

NBER WORKING PAPER SERIES

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

Melvin Stephens, Jr.
Takashi Unayama

Working Paper 21248
<http://www.nber.org/papers/w21248>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2015

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). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2015 by Melvin Stephens, Jr. and Takashi Unayama. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

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

Melvin Stephens, Jr. and Takashi Unayama

NBER Working Paper No. 21248

June 2015

JEL No. C80,E21,H53,H55

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.

Melvin Stephens, Jr.

University of Michigan

Department of Economics

341 Lorch Hall

611 Tappan St.

Ann Arbor, MI 48109-1220

and NBER

mstep@umich.edu

Takashi Unayama

Institute of Economic Research

Hitotsubashi University

Naka 2-1

Kunitachi, Tokyo, Japan

unayama@ier.hit-u.ac.jp

1 Introduction

Vast economic literatures estimate the impacts of government benefits (e.g., Social Security, unemployment insurance, welfare) on an array of household-level and individual-level outcomes (e.g., see the surveys of Krueger and Meyer 2002; and Currie 2004). Since program participation is quite often a choice, a typical approach for estimating the causal effect of transfer benefits on an outcome of interest is to treat the amount of benefits received as an endogenous regressor and then implement an estimation method such as instrumental variables (IV) that exploits an exogenous source of variation in benefits. While benefit amounts reported on household surveys are typically thought to be measured with error, it is well-known that IV estimates are consistent if the measurement error in the endogenous regressor is classical. However, it is quite common for benefit amounts on surveys to be underreported or for the data provider to replace missing information with imputed values. In this paper, we demonstrate both analytically and through the use of two empirical examples that IV estimation with underreported or imputed endogenous variables will tend to overstate, and in some instances quite substantially, the causal effect of program benefits.

Recorded values of program benefits received on major surveys such as the Current Population Survey (CPS), the Decennial Census, the Panel Study of Income Dynamics (PSID), and the Survey of Income and Program Participation (SIPP) suffer from measurement error that is non-classical. First, when households acknowledge receiving a benefit but do not recall the amount, these benefits routinely are imputed. While it is well-known that a substantial and increasing share of earnings reports are imputed (Lillard, Smith, and Welch 1986; Hirsch and Schumacher 2004; Bollinger and Hirsch 2006; Heckman and LaFontaine 2006), the extent of benefit imputation has received far less attention. More generally, researchers typically fail to even acknowledge the use of imputed data. Surveying articles published in eight journals between 2004 and 2013, inclusive, and that used the CPS as the primary data source, we find that only 19 percent (16 out of 86 articles) mention that the data contains imputed values.¹

Figure 1 displays the share of imputed income measures among respondents with non-zero

¹The journals we surveyed are the *American Economic Review* (excluding the *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*. Even when acknowledged, some of these studies still treat the imputed values as actual data.

amounts in the March CPS from 1980-2013.² The Figure shows that benefit imputations have increased just as dramatically as earnings imputations over this period.³ The imputation rates of positive earnings reports are only slightly larger than those of positive Supplemental Security income reports while Social Security imputation rates exceed those of earnings.⁴

A second issue is that benefit levels are underreported in surveys. For example, the Consumer Expenditure Survey (CE) did not construct imputations prior to 2004 although roughly one in six of the consumer units were flagged as “incomplete income reporters.” Interestingly, only one valid non-zero report for a single major source of income was needed to qualify as a complete income reporter which means this designation is likely a misnomer for a number families (Paulin and Ferraro 1994). Meyer, Mok, and Sullivan (2009) find across a number of benefit categories that total benefits received, computed by aggregating and appropriately weighting responses in a given survey, fall short of administrative records detailing the total amount of benefits paid. This total dollars underreporting occurs even when including imputations. Thus, benefit imputations may fall short of actual benefits received for those individuals that acknowledge receipt but forget amounts received or some households may fail to report the receipt of benefits (or both).

In this paper, we show that underreporting and imputation can lead instrumental variables estimates to dramatically overstate the impacts of transfer programs. Hirsch and Schumacher (2004) and Bollinger and Hirsch (2006) demonstrate that OLS regression coefficients typically will be attenuated when using imputed outcomes if the imputation procedure does not explicitly condition on the regressor of interest. For example, they note that the CPS imputations do not account for union status when assigning earnings to those that fail to report an earnings amount.

²Meyer, Mok, and Sullivan (2009) present a similar set of results although in terms of dollars imputed rather than individuals. Prior to 1988, unemployment insurance and worker’s compensation benefits are combined with other government benefits and cannot be separately examined. Prior to 1982, the imputation codes for AFDC/TANF in the CPS data do not match the values listed in the codebook and thus not shown here.

³A sharp rise in earnings imputations is also found during the 1990s in the CPS Outgoing Rotation Group which collects earnings information from households at the end of their four month CPS rotation (Bollinger and Hirsch 2006). Determining which earnings values are imputed in the March CPS becomes less transparent beginning in 1988. A discussion of the issues involved along with the Stata and SAS code to correctly determine imputed earnings values can be found at <http://www.psc.isr.umich.edu/dis/data/kb/answer/1349>.

⁴These imputation rates not only account for the item non-response but also for the fact that each year roughly 10% of households do not answer (or only provided limited data for) the March CPS supplement. Whole supplement non-response information is contained in the variable FL-665 which is available on the public-use CPS data files although, notably, it is not currently provided with the IPUMS CPS data. While the whole supplement non-response flag does not appear on the public-use CPS data until 1991, 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.

Intuitively, the lack of correlation between imputed earnings and union status should attenuate the estimated union wage gap. They indeed find that that union wage gaps are substantially understated when using the full CPS sample, which includes imputed earnings values, as compared to the sub-sample of observations with non-imputed earnings information.

To estimate the impact of transfer programs, the amount of benefits received is typically treated as an endogenous regressor and, thus, is the outcome in the first stage equation when applying IV methods. If the imputation method used to replace to missing benefit amounts does not condition on the instrumental variable, the estimated first stage coefficient on the instrument generally converges to a value smaller than the true parameter as in the case of union wage gaps. For example, if the instrument relies on changes in benefit program rules that vary across states and over time but the imputation procedure does not condition on state of residence, then the imputed values will not be correlated with the instrumental variable.

Since the IV estimate can be computed as the ratio of the reduced form to the first stage coefficients on the instrument, the aforementioned inconsistency of the first stage estimator will yield an IV estimate that exceeds its true parameter value. We show analytically that when the instrument is uncorrelated with the imputed values and missing observations are randomly assigned, the IV estimator will converge to a value exceeding the true IV parameter by a factor of $1/p$ where p is the fraction of households correctly reporting benefit values. Given that only two-thirds of CPS households correctly report benefits in recent years, IV estimates generated using imputed benefits will be biased upwards by 50 percent.

The underreporting of benefits yields a similar impact on the IV estimator. If underreporting takes the form of not reporting benefits at all then the resulting IV estimator is the same as in the situation when the imputations are uncorrelated with the instrumental variable. The extent of this inconsistency falls as the degree of underreporting diminishes as we demonstrate below.

If the random assignment of non-reporting which underlies our analytical results holds, then a straightforward empirical solution is to perform the analysis only using the non-imputed sub-sample. If values instead are missing at random, then methods which account for selection on observable characteristics such as inverse propensity score weighting can be applied. However, when there is selection on unobservables, the estimates using the non-imputed sample will be inconsistent as well.

