

An Evaluation of Retirement Income in the CPS ASEC Using Form 1099-R Microdata*

C. Adam Bee, *US Census Bureau*

Last modified November 6, 2013

ABSTRACT

In the past several decades, individuals 65 and over have experienced remarkable declines in their official poverty rate, from 35.2 percent in 1959 to 9.1 percent in 2012. These poverty rates, however, are based on self-reported income data from the Current Population Survey Annual Social and Economic Supplement (CPS ASEC), also known as the March Supplement. Retirement income (other than Social Security or Veteran's benefits) is a significant portion of income reported by the elderly, but some analysts have suggested that the shift from pensions to individual retirement accounts may cause surveys like the CPS ASEC to mismeasure retirement income. In this paper I evaluate the quality of the retirement income data in the 2010 CPS ASEC by individually matching respondents to IRS 1099-R forms filed with tax returns for the 2009 tax year. I find that reciprocity of retirement income may be underreported on the CPS ASEC, as only a third of individuals with matched 1099-R forms reported receiving any retirement income. Much of this discrepancy, however, may be due to differences in scope, as the 1099-R covers a wider set of payments than is intended to be captured in the CPS ASEC retirement income items. In cases where positive amounts are reported on the CPS ASEC, however, those amounts appear to be quite accurate. Among cases with positive reported income, the correlation between CPS ASEC and 1099-R log amounts is 0.769. Upon modeling 1099-R reciprocity as a function of age with discontinuities at various policy-relevant age cutoffs, I find a 14.0 percentage-point jump in reciprocity at age 59½, the age at which IRA withdrawals are allowed without penalty. Taken together, these results provide suggestive evidence that retirement income may represent an underestimated contribution to the material well-being of the elderly.

* Contact: Charles Adam Bee, US Census Bureau, SEHSD, HQ-7H166D, 4600 Silver Hill Road, Washington, DC 20233-8500. E-mail: Charles.A.Bee@census.gov. Phone: 301-763-9001. The author thanks Martha Stinson and Michael Berning for invaluable assistance in obtaining access to the Form 1099-R files; Edward Welniak, Charles Nelson, and Charles Hokayem for helpful comments; and Stephen Clark for statistical review. The views expressed in this research, including those related to statistical, methodological, technical, or operational issues, are solely those of the author and do not necessarily reflect the official positions or policies of the Census Bureau. The author accepts responsibility for all errors. This paper is released to inform interested parties of ongoing research and to encourage discussion of work in progress. This paper reports the research and analysis undertaken by Census Bureau staff. It has undergone more limited review than official publications.

INTRODUCTION

In the past several decades, individuals 65 and over have experienced remarkable declines in poverty, from 35.2 percent in 1959 to 9.1 percent in 2012 (DeNavas, et al., 2013). As Figure 1 illustrates, the elderly have gone from being the poorest age group to the least poor group, outperforming even working-age adults. These declines in official poverty statistics, however, are based on self-reported household income data from the Current Population Survey Annual Social and Economic Supplement (CPS ASEC), also known as the March Supplement.

Some analysts, however, have expressed concern that the retirement income items on the CPS ASEC are designed to capture payments from defined benefit retirement plans, such as pensions. Such payments are legitimately counted as income under almost any definition of the income concept. During the late 1980s and 1990s, however, retirement income shifted toward defined contribution plans and tax-advantaged savings accounts like Individual Retirement Accounts (IRAs). These payments are a combination of dissaving, asset income, and realized capital gains.

CPS ASEC interviewers are instructed to count retirement account distributions as income if they are received on a regular basis, though contributions to such plans are also counted at the time those wages are earned. This double-counting could potentially cause overstatement of retiree income, overestimation of median incomes in areas with high concentrations of retirees (*e.g.*, Florida, Arizona), and downwardly bias poverty rates for the elderly.

On the other hand, to the extent that retirees withdraw funds from personal retirement accounts in lump sums or otherwise do not report the accumulated earnings in these accounts, then the resources available to seniors may be understated in official statistics based on the CPS.

One might think that distributions from personal retirement accounts are mainly the domain of high-income households, and thus measurement error would be unlikely to affect the poverty rate. As Table 1 shows, Sabelhaus and Schrass (2009) find that IRA distributions are proportionally largest for households in the bottom two deciles of the income distribution, suggesting that these funds may have the potential to alter the observed poverty status of a substantial fraction of families.

In this paper I evaluate the quality of the retirement income data in the 2010 CPS ASEC by matching it to individual microdata from IRS 1099-R forms filed with tax returns in the tax year 2009. Overall, I find that reported CPS ASEC values match 1099R values well.

