Estimating the Impacts of Program Benefits: Using Instrumental Variables with Underreported and Imputed Data*

Melvin Stephens Jr.
University of Michigan and NBER

Takashi Unayama
Hitotsubashi University

This Version: July 16, 2018

---

*Stephens: Department of Economics, University of Michigan, 611 Tappan St., 341 Lorch Hall, Ann Arbor, MI 48109-1220, e-mail: mstep@umich.edu and National Bureau of Economic Research, Cambridge, MA. Unayama: Institute of Economic Research, Hitotsubashi University, Naka 2-1, Kunitachi, Tokyo, Japan, e-mail: unayama@ier.hit-u.ac.jp. Tom Eldridge provided excellent research assistance. We thank Gary Engelhardt as well as seminar participants at Illinois, Michigan, Virginia, Williams, and the NBER Summer Institute for helpful suggestions. We also thank the Statistical Bureau of the Japanese Government for allowing access to the Family Income and Expenditure Survey data. A part of this project is financially supported by KAKENHI (15H03357).

Estimating the Impacts of Program Benefits:

Using Instrumental Variables with Underreported and Imputed Data

Abstract

Survey non-response has risen in recent years which has increased the share of imputed and underreported values found on commonly used datasets. While this trend has been well-documented for earnings, the growth in non-response to government transfers questions has received far less attention. We demonstrate analytically that the underreporting and imputation of transfer benefits can lead to program impact estimates that are substantially overstated when using instrumental variables methods to correct for endogeneity and/or measurement error in benefit amounts. We document the importance of failing to account for these issues using two empirical examples.

1 Introduction

Vast economic literatures estimate the impacts of government benefits, typically using instrumental variables (IV) methods that treat benefit amounts as an endogenous regressor since program participation is often a choice (e.g., see surveys by Krueger and Meyer 2002; Currie 2004). Benefits reported on household surveys are typically measured with error and these errors are not likely to be classical as it is quite common for benefit amounts to be underreported (understated) or contain imputed values. We demonstrate analytically and via two empirical examples that IV estimation in such cases tends to overstate, sometimes substantially, the causal effect of program benefits.

Benefits are routinely imputed when households acknowledge receiving a benefit but do not recall the amount.[1] The top left corner of Figure 1 shows the well-known substantial increase in earnings imputations in the CPS (Lillard, Smith, and Welch 1986; Hirsch and Schumacher 2004; Bollinger and Hirsch 2006; Heckman and LaFontaine 2006).[2] The Figure also shows the far less appreciated fact that benefit imputations have increased just as dramatically over this period.[3]

A related issue is the underreporting of benefits in surveys. For example, the Consumer Expenditure Survey requires a single valid non-zero report from a major income source to deem a household as a "complete income reporter," potentially ignoring many other income sources (Paulin and Ferraro 1994). Meyer, Mok, and Sullivan (2009) find that total benefits received, computed by aggregating and appropriately weighting survey responses, fall short of administrative records of total benefit disbursements even when including imputations.

These types of measurement error yield important inconsistencies in empirical analysis. E.g.,

[1]Many researchers fail to acknowledge imputed data. Surveying articles that use the Current Population Survey (CPS) as the primary data source and were published in 2004 and 2013, inclusive, we find only 19 percent (16 out of 86 articles) mention imputed values. Even when acknowledged, some studies still treat imputed values as actual data. The eight journals surveyed are the *American Economic Review* (except *Papers and Proceedings* issues), *Industrial and Labor Relations Review*, the *Journal of Human Resources*, the *Journal of Labor Economics*, the *Journal of Public Economics*, *Labour Economics*, the *Quarterly Journal of Economics*, and the *Review of Economics and Statistics*.

[2]A sharp rise in earnings imputations is also found in the 1990s in the CPS Outgoing Rotation Group (Bollinger and Hirsch 2006). Determining which earnings values are imputed in the March CPS becomes less transparent beginning in 1988. See http://www.psc.isr.umich.edu/dis/data/kb/answer/1349.

[3]Meyer, Mok, and Sullivan (2009) present a related set of results in terms of dollars imputed rather than individuals. Prior to 1988, unemployment insurance and worker's compensation benefits are combined with other benefits. Prior to 1982, the CPS imputation codes for AFDC/TANF do not match the codebook values. Figure 1 imputation rates account for item non-response and whole supplement non-response, the roughly 10% of households that do not provide sufficient data for the March supplement. The variable FL-665, a flag for whole supplement non-response, does not appear on the public-use CPS data until 1991 although it does appear on the Unicon CPS files beginning in 1988. We thank Jay Stewart of the Bureau of Labor Statistics for directing us to these pre-1991 data.

since CPS earnings imputations do not account for union status, imputed earnings are uncorrelated with union status. As a result, Hirsch and Schumacher (2004) and Bollinger and Hirsch (2006) find that OLS estimates of the union wage gap are substantially understated (attenuated) when including imputed earnings as compared to only using non-imputed earnings observations.

We show that underreporting and imputation can lead IV estimates to dramatically overstate the impacts of transfer programs. For example, if the instrument is based on program rules that vary across states and over time, imputed benefit values are not correlated with the instrumental variable if the imputation procedure does not condition on state of residence. The first stage estimated impact of the instrument on benefit amounts generally will be attenuated when using imputed benefits and, since the IV estimate is the ratio of the reduced form to the first stage coefficients on the instrument, the IV estimate will exceed is true value. When the instrument is uncorrelated with the imputed values and missing observations are randomly assigned, we show that the probability limit of the IV estimator exceeds the true IV parameter by a factor of $1/p$ where p is the fraction of households correctly reporting benefits. Since only two-thirds of recent CPS households correctly report benefits, IV estimates generated using imputed benefits are biased upwards by 50 percent. Benefit underreporting has a similar impact on the IV estimator.