We briefly discuss possible solutions in such instances although it is worth noting that institutional features of a given transfer program may be exploited to yield consistent estimates.

We present two examples to demonstrate the empirical importance of these estimation issues. The first example makes use of the well-known U.S. Social Security “notch.”⁵ Social Security benefits were indexed to inflation beginning in 1975 whereas previously the U.S. Congress made such adjustments on an intermittent basis (Wilcox 1989). However, the first attempt at automatic cost of living adjustments actually created a “double indexation” of benefits which led to a rapid growth in payments for new benefit recipients in the years that followed. Congress ultimately decided not to reduce benefits for those already receiving payments but rather adjusted the formula so that benefits were reduced beginning with the 1917 birth cohort and culminated with the elimination of double indexation with the 1921 birth cohort.

We follow Englehardt, Gruber, and Perry (2005) by using the notch to examine the impact of Social Security income on the ability of elderly households to live independently. However, since Social Security benefits are not imputed with respect to exact age in the CPS, but rather by using broad age categories, the impact of the notch-based instrument on Social Security benefits is four times larger for families with non-imputed benefits than those with imputed benefits. We find that the IV estimate of the impact of Social Security benefits on living arrangements is biased upwards by 20 to 30 percent.

Our second example examines the consumption response of Japanese households to child benefit payments which arrive once every four months. To test whether consumption exhibits “excess sensitivity,” the typical approach based on an Euler Equation framework is to regress changes in consumption on changes in income. Since income changes can be comprised of both predictable and unpredictable elements, it is common to instrument for income changes using a predictable income change measure. We use the predictable pattern of Japanese child benefit payments, which depend upon the age and number of children in the household, as an instrument for monthly household income changes.

Since only one quarter of eligible households report receiving these child benefit payments in survey data, the first stage slope coefficient from regressing changes in reported income on the

⁵Krueger and Pischke (1992) discuss the details of the legislative changes surrounding the notch.

programmatic benefit changes is severely attenuated. We find that the IV estimate is dramatically inconsistent in the upward direction and is overstated by more than a factor of three. However, we know the first stage coefficient should equal one since benefit payments should lead to a one for one increase in monthly income. Thus, the reduced form estimate, as the numerator of the IV estimate, yields the correct excess sensitivity estimate and is substantially smaller than the IV estimate.

The measurement error induced by underreporting and imputations is akin to “mean reverting” measurement error (Bound and Krueger 1991; Bound, Brown, Duncan, and Rodgers 1994). Berger, Black, and Scott (2000) analyze the inconsistency of the IV estimator when using one noisy measure to instrument for another noisy measure when both are contaminated with mean reverting measurement error.⁶ In our analysis, the inconsistency of the IV estimator arises when the endogenous regressor is either underreported or imputed even when the instrument is perfectly measured. Our findings are appropriate for many settings where program benefit rules vary by well-measured observable characteristics such as age and state of residence such that the instrumental variable is correctly reported.⁷ Our results can easily be extended to situations in which the outcome of interest is also underreported and/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 $Cov(x, u) \neq 0$. It is straightforward to extend our analysis to include additional exogenous covariates, as we show below, as well as to binary variables.

Suppose that z is a valid, continuous instrumental variable for x such that $Cov(x, z) \neq 0$ and

⁶See Kane, Rouse, and Staiger (1999) for a related analysis and Card (1996) for the case of with a binary endogenous regressor and a binary instrument.

⁷Mogstad and Wiswall (2012) examine the consistency of the IV estimator when the instrument is only observed for a subset of observations.

$Cov(z, u) = 0$ and the population regression model for x is

$$x = \pi_0 + \pi_1 z + \epsilon \quad (2)$$

while the corresponding reduced form equation is

$$y = \delta_0 + \delta_1 z + \varepsilon \quad (3)$$

When the data are free of measurement error, the OLS estimators for the coefficients on z in equations (2) and (3), $\hat{\pi}_1$ and $\hat{\delta}_1$, respectively, are consistent since z is assumed to be exogenous.

It is well-known that the instrumental variables (IV) estimator for β_1 can be written as the ratio of the OLS estimators of the slopes in the reduced form to the first stage equations

$$\begin{aligned} \hat{\beta}_1^{IV} &= \frac{\sum_i (z_i - \bar{z})(y_i - \bar{y})}{\sum_i (z_i - \bar{z})(x_i - \bar{x})} \\ &= \frac{\left(\sum_i (z_i - \bar{z})(y_i - \bar{y}) \right) / \left(\sum_i (z_i - \bar{z})^2 \right)}{\left(\sum_i (z_i - \bar{z})(x_i - \bar{x}) \right) / \left(\sum_i (z_i - \bar{z})^2 \right)} \\ &= \frac{\hat{\delta}_1}{\hat{\pi}_1} \end{aligned} \quad (4)$$

Typically the amount of government benefits received is the endogenous regressor when examining the impact of a program on an outcome of interest. However, the instrumental variables used in the analysis typically are based on demographic characteristics that are far less likely to be misreported. For example, Medicaid eligibility is driven primarily by the age of a child while the maximum earned income tax credit (EITC) depends upon the number of children in the household. As such, the scenarios examined here in which the endogenous regressor is underreported or imputed but the instrumental variable is not are relatively common.

Suppose the data contain an indicator, s , which allows us to group observations based upon whether the endogenous regressor, x , is an actual report, $s = 1$, or an underreport/imputed values,

$s = 0$.⁸ As is well-known, we can write the OLS estimator as a weighted average of the OLS estimators for each sub-group. Applying this result to the first stage slope coefficient yields

$$\begin{aligned}
\hat{\pi}_1 &= \frac{\sum_i (z_i - \bar{z})(x_i - \bar{x})}{\sum_i (z_i - \bar{z})^2} \\
&= \frac{\sum_{s=1} (z_i - \bar{z})(x_i - \bar{x})}{\sum_i (z_i - \bar{z})^2} + \frac{\sum_{s=0} (z_i - \bar{z})(x_i - \bar{x})}{\sum_i (z_i - \bar{z})^2} \\
&= \frac{\sum_{s=1} (z_i - \bar{z})^2}{\sum_i (z_i - \bar{z})^2} \cdot \frac{\sum_{s=1} (z_i - \bar{z})(x_i - \bar{x})}{\sum_{s=1} (z_i - \bar{z})^2} + \frac{\sum_{s=0} (z_i - \bar{z})^2}{\sum_i (z_i - \bar{z})^2} \cdot \frac{\sum_{s=0} (z_i - \bar{z})(x_i - \bar{x})}{\sum_{s=0} (z_i - \bar{z})^2} \\
&= \frac{SS_{z,s=1}}{SS_z} \cdot \hat{\pi}_{1,s=1} + \frac{SS_{z,s=0}}{SS_z} \cdot \hat{\pi}_{1,s=0}
\end{aligned} \tag{5}$$

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, $\hat{\pi}_{1,s=1}$ and $\hat{\pi}_{1,s=0}$, where the weights are the share of the variation in the instrument, SS_z , belonging to each group.

