefer the simpler weighted results only because our analysis combines SIPP panels and it is not obvious that the use of replicate weights across panels with different numbers of replicate weights yields valid standard errors (personal communication with the US Census Bureau).

Figure 1 Households reporting food insufficiency over previous four months



Source: Authors' calculations from the 1996, 2001, 2004, and 2008 panels of the Survey of Income and Program Participation.
Note: FPL = federal poverty guideline.

Results

Descriptive Statistics

The rates of food insufficiency for each panel of the SIPP are shown in figure 1. Over a four-month period in the 1996 panel, 2.4% of all households reported food insufficiency. This percentage increased to 2.7% in the 2004 panel and to 3.3% in the 2008 panel. Food insufficiency rates were substantially higher among lower-income households. The share of lower-income households reporting food insufficiency rose slightly from 6.6% in 1996 to 7.1% in 2008.

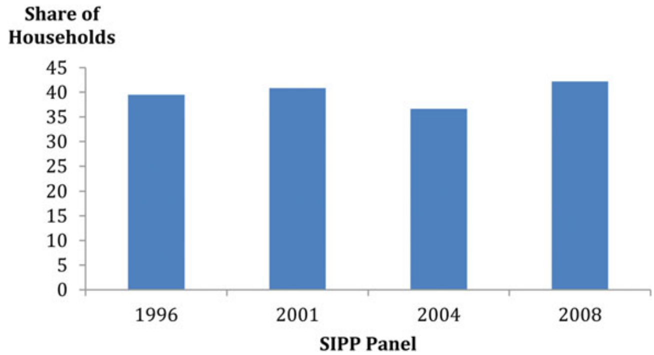
A substantial share of households in the SIPP have imputed income (see figure 2), which potentially complicates any analysis of the effect of income changes on an outcome of interest.[11] In the 1996 panel, 40% of household records had imputed income in the wave food insufficiency was measured or the previous wave. This dipped to 37% in the 2004 panel and grew to 42% in the 2008 panel. Using a longer time period (for example, using three waves of data to calculate annual income) would result in higher estimates of the share of households with imputed income.

Households with imputed income appear similar to households with reported income (see table 1). Both types of households have heads of households of similar ages, have similar educational attainments, and have comparable numbers of children. Households with imputed income have relatively lower household income (on average, 363% of FPL compared with 377% of FPL among households without imputed income). However, households with imputed data are significantly more likely than other households to have changes in income (increases or decreases) exceeding 25%. Of course, these measured changes are likely not real.

The percentage of households experiencing large changes in total household income (income drops or increases greater than or equal to 25%) is displayed in figure 3. Including all observations, the share of households that experienced changes in income between waves that exceeded 25% ranges

[11] *Because of item nonresponse in recent years, other household surveys, including the Current Population Survey, have also experienced a rise in imputation rates among income variables. See Bollinger and Hirsch (2013), Czajka (2009), and Meyer, Mok, and Sullivan (2009).*

Figure 2 Income imputation over a two-wave period



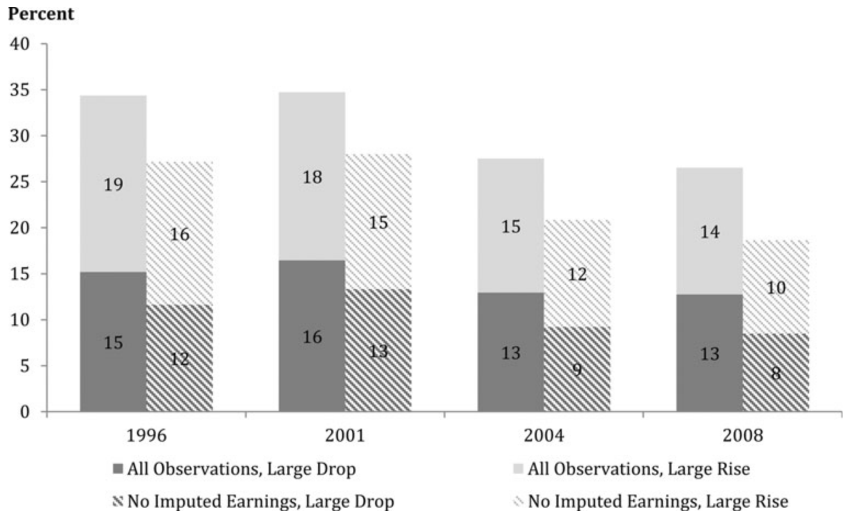
Source: Authors' calculations from the 1996, 2001, 2004, and 2008 panels of the Survey of Income and Program Participation.