BACKGROUND

This paper contributes to two different literatures. First, it adds to a long-running, extensive set of papers that evaluate CPS income data quality by individually matching administrative records and tax records. In a 1978 history of the landmark 1973 CPS-IRS-SSA Exact Match Study, Kilss

and Scheuren note that the first administrative data linkage to the CPS was accomplished with the March 1964 CPS, and several have been conducted intermittently since then. Nelson (1983) compares IRS interest amounts to those in the March 1983 CPS. David, *et al.* (1986) match 1981 CPS ASEC earnings to 1980 IRS tax records. Coder (1990) compares IRS and CPS incomes for married couples who reported their earnings for 1985. Roemer (2002) compares reported wage earnings collected in the CPS to the Social Security Administration's Detailed Earnings Records for the reference years 1990, 1993, and 1996.

One previous paper has compared 1099-R values to the CPS ASEC. In a 2012 IRS study estimating tax filing compliance, Erard, Payne, and Plumley compare the distributions of tax-year 2010 1099-R values to those in the public-use CPS ASEC. They find that the 1099-R has many more people with pension income than the CPS ASEC, at every level of pension income but especially at the tails in the "\$5,000 and under" category and the "over \$45,000" category. They conclude that pension income is considerably underreported the CPS ASEC, leading to an under-counting of taxpayers required to file and thus an over-estimation of the voluntary filing rate, a measure of income tax compliance. However, Erard, Payne, and Plumley did not match respondents on an individual basis; their results are based on differences in income distributions, and so they cannot clearly distinguish between potential underreporting of reciprocity and underreporting of amounts. This present paper may be the first and only study to evaluate the CPS using an individual match to the 1099R.

Second, this paper contributes to a small but growing literature focused on assessing the potential impact of the shift from DB to DC retirement plans on survey-based measures of resources available to the elderly. Iams and Purcell (2013) find that among the one-fifth of families that report receiving distributions from retirement accounts in the SIPP in 2009, median family income would be 18 percent higher if those distributions were counted as income. Meyer and Sullivan (2010) report that consumption poverty, as they measure using the Consumer Expenditure Survey, decreases even more dramatically than the drop in income poverty from the CPS ASEC, although this may be partially due to changes in insurance and Medicare as well as measurement issues arising from changes in pensions.

Sabelhaus and Schrass (2009) document the shift in the institutional frameworks around retirement income away from pensions and defined benefit (DB) plans and toward individual retirement accounts and defined contribution (DC) plans. For example, they show that full-time workers covered by employer-sponsored retirement plans have been steadily from DB plans toward DC plans, as illustrated in Figure 2. They also compare IRA withdrawals reported in the CPS to the values of withdrawals reported on tax returns, finding that only \$6 billion were reported in the CPS compared to \$166 billion reported in tax returns for 2006. Sabelhaus and Schrass suggest this might be partly due to the CPS instruction to only report "regular" withdrawals. The Survey of Consumer Finances (SCF) and the Investment Company Institute IRA Owners Survey both do not contain this instruction, and they obtain estimates of withdrawals of \$95 billion and \$72 billion, respectively. The authors also credit a 2004 change in

the question order of the SCF for increasing the amount of IRA withdrawals detected in the survey. Moving the question about withdrawals from the income section to a section specifically dedicated to aspects of the IRA increased the measured proportion of households taking withdrawals by an order of magnitude and quadrupled the estimated aggregate IRA withdrawals.

DATA

The Current Population Survey Annual Social and Economic Supplement

The Current Population Survey (CPS) is a survey of households collected monthly by the Census Bureau, sponsored by the Bureau of Labor Statistics. The CPS provides data on basic labor market outcomes such as the monthly unemployment rate. Each year in months February through April, CPS respondents also complete the Annual Social and Economic Supplement (CPS ASEC) which asks detailed questions about employment, income, and health insurance status in the previous year.²

In its income section, the CPS ASEC asks about the sources and amounts of retirement income received by household members during the previous calendar year.³ The questions are as follows:

Did [*name of household member*] receive any pension or retirement income?

- 0 = not in universe
- 1 = yes
- 2 = no

[This repeats for four different income sources. The first two sources can have income up to \$3,000,000 each, and the second two sources can have up to \$100,000.]

What was the source of income?

- 0 = none or not in universe
- 1 = company or union pension
- 2 = federal government retirement
- 3 = US military retirement
- 4 = state or local government retirement

² Data are subject to error arising from a variety of sources. For more information on sampling and non-sampling error, see www.census.gov/apsd/techdoc/cps/cpsmar11.pdf. For the purposes of the analyses in this paper, however, the relevant universe is not the U.S. civilian non-institutionalized population from which the CPS sample was drawn, but rather the actual respondents of the 2010 CPS ASEC. Values reported herein thus are not “estimates”; they are exact population counts.

³ The instrument also asks about widow or survivor income (including company or union survivor pensions, US military retirement survivor pensions, other types of survivor pensions, and regular payments from estates, trusts, annuities, or life insurance. It is unclear whether these types of income require the filing of Form 1099-R, so I do not include widow or survivor income. I also do not include disability income, for the same reason. Future research using these files may include analysis of these variables.