2.2 Interpreting the IV estimator When Non-Response is Random

Suppose that whether x is an actual or an under/imputed report is randomly assigned where $p = \text{Prob}[s = 1]$ is the probability of providing an actual report. The first stage slope estimator for the sample of actual reporters, $\hat{\pi}_{1,s=1}$, will be a consistent estimator for π_1 . In addition, the weights in the final line of (5), $\frac{SS_{z,s=1}}{SS_z}$ and $\frac{SS_{z,s=0}}{SS_z}$, are consistent estimators of p and $1 - p$, respectively. However, the corresponding estimator for the under/imputed reporters, $\hat{\pi}_{1,s=0}$, depends upon the corresponding underreporting or imputation process.

When the endogenous regressor is imputed, the impact on $\hat{\pi}_1$ depends upon the imputation process. For example, the ‘‘hot deck’’ imputation procedure used by the U.S. Census Bureau to allocate values when items are missing due to non-response selects a replacement amount from a ‘‘donor’’ who has the same values for a small set of observed characteristics. While this procedure retains the covariance between the allocated variable and the characteristics used as part of the

⁸Most datasets contain flags to indicate which observations are imputed and which are not. Although underreporting is less likely to be flagged, this dichotomy is useful for understanding how this type of survey response can affect the estimates.

matching process, Hirsch and Schumacher (2004) and Bollinger and Hirsch (2006) note that the covariance between the allocated variable and other characteristics is not preserved. If the procedure to impute missing values for the endogenous regressor x does not depend upon the instrumental variable, then there will fail to be a correlation between x and z among the imputed observations. Thus, $\hat{\pi}_{1,s=0} \approx 0$ and the probability limit of the first stage slope estimator $\hat{\pi}_1$ will equal $p\pi_1 + (1 - p)\pi_1 = p\pi_1$.⁹

In the case of underreporting, suppose that the underreported value of x is a constant fraction θ of actual x among all underreporters. It is straightforward to show that the probability limit of the first stage slope coefficient for underreporters is $\theta \cdot \pi_1$. Thus, with underreporting, the probability limit of the first stage estimator is $p\pi_1 + (1 - p)\theta\pi_1 = \pi_1(p + [1 - p]\theta) < \pi_1$.

In the extreme case where individuals simply forget to report x (i.e., $\theta = 0$), the probability limit of $\hat{\pi}_1$ falls to $p\pi_1$. Failure to report benefits might arise when benefits are small in value or are received infrequently. Our analysis directly translates to some instances where benefits are not reported, e.g., if program benefits are universal but yet respondents fail to report receipt of these payments. In other instances, we may simply lack knowledge as to whether or not recipients in fact took up these benefits. For example, Blank and Card (1991) find take-up rates of roughly two-thirds among those eligible for Unemployment Insurance (UI) benefits. However, Meyer, Mok, and Sullivan (2009) find that roughly 75% of UI benefit payments are reported in many prominent national surveys even when including imputed benefit payments. If underreporters are treated as failing to take up benefits, then the issues raised here may be quite important for interpreting the findings.

The impact of underreported and imputed values of x on the IV estimator can be seen by

⁹In general, whether or not the probability limit of $\hat{\pi}$ exceeds π depends upon the exact imputation procedure. Bollinger and Hirsch (2006) and Heckman and LaFontaine (2006) show that the returns to a GED are overstated in the CPS. The CPS pools GED recipients with high school graduates as well those attending a post secondary institution (but not completing a four year degree). Regressions using these data yield larger GED returns among those with imputed wages relative to those who provide wage information. In this case, $plim(\hat{\pi}) > \pi$.

substituting in (5) and an analogous expression for the reduced form estimator into (4)

$$\begin{aligned}
\hat{\beta}_1^{IV} &= \frac{\hat{\delta}_1}{\hat{\pi}_1} \\
&= \frac{\frac{SS_{z,s=1}}{SS_z} \cdot \hat{\delta}_{1,s=1} + \frac{SS_{z,s=0}}{SS_z} \cdot \hat{\delta}_{1,s=0}}{\frac{SS_{z,s=1}}{SS_z} \cdot \hat{\pi}_{1,s=1} + \frac{SS_{z,s=0}}{SS_z} \cdot \hat{\pi}_{1,s=0}}
\end{aligned} \tag{6}$$

As we discussed above, the denominator will converge in probability to values smaller than π_1 when the endogenous regressor is either underreported or imputed. As long as the dependent variable in the main equation, y , is not imputed or underreported, the reduced form slope coefficients, $\hat{\delta}_{1,s=1}$ and $\hat{\delta}_{1,s=0}$, for the actual and under/imputed reporters, respectively, will yield consistent estimates of δ_1 . Thus, the probability limit of the IV estimator will exceed β_1 due to the underreporting or imputation of the endogenous regressor. In cases where either underreporters all state that they received no benefit or where the imputations are uncorrelated with the instrumental variable, the probability limit of $\hat{\beta}_1^{IV}$ becomes $\beta_1/p > \beta_1$.

Extending the analysis to include exogenous regressors is straightforward through an application of the Frisch-Waugh-Lovell Theorem. As is well-known, the OLS estimate for the coefficient on x when regressing y on x plus a vector of covariates \mathbf{w} is numerically equivalent to performing the following steps: a) regressing y on \mathbf{w} and saving the residuals, y^* , b) regressing x on \mathbf{w} and saving the residuals, x^* , and c) regressing y^* on x^* . Similarly, the 2SLS estimate for the coefficient on x when regressing y on x plus a vector of covariates \mathbf{w} using z as an instrument for x is numerically equivalent to first separately regressing y , x , and z on \mathbf{w} and saving the residuals, y^* , x^* , and z^* , respectively, and then regressing y^* on x^* using z^* as an instrument. Thus, we can continue to apply (6) except that the reduced form and first stage coefficients are based on the regressions using y^* , x^* , and z^* and the weights depend upon the shares of the variation in z^* between the actual and underreported/imputed sub-samples.

If the non-reporting of values is randomly assigned across observations then the practical solution is to simply restrict the analysis to only observations with non-imputed/non-underreported values. IV estimates generated by using these samples will yield consistent estimates of the underlying parameters of interest. Of course, the most likely scenario is that non-reporting is not

generated by simple random assignment.

2.3 Accounting for Sample Selection

A number of straightforward approaches are available when the non-reporting of values follows the selection on observables assumption. One approach is to estimate the equations of interest using the non-imputed sample only but applying inverse propensity score weighting where the weights are generated by first estimating a binary selection equation which includes the observable determinants of selection (Bollinger and Hirsch 2006). Another approach is to construct a new set of imputations which use the instruments as part of imputation process and then estimate the models using the full sample (Hirsch and Schumacher 2004; Heckman and Lafontaine 2006). This approach may be unattractive depending on the complexity of the imputation procedure. A third option is to implement a modified version of the “general correction” formula derived by Bollinger and Hirsch (2006) which adjusts the parameters for differences in the distributions of observable characteristics between actual reporters and imputed reporters. These studies find no substantive differences between the results using the non-imputed sample only and when implementing these methods to account for selection on observables.

Multiple methods can be implemented to address non-reporting which results from selection on unobservables. One approach is to jointly estimate a selection equation along with the first stage equation and outcome of interest. Of course, outside of relying on functional form assumptions to achieve identification, a plausible exclusion is required for the selection equation. Given that in many circumstances the availability of a plausible instrument to account for the endogeneity of benefits may have been a struggle to find, the requirement to find a valid exclusion restriction to achieve identification may be quite limiting.