Table 1 Characteristics of Households in the Sample

	All observations	Not imputed	Imputed
Percent of households unless indicated otherwise			
Household income as percent of federal poverty guideline (FPL)*	371	377	363
Number of children under age 17 years*	1.0	1.0	1.1
Less than high school*	10	9	12
High school graduate*	25	24	27
Some college	35	35	35
College degree*	30	32	26
Non-Hispanic white*	69	71	67
Non-Hispanic black*	13	11	15
Hispanic*	13	13	13
Other	5	5	5
Married with children	39	39	39
Married without children*	18	16	21
Single with children*	15	14	16
Single without children*	28	31	25
Male sex*	52	53	51
Food insufficient in last month*	2	1	2
Food insufficient in any month*	3	2	3
Income drop $\geq$25%*	14	11	20
Income rise $\geq$25%*	16	13	22
Income drop relative to FPL $\geq$10%*	23	19	30
Income rise relative to FPL $\geq$10%*	29	25	33
Percentage change in income*	2	2	1
Variance of income, in thousands*	32	23	45
Age, in years*	41	40	42
Number of households	55,527	33,018	22,509

Source: Authors' calculations from the 1996, 2001, 2004, and 2008 panels of the Survey of Income and Program Participation.
Notes: Data are percentage of households unless indicated otherwise. Characteristic as measured in last wave. Weighted by person weights.
**Difference is significant at the 1% level.*

Figure 3 Share of households experiencing a 25% or greater change in income



Source: Authors' calculations from the 1996, 2001, 2004, and 2008 panels of the Survey of Income and Program Participation.
Note: The category "All Observations" includes observations with reported income and observations with imputed income.

from 35% in the 2001 panel to 27% in the 2008 panel. Households that experienced large changes in income were slightly more likely to experience large increases rather than large drops. When we exclude households with imputed income, the share of households that experienced large changes in income is about 7 percentage points lower in each panel and exhibits the same downward trend.

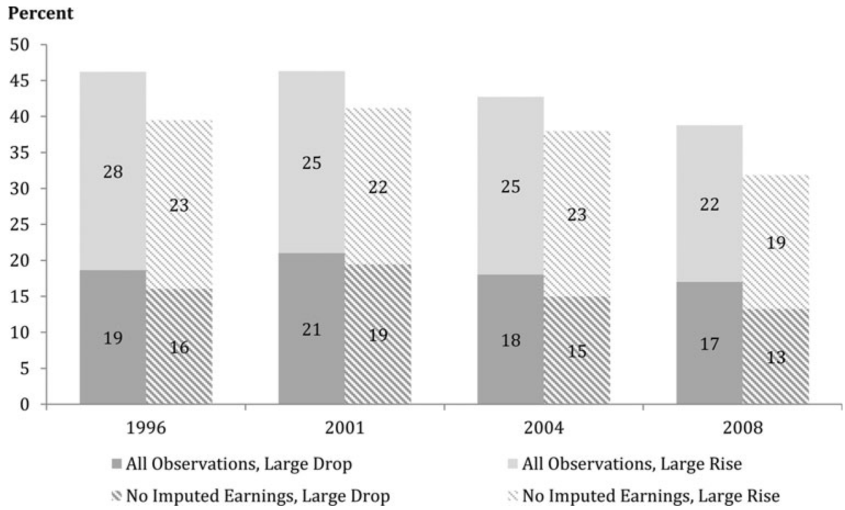
A similar pattern is observed when we restrict the sample to households at less than 200% of the poverty line (see figure 4). Large changes in income are more common among low-income households than all households. The share of lower-income households experiencing large income changes also declines over time and is lower when observations with imputed income are dropped.

These descriptive results suggest that imputation in the SIPP likely leads to a large number of households being incorrectly identified as having volatile incomes. This misclassification likely leads to an understatement of the association between rates of food insufficiency and income volatility. In the next section, we test those propositions.

