ESTIMATING THE IMPACTS OF PROGRAM BENEFITS: USING INSTRUMENTAL VARIABLES WITH UNDERREPORTED AND IMPUTED DATA

Melvin Stephens Jr. and Takashi Unayama*

Abstract—Survey nonresponse has risen in recent years, which has increased the share of imputed and underreported values found on commonly used data sets. While this trend has been well documented for earnings, the growth in nonresponse 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 or measurement error in benefit amounts. We document the importance of failing to account for these issues using two empirical examples.

I. 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 (see surveys by Krueger & 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 with 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, & Welch 1986; Hirsch & Schumacher, 2004; Bollinger & Hirsch, 2006; Heckman & 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 nonzero report from a major income source to deem a household as a “complete income reporter,” potentially ignoring many other income sources (Paulin & 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. For example, 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 nonimputed 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 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 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.

1Meyer, 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 nonresponse and whole supplement nonresponse, the roughly 10% of households that do not provide sufficient data for the March supplement. The variable FL-665, a flag for whole supplement nonresponse, 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.
households correctly report benefits, IV estimates generated using imputed benefits are biased upward by 50%. Benefit underreporting has a similar impact on the IV estimator.

If the nonreporting is randomly assigned, a straightforward empirical solution is to use only the nonimputed subsample. If values instead are missing at random (i.e., random after conditioning on covariates), methods that account for selection on observable characteristics such as inverse propensity score weighting can be applied. With selection on unobservables, estimates using the nonimputed 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 upward by 20% to 30%. 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 & Krueger 1991; Bound et al., 1994). Black, Berger, 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. 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.

II. Econometric Framework

A. Model Setup

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

$$ y = \beta_0 + \beta_1 x + u, \quad (1) $$

where $x$ is an endogenous, continuous regressor such that $\text{Cov}(x, u) \neq 0$.\footnote{Gibson and Kim (2010) discuss a related issue for errors from using long-term retrospective recall data.}

Suppose that $z$ is a valid, continuous instrumental variable for $x$ such that $\text{Cov}(x, z) \neq 0$ and $\text{Cov}(z, u) = 0$. The
first-stage and reduced-form equations are, respectively,
\begin{align*}
x &= \pi_0 + \pi_1 z + \epsilon, \quad (2) \\
y &= \delta_0 + \delta_1 z + \epsilon. \quad (3)
\end{align*}

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 as 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 \(\hat{\pi}_1\), which can be written as \(\hat{\delta}_1 / \hat{\pi}_1\), is also consistent.

Dividing the sample into two groups, \(g = \{0, 1\}\), based on whether the endogenous regressor, \(x_{ig}\), is an actual report (\(g = 1\)) or an underreport or imputed value (\(g = 0\)), and denoting \(S_{ZX}^{Total} = \sum_{i=1}^{N_g} S_{ZX}^{g}\), \(S_{ZX}^{Between} = \sum_{i=1}^{N_g} S_{ZX}^{g} (x_i - \bar{x}) (z_i - \bar{z})\), \(S_{ZX}^{Within} = \sum_{i=1}^{N_g} S_{ZX}^{g} (z_i - \bar{z}) (x_i - \bar{x})\), and \(\bar{x}_g\) and \(\bar{z}_g\) are group and overall means, respectively, we can use the within-between decomposition (Greene, 2008) to rewrite \(\hat{\pi}_1\),
\[
\hat{\pi}_1 = \frac{S_{ZX}^{Total}}{S_{ZZ}^{Total}} = \frac{S_{ZX}^{g=1} + S_{ZX}^{g=0} + S_{ZX}^{Between}}{S_{ZZ}^{Total}}
\]
\[
= \frac{S_{ZZ}^{g=1}}{S_{ZZ}^{Total}} \times \frac{S_{ZX}^{g=1}}{S_{ZZ}^{g=1}} + \frac{S_{ZZ}^{g=0}}{S_{ZZ}^{Total}} \times \frac{S_{ZX}^{g=0}}{S_{ZZ}^{g=0}} + \frac{S_{ZZ}^{Between}}{S_{ZZ}^{Total}} \times \frac{S_{ZX}^{Between}}{S_{ZZ}^{Between}}
\]
\[
= \frac{S_{ZZ}^{g=1}}{S_{ZZ}^{Total}} \times \hat{\pi}_{1,g=1} + \frac{S_{ZZ}^{g=0}}{S_{ZZ}^{Total}} \times \hat{\pi}_{1,g=0} + \frac{S_{ZZ}^{Between}}{S_{ZZ}^{Total}} \times \hat{\pi}_{1,Between}, \quad (4)
\]

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 on within- and between-group variation in the instrument.\(^8\)

\(^7\)Most data sets contain flags to indicate which observations are imputed and which are not. This setup is also useful for understanding the impact of underreporting even though this behavior typically is not explicitly flagged.