A number of alternative methods may also 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. Another approach is to derive bounds on β_1 . Although estimation using the non-imputed sample is inconsistent when the sample is subject to selection on unobservables, using the

full sample to estimate the reduced form equation (3) still yields a consistent estimate of δ_1 . Thus, the problem which arises is to generate a consistent estimate of the first stage slope parameter, π_1 . While in general it is challenging to produce bounds on $E[x|z]$ since x may be unbounded, most government benefits have a natural set of bounds due to programmatic rules. As such, 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 .¹⁰ We do not pursue these approaches in the current paper but simply note that their use may be valuable in some contexts.

3 Empirical Examples

3.1 The Impact of the Social Security Notch Using Imputed Social Security Benefits

The U.S. Social Security “notch” generated a sizable change in Social Security benefits for the affected birth cohorts. Social Security benefits were indexed to inflation beginning in 1975 rather than having Congress make such adjustments on an intermittent basis as had been done previously. However, the first attempt at automatic cost of living adjustments actually led to a “double indexation” of benefits. The double indexation resulted from using both a non-indexed measure of average monthly earnings that increased with inflation along with having a replacement rate that increased with inflation.¹¹ Congress ultimately decided not to reduce benefits for those already receiving benefit payments but rather adjusted the formula so that benefits were reduced beginning with the 1917 birth cohort and culminated with the elimination of double indexation with the 1921 birth cohort.¹²

Englehardt, Gruber, and Perry (2005) (hereafter EGP) use the notch to investigate the impact of Social Security income on the probability that elderly-headed families are living independently using data from the 1980-1999 March CPS supplements. OLS regressions of this relationship are

¹⁰See Kreider et al (2012) for a recent application of bounding when SNAP benefits are mis-reported.

¹¹The current formula first indexes average monthly earnings before computing the benefit amount but does not index the replacement rate.

¹²The Social Security Administration’s explanation of the notch is found at <http://www.ssa.gov/history/pdf/notch.pdf>.

almost certainly inconsistent. Since Social Security benefits are a function of lifetime earnings, differences in living arrangements are more likely to reflect permanent income differences across households than the causal influence of Social Security income. Therefore, EGP use the variation across birth cohorts driven by the notch to instrument for Social Security benefits. However, the issues raised in our analytical discussion could be particularly important for the magnitude of the findings since, as shown previously in Figure 1, roughly 20% of Social Security benefit recipients in the CPS have imputed benefits during the 1980s while this figure steadily rises to nearly 30% of recipients by 1999.

EGP limit their analysis to families where the Social Security recipient is age 65 and older in the case where this recipient is male or a never married female and ages 62 and up in the case of a widowed or divorced female.¹³ They also limit the analysis to the 1900-1933 birth cohorts which are roughly symmetric around the 1916 cohort for whom the Social Security reached a peak before the laws generating the notch became effective.

EGP create their instrument for the impact of the notch on Social Security benefits by computing benefits for a fixed earnings profile across different birth cohorts. Specifically, they fix the earnings profile to reflect the median male earner in the 1916 birth cohort. They then use this profile not only to compute the Social Security benefit of someone born in 1916, but also for every birth cohort from 1900-1933 by using the Consumer Price Index (CPI) to inflate earnings across time.¹⁴ By using the same earnings profile for each birth cohort, the idea is that the instrumental variable will only reflect changes in the programmatic rules across birth cohorts. When the Social Security recipient is a widowed or divorced female, the benefit is assumed to be based on her deceased/ex-husband's earnings record. Since they find through other data sources that these husbands were on average three years older than their spouses, they base the instrument on the birth cohort three years older than the woman's birth cohort.

The solid line in Figure 2 shows the instrument by year of birth cohort.¹⁵ Benefits rise rapidly for birth cohorts up to and including the 1916 birth cohort. Due to the correction for double indexation

¹³They do not use the CPS definition of families within households but rather treat the primary CPS family as one family and all other adults in the household as individual families. See their paper for more details of this procedure.

¹⁴EGP assume that benefits are claimed beginning at age 65. They provide more details regarding the construction of the instrument in their paper.

¹⁵We thank Gary Englehardt for sharing the values of the instrument by birth cohort.

discussed above, the benefits fall rapidly for the 1917-1921 birth cohorts before remaining relatively flat for the remaining birth cohorts displayed in the figure.

Imputations in the March CPS arise from two sources: item non-response and whole supplement non-response. Item non-response occurs when the respondent indicates that they have received, e.g., Social Security benefits, but do not recall or disclose the amount of benefits received. Whole supplement non-response occurs when households finish the basic CPS interview but then refuse to participate in the March Supplement. Whole supplement non-response has remained constant at roughly ten percent such that the increase in non-reporting in recent years is driving by item non-response.¹⁶

The CPS uses the “hot deck” imputation method to allocate missing values by taking a value from a donor observation with the exact same set of characteristics. For whole supplement non-response, the entire set of responses to the supplement is taken from a donor. For the March CPS supplement, all donors are drawn from the same year. A donor observation has the exact same values for a subset of observable characteristics. 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).

In the case of item non-response for Social Security benefits, the characteristics used in the match are age (7 categories), sex (2), marital status (4), race/ethnicity (3), education (3), work status (2), and pension type (6) resulting in 6,048 possible combinations.¹⁷ If a donor cannot be found within the same combination, the race/ethnicity and work status categories are dropped from the procedure and a donor is pursued from the resulting 1,008 combinations.¹⁸

Most relevant for our analysis is that the method used to allocate missing Social Security benefits due to item non-response is not based on exact year of birth/age. Instead, the seven age categories used in this procedure are: less than 35, 35 to 54, 55 to 61, 62 to 64, 65 to 69, 70 to 74, and 75 and

¹⁶Prior to 1988, information on whole supplement imputation was contained in the data allocation flag for each income measure.

¹⁷We thank Ed Welniak for providing us with the internal Census Bureau documents detailing the characteristics used in 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.

¹⁸According to the internal Census documents, in the “rare occurrence” that a donor is not found from this second pass at generating an allocated value, an amount “will be imputed from a matrix.”

over. Since our analysis is restricted to those ages 65 and older (62 and older in the case of widowed and divorced women), there are only three (four) relevant age categories used in the imputation procedure. As such, the imputed benefits due to item non-response will fail to capture the sharp spike in Social Security income by birth cohort that is exhibited by the notch instrument.

When the entire supplement is missing, the matching procedure depends first upon marital status and labor force status. Within the five possible marital status/employment groupings, age is again used a match category. However, ages 65 and up are always grouped together, and in some cases ages 55 and up pooled together if a donor is not initially found. Thus, whole supplement non-response will also fail to capture the year of birth based movements in Social Security benefits due to the notch.