- 5 = US railroad retirement
- 6 = regular payments from annuities or paid insurance policies
- 7 = regular payments from Keogh or 401(k) accounts
- 8 = other sources including IRA or Keogh or don't know

How much did [*name of household member*] receive from [*source type*] during 2009?

- 0 = not in universe
- 1-3,000,000 = retirement income, first source

Like most of the items collected in the CPS ASEC, the income data are subject to post-collection editing and imputation. Editing replaces contradictory values with valid values. Imputation fills in each missing value with a valid value taken from another respondent who matches on various demographic characteristics. The table below lists the specific variables used in imputing retirement income by hot-deck procedure. If a valid value is not located from a donor that matches on the Level 1 variables, then the algorithm attempts to find a donor that matches on the less restrictive Level 2 variables.

Match variables for imputing RET_VAL(1)

Variable	Level 1	Level 2
1. Rental income source	RETSC1	RETSC1
2. Age	AGE2	AGE2
3. Sex	SEX1	SEX1
4. Race	RACE1	RACE1
5. Education	ED2	ED2
6. Social security reciprocity	SS1	
7. Interest/dividend reciprocity	ID1	
8. Worker status	WS1	

The IRS Form 1099-R extract file

The Form 1099-R (Figure 3) is an information return prepared by payers (e.g., financial institutions) of pensions, annuities, retirement or profit-sharing plans, IRA distributions or conversions, distributions from certain types of insurance contracts, or a few other related transactions. The 1099-R covers a broader set of payment types than the CPS ASEC retirement items, which may explain some of the apparent underreporting documented below. The payer sends a copy to the taxpayer/recipient and another copy to the IRS. The form is only required when the proceeds of payments are more than \$10.

The 2009 Form 1099-R file is an extract from the Information Returns Master File. It is provided to the Census Bureau by the IRS and contains almost 70 million records. Each record represents one form filed by a payer on behalf of a potential taxpayer regarding income paid during 2009. Use of the file is restricted to only a few specific statistical purposes, and each individual project

must be approved on a case-by-case basis. The file (after processing to remove personally identifiable information, as described below) includes the following fields: the Protected Identification Key (PIK, described below), the gross distribution amount in Box 1 of the form, and the distribution type, which is collapsed into two values: 1. payments from IRAs, and 2. payments from pensions, annuities, and other employer-sponsored plans.⁴ In this paper I only analyze the gross distributions, summed by PIK and matched by PIK to the CPS ASEC.

METHODS

Each observation in the CPS ASEC is assigned a sample weight according to the number of people that person represents in the U.S. civilian non-institutional population. Since this study is not concerned with estimating parameters of the U.S. population at large, however, and instead focuses on describing the properties of the CPS ASEC microdata, I do not use these sample weights. All results below are based on unweighted data.

In order to match the CPS ASEC to the Form 1099-R file, both files must first undergo Person Identification Validation System (PVS) processing, which is undertaken by the Center for Administrative Records Research and Applications (CARRA) at the U.S. Census Bureau. CARRA specializes in data integration and record linkage. The PVS process is quite complicated, and a full treatment of it is beyond the scope of this paper.⁵ In general, PVS can be thought of as a probabilistic matching process that assigns a Social Security Number to each person based on name, address, date of birth, and gender. The PVS then encrypts the Social Security Number, thereby generating anonymous person identifiers called Protected Identification Keys (PIKs). Since each person is (ideally) assigned a unique PIK, the PIK can be used to match people across different surveys, public and private administrative records, and other person-level data sources.

For the purposes of this study, a few unique features of the PVS are relevant. First, since the PVS is necessarily a probabilistic match, it introduces non-sampling error to estimates. NORC's (2011) analysis of ACS 2009 unmatched records finds that unmatched records differ from matched records on many dimensions, including reported income and employment. This suggests that the PVS process will bias income estimates. Even if the relevant characteristics of the matching records equaled the unmatched records, the PVS necessarily introduces noise which should increase the variance of estimates. Appropriate methods of accounting for this increased variance remain an open research question. Below I further discuss the specific implications of PVS matching error in relation to certain results.

⁴ Payments that do not fit into either of these categories are excluded from the extract file provided to the Census Bureau.

⁵ For details, see NORC (2011):

<http://www.norc.org/PDFs/May%202011%20Personal%20Validation%20and%20Entity%20Resolution%20Conference/PVS%20Assessment%20Report%20FINAL%20JULY%202011.pdf>

Some respondents have multiple matching Form 1099-R records. One reason for this is that a person will receive a Form 1099-R for each source of income, so a person can have multiple Forms 1099-R in a tax year. I sum the amounts from each person's matching forms.

Table 3 summarizes the results of the PVS matching process. More than 99 percent of all records in the tax year 2009 Form 1099-R file are successfully assigned a PIK, while 88 percent of records in the CPS ASEC are successfully matched to a PIK. The resulting 184,883 records in the CPS form the analysis sample, among which 23,555 records are matched to at least one Form 1099-R.

RESULTS

Comparison of unconditional