\(^8\)The analysis focuses on imputed or underreported values of \(x\) but can be extended to either \(y\) or \(z\). Typically the instruments for benefits depend on well-measured demographic characteristics. For example, Medicaid eligibility may depend on a child’s age, and the earned income tax credit (EITC) depends on 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.

\(B.\) 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 \(\pi_1\). The final term in equation (4) converges to 0 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 on 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 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 on \(z\), \(\hat{\pi}_{1,g=0} \approx 0.9\) Thus, by equation (4), the probability limit of \(\hat{\pi}_1\) will equal \(p\pi_1 + (1 - p) \times 0 = p\pi_1\).\(^9\)

For underreporting, suppose that observed \(x\) is a constant fraction, \(\theta\), of actual \(x\). It is straightforward to show that \(\text{plim}(\hat{\pi}_1)\) for underreporters is \(\theta \times \pi_1\), and, thus, for the full sample \(\text{plim}(\hat{\pi}_1) = p\pi_1 + (1 - p) \theta \pi_1 < \pi_1\). Alternatively, when failing to report benefits (\(\theta = 0\)), 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 equation (4) and an analogous expression for the reduced-form estimator into the IV estimator:

\[
\hat{\beta}_1^{IV} = \frac{\hat{\delta}_1}{\hat{\pi}_1} = \frac{S_{ZZ}^{g=1}}{S_{ZZ}^{Total}} \times \hat{\delta}_{1,g=1} + \frac{S_{ZZ}^{g=0}}{S_{ZZ}^{Total}} \times \hat{\delta}_{1,g=0} + \frac{S_{ZZ}^{Between}}{S_{ZZ}^{Total}} \times \hat{\delta}_{1,Between}
\]

\[
\frac{S_{ZZ}^{g=1}}{S_{ZZ}^{Total}} \times \hat{\pi}_{1,g=1} + \frac{S_{ZZ}^{g=0}}{S_{ZZ}^{Total}} \times \hat{\pi}_{1,g=0} + \frac{S_{ZZ}^{Between}}{S_{ZZ}^{Total}} \times \hat{\pi}_{1,Between}.
\]

As discussed above, the denominator converges to values smaller than \(\pi_1\) 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/impacted reporters, \(\hat{\pi}_{1,g=1}\) and \(\hat{\pi}_{1,g=0}\), may not exactly equal 0.

\(^9\)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 0.

\(^{10}\)In general, whether the probability limit of \(\hat{\pi}_1\) exceeds \(\pi_1\) depends on 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 postsecondary institution. Regressions yield larger GED returns among those with imputed wages relative to those who provide wage information (\(\text{plim}(\hat{\pi}_1) > \pi_1\)).
If the nonreporting 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 nonimputed or nonunderreported observations. Alternatively, the availability of administrative data can provide a straightforward rescaling of the first-stage estimate when nonreporting is random. If the nonreporting of values follows the selection on observables assumption, a number of straightforward methods are applicable: apply inverse propensity score weighting to the nonimputed sample (Bollinger & Hirsch, 2006), construct imputations using the instruments in the imputation process (Hirsch & Schumacher, 2004; Heckman & 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 nonimputed sample only and correcting for selection on observables.

Additional methods may prove useful when confronted with nonrandom nonresponse. Recent estimation methods have focused on providing consistent point estimates when data are missing as a function of the outcomes only (Tang, Little, & Raghunathan 2003; Ramalho & Smith, 2013) rather than as a function of the regressors as with the selection on observables assumption. Another option is to construct bounds for $\pi_1$, which, since using the full sample yields consistent estimates of $\beta_1$, will help produce bounds on $\beta_1$. 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 $\pi_1$ or use the approach of Manski and Pepper (2000) to derive bounds on $\beta_1$. We do not pursue these approaches in this paper.

A. The Impact of the Social Security Notch Using Imputed Benefits

The U.S. Social Security “notch” generated a sizable change in Social Security (hereafter SS) benefits for the affected birth cohorts (Krueger & Pischke 1992). Englehardt, Gruber, and Perry (2005, hereafter EGP), using data from the 1980–1999 March CPS supplements, investigate the impact of SS income on the probability that elderly-headed families live independently. As OLS estimates of this relationship are likely inconsistent because SS 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 SS benefits. Our analysis is likely important in this case since the share of SS 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 SS benefit for every birth cohort from 1900 to 1933, using the Consumer Price Index (CPI) to deflate earnings across time. By fixing the earnings profile, the instrumental variable reflects only changes in the programmatic rules across birth cohorts. The solid line in figure 2 shows the instrument by birth cohort.