The long and short dashed lines in Figure 2 show the average Social Security income by birth cohort for non-imputed and imputed values, respectively. Whereas the actual reports of Social Security income in the CPS exhibit strong evidence of the notch, there is no evidence of a notch among the imputed values. There certainly is evidence of an increase in benefits for the pre-notch cohorts which is consistent with the fact that the imputations are based on somewhat close ages. However, the imputation procedure does not capture the rapid decline in benefits following the implementation of the notch.¹⁹

EGP estimate a model of the form

$$P_{i,t} = \theta SSIncome_{i,t} + \beta \mathbf{X}_{i,t} + \gamma_i + \alpha_t + \phi_i + u_{i,t} \quad (7)$$

where $P_{i,t}$ is an indicator for having a shared living arrangement; $SSIncome_{i,t}$ is family Social Security income measured in thousands of dollars; $\mathbf{X}_{i,t}$ includes four categories each for the education of the head and the spouse (if present), age of the spouse (if present), marital status (married, widowed, and divorced), white, and female; γ_i is a full set of indicators for the age (age+3 for widowed and divorced women) from ages 65 to 90; α_t is a set of survey year indicators, and ϕ_i is a set of indicators for the nine Census divisions. The legislative Social Security benefits based on the median earner from the 1916 birth cohort shown in Figure 2 is used as an instrument for

¹⁹Although not shown here, when the imputed values are separated between item non-response and whole imputation non-response, neither series shows any evidence of a decline in benefits for those born after 1916.

$SSIncome_{i,t}$.

We should note two differences between our analysis and EGP's. First, we only have the Social Security instrument for the 1900-1930 birth cohorts rather than through the 1933 birth cohort. Given the limited number of missing birth cohorts, we do not believe that this omission has any important impact on our analysis. Second, whereas EGP create age-by-year of birth cells for each survey year, our analysis uses individual-level data. The results from using cell-level data, shown in Appendix Table 1, are quite similar to our main results shown in Table 1.

Table 1 presents our results.²⁰ Based on our discussion above, we apply the Frisch-Waugh-Lovell theorem by first regressing the shared living indicator, Social Security income, and the instrument on the remaining exogenous covariates in the model and saving the residuals. We then use these residuals to estimate the first stage and reduced form models across both the full sample and the two sub-samples to be consistent with the decomposition shown in equation (4). The OLS and 2SLS results do not use the residuals but rather are estimated separately for the full sample as well as for each sub-sample.

Panel A in Table 1 presents our findings for the full sample of 256,710 families in column (1) whereas are findings for those without and with imputed Social Security income are shown in columns (2) and (3), respectively. Applying OLS to equation (7) finds that the probability of living in a shared arrangement falls with increases in Social Security. The impact is more than twice as large in the non-imputed sample than in the imputed sample which is consistent with attenuation bias due to imputation reducing the estimate for those with allocated benefits.

The first stage estimates vary across the columns as predicted by our analytical results. The estimated effect of the instrument on Social Security income is nearly 20% larger in the non-imputed sample than in the full sample, consistent with the share of Social Security benefits that are imputed in the CPS during this period. In addition, the estimated first stage relationship is more than three times larger for the non-imputed sample than the imputed sample which is consistent with the inability of the imputation proceed to match the sharp changes in benefits due to the notch.

There is some variability in the reduced form estimates across the three columns. The shared

²⁰All of our estimates are weighted by the individual sampling weight for the Social Security recipient. The standard errors are clustered at the year of birth level.

living arrangements measure is based strictly on the household roster which is far less likely to be affected by measurement error than Social Security income. There is no reason to expect that the relationship between this variable and the instrument would systematically vary based on whether or not Social Security income is imputed. Although the reduced form estimate for those with imputed Social Security income sample is somewhat larger than our result for the non-imputed sample, the standard errors are sufficiently large that these differences are not statistically meaningful.

The final row of Panel A shows the 2SLS estimates of the impact of Social Security income for the full sample as well as separate estimates for the non-imputed and imputed samples. We see that the 2SLS estimate of -0.023 is over 25% larger when using the full sample as opposed to the estimate of -0.018 using the non-imputed sample only. Assuming that Social Security benefits are missing at random, the results from the full sample substantially overstate the efficacy of Social Security benefits in reducing shared living arrangements.

EGP also compute elasticities by combining the point estimate along with the sample fraction living in shared arrangements and average Social Security income. Given that 25.1% of the full sample is in shared living arrangements and average Social Security income is \$5,769, and remembering that the income is measured in thousands of dollars in the regression, yields an elasticity of -0.53.²¹ Doing the same calculation for the non-imputed sample lowers the elasticity to -0.41 which means that the inclusion of the imputed values raises the estimated elasticity by over 25%.²²

In the final column of Table 1, we present results that account for the fact that the non-imputed observations may constitute a selected sample. Following Bollinger and Hirsch (2006), we use inverse propensity score weighting (IPW) to correct for sample selection based on observable characteristics. We estimate a probit for using an indicator for reporting an actual Social Security value as the dependent variable using the same regressors as we include in (7). We then re-estimate all of the equations for the non-imputed sample using the IPW based on the probit estimates. As shown in column (4), all of the results are nearly identical to those found when using the non-imputed sample (column (2)).

²¹EGP find an elasticity of -0.41. Our 2SLS estimate of -0.023 is quite close to their 2SLS estimate is -0.0205. However, in their sample 26.5% have share living arrangements and average Social Security income is \$5,323.

²²The share of the non-imputed sample living in shared arrangements is 0.248 while average Social Security income is \$5,627.

Panel B presents results for the sub-sample of widowed families for whom EGP find an elasticity of -1.30. We find that the first stage estimate for the non-imputed sample is nearly 20% larger than the pooled sample and is nearly 150% larger than in the imputed benefits sample. The reduced form is somewhat larger for the imputed households relative to the non-imputed households although the estimates for the entire sample of widows and those with non-imputed benefits are quite similar.

The 2SLS estimate of -0.112 for the full sample of widows shown in the final row of Panel B is 25% larger than the corresponding estimate for the sample of families without imputed Social Security benefits.²³ The corresponding elasticities for the entire sample of widows and those with non-imputed benefits are -1.71 and -1.37, respectively.²⁴ Again, including families with imputed benefits greatly overstates the impact of Social Security benefits on shared living arrangements.

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

Public transfers to families on the basis of the age and number of their children are prevalent in a number of developed countries (OECD 2011). In Japan, child benefits are paid three times a year, in equal amounts, in February, June, and October. The Life-Cycle/Permanent Income Hypothesis (LCPIH) predicts that households will smooth consumption in response to predictable changes in income including the receipt of regular 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).

We use monthly panel data from the Japanese Family Income and Expenditure Survey (JFIES) to test whether consumption exhibits “excess sensitivity” to the receipt of child benefit payments. Following the prior literature, we regress changes in consumption on changes in income but, since observed changes in income may reflect unexpected information, instrument for income using the changes in the child benefit payments based on the programmatic rules. However, child benefits are greatly underreported in the JFIES with only one-quarter of eligible households reporting the

²³Our full widow sample estimate of -0.112 is slightly higher than their estimate of -0.086.

²⁴In the entire sample of widows, 30.0% are in shared living arrangements and average Social Security income is \$4,568. Among widow families with non-imputed benefit data, 29.5% are in shared living arrangements and average Social Security income is \$4,482.

receipt of these payments. As such, our analytical analysis indicates that excess sensitivity tests will dramatically overstate the consumption response to child benefit receipt.

Japan introduced its child benefit system in 1972 although only households with three or more children initially received benefits from this program. Families with two or children became eligible in 1986 while eligibility was extended to families with one child in 1992. Benefits were initially available until the child was fifteen although the eligibility age was reduced to three when the second child became eligible in 1986.²⁵ In subsequent years, the age limit has been raised repeatedly, rising to six in 2000, to nine in 2004, to twelve in 2006, and to fifteen in 2009. Child benefits were means tested until 2009 after which only the age and number of children are the criteria for the receiving these payments.