Association of Income Volatility and Food Insufficiency

We confirm that, according to the SIPP, households with imputed income are more likely to experience large changes in income (see table 2). After controlling for demographic and income characteristics, households with imputed income have an 18–percentage point higher probability of experiencing a large percentage change in income. Households with lower incomes relative to FPL and households headed by younger workers and by those who are single are more likely to experience large changes in income, which is consistent with other findings (Congressional Budget Office 2008).

Figure 4 Households below 200% of the poverty level experiencing a 25% or greater change in income



Source: Authors' calculations from the 1996, 2001, 2004, and 2008 panels of the Survey of Income and Program Participation.
Note: The category "All Observations" includes observations with reported income and observations with imputed income.

Table 2 Marginal Effects on 25% or Greater Change in Income

Regressor	Marginal effect	Standard error
Imputed income	0.178***	0.004
Household income as percentage of federal poverty guideline	−0.0003***	0.00001
Number of children under age 17 years	0.004	0.003
Less than high school	0.016*	0.009
High school	−0.004	0.006
College	−0.012**	0.006
White	−0.005	0.009
Black	−0.004	0.011
Hispanic	−0.004	0.011
Married with children	−0.048***	0.006
Married without children	−0.003	0.009
Single without children	0.001	0.008
Male sex	0.002	0.004
Age	−0.002***	0.0003
Number of observations	55,527	

Source: Authors' calculations from the 1996, 2001, 2004, and 2006 panels of the Survey of Income and Program Participation.
Notes: The dependent variable equals 1 if the household experiences a percentage change in income of 25% or more and zero otherwise. Model estimated using logistic regression, weighted by person final weight.
$^*p < 0.10$.
$^{**}p < 0.05$.
$^{***}p < 0.01$.

We report the marginal effects from estimating equation (4) for households with incomes of less than 200 percent of FPL using a logistic regression in table 3. The marginal effect is calculated as the change in the

Table 3 Marginal Effects on Food Insufficiency in Any Month

Regressor	**All observations**	**Excludes observations with any imputations**	**Excludes observations with current-wave imputation**	**Excludes observations with logical imputation**	**Excludes observations with previous wave imputation**
White non-Hispanic	−0.0154*	−0.0146	−0.0165*	−0.0154*	−0.0152
	(0.00824)	(0.0106)	(0.00851)	(0.00841)	(0.0104)
Black non-Hispanic	−0.00584	−0.00348	−0.00485	−0.00703	−0.00654
	(0.00829)	(0.0110)	(0.00868)	(0.00836)	(0.0105)
Hispanic	0.00255	0.00767	0.00393	0.00441	0.00385
	(0.00902)	(0.0122)	(0.00948)	(0.00937)	(0.0114)
Less than high school	0.0948***	0.0767***	0.0928***	0.0862***	0.0852***
	(0.0167)	(0.0203)	(0.0177)	(0.0163)	(0.0203)
High school graduate	0.0611***	0.0479***	0.0583***	0.0579***	0.0512***
	(0.0126)	(0.0153)	(0.0131)	(0.0125)	(0.0150)
Some college	0.0698***	0.0630***	0.0678***	0.0679***	0.0617***
	(0.0125)	(0.0155)	(0.0129)	(0.0125)	(0.0150)
Married with kids	−0.0435***	−0.0456***	−0.0434***	−0.0479***	−0.0473***
	(0.00653)	(0.00851)	(0.00678)	(0.00657)	(0.00820)
Married without kids	−0.0338***	−0.0266***	−0.0314***	−0.0328***	−0.0289***
	(0.00510)	(0.00743)	(0.00534)	(0.00524)	(0.00722)
Single with kids	−0.0191***	−0.0226***	−0.0222***	−0.0227***	−0.0216***
	(0.00589)	(0.00716)	(0.00601)	(0.00588)	(0.00711)
Number of children under age 17 years	1.45e-05	0.00114	0.000125	0.00113	0.00272
	(0.00214)	(0.00269)	(0.00228)	(0.00216)	(0.00255)

Continued

Table 3 *Continued*

Regressor	All observations	Excludes observations with any imputations	Excludes observations with current-wave imputation	Excludes observations with logical imputation	Excludes observations with previous wave imputation
Household income relative to federal poverty guideline	−0.0213***	−0.0269***	−0.0206***	−0.0222***	−0.0261***
	(0.00386)	(0.00483)	(0.00409)	(0.00393)	(0.00475)
Income rises ≥25% between waves	0.0107**	0.00887	0.00840	0.00911*	0.0113*
	(0.00516)	(0.00680)	(0.00564)	(0.00522)	(0.00663)
Income drops ≥25% between waves	0.0165***	0.0207**	0.0138**	0.0179***	0.0213***
	(0.00598)	(0.00851)	(0.00625)	(0.00622)	(0.00813)
Observations	15,910	8,955	13,849	15,126	9,837
Mean of dependent variable	0.071	0.067	0.069	0.071	0.069

Source: 1996, 2001, 2004, and 2008 panels of the Survey of Income and Program Participation.
Notes: Standard errors are in parentheses. Regressions weighted by person final weight and include indicators for each panel. Marginal effects are estimated relative to the omitted group – unmarried householders without children, who are non-Hispanic of other race, college graduates, and appear in the 2008 panel.
$^*p < 0.10$.
$^{**}p < 0.05$.
$^{***}p < 0.01$.

Table 4 Marginal Effects on Food Insufficiency in Last Month

Regressor	All observations	Excludes observations with any imputations	Excludes observations with current-wave imputation	Excludes observations with logical imputation	Excludes observations with previous wave imputation
White non-Hispanic	−0.00759	−0.0104	−0.0120**	−0.00666	−0.00854
	(0.00593)	(0.00755)	(0.00599)	(0.00613)	(0.00740)
Black non-Hispanic	−0.00985*	−0.0123*	−0.0132***	−0.00942*	−0.0131**
	(0.00546)	(0.00658)	(0.00508)	(0.00569)	(0.00644)
Hispanic	−0.00363	−0.000254	−0.00552	−0.00158	−0.00152
	(0.00621)	(0.00828)	(0.00604)	(0.00663)	(0.00791)
Less than high school	0.0583***	0.0422***	0.0569***	0.0544***	0.0475***
	(0.0138)	(0.0152)	(0.0145)	(0.0134)	(0.0153)
High school graduate	0.0357***	0.0241**	0.0340***	0.0332***	0.0248**
	(0.00990)	(0.0112)	(0.0103)	(0.00978)	(0.0109)
Some college	0.0448***	0.0362***	0.0447***	0.0427***	0.0353***
	(0.0102)	(0.0116)	(0.0106)	(0.0101)	(0.0112)
Married with kids	−0.0251***	−0.0256***	−0.0263***	−0.0269***	−0.0257***
	(0.00496)	(0.00637)	(0.00517)	(0.00501)	(0.00618)
Married without kids	−0.0202***	−0.0161***	−0.0169***	−0.0199***	−0.0175***
	(0.00383)	(0.00564)	(0.00414)	(0.00395)	(0.00545)
Single with kids	−0.00617	−0.00868	−0.00912*	−0.00814*	−0.00595
	(0.00472)	(0.00580)	(0.00480)	(0.00476)	(0.00576)
Number of children under age 17 years	−0.00138	−0.00124	−0.00107	−0.000871	−0.000232
	(0.00166)	(0.00209)	(0.00180)	(0.00168)	(0.00195)

Continued

Table 4 *Continued*

Regressor	All observations	Excludes observations with any imputations	Excludes observations with current-wave imputation	Excludes observations with logical imputation	Excludes observations with previous wave imputation
Household income relative to federal poverty guideline	−0.0161***	−0.0193***	−0.0154***	−0.0165***	−0.0197***
	(0.00287)	(0.00365)	(0.00306)	(0.00291)	(0.00356)
Income rises ≥25% between waves	−0.00111	−0.00527	−0.00473	−0.00269	−0.00230
	(0.00360)	(0.00453)	(0.00371)	(0.00365)	(0.00450)
Income drops ≥25% between waves	0.0115**	0.0126*	0.00874*	0.0120**	0.0145**
	(0.00460)	(0.00657)	(0.00473)	(0.00479)	(0.00633)
Observations	15,910	8,955	13,849	15,126	9,837
Mean of dependent variable	0.045	0.042	0.043	0.045	0.043

Source: 1996, 2001, 2004, and 2008 panels of the Survey of Income and Program Participation.
Notes: Standard errors are in parentheses. Regressions weighted by person final weight and include indicators for each panel. Marginal effects are estimated relative to the omitted group – unmarried householders without children, who are non-Hispanic of other race, college graduates, and appear in the 2008 panel.
$^{*}p < 0.10$.
$^{**}p < 0.05$.
$^{***}p < 0.01$.

probability of food insufficiency evaluated at the mean values of the covariables. For binary variables such as a large drop in income, the marginal effect is the difference between the probability associated with individuals who experienced a large drop in income and the probability associated with individuals who did not experience a large drop. The first column reports the association between large percentage changes in income and food insufficiency among all households, including those with imputed income. The following columns report the association when observations with imputed income are excluded—those with any type of imputation, those who are hot-deck imputed using current wave characteristics, those who have missing values derived using information from the rest of their interviews, and those who are hot-deck imputed using previous wave characteristics.

Estimates using all observations, including those with imputed income, show that households that experienced a 25% or greater drop in income between waves had a 1.7–percentage point higher chance of reporting food insufficiency in any month than households with small changes in income. This represents a 23% increase in food insufficiency. Surprisingly, both large rises and large drops in income are associated with a higher likelihood of food insufficiency.

These estimates are biased downward, however, because of the inclusion of observations with imputed data. The exclusion of observations with any type of imputation yields a stronger association between income drops and food insufficiency—households that experience large income drops have a 2.1–percentage point greater chance of food insufficiency, a 31% increase (see column 2 of table 3). Large income rises, however, become statistically insignificant. Although the association between large income drops and food insufficiency is almost 25% greater when imputed observations are excluded, the difference between these estimates is not statistically significant.

The various methods used to impute missing income values differ in their effects on the relationship between income volatility and food insufficiency. Approximately 13% of observations have income imputed using only current wave characteristics; dropping these observations results in very little change to the estimates of marginal effects. Even fewer observations have missing values derived from the rest of their interview, so dropping observations with those imputations does not change the estimates substantially either. Most income imputations use previous wave characteristics; dropping those observations increases the estimated marginal effect of a large income drop to 2.1 percentage points.

Using reported food insufficiency in the last month as the dependent variable results in the same pattern of estimates (see table 4). Large drops in income are associated with a 1.2–percentage point, or 26%, higher probability of food insufficiency when all observations are included. Dropping observations with any imputation increases the association to 1.3 percentage points, with imputations using previous wave characteristics underlying most of that change.

We check the robustness of our results using several alternative specifications. First, we include the percentage change in income as a continuous variable (see table 5). We find that a 10% drop in income is associated with a 1–percentage point increase in food insufficiency. Dropping observations with any imputed income data yields larger estimates—a 10% drop in income is associated with a 2–percentage point increase in food

Table 5 Marginal Effect of Income Change on Food Insufficiency in Any Month Using Logistic Regression

Regressor	**All observations**	**Excludes observations with any imputations**	**Excludes observations with current-wave imputation**	**Excludes observations with logical imputation**	**Excludes observations with previous wave imputation**
Income rises ≥25% between waves	0.0107**	0.00887	0.00840	0.00911*	0.0113*
	(0.00516)	(0.00680)	(0.00564)	(0.00522)	(0.00663)
Income drops ≥25% between waves	0.0165***	0.0207**	0.0138**	0.0179***	0.0213***
	(0.00598)	(0.00851)	(0.00625)	(0.00622)	(0.00813)
Arc percentage change in household income between waves	−0.00102	−0.00183	−0.000210	−0.00209	−0.00131
	(0.00281)	(0.00337)	(0.00300)	(0.00284)	(0.00330)
Income change relative to federal poverty guideline (FPL) rises ≥10%	0.00856	0.00752	0.00901	0.00683	0.00618
	(0.00541)	(0.00733)	(0.00602)	(0.00552)	(0.00689)
Income change relative to FPL drops ≥10%	0.0199***	0.0285***	0.0177***	0.0218***	0.0262***
	(0.00643)	(0.00944)	(0.00684)	(0.00676)	(0.00870)
Variance in household income (income in $1,000 s)	8.66e-05	3.32e-05	0.000164*	1.32e-05	0.000150
	(7.95e-05)	(0.000131)	(9.86e-05)	(7.53e-05)	(0.000127)
Observations	15,910	8,955	13,849	15,126	9,837
Mean of dependent variable	0.071	0.067	0.069	0.071	0.069

Source: 1996, 2001, 2004, and 2008 panels of the Survey of Income and Program Participation.
Notes: Standard errors are in parentheses. Regressions weighted by person final weight. Marginal effects are estimated relative to the omitted group – unmarried householders without children, who are non-Hispanic of other race, college graduates, and appear in the 2008 panel.
*$p < 0.10$.
**$p < 0.05$.
***$p < 0.01$.

insufficiency. Although these estimates are consistent with the preferred specification using indicator variables for large income drops and rises, they are not statistically significant. Next, we measure income changes relative to FPL and define large income changes as rises or drops exceeding 10%. Using this specification, income drops exceeding 10% of FPL are associated with a 2–percentage point higher probability of food insufficiency. Excluding observations with any imputed income data increases that association to 3 percentage points. Finally, we include the variance of income to see whether income variability is associated with food insufficiency. Including all observations results in estimates that suggest that higher income variability is associated with a higher probability of food insufficiency. However, these effects are not precisely estimated and are not statistically different from zero.

The estimated marginal effects are largely unchanged when food insufficiency in the last month is the outcome instead (see table 6). When observations imputed using previous wave characteristics are dropped, a 10% drop in income is associated with a statistically significant 6–percentage point increase in food insufficiency. Estimates using the variance of income remain imprecisely estimated.

Discussion and Conclusions

This article explores the impact of using imputed data in calculating income volatility and then measuring the association between income volatility and food insufficiency. Imputed data, particularly hot-deck imputed data, can contribute to misleading estimates when calculating changes in the potentially imputed variable over time because the calculated changes in effect capture cross-sectional variation instead of variation over time.

We show that including imputed observations leads to a substantial understatement of the association between income drops and food insufficiency among lower-income households. When all observations are included, a large drop in income is associated with a 1.7–percentage point higher likelihood of food insufficiency in any month or a 1.2–percentage point higher likelihood of food insufficiency in the last month. These estimates increase to 2.1 percentage points and 1.3 percentage points, respectively, when observations with any type of imputation are dropped. Approximately 7% of lower-income households had food insufficiency in any month, so the marginal effect of a large income drop when imputed observations are excluded represents a 31% increase in food insufficiency. Although the direction of the association between income volatility and food insufficiency are unchanged when imputed observations are excluded and the estimates are not statistically different, the magnitude of the association between income changes and food insufficiency is substantially understated when imputed observations are included.

Including hot-deck imputed data tends to overstate income volatility. This poses challenges for policymakers and researchers who rely on measures of income volatility. When imputed income data are included, income volatility may not appear to be strongly associated with nutritional outcomes. Because income volatility has a larger impact on nutritional outcomes, such as food sufficiency, than previous estimates that used imputed data imply, programs with a documented ability to reduce income volatility

Table 6 Marginal Effect of Income Change on Food Insufficiency in Last Month Using Logistic Regression

Regressor	All observations	Excludes observations with any imputations	Excludes observations with current-wave imputation	Excludes observations with logical imputation	Excludes observations with previous wave imputation
Income rises ≥25% between waves	−0.00111	−0.00527	−0.00473	−0.00269	−0.00230
	(0.00360)	(0.00453)	(0.00371)	(0.00365)	(0.00450)
Income drops ≥25% between waves	0.0115**	0.0126*	0.00874*	0.0120**	0.0145**
	(0.00460)	(0.00657)	(0.00473)	(0.00479)	(0.00633)
Arc percentage change in household income between waves	−0.00434**	−0.00561**	−0.00421**	−0.00499**	−0.00565**
	(0.00202)	(0.00239)	(0.00206)	(0.00207)	(0.00233)
Income change relative to federal poverty guideline (FPL) rises ≥10%	2.93e-05	−0.00292	−0.00221	−0.00141	−0.00339
	(0.00390)	(0.00515)	(0.00412)	(0.00402)	(0.00479)
Income change relative to FPL drops ≥10%	0.0163***	0.0239***	0.0136**	0.0173***	0.0239***
	(0.00511)	(0.00775)	(0.00534)	(0.00539)	(0.00712)
Variance in household income, income in thousands	6.04e-06	−3.59e-05	2.33e-05	3.57e-06	−5.28e-05
	(6.43e-05)	(0.000128)	(8.09e-05)	(6.80e-05)	(0.000112)
Observations	15,910	8,955	13,849	15,126	9,837
Mean of dependent variable	0.045	0.042	0.043	0.045	0.045

Source: 1996, 2001, 2004, and 2008 panels of the Survey of Income and Program Participation.

Notes: Standard errors are in parentheses. Regressions weighted by person final weight. Marginal effects are estimated relative to the omitted group – unmarried householders without children, who are non-Hispanic of other race, college graduates, and appear in the 2008 panel.

$^{*}p < 0.10$.

$^{**}p < 0.05$.

$^{***}p < 0.01$.

(Unemployment Insurance, progressive taxation), certainly have a greater ability to improve these outcomes than previously believed. Similarly, food assistance programs (such as SNAP) may provide a larger buffer against the effects of income volatility than previously thought. Similar concerns to those discussed here exist when examining the relationship between income volatility and outcomes such as a child's education, physical and mental health, the loss of housing, divorce, or other nutritional outcomes beyond food insecurity. As a result, we strongly advise caution when examining changes in income in the SIPP (and in similar panel surveys) that include imputed observations.

References

Allen, R.G.D., and A.P. Lerner. 1934. The Concept of Arc Elasticity of Demand. *Review of Economic Studies* 1 (3): 226–30.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney. 2008. Trends in U.S. Wage Inequality: Revising the Revisionists. *Review of Economics and Statistics* 90 (2): 300–23.

Blundell, Richard, and Luigi Pistaferri. 2003. Income Volatility and Household Consumption: The Impact of Food Assistance Programs. *Journal of Human Resources* 38 (Suppl.): 1032–50.

Bollinger, Christopher R. 1998. Measurement Error in the CPS: A Nonparametric Look. *Journal of Labor Economics* 16 (3): 576–94.

Bollinger, Christopher R., and Barry T. Hirsch. 2006. Match Bias from Earnings Imputation in the Current Population Survey: The Case of Imperfect Matching. *Journal of Labor Economics* 24 (3): 483–519.

——. 2013. Is Earnings Nonresponse Ignorable? *Review of Economics and Statistics* 95 (2): 407–16.

Bound, John, Charles Brown, Greg J. Duncan, and Willard L. Rodgers. 1994. Evidence on the Validity of Cross-Sectional and Longitudinal Labor Market Data. *Journal of Labor Economics* 12 (3): 345–68.

Bound, John, and Alan B. Krueger. 1991. The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right? *Journal of Labor Economics* 9 (1): 1–24.

Coleman-Jensen, Alisha, Mark Nord, Margaret Andrews, and Steven Carlson. 2011. Household Food Security in the United States in 2010. Washington DC: US Department of Agriculture, Economic Research Report No. 125.

Congressional Budget Office. 2008. *Recent Trends in the Variability of Individual Earnings and Household Income*. Washington DC: Congressional Budget Office.

Czajka, John L. 2009. SIPP Data Quality. In: *Reengineering the Survey of Income and Program Participation*, ed. Constance F. Citro, and John Karl Scholz, 127–57. Washington DC: National Research Council.

Davis, Steven J., John C. Haltiwanger, and Scott Schuh. 1996. *Job Creation and Destruction*. Cambridge, MA: MIT Press.

DiNardo, John, Nicole M. Fortin, and Thomas Lemieux. 1996. Labor Market Institutions and the Distribution of Wages, 1973–1992: A Semiparametric Approach. *Econometrica* 64 (5): 1001–44.

Dooley, Martin D., and Peter Gottschalk. 1984. Earnings Inequality among Males in the United States: Trends and the Effect of Labor Force Growth. *Journal of Political Economy* 92 (1): 59–89.

Gruber, Jonathan. 1997. The Consumption Smoothing Benefits of Unemployment Insurance. *American Economic Review* 87 (1): 192–205.

Gundersen, Craig, and Joseph Gruber. 2001. The Dynamic Determinants of Food Insufficiency. In: *Second Food Security Measurement and Research Conference,*