Imputations in the March CPS arise from two types of nonresponse. Item nonresponse arises when the respondent reports receiving a benefit but does not provide an amount. Whole supplement nonresponse occurs when households finish the basic CPS interview but do not participate in the March supplement. As whole supplement nonresponse has remained constant at roughly 10%, the recent increase in nonreporting is driven by item nonresponse. 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 variables (e.g., age, while some values of a single categorical characteristic are combined (e.g., race/ethnicity).

11The between-group estimate $\delta_{1,\text{between}} = 0$ if group membership is randomly assigned.

12Extending the analysis to include exogenous regressors, $w$, 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 $w$ is numerically equivalent to first separately regressing $y$ and $x$ on $w$ and then using the resulting residuals in a simple regression. As analogous procedure is available for 2SLS, we can again apply equation (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 subsamples.

13As a referee noted, two sample IV (Angrist & Krueger, 1992) using the full sample for the reduced form and the nonimputed sample for the first stage may provide an efficiency gain over 2SLS with only nonimputed data.

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

15See Kreider et al. (2012) for a recent application of bounding when SNAP (food stamp) benefits are misreported.


17EGP limit their analysis to families containing a SS 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.

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

19Prior 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 nonresponse.

20We 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
Age is used in the hot deck procedure to impute missing SS benefits. For item nonresponse, 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 and over. For whole supplement nonresponse, the procedure always groups those ages 65 and up. The long and short dashed lines in figure 2 show average SS income by birth cohort for nonimputed and imputed values, respectively. Actual SS income reports exhibit strong evidence of the notch while the imputed values do not.

ECP estimate the equation,

\[ P_{i,t} = \alpha SSIncome_{i,t} + \beta X_{i,t} + \gamma_i + \phi_i + u_{i,t} \] 

(6)

where \( P_{i,t} \) is an indicator for having a shared living arrangement; \( SSIncome_{i,t} \) is family SS income in thousands of dollars; \( X_{i,t} \) includes indicators of the head’s and spouse’s (if present) education, spouse’s age (if present), marital status (married, widowed, and divorced), white, and female; \( \gamma_i \) is a full set of indicators for the age (age + 3 for widowed and divorced women) from ages 65 to 90; \( \alpha \) is a set of survey year indicators, and \( \phi_i \) 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 arrange-ment falls as SS income increases. The impact is over twice as large in the nonimputed 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 SS income is nearly 20% larger in the nonimputed sample than in the full sample, consistent with the share of SS benefit imputations during this period. Relatedly, the first-stage relationship is more than three times larger for the nonimputed 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 on whether SS income is imputed.

---

**Table 1.** Impact of Social Security Benefits on Shared Living Arrangements

<table>
<thead>
<tr>
<th>Sample:</th>
<th>Pooled</th>
<th>Nonimputed</th>
<th>Imputed</th>
<th>Nonimputed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SS Income</td>
<td>SS Income</td>
<td>SS Income</td>
<td>IPW</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>OLS</td>
<td>-0.010</td>
<td>-0.012</td>
<td>-0.005</td>
<td>-0.011</td>
</tr>
<tr>
<td></td>
<td>(0.0005)</td>
<td>(0.0005)</td>
<td>(0.0008)</td>
<td>(0.0005)</td>
</tr>
<tr>
<td>First stage</td>
<td>0.227</td>
<td>0.267</td>
<td>0.070</td>
<td>0.270</td>
</tr>
<tr>
<td>(residuals)</td>
<td>(0.050)</td>
<td>(0.056)</td>
<td>(0.040)</td>
<td>(0.059)</td>
</tr>
<tr>
<td>Reduced form</td>
<td>-0.0052</td>
<td>-0.0048</td>
<td>-0.0071</td>
<td>-0.0049</td>
</tr>
<tr>
<td>(residuals)</td>
<td>(0.0026)</td>
<td>(0.0023)</td>
<td>(0.0045)</td>
<td>(0.0022)</td>
</tr>
<tr>
<td>2SLS</td>
<td>-0.023</td>
<td>-0.018</td>
<td>-0.097</td>
<td>-0.018</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.010)</td>
<td>(0.068)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>N</td>
<td>256,710</td>
<td>203,983</td>
<td>52,727</td>
<td>203,983</td>
</tr>
</tbody>
</table>

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 SS income and also include controls listed in the text. The first-stage and reduced-form estimates are the coefficient on the SS instrument based on the Frisch-Waugh-Lovell decomposition. Standard errors are clustered at the year of birth.