While benefit eligibility changed with respect to parity and child age over this period, during the 1970s and early 1980s, the benefit level per child remained relatively stable in real terms. Beginning in 1992, benefits were set at five thousand yen per month for first and second child and at ten thousand yen for each additional child. In 2006, the monthly benefit amount was set at ten thousand yen regardless of parity but only until age three. Benefits were significantly increased between 2009 and 2012, but then subsequently decreased. Stephens and Unayama (2014) find that child benefits rise as a share of family income during this period, peaking at over 3.5 percent in the late 1990s and remaining above 3 percent in the years that followed.

Using monthly JFIES data from 1992-2009, we test whether monthly household consumption responds to the receipt of child benefit payments. We limit our sample to the period where families with one child are eligible to receive child benefits but before the means test was eliminated. Families are surveyed in the JFIES for six consecutive months. Each day during the survey period, households are instructed to enter all expenditures and income into diary. Our sample contains monthly summaries of expenditure and income in a detailed set of categories.²⁶ We construct a measure of monthly non-durable expenditure to test the LCPIH as this category is commonly used in the literature testing excess sensitivity (e.g., see Parker 1999; and Souleles 1999).

Child benefits are included in the data as part of a variable titled “other social security.”

²⁵The effective eligible age was 5 in 1972 and 10 in 1973 as a transition.

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

This variable contains benefits received from any social welfare program except for public pension payments. In the months that benefits are distributed, we find that 24% of eligible households report positive benefits for this variable. Among these positive reporters, 70% report values which exactly match what we expect they would receive based on programmatic rules. Two-thirds of the remaining positive reporting households give values that exceed the eligible value, consistent with additional benefits being received and summed into this variable. In the months in which benefits are not distributed, only 4% of households report positive benefit receipt which most likely is attributable to other social welfare programs.

To examine whether non-durable consumption is excessively sensitive to child benefits, we estimate the equation

$$\Delta C_{i,t} = \alpha_0 + \alpha \Delta HHI_{i,t} + \gamma \mathbf{X}_{i,t} + u_{i,t} \quad (8)$$

where $\Delta C_{i,t}$ is the change in consumption from month $t - 1$ to month t , $\Delta HHI_{i,t}$ 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.

Of course, simply estimating (8) using OLS does not yield a valid test of the LCPIH since any unanticipated changes in income might also influence consumption decisions. Since consumption should not be correlated with predictable income changes, we can use the anticipated change in income due to the timing of child benefit payments as an instrumental variable for income. We construct the child benefit instrument based on the programmatic rules which are a function of the age and number of children as well as the means test, i.e., whether or not household income exceeds the annual income threshold to receive child benefits.²⁷

Since we observe that roughly three-fourths of households do not report receipt of the child benefit payments even though they are eligible for these transfers, we know that the IV estimator will suffer from the inconsistency discussed in the previous section. One potential concern with this interpretation might be that non-reporters may, in fact, be ineligible for child benefits. If we

²⁷Although households only appear in the JFIES for six months, upon entry into the sample households are asked to report total household income for the twelve months prior to the survey period. We use this income measure to determine whether or not households are above or below the means test threshold.

mis-measured benefit eligibility in this way, we would expect higher rates of zero benefit reporting among higher income benefit eligible households since they are more likely to be misclassified. However, as we show below, when we split the sample between high and low income households to examine whether the response can be attributed to liquidity constraints (e.g., Zeldes 1989), we find nearly identical first stage estimates of the impact of child benefits on income for both above and below median income households.

Our measure of household income, $HHincome_{i,t}$, includes all sources of household income including regular earnings, interest payments, government transfers, etc. However, we exclude bonus income from our measure of household income. The two months in which bonuses are typically received in Japan are December and June, with June also being a month in which child benefits are received. Since bonuses can be substantial relative to monthly household income, our first stage estimates are sensitive to the inclusion of these payments.²⁸

Table 2 reports the estimates of the excess sensitivity of consumption with respect to income.²⁹ Our full sample estimates are reported in the first column. Our OLS estimates of the impact of the marginal propensity of consume out of income, 0.086, while statistically significant, is somewhat small.

Since the $HHincome$ variable also contains many sources of income, some of which may have been unanticipated, we instead instrument for the change in monthly benefits received with the programmatic child benefit amount. We find a relatively large and significant estimate of 0.190 which typically would be interpreted as a marginal propensity to consume out of the income. A finding of this magnitude typically is considered to be evidence of a substantial violation of the LCPIH.

Turning to the first stage estimate, we see that the first stage estimate is 0.302. In the absence of underreporting, we would expect this coefficient to equal one. That is, actual income should increase one for one with programmatic benefits. However, the large degree of underreporting attenuates the first stage estimate in the manner predicted by our analytical results. It is useful to

²⁸Our first stage estimate when including bonus income is nearly twice the size of the corresponding estimate shown in Table 2. However, the result is still substantially less than one meaning that the IV estimates would still be biased upwards if we were to include bonus income in our measure of household income.

²⁹The standard errors are clustered at the household level.

note that if reported benefits were only affected by classical measurement error, we would anticipate that the estimated coefficient on the programmatic benefit would still equal one since left-hand side measurement error does not affect the slope parameters in the classical errors-in-variables model.

The reduced form estimate of 0.057 for the full sample shown in the final row of Table 2 is the consumption response to the programmatic change in benefits. Since the IV estimate is the ratio of the reduced form estimate to the first stage estimate and the first stage estimate should equal one if there is no underreporting, we would expect the IV estimate to simply equal the reduced form estimate if benefits were not underreported. Thus, the underreporting of child benefits inflates the causal estimate by more than a factor of three. While the reduced form estimate still rejects the most basic LCPIH, the magnitude of the difference between this result and the above IV estimate yields a quite different substantive interpretation as to the deviation of behavior from the standard model.

One theoretical mechanism for why we might see excess sensitivity is the presence of liquidity constraints. Households that wish to borrow from future income but are unable to do so will respond to anticipated income changes since the marginal utility of a dollar today exceeds that of a dollar tomorrow. A common approach to testing for the importance of liquidity constraints is the split the sample between those that are likely to be constrained and those that are not (e.g., Zeldes 1989). Following in this tradition, we split the sample based upon whether the household is above or below the median income of child benefit eligible families in each survey year.

The 2SLS estimates shown in columns (2) and (3) of Table 2 show a large consumption response for both constrained (below median income) and unconstrained households. However, since the first stage estimate is only slightly larger for constrained households relative to unconstrained households, the reduced form findings are estimates of interest for the reasons discussed above. The reduced form estimate of 0.063 is significant for constrained group while the estimate of 0.049 is insignificant for the unconstrained group. However, these estimates are quite similar across the two groups. The key takeaway, however, is that the 2SLS estimates are dramatically overstated for both groups due to underreporting of benefit receipt.