If the non-reporting is randomly assigned, a straightforward empirical solution is to only use the non-imputed sub-sample. If values instead are missing at random (i.e., random after conditioning on covariates), methods which account for selection on observable characteristics such as inverse propensity score weighting can be applied. With selection on unobservables, estimates using the non-imputed sample are also inconsistent. We briefly discuss possible solutions in such instances.

We present two examples to demonstrate the empirical importance of these estimation issues. The first example uses the U.S. Social Security "notch" which Englehardt, Gruber, and Perry (2005) exploit to examine the impact of Social Security income on the propensity of the elderly to live independently. Since Social Security benefit imputations in the CPS use broad age categories rather than exact age, we find that the IV estimates are biased upwards by 20 to 30 percent. Our second example is a test for "excess sensitivity" among Japanese households in which monthly consumption changes are regressed on monthly income changes using the predictable pattern of child benefit payments as an instrument. Since only one quarter of eligible households report receiving these payments, the IV estimate is overstated by more than a factor of three.

The measurement error induced by underreporting and imputation is akin to "mean reverting" measurement error (Bound and Krueger 1991; Bound, Brown, Duncan, and Rodgers 1994).[4] Berger, Black, and Scott (2000) analyze the inconsistency of the IV estimator when using a noisy measure to instrument for another noisy measure when both suffer from mean reverting measurement error.[5] In our analysis, this inconsistency arises even when the instrument is correctly measured as is typical when benefit rules vary by well-measured characteristics such as age and state of residence. Our results easily extend to situations where the outcome of interest is underreported or imputed.

2 Econometric Framework

2.1 Model Setup

We focus on the population regression model for a continuous outcome y

$$y = \beta_0 + \beta_1 x + u \tag{1}$$

where x is an endogenous, continuous regressor such that $Cov(x, u) \neq 0$.[6]

Suppose that z is a valid, continuous instrumental variable for x such that $Cov(x, z) \neq 0$ and $Cov(z, u) = 0$. The first stage and reduced form equations are, respectively,

$$x = \pi_0 + \pi_1 z + \epsilon \tag{2}$$

$$y = \delta_0 + \delta_1 z + \varepsilon \tag{3}$$

Since z is assumed to be exogenous and free from measurement error, the OLS estimators for the coefficients on z in equations (2) and (3) are consistent as long the left-hand side variables in each equation are free of measurement error or suffer from classical measurement error. In addition, under these conditions, the IV estimator for β_1, which can be written as $\hat{\delta}_1/\hat{\pi}_1$, is also consistent.

[4] Gibson and Kim (2010) discuss a related issue for errors from using long-term retrospective recall data.

[5] See Card (1996) and Kane, Rouse, and Staiger (1999) for related analyses.

[6] It is straightforward to extend our analysis to include exogenous covariates and to account for binary variables.

Dividing the sample into two groups, $g = \{0,1\}$, based upon whether the endogenous regressor, x_{ig}, is an actual report ($g = 1$) or an underreport/imputed value ($g = 0$), and denoting $S_{ZX}^{Total} = \sum_{g=0}^{1}\sum_{i=1}^{N_g}\left(z_{ig} - \overline{\overline{z}}\right)\left(x_i - \overline{\overline{x}}\right)$, $S_{ZX}^{Between} = \sum_{g=0}^{1} N_g\left(\overline{z}_{i.} - \overline{\overline{z}}\right)\left(\overline{x}_{i.} - \overline{\overline{x}}\right)$, and $S_{ZX}^{Within} = \sum_{g=0}^{1}\sum_{i=1}^{N_g}\left(z_{ig} - \overline{z}_{i.}\right)\left(x_i - \overline{x}_{i.}\right) = S_{ZX}^{g=1} + S_{ZX}^{g=0}$, where $\overline{x}_{i.}$ and $\overline{\overline{x}}$ are group and overall means, respectively, we can use the within-between decomposition (e.g., Greene 2008, p.191-2) to re-write $\hat{\pi}_1$

$$\begin{aligned}
\hat{\pi}_1 = \frac{S_{ZX}^{Total}}{S_{ZZ}^{Total}} &= \frac{S_{ZX}^{g=1} + S_{ZX}^{g=0} + S_{ZX}^{Between}}{S_{ZZ}^{Total}} \\
&= \frac{S_{ZZ}^{g=1}}{S_{ZZ}^{Total}} \cdot \frac{S_{ZX}^{g=1}}{S_{ZZ}^{g=1}} + \frac{S_{ZZ}^{g=0}}{S_{ZZ}^{Total}} \cdot \frac{S_{ZX}^{g=0}}{S_{ZZ}^{g=0}} + \frac{S_{ZZ}^{Between}}{S_{ZZ}^{Total}} \cdot \frac{S_{ZX}^{Between}}{S_{ZZ}^{Between}} \qquad (4) \\
&= \frac{S_{ZZ}^{g=1}}{S_{ZZ}^{Total}} \cdot \hat{\pi}_{1,g=1} + \frac{S_{ZZ}^{g=0}}{S_{ZZ}^{Total}} \cdot \hat{\pi}_{1,g=0} + \frac{S_{ZZ}^{Between}}{S_{ZZ}^{Total}} \cdot \hat{\pi}_{1,Between}
\end{aligned}$$

where N_g is the number of observations in group g.[7] Thus, the OLS estimator for the first stage slope coefficient, $\hat{\pi}_1$, is a weighted average of the corresponding estimators when the model is estimated separately for each group, $\hat{\pi}_{1,g=1}$ and $\hat{\pi}_{1,g=0}$ and the between group estimator $\hat{\pi}_{1,Between}$ where the weights depend upon within and between group variation in the instrument.[8]

2.2 Interpreting the IV estimator

Suppose that group membership, g, is randomly assigned and $p = P[g = 1]$ is the probability of providing an actual report. The first stage slope estimator using the sample of actual reporters, $\hat{\pi}_{1,g=1}$, is a consistent estimator of π_1. The final term in (4) converges to zero since the group means converge to the overall means. The within group weights, $S_{ZZ}^{g=1}/S_{ZZ}^{Total}$ and $S_{ZZ}^{g=0}/S_{ZZ}^{Total}$, are consistent estimators of p and $1-p$, respectively. However, the first stage estimator for the under/imputed reporters, $\hat{\pi}_{1,g=0}$, depends upon the underreporting or imputation process.

A common imputation procedure, the "hot deck," selects a replacement amount from a "donor" with the same values for a small set of characteristics. Hirsch and Schumacher (2004) and Bollinger

[7] Most datasets contain flags to indicate which observations are imputed and which are not. This set-up is also useful for understanding the impact of underreporting even though this behavior typically is not explicitly flagged.

[8] The analysis focuses on imputed/underreported values of x but can be extended to either y or z. Typically, the instruments for benefits depend on well-measured demographic characteristics. E.g., Medicaid eligibility may depend on a child's age and the earned income tax credit (EITC) depends upon the family's number of children. Thus, it is likely that the endogenous regressor will be underreported or imputed while the instrumental variable is not.

and Hirsch (2006) note that this procedure does not preserve the covariance between the allocated variable and the characteristics in the data that are left out of the imputation procedure. If the imputed value of x does not depend upon z, $\hat{\pi}_{1,g=0} \approx 0$.[9] Thus, by equation (4), the probability limit of $\hat{\pi}_1$ will equal $p\pi_1 + (1-p) \cdot 0 = p\pi_1$.[10]

For underreporting, suppose that observed x is a constant fraction, θ, of actual x. It is straightforward to show that $plim\,(\hat{\pi}_1)$ for underreporters is $\theta \cdot \pi_1$ and, thus, for the full sample $plim\,(\hat{\pi_1}) = p\pi_1 + (1-p)\,\theta\pi_1 < \pi_1$. Alternatively, when failing to report benefits (i.e., $\theta = 0$), perhaps when payments are small or received infrequently, the probability limit of $\hat{\pi}_1$ falls to $p\pi_1$.

The impact of underreported or imputed values of x on the IV estimator can be seen by substituting (4) and an analogous expression for the reduced form estimator into the IV estimator

$$\hat{\beta}_1^{IV} = \frac{\hat{\delta}_1}{\hat{\pi}_1} = \frac{\frac{S_{ZZ}^{g=1}}{S_{ZZ}^{Total}} \cdot \hat{\delta}_{1,g=1} + \frac{S_{ZZ}^{g=0}}{S_{ZZ}^{Total}} \cdot \hat{\delta}_{1,g=0} + \frac{S_{ZZ}^{Between}}{S_{ZZ}^{Total}} \cdot \hat{\delta}_{1,Between}}{\frac{S_{ZZ}^{g=1}}{S_{ZZ}^{Total}} \cdot \hat{\pi}_{1,g=1} + \frac{S_{ZZ}^{g=0}}{S_{ZZ}^{Total}} \cdot \hat{\pi}_{1,g=0} + \frac{S_{ZZ}^{Between}}{S_{ZZ}^{Total}} \cdot \hat{\pi}_{1,Between}} \tag{5}$$

As discussed above, the denominator converges to values smaller than π_1 when the endogenous regressor is underreported or imputed. The reduced form within-group estimates, $\hat{\delta}_{1,g=1}$ and $\hat{\delta}_{1,g=0}$, for the actual and under/imputed reporters, respectively, are consistent if y is neither imputed nor underreported.[11] Thus, the probability limit of the IV estimator exceeds β_1 and equals $\beta_1/p > \beta_1$ if underreporters all report no benefits or if the imputations are uncorrelated with the instrument.[12]

If the non-reporting of values is randomly assigned across observations (i.e., missing completely at random), then a practical solution to generate consistent IV estimates is to simply restrict the analysis to only non-imputed/non-underreported observations.[13] Alternatively, the availability of

[9] Bollinger and Hirsch (2006) note that some correlation between the imputed x and z will occur if the covariates used in the imputation process for x are correlated with the instrument and hence $\hat{\pi}_{1,g=0}$ may not exactly equal zero.

[10] In general, whether or not the probability limit of $\hat{\pi}$ exceeds π depends upon the imputation procedure. Bollinger and Hirsch (2006) and Heckman and LaFontaine (2006) show that CPS earnings imputations pool GED recipients with high school graduates and those attending, but not graduating from, a post secondary institution. Regressions yield larger GED returns among those with imputed wages relative to those who provide wage information ($plim\,(\hat{\pi_1}) > \pi_1$.)

[11] The between group estimate $\hat{\delta}_{1,Between} = 0$ if group membership is randomly assigned.

[12] Extending the analysis to include exogenous regressors, $\mathbf{w}$, is straightforward using the Frisch-Waugh-Lovell Theorem. The OLS estimate for the coefficient on x when regressing y on x plus a vector of covariates $\mathbf{w}$ is numerically equivalent to first separately regressing y and x on $\mathbf{w}$ and then using the resulting residuals in a simple regression. As analogous procedure is available for 2SLS, we can again apply (5) but must compute the weights using the shares of the variation in the residualized values of the instrument z between the actual and underreported/imputed sub-samples.

[13] As a referee noted, Two Sample IV (Angrist and Krueger 1992) using the full sample for the reduced form and the non-imputed sample for the first stage may provide an efficiency gain over 2SLS with only non-imputed data.

administrative data can provide a straightforward re-scaling of the first stage estimate when non-reporting is random. If the non-reporting of values follows the selection on observables assumption, a number of straightforward methods are applicable: apply inverse propensity score weighting to the non-imputed sample (Bollinger and Hirsch 2006), construct imputations using the instruments in the imputation process (Hirsch and Schumacher 2004; Heckman and Lafontaine 2006), and implement the "general correction" formula of Bollinger and Hirsch (2006) to adjust the estimates for observable differences between the groups defined by s_i. When implementing these methods, the studies listed above do not find substantively different results between using the non-imputed sample only and correcting for selection on observables.

Additional methods may prove useful when confronted with non-random non-response. Recent estimation methods have focused on providing consistent point estimates when data are missing as a function of the outcomes only (Tang, Little, and Raghunathan 2003; Ramalho and Smith 2013) rather than as a function of the regressors as with the selection on observables assumption.[14] Another option is to construct bounds for π_1 which, since using the full sample yields consistent estimates of δ_1, will help produce bounds on β_1. Since most government benefits have a natural set of bounds due to programmatic rules, it may be possible to adapt methods developed by Manski (1997) and Kline and Santos (2013) to generate bounds on π_1 or use the approach of Manski and Pepper (2000) to derive bounds on β_1.[15] We do not pursue these approaches in the current paper.

3 Empirical Examples

3.1 The Impact of the Social Security Notch Using Imputed Benefits

The U.S. Social Security "notch" generated a sizable change in Social Security (hereafter S.S.) benefits for the affected birth cohorts (Krueger and Pischke 1992).[16] Englehardt, Gruber, and Perry (2005) (hereafter EGP), using data from the 1980-1999 March CPS supplements, investigate the impact of S.S. income on the probability that elderly-headed families live independently.[17]

[14] These methods are related to those used in the literature on choice based sampling.

[15] See Kreider et al (2012) for a recent application of bounding when SNAP (food stamp) benefits are mis-reported.

[16] The Social Security Administration provides details of the notch at http://www.ssa.gov/history/pdf/notch.pdf

[17] EGP limit their analysis to families containing a S.S. recipient who is a male or never married female age 65 and up or is a widowed or divorced female age 62 and up. Their paper provides details of sample construction.

As OLS estimates of this relationship are likely inconsistent because S.S. benefits are a function of lifetime earnings (e.g., wealthier individuals have higher benefits and are more likely to live independently), EGP use the variation across birth cohorts driven by the notch to instrument for S.S. benefits. Our analysis is likely important in this case since the share of S.S. benefit recipients with imputed benefits in the CPS rises from 20% to nearly 30% during this period.

To create their instrument, EGP construct a lifetime earnings profile based on the median male earner in the 1916 birth cohort. They use this profile to compute the S.S. benefit for every birth cohort from 1900-1933, using the Consumer Price Index (CPI) to deflate earnings across time. By fixing the earnings profile, the instrumental variable only reflects changes in the programmatic rules across birth cohorts. The solid line in Figure 2 shows the instrument by birth cohort.[18]

Imputations in the March CPS arise from two types of non-response. Item non-response arises when the respondent reports receiving a benefit but does not provide an amount. Whole supplement non-response occurs when households finish the basic CPS interview but do not participate in the March Supplement. As whole supplement non-response has remained constant at roughly ten percent, the recent increase in non-reporting is driven by item non-response.[19]

The CPS uses the hot deck imputation method to allocate missing values by taking a value from a donor observation with the same values for a subset of observable characteristics. For the March CPS supplement, all donors are drawn from the same year. To broaden the scope of potential matches, continuous match characteristics are collapsed into categorical values (e.g., age) while some values of a single categorical characteristic are combined (e.g., race/ethnicity).[20]

Age is used in the hot deck procedure to impute missing S.S. benefits. For item non-response, the imputation procedure uses seven age categories for selecting a donor: less than 35, 35 to 54, 55 to 61, 62 to 64, 65 to 69, 70 to 74, and 75+. For whole supplement non-response, the procedure always groups those ages 65 and up. The long and short dashed lines in Figure 2 show average S.S. income by birth cohort for non-imputed and imputed values, respectively. Actual S.S. income

[18] We thank Gary Englehardt for sharing the values of the instrument by birth cohort.

[19] Prior to 1988, information on whole supplement imputation was contained in the data allocation flag for each income measure. In subsequent years, there is a single flag indicating whole supplement non-response.

[20] We thank Ed Welniak for providing us with the internal Census Bureau documents detailing the hot deck procedure beginning with the March 1989 CPS and retroactively applied to the March 1988 CPS. These documents are available from the authors. We have not been able to determine the imputation procedures used in earlier years.

reports exhibit strong evidence of the notch while the imputed values do not.

EGP estimate the equation

$$P_{i,t} = \theta SSIncome_{i,t} + \boldsymbol{\beta}\mathbf{X}_{i,t} + \boldsymbol{\gamma}_i + \boldsymbol{\alpha}_t + \boldsymbol{\phi}_i + u_{i,t} \tag{6}$$

where $P_{i,t}$ is an indicator for having a shared living arrangement; $SSIncome_{i,t}$ is family S.S. income in thousands of dollars; $\mathbf{X}_{i,t}$ includes indicators the head's and spouse's (if present) education, spouse's age (if present), marital status (married, widowed, and divorced), white, and female; $\boldsymbol{\gamma}_i$ is a full set of indicators for the age (age+3 for widowed and divorced women) from ages 65 to 90; $\boldsymbol{\alpha}_t$ is a set of survey year indicators, and $\boldsymbol{\phi}_i$ is a set of indicators for the nine Census divisions.[21]

Table 1 presents our results.[22] Applying OLS to equation (6) shows that the probability of living in a shared arrangement falls as S.S. income increases. The impact is over twice as large in the non-imputed sample (column (2)) than in the imputed sample (column (3)), consistent with an attenuation bias due to including those with allocated benefits.

The first stage estimates vary as predicted by our analytical results. The estimated effect of the instrument on S.S. income is nearly 20% larger in the non-imputed sample than in the full sample, consistent with the share of S.S. benefit imputations during this period. Relatedly, the first stage relationship is more than three times larger for the non-imputed sample than the imputed sample.

As the shared living arrangements measure is based on the household roster, there is no reason to expect the reduced form estimate to depend upon whether S.S. income is imputed. While the reduced form estimate in the imputed sample is larger than the non-imputed sample estimate, the standard errors are sufficiently large that these differences are not statistically meaningful.

The final row of Table 1 presents the 2SLS estimates. The 2SLS estimate of -0.023 when using the full sample is over 25% larger than the estimate of -0.018 using the non-imputed sample only.[23]

[21] Our analysis differs from EGP's in two ways. First, we use the 1900-1930 birth cohorts rather than the 1900-1933 cohorts. Second, whereas EGP use age-by-year of birth cells, we use individual-level data to match our analytical results. Cell-level results are quite similar to our findings shown here (see Stephens and Unayama 2015a).

[22] Our estimates are weighted by the individual sampling weight for the S.S. recipient. The standard errors are clustered at the year of birth level. As equation (6) includes a number of exogenous covariates, we apply the Frisch-Waugh-Lovell theorem as described earlier and use the resulting residuals to estimate the first stage and reduced form models in order to be consistent with the decomposition shown in equation (5).

[23] Relatedly, the estimated weights based on the decompositions shown in (4) and (5) are 0.791, 0.209, and ≈ 0 for the non-imputed, imputed, and between estimators, respectively.

Assuming that S.S. benefits are missing at random, the results from the full sample substantially overstate the efficacy of S.S. benefits in reducing shared living arrangements.[24]

Finally, we use inverse propensity score weighting (IPW) to correct for selection on observable characteristics (Bollinger and Hirsch 2006). We estimate a probit using an indicator for reporting an actual S.S. value as the outcome and use the same regressors as in equation (6). The IPW estimates (column (4)) are nearly identical to the non-imputed sample only estimates (column (2)).

3.2 Excess Sensitivity and the Underreporting of Japan's Child Benefit

A number of developed countries provide transfers based the age and number of children in a family (OECD 2011). In Japan, child benefits are paid three times a year, in equal amounts, in February, June, and October. While the Life-Cycle/Permanent Income Hypothesis (LCPIH) predicts that households will smooth consumption in response to predictable changes in income, a number of papers find that consumption is sensitive to the timing of income receipt including various types of government transfer payments (Stephens 2003; Shapiro 2005; Mastrobouni and Weinberg 2009; Stephens and Unayama 2011) and paychecks (Stephens 2006).

Japan introduced its child benefit system in 1972 by providing benefits to households with three or more children, extending to two child families in 1986, and to one child families in 1992.[25] Child benefits were means tested until 2009. Benefits initially continued until the child was fifteen but this age limit was lowered to three in 1986 before being incrementally raised over multiple years and again reached age fifteen in 2009. Benefits were relatively stable in real terms in the 1970s and 1980s, increased in 1992 and again in the mid and late 2000s, before subsequently decreasing. These benefits constitute over three percent of family income (Stephens and Unayama 2015b).

We test whether consumption exhibits "excess sensitivity" using monthly panel data from the Japanese Family Income and Expenditure Survey (JFIES) during 1992-2009 when all families with children are eligible for benefits and benefits are means tested. Families are surveyed in the JFIES for six consecutive months and are instructed to enter all expenditures and income into a daily

[24] When converted to elasticities, following EGP, we find a full sample elasticity of -0.53 which is 25% larger than our elasticity of -0.41 for the non-imputed sample. Stephens and Unayama (2015a) provide details of these calculations.

[25] Stephens and Unayama (2015b) provide a more detailed discussion of Japan's child benefit system.

diary. Our data contains detailed expenditure and income categories at a monthly frequency.[26]

Child benefits are recorded as part of an "other social security" variable which contains all social welfare benefits except public pension payments. In benefit distribution months, only 24% of eligible households report positive benefits amounts with 70% of positive reports exactly matching the child benefit value predicted by programmatic rules (i.e., based on age and number of children) and 20% of positive reports being too high, likely due to receiving additional transfer benefits. Only 4% of households report benefit receipt in non-benefit distribution months.

We regress monthly non-durable consumption changes on monthly income changes and, since income changes may reflect unexpected information (e.g., job loss), instrument for income changes using the monthly child benefit disbursement pattern.[27] Specifically, we estimate the equation

$$\Delta C_{i,t} = \alpha_0 + \alpha \Delta HHincome_{i,t} + \boldsymbol{\gamma} \mathbf{X}_{i,t} + u_{i,t} \tag{7}$$

where $\Delta C_{i,t}$ is the change in non-durable consumption from month $t-1$ to month t, $\Delta HHincome_{i,t}$ is the change in household income between adjacent months, and $\mathbf{X}_{i,t}$ are additional controls for monthly consumption growth including calendar year and month indicators, survey month indicators, the change in the number of household members, and the age of the household head and its square. The substantial amount of child benefit underreporting reduces the endogenous variable, $\Delta HHincome_{i,t}$, which makes our analytical results relevant for this analysis.[28]

Table 2 reports the tests of excess sensitivity.[29] Using our full sample (column (1)), the OLS estimate of the marginal propensity of consume out of income is 0.087. After instrumenting for income changes, we find a relatively large and significant estimate of 0.181. A finding of this magnitude typically is considered to be evidence of a substantial violation of the LCPIH.

The full sample first stage estimate is 0.301 although, in the absence of underreporting, we

[26] Additional details regarding the JFIES are given in Stephens and Unayama (2011).

[27] Non-durable expenditure is the outcome commonly used in the literature (e.g., Stephens and Unayama 2011). Upon entry into the JFIES, households report total household income for the twelve months prior to the survey period. We use this measure to determine whether households are above or below the means test threshold.

[28] $HHincome_{i,t}$ includes all monthly household income sources except bonus income. Bonuses are typically received in June (a child benefit month) and/or December. The first stage estimates are sensitive to including bonuses although they remain substantially less than one and the corresponding IV estimates are still biased upwards.

[29] The standard errors are clustered at the household level.

would expect this coefficient to equal one, i.e., income increasing one for one with benefits.[30] Thus, the large degree of underreporting severely attenuates the first stage estimate. Furthermore, since the IV estimate is the ratio of the reduced form estimate to the first stage estimate, in this example we would expect the IV estimate to simply equal the reduced form estimate in the absence of underreporting. A comparison of the IV and reduced form estimates indicates that underreporting inflates the causal estimate by more than a factor of three and yields a quite different substantive interpretation of the deviation of behavior from the standard model.[31]

One theoretical mechanism for excess sensitivity is that liquidity constrained households respond to anticipated income changes. Following Zeldes (1989), we split the sample based upon whether the household is above (unconstrained) or below (constrained) current year median sample income. For both types of households we find evidence of excess sensitivity in the reduced form estimates in Table 2. However, we find similar attenuated first stage estimates and dramatically overstated 2SLS estimates for both groups due to benefit underreporting.

Assuming that underreporting occurs randomly, we examine a "correct reports" sample defined as observations where the "other social security" amount matches the amount computed for the instrument. We only lose roughly one-third of the sample as benefit changes are zero for the majority of months. The OLS estimate for the correct reports sample (column (4)) is similar to the full sample estimate. Moreover, consistent with the prediction that one yen of child benefits raises family income by one yen, we cannot reject the null that the first stage estimate is one.

However, the remaining estimates in column (4) suggest that the correct reports sample estimates are subject to selection on unobservables. The reduced form estimate for this sample is twice as large as the corresponding full sample estimate. Adjusting for selection on observables (column (5)) yields similar results. One possibility is households that are most likely to report child benefits are more likely to change their spending due to benefits. Clearly, only using non-missing observations is not a universal panacea for addressing underreported and imputed data.

[30] One possibility is that child benefits crowd out other sources of income, e.g., earnings are reduced as a behavioral response to child benefits. Even if benefits lead households to work less, it seems very unlikely that these work effort reductions would exactly coincide with the months of benefit receipt rather than be spread throughout the year.

[31] One concern is that we may misclassify ineligible households as being eligible due to mis-measured household income. However, we split the sample between high and low income households below in order to examine whether the response can be attributed to liquidity constraints, we find nearly identical first stage estimates for both samples.

4 Discussion

The continuing rise in survey non-response has increased the share of observations with imputed and underreported values for government benefits. We demonstrate analytically that the underreporting and imputation of government transfers can lead to a substantial overstatement of the causal effect of government transfers when applying instrumental variables methods to correct for the endogeneity and/or measurement error. Our empirical findings confirm these concerns.

We conclude with some observations for empirical research. First, researchers should pay close attention to the magnitude of the first stage estimates in addition to the strength of the instruments. Second, when non-reporting is not random, caution needs to be used when dropping non-responders as illustrated by our child benefit example. Third, researchers should take care to construct correct variance estimates when using imputed data, possibly through adjusted variance formulas (Abadie and Imbens 2012) or bootstrap methods (Shao and Sitter 1996). Finally, it is important to understand the imputation procedures a data provider uses. For example, for the four benefit items in the March CPS that use state of residence for imputations, this information is collapsed into five broad groupings which do not reflect geographic location, are constant over time and, thus, are unlikely to be correlated with the state-year variation used in many IV applications.

Bibliography

Abadie, Alberto and Guido W. Imbens (2012) "A Martingale Representation for Matching Estimators," *Journal of the American Statistical Association*, 107(498):833-43.

Angrist, Joshua D. and Alan B. Krueger (1992) "The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples," *Journal of the American Statistical Association*, 87(418):328-36.

Black, Dan A., Mark C Berger, and Frank A Scott (2000) "Bounding Parameter Estimates with Nonclassical Measurement Error," *Journal of the American Statistical Association*, 95(451):739-48.

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.

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.

Card, David (1996) “The Effect of Unions on the Structure of Wages: A Longitudinal Analysis,” *Econometrica*, 64(4):957-79.

Currie, Janet (2004) “The Take Up of Social Benefits,” *National Bureau of Economic Research Working Paper No. 10488.*

Engelhardt, Gary V., Jonathan Gruber, and Cynthia D. Perry (2005) “Social Security and Elderly Living Arrangements: Evidence from the Social Security Notch,” *The Journal of Human Resources*, 40(2):354-72.

Gibson, John, and Bonggeun Kim (2010) ”Non-Classical Measurement Error in Long-Term Retrospective Recall Surveys,” *Oxford Bulletin of Economics and Statistics,* 72(5): 687-95.

Greene, William H. (2008) *Econometric Analysis*, Upper Saddle River: Pearson Prentice Hall.

Heckman, James J., and Paul A. LaFontaine (2006) “Bias-Corrected Estimates of GED Returns,” *Journal of Labor Economics*, 24(3):661-700.

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.

Kane, Thomas J., Cecilia Elena Rouse, and Douglas Staiger (1999) “Estimating Returns to Schooling When Schooling is Misreported,” *National Bureau of Economic Research Working Paper*

No. 7235.

Kline, Patrick and Andres Santos (2013) "Sensitivity to Missing Data Assumptions: Theory and an Evaluation of the U.S. Wage Structure," *Quantitative Economics*, 4(2):231-67.

Kreider, Brent, John V. Pepper, Craig Gundersen, and Dean Jolliffe (2012) "Identifying the Effects of SNAP (Food Stamps) on Child Health Outcomes When Participation Is Endogenous and Misreported," *Journal of the American Statistical Association,* 107:499, 958-75.

Krueger, Alan B. and Bruce D. Meyer. (2002) "Labor Supply Effects of Social Insurance," *Handbook of Public Economics, Vol. 4, Part 7*, Chapter 33.

Krueger, Alan B. and Jörn-Steffen Pischke. (1992) "The Effect of Social Security on Labor Supply: A Cohort Analysis of the Notch Generation," *Journal of Labor Economics*, 10(4):412-37.

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.

Manski, Charles F. (1997) "Monotone Treatment Response," *Econometrica*, 65(6):1311-34.

Manski, Charles F. and John V. Pepper (2000) "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, 68(4):997-1010.

Mastrobuoni, Giovanni and Matthew Weinberg (2009) "Heterogeneity in Intra-Monthly Consumption Patterns, Self-Control, and Savings at Retirement," *American Economic Journal: Economic Policy*, 1(2):163-89.

Meyer, Bruce D., Wallace K. C. Mok and James X. Sullivan (2009) "The Under-Reporting of Transfers in Household Surveys: Its Nature and Consequences," *National Bureau of Economic Research Working Paper No. 15181.*

Organisation for Economic Co-operation and Development (OECD). 2011. Doing Better for Families. OECD Publishing.

Paulin, Geoffrey D. and David L. Ferraro (1994) "Imputing income in the Consumer Expenditure Survey," *Monthly Labor Review*, 117(12):23-31.

Ramalho, E.A. and Smith, R.J. (2013) "Discrete choice nonresponse", *Review of Economic Studies*, 80(1), 343-64.

Shao, Jun and Randy R. Sitter (1996) "Bootstrap for Imputed Survey Data," *Journal of the American Statistical Association*, 91(435):1278-88.

Shapiro, Jesse M. (2005) "Is There a Daily Discount Rate? Evidence from the Food Stamp Nutrition Cycle," *Journal of Public Economics*, 89(2-3):303-25.

Stephens Jr., Melvin (2003) "'3rd of tha Month': Do Social Security Recipients Smooth Consumption Between Checks?" *American Economic Review*, 93(1):406-22.

Stephens Jr., Melvin (2006) "Paycheque Receipt and the Timing of Consumption," *The Economic Journal*, 116(513):680-701.

Stephens Jr., Melvin and Takashi Unayama (2011) "The Consumption Response to Seasonal Income: Evidence from Japanese Public Pension Benefits," *American Economic Journal: Applied Economics*, 3(4): 86-118.

Stephens Jr., Melvin and Takashi Unayama (2015a) "Estimating the Impacts of Program Benefits: Using Instrumental Variables with Underreported and Imputed Data," *National Bureau of Economic Research Working Paper No. 21248.*

Stephens Jr., Melvin and Takashi Unayama (2015b) "Child Benefit Payments and Household Wealth Accumulation," *Japanese Economic Review*, 66(4):447-65.

Tang, Gong, Little, Roderick J.A., and Trivellore E. Raghunathan (2003) "Analysis of multivariate missing data with nonignorable nonresponse," *Biometrika,* 90(4):747-64.

Van Lancker, Wim, Joris Ghysels, and Bea Cantillon (2012) "An International Comparison of the Impact of Child Benefits on Poverty Outcomes for Single Mothers," *Centre for Social Policy*

Working Paper No. 12/03.

Zeldes, Stephen P. (1989) "Consumption and Liquidity Constraints: An Empirical Investigation," *Journal of Political Economy*, 97(2):305-46.

Table 1 - The Impact of Social Security Benefits on Shared Living Arrangements

Sample:	Pooled	Non-Imputed S.S. Income	Imputed S.S. Income	Non-Imputed S.S. Income IPW
	(1)	(2)	(3)	(4)
OLS	-0.010	-0.012	-0.005	-0.011
	(0.0005)	(0.0005)	(0.0008)	(0.0005)
First Stage	0.227	0.267	0.070	0.270
(Residuals)	(0.050)	(0.056)	(0.040)	(0.059)
Reduced Form	-0.0052	-0.0048	-0.0071	-0.0049
(Residuals)	(0.0026)	(0.0023)	(0.0045)	(0.0022)
2SLS	-0.023	-0.018	-0.097	-0.018
	(0.014)	(0.010)	(0.068)	(0.010)
N	256,710	203,983	52,727	203,983

Notes: Each estimate in the Table is from a separate regression. The dependent variable is an indicator whether the family is living in a shared arrangement. The OLS and 2SLS estimates are the coefficients on family S.S. income and also include controls listed in the text. The first stage and reduced form estimates are the coefficient on the S.S. instrument based on the Frisch-Waugh-Lovell decomposition. Standard errors are clustered at the year of birth.

Table 2 - The Impact of Japanese Child Benefits on Consumption

Sample:	Full	Below Median Income	Above Median Income	Correct Reports	Correct Reports IPW
	(1)	(2)	(3)	(4)	(5)
OLS	0.087	0.075	0.094	0.085	0.087
	(0.010)	(0.007)	(0.016)	(0.009)	(0.010)
2SLS	0.181	0.185	0.163	0.103	0.107
	(0.074)	(0.085)	(0.120)	(0.035)	(0.037)
First Stage	0.301	0.329	0.282	1.08	1.10
	(0.032)	(0.044)	(0.046)	(0.047)	(0.047)
Reduced Form	0.055	0.061	0.046	0.111	0.118
	(0.023)	(0.028)	(0.034)	(0.039)	(0.042)
N	217,312	108,391	108,921	144,595	144,595

Notes: Each estimate in the Table is from a separate regression. The dependent variable is the change in in non-durable consumption from month $t-1$ to t. The OLS and 2SLS estimates are the coefficients on the change in reported other social security income from month $t-1$ to t. The first stage and reduced form estimates are the coefficient on the programmatic child benefit change from month $t-1$ to t. Additional controls are listed in the text.

Figure 1: March CPS Share Imputed Among Those With Positive Amounts 1988-2013

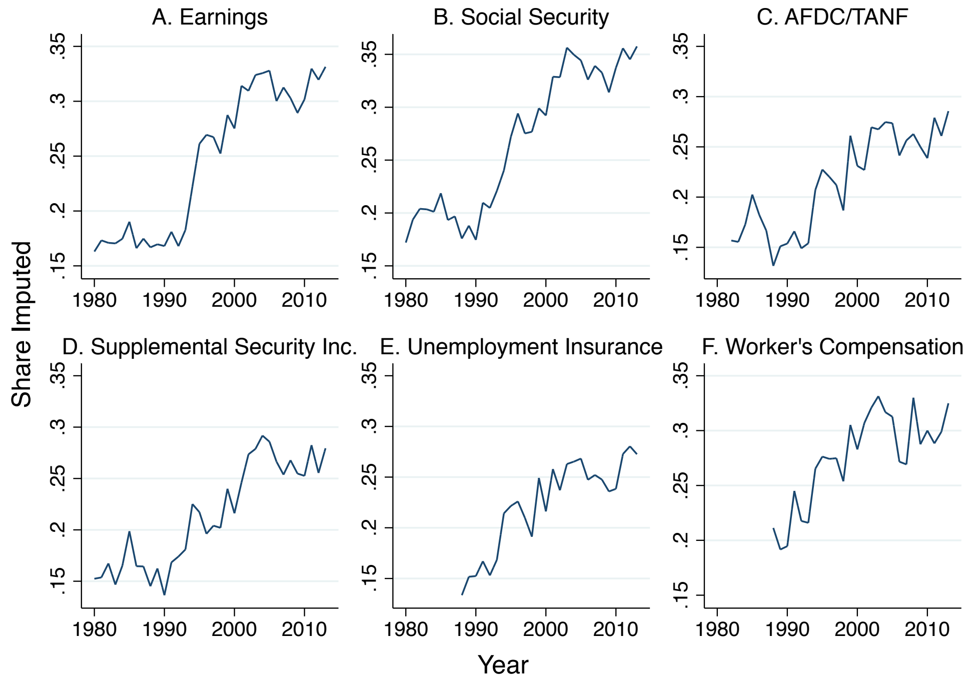

Figure 2: CPS Social Security Income and the Instrument by Year of Birth