Volume II: Papers, ed. Andrews Margaret, and Mark Prell, 92–110. Washington DC: US Department of Agriculture.
Gundersen, Craig, and Brent B. Kreider. 2008. Food Stamps and Food Insecurity: What Can Be Learned in the Presence of Non-classical Measurement Error? *Journal of Human Resources* 43 (2): 352–82.
Gundersen, Craig, and James P. Ziliak. 2003. The Role of Food Stamps in Consumption Stabilization. *Journal of Human Resources* 38 (Suppl.): 1051–79.
Hensher, David A., John M. Rose, and William H. Greene. 2005. *Applied Choice Analysis: A Primer*. Cambridge, UK: Cambridge University Press.
Hirsch, Barry T., and Edward J. Schumacher. 2004. Match Bias in Wage Gap Estimates Due to Earnings Imputation. *Journal of Labor Economics* 22 (3): 689–722.
Kim, Bonggeun, and Gary Solon. 2005. Implications of Mean-Reverting Measurement Error for Longitudinal Studies of Wages and Employment. *Review of Economics and Statistics* 87 (1): 193–6.
Leete, Laura, and Neil Bania. 2010. The Effect of Income Shocks on Food Insufficiency. *Review of Economics of the Household* 8 (4): 505–26.
Lillard, Lee, James P. Smith, and Finis Welch. 1986. What Do We Really Know about Wages? The Importance of Nonreporting and Census Imputation. *Journal of Political Economy* 94 (3): 489–506.
Mellow, Wesley, and Hal Sider. 1983. Accuracy of Response in Labor Market Surveys: Evidence and Implications. *Journal of Labor Economics* 1 (4): 331–44.
Meyer, Bruce D., Wallace K.C. Mok, and James X. Sullivan. 2009. The Under-Reporting of Transfers in Household Surveys: Its Nature and Consequences. Working Paper 15181, National Bureau of Economic Research.
Mykerezi, Elton, and Bradford Mills. 2010. The Impact of Food Stamp Program Participation on Household Food Insecurity. *American Journal of Agricultural Economics* 92 (5): 1379–91.
Nord, Mark, and Anne Maria Golla. 2009. Does SNAP Decrease Food Insecurity? Untangling the Self-Selection Effect. US Department of Agriculture, Economic Research Report No. 85.
Ratcliffe, Caroline, Signe-Mary McKernan, and Sisi Zhang. 2011. How Much Does the Supplemental Nutrition Assistance Program Reduce Food Insecurity? *American Journal of Agricultural Economics* 93 (4): 1082–98.
Ribar, David C., and Karen S. Hamrick. 2003. Dynamics of Poverty and Food Sufficiency. US Department of Agriculture, Food Assistance and Nutrition Research Report No. 36.
Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: J. Wiley & Sons.
Tornqvist, Leo, Pentti Vartia, and Yrjo O. Vartia. 1985. How Should Relative Changes Be Measured? *American Statistician* 39 (1): 43–46.
US Census Bureau. 2001. *Survey of Income and Program Participation Users' Guide (Supplement to the Technical Documentation)*, 3rd ed. Washington DC: US Census Bureau.
Welch, Finis. 1979. Effects of Cohort Size on Earnings: The Baby Boom Babies' Financial Bust. *Journal of Political Economy* 87 (5): S65–S97.
Williams, Roberton. 1992. Analyzing the Dynamics of Program Participation with Data from the SIPP. *Journal of Economic and Social Measurement* 18 (1–4): 285–301.

Appendix

This appendix provides details of the sample and income variables used for the analysis. The analysis was based on data from the 1996, 2001, 2004, and 2008 SIPP panels. For each panel, we observed income over a twelve-month period ending when food insufficiency was measured. The waves included in the analysis were as follows: waves 6, 7, and 8 of the 1996 panel;

waves 6, 7, and 8 of the 2001 panel; waves 3, 4, and 5 of the 2004 panel; and waves 7, 8, and 9 of the 2008 panel.

We calculated household total income in each wave as the sum of earnings (TPMSUM1 and TPMSUM2), business income (TBMSUM1 and TBMSUM2), severance (T15AMT), income from other jobs (TMLMSUM), and unearned income (the difference between TPTOTINC and all earned income). Wave-based household income from all three waves was used to calculate the average and variance of income. Wave-based household income from the latter two waves of each panel was used to calculate the APC. Observations were weighted by person weights (WPFINWGT) in the fourth reference month of the last wave included for each panel to correspond with the timing of the food insufficiency question.

Observations with imputed income data were identified using the allocation flags for the income sources listed herein and various sources of unearned income—Social Security, railroad retirement, federal Supplemental Security Income for adults and children, pensions (company, union, local, state, or federal government), military retirement pay, unemployment compensation, veteran's compensation, worker's compensation, private disability insurance, public assistance, general assistance, foster care, and other welfare.