Under the assumption that underreporting occurs randomly, we examine the sample of “correct reports” in column (4). We define correct reports as those observations in which the amount in the

“other social security” variable matches the amount found computed for the instrument based on the structure of the child benefit program. Since benefits are only provided in a limited number of months, we only lose roughly one-third of the overall sample (i.e., benefit changes are zero for the majority of months and the vast majority of households report no benefits during these months). We do, however, lose a far higher share of observations which include a month of benefit receipt. The OLS estimate for the correct reports sample is quite similar to the full sample estimate. Moreover, consistent with the prediction that each additional yen of child benefits should raise family income by the same amount, we cannot reject the null that the first stage estimate is one.

However, the remaining estimates in column (4) suggest that the results from using the correct reporters are subject to selection on unobservables. While it is true that the 2SLS estimate for the correct reporters is roughly half as large as the corresponding full sample estimate, it is still much larger than the full sample reduced form estimate of 0.057. In fact, the reduced form estimate for the correct reports is twice as large as the full sample reduced form. One potential reason for this discrepancy is that the households which are most likely to report receiving child benefits are much more likely to change their spending due to these benefits. Adjusting the correct reports results for selection on observables through inverse propensity score weighting (column (5)), does not alter this conclusion. These results suggest that simply limiting the sample to non-missing observations is not a universal panacea for addressing underreported and imputed data.

4 Discussion

Survey non-response has continued to rise in recent years which has generated an increase in the share of observations with imputed values on a number of commonly used datasets. While the importance of this trend has been long recognized for earnings data, the share of imputed values for transfer benefits has received far less attention. In addition, benefits from a number of government programs are substantially underreported by respondents even after imputations are taken into account.

We demonstrate analytically that, under the assumption that non-reporting is random, the underreporting and imputation of government transfers can lead to a substantial overstatement

of the impact of government transfers when applying instrumental variables methods to correct for the endogeneity and/or measurement error in these measures. Our examples demonstrate the empirical relevance of these results. Our findings suggest that IV estimates generated by analyses that fail to account for imputed or underreported data may substantially overstate the impact of the endogenous regressor.

Assuming that non-reporting is (conditionally) random provides straightforward adjustments to yield consistent estimates. However, as our Japanese child benefit findings based on the correct reports sample indicate, caution needs to be used when dropping non-responders. When benefits are non-randomly missing, we still can generate consistent estimates of the reduced form equation which means the issue shifts to finding a consistent estimate of the first stage parameters. In some cases, such as with Japanese child benefits, we can rely on programmatic rules to determine what the first stage estimate should be. In other cases, plausible bounds on the first stage estimate (or even the IV estimate) may be available since nearly all transfers programs have natural bounds on benefit amounts.

Our findings also suggest that researchers should pay close attention to the magnitude of the first stage estimates in addition to carefully examining the strength of the instruments. While the predicted magnitude of the first stage estimate is not always clear, as in the Social Security notch example, in some instances we may have a strong prior regarding the first stage coefficients as in our child benefit example. Deviations from these priors should not be assumed to be driven by classical measurement error but rather carefully investigated as the consequences for the consistency of the IV estimator could be quite severe.

Our review of recent published papers that we discussed in the introduction finds that the vast majority of studies fail to acknowledge the presence of imputed values. We note that even in the absence of the inconsistency issues discussed here, the usual standard errors are incorrect. Abadie and Imbens (2012) derive analytic variances for estimators using imputed data and use a Monte Carlo experiment to demonstrate that the usual variance formulas can be substantially biased. Bootstrap methods are available which yield the correct variances if the researchers are able to replicate the imputation method (Shao and Sitter 1996) although this approach has rarely been implemented in the economic literature (see Heckman and LaFontaine 2006, for an example).

We do not apply such methods here since the emphasis of our analysis is the consistency of the estimators.

Finally, the issues raised here may affect a broad range of studies which estimate the causal effect of program benefits. A common empirical practice is to rely on variation in benefits across U.S. states and over time within states. However, imputation procedures typically do not account for state of residence. For example, only four items in the March CPS use state of residence as part of the imputation process (unemployment insurance, worker's compensation, Supplement Security income, and welfare). Within these four items, state of residence is collapsed into five groups, which do not reflect geographic location, and these groupings have not changed since the 1980s. As such, these imputations will not reflect the programmatic variation being used to identify the causal impact of benefits.

Bibliography

- Abadie, Alberto and Guido W. Imbens (2012) “A Martingale Representation for Matching Estimators,” *Journal of the American Statistical Association*, 107(498):833-843.
- 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):739-748.
- 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):345-368.
- 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):957-979.
- 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.
- 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):231-267.
- 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, 958-975.
- 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):412-437.
- 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):1311-1334.
- 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.

Mogstad, Magne and Matthew Wiswall (2012) "Instrumental Variables Estimation With Partially Missing Instruments," *Economics Letters*, 114:186-189.

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

Parker, Jonathan A. (1999) "The Reaction of Household Consumption to Predictable Changes in Social Security Taxes," *American Economic Review*, 89(4):959-73.

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), 343-364.

Shao, Jun and Randy R. Sitter (1996) "Bootstrap for Imputed Survey Data," *Journal of the American Statistical Association*, 91(435):1278-1288.

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.

Souleles, Nicholas S. (1999) "The Response of Household Consumption to Income Tax Refunds," *American Economic Review*, 89(4):947-58.

Stephens Jr., Melvin (2003) "'3rd of the 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 (2014) "Child Benefit Payments and Household Wealth Accumulation," manuscript.
- Tang, Gong, Little, Roderick J.A., and Trivellore E. Raghunathan (2003) "Analysis of multivariate missing data with nonignorable nonresponse," *Biometrika*, 90, 4, pp. 747-764.
- 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.
- Wilcox, David W (1989) "Social Security Benefits, Consumption Expenditure, and the Life Cycle Hypothesis," *Journal of Political Economy*, 97(2):288-304.
- Zeldes, Stephen P. (1989) "Consumption and Liquidity Constraints: An Empirical Investigation," *Journal of Political Economy*, 97(2):305-46.

Table 1 - Social Security Notch Regressions

Sample:	Pooled	Non-Imputed S.S. Income	Imputed S.S. Income	Non-Imputed S.S. Income IPW
	(1)	(2)	(3)	(4)
<i>A. Full Sample</i>				
OLS	-0.010 (0.0005)	-0.012 (0.0005)	-0.005 (0.0008)	-0.011 (0.0005)
First Stage (Residuals)	0.227 (0.050)	0.267 (0.056)	0.070 (0.040)	0.270 (0.059)
Reduced Form (Residuals)	-0.0052 (0.0026)	-0.0048 (0.0023)	-0.0071 (0.0045)	-0.0049 (0.0022)
2SLS	-0.023 (0.014)	-0.018 (0.010)	-0.097 (0.068)	-0.018 (0.010)
N	256,710	203,983	52,727	203,983
<i>B. Widow Sample</i>				
OLS	-0.021 (0.001)	-0.023 (0.001)	-0.012 (0.002)	-0.023 (0.001)
First Stage (Residuals)	0.139 (0.056)	0.162 (0.064)	0.052 (0.042)	0.164 (0.064)
Reduced Form (Residuals)	-0.016 (0.003)	-0.015 (0.003)	-0.020 (0.006)	-0.015 (0.003)
2SLS	-0.112 (0.049)	-0.090 (0.037)	-0.306 (0.213)	-0.089 (0.036)
N	112,073	89,155	22,918	89,155

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 Social Security income and also include as controls: four categories each for the education of the head and the spouse (if present), age of the spouse (if present), marital status (married, widowed, and divorced), white, and female; a full set of indicators for the age (age+3 for widowed and divorced women) from ages 65 to 90; survey year indicators, and indicators for the nine Census divisions. The first stage and reduced form estimates are the coefficient on the Social Security instrument described in the text where the variables used in estimation are the residuals from regressions of the measures on the controls used in the OLS and 2SLS specifications. Standard errors are clustered at the year of birth.

Table 2 - Excess Sensitivity Regressions

Sample:	Full	Below Median Income	Above Median Income	Correct Reports	Correct Reports IPW
	(1)	(2)	(3)	(4)	(5)
OLS	0.086 (0.010)	0.073 (0.007)	0.093 (0.015)	0.083 (0.009)	0.085 (0.010)
2SLS	0.190 (0.074)	0.189 (0.084)	0.176 (0.121)	0.104 (0.035)	0.110 (0.037)
First Stage	0.302 (0.032)	0.331 (0.044)	0.281 (0.046)	1.08 (0.047)	1.10 (0.047)
Reduced Form	0.057 (0.023)	0.063 (0.028)	0.049 (0.034)	0.113 (0.039)	0.121 (0.042)
N	217,312	108,391	108,921	144,865	144,865

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 month t . The OLS and 2SLS estimates are the coefficients on the change in reported other social security income from month $t - 1$ to month t . The first stage and reduced form estimates are the coefficient on the programmatic change in child benefits from month $t - 1$ to month t which are computed based on the age and number of children in the household. Each regression also includes 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.

**Appendix Table 1 - Social Security Notch Regressions
Cell Mean Regressions**

Sample:	Pooled	Non-Imputed S.S. Income	Imputed S.S. Income
	(1)	(2)	(3)
<i>A. Full Sample</i>			
OLS	-0.016 (0.005)	-0.015 (0.004)	-0.004 (0.009)
First Stage	0.260 (0.016)	0.311 (0.015)	0.049 (0.026)
Reduced Form	-0.0068 (0.0022)	-0.0064 (0.0016)	-0.0088 (0.0039)
2SLS	-0.026 (0.008)	-0.021 (0.005)	-0.178 (0.104)
N	456	456	456
<i>B. Widow Sample</i>			
OLS	-0.041 (0.009)	-0.039 (0.009)	-0.0015 (0.011)
First Stage	0.137 (0.010)	0.155 (0.012)	0.065 (0.024)
Reduced Form	-0.015 (0.003)	-0.014 (0.003)	-0.019 (0.005)
2SLS	-0.111 (0.021)	-0.088 (0.018)	-0.297 (0.117)
N	456	456	456

Notes: Each estimate in the Table is from a separate regression. The data used in this table are cell means which are averaged at the year of birth-age level using sample weights to construct the averages. The dependent variable is the share of families in a cell living in a shared arrangement. The OLS and 2SLS estimates are the coefficients on family Social Security income and also include as controls: shares in four education for the education of the head and the spouse (if present), average age of the spouse (if present), share of heads in each marital status category (married, widowed, and divorced), share white, and share female; a full set of indicators for the age (age+3 for widowed and divorced women) from ages 65 to 90; survey year indicators, and share of the cell residing in each of the nine Census divisions. The first stage and reduced form estimates are the coefficient on the Social Security instrument described in the text where the variables used in estimation are the residuals from regressions of the measures on the controls used in the OLS and 2SLS specifications. Standard errors are clustered at the year of birth.

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

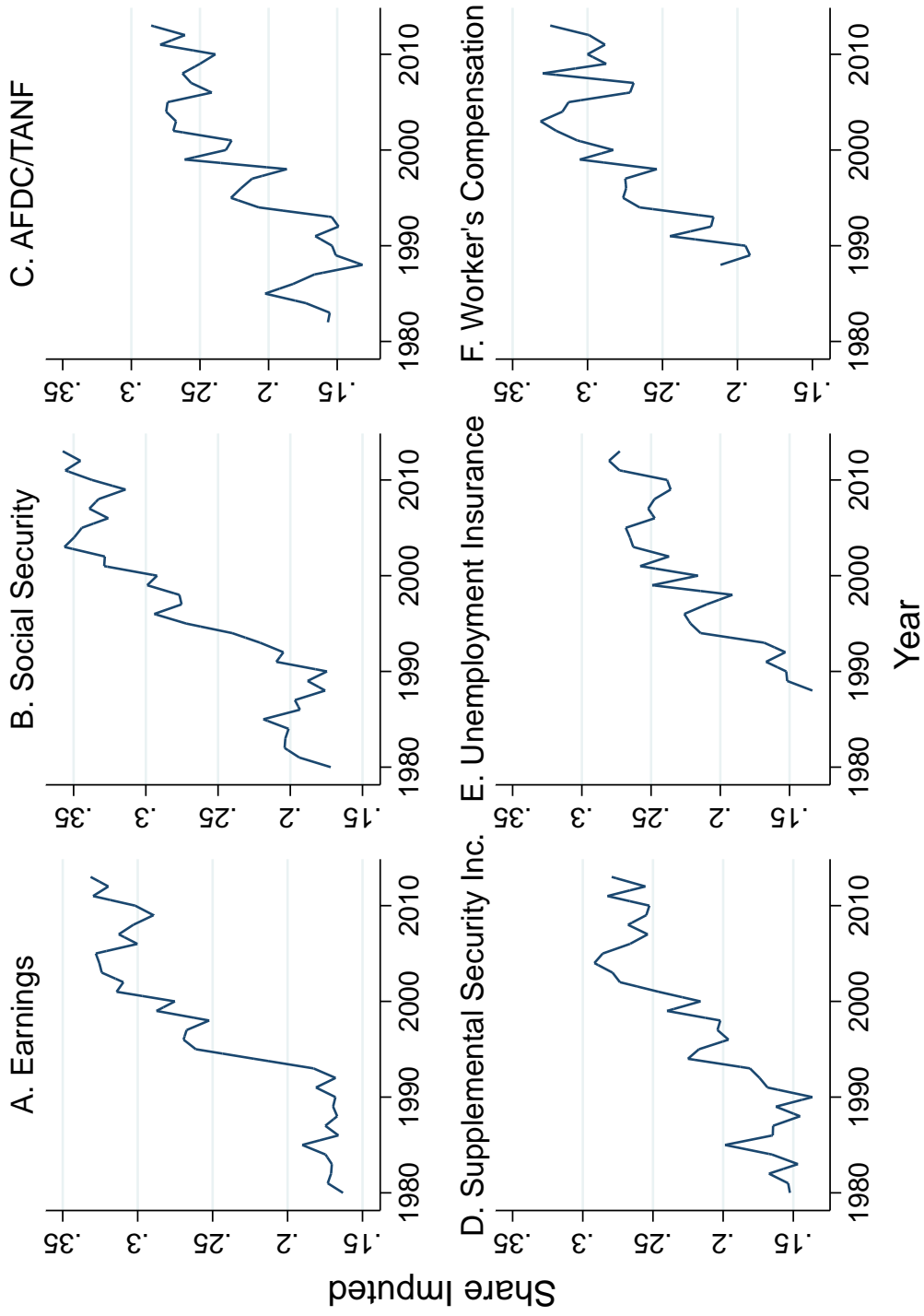


Figure 2: Social Security Income and the Instrument by Year of Birth