---

21Our 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 & Unayama, 2015a).

22Our estimates are weighted by the individual sampling weight for the SS 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).
While the reduced-form estimate in the imputed sample is larger than the nonimputed 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 nonimputed sample only.\(^{23}\) Assuming that SS benefits are missing at random, the results from the full sample substantially overstate the efficacy of SS benefits in reducing shared living arrangements.\(^{24}\)

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

B. Excess Sensitivity and the Underreporting of Japan’s Child Benefit

A number of developed countries provide transfers based on 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 & Weinberg, 2009; Stephens & 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 15, but this age limit was lowered to 3 in 1986 before being incrementally raised over multiple years and again reached age 15 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 3% of family income (Stephens & Unayama, 2015b).

We test whether consumption exhibits “excess sensitivity” using monthly panel data from the Japanese Family Income and Expenditure Survey (JFIES) from 1992 to 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 diary. Our data contain detailed expenditure and income categories at a monthly frequency.\(^{26}\)

Child benefits are recorded as part of an “other social security” variable that 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 nonbenefit distribution months.

We regress monthly nondurable 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

$$\Delta C_{i,t} = \alpha_0 + \alpha \Delta HHincome_{i,t} + \gamma X_{i,t} + u_{i,t}$$

(7)

where $\Delta C_{i,t}$ is the change in nondurable consumption from month $t - 1$ to month $t$, $\Delta HHincome_{i,t}$ is the change in household income between adjacent months, and $X_{i,t}$ 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 would expect this coefficient to equal 1 (i.e., income increasing 1 for 1 with benefits).\(^{30}\)

\(^{23}\) Relatedly, the estimated weights based on the decompositions shown in equations (4) and (5) are 0.791, 0.209, and $\approx 0$ for the nonimputed, imputed, and between estimators, respectively.

\(^{24}\) 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 nonimputed 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.

\(^{26}\) Additional details regarding the JFIES are given in Stephens and Unayama (2011).

\(^{27}\) Nondurable expenditure is the outcome commonly used in the literature (e.g., Stephens & 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) or December, or both months. The first-stage estimates are sensitive to including bonuses, although they remain substantially less than 1, and the corresponding IV estimates are still biased upward.

\(^{29}\) The standard errors are clustered at the household level.

\(^{30}\) One possibility is that child benefits crowd out other sources of income—for example, earnings are reduced as a behavioral response to...
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 3 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 on 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 lose roughly one-third of 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.

The continuing rise in survey nonresponse 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 nonreporting is not random, caution needs to be used when dropping nonresponders, 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 & Imbens, 2012) or bootstrap methods (Shao & 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 that 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.

### References


---

### Table 2.—The Impact of Japanese Child Benefits on Consumption

<table>
<thead>
<tr>
<th>Sample:</th>
<th>Full</th>
<th>Below Median Income</th>
<th>Above Median Income</th>
<th>Correct Reports</th>
<th>Correct Reports IPW</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
</tr>
<tr>
<td>OLS</td>
<td>0.087</td>
<td>0.075</td>
<td>0.094</td>
<td>0.085</td>
<td>0.087</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.007)</td>
<td>(0.016)</td>
<td>(0.009)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>2SLS</td>
<td>0.181</td>
<td>0.185</td>
<td>0.163</td>
<td>0.103</td>
<td>0.107</td>
</tr>
<tr>
<td></td>
<td>(0.074)</td>
<td>(0.085)</td>
<td>(0.120)</td>
<td>(0.035)</td>
<td>(0.037)</td>
</tr>
<tr>
<td>First stage</td>
<td>0.301</td>
<td>0.329</td>
<td>0.282</td>
<td>1.08</td>
<td>1.10</td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td>(0.044)</td>
<td>(0.046)</td>
<td>(0.047)</td>
<td>(0.047)</td>
</tr>
<tr>
<td>Reduced form</td>
<td>0.055</td>
<td>0.061</td>
<td>0.046</td>
<td>0.111</td>
<td>0.118</td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td>(0.026)</td>
<td>(0.034)</td>
<td>(0.039)</td>
<td>(0.042)</td>
</tr>
<tr>
<td>N</td>
<td>217,312</td>
<td>108,391</td>
<td>108,921</td>
<td>144,595</td>
<td>144,595</td>
</tr>
</tbody>
</table>

Each estimate in the table is from a separate regression. The dependent variable is the change 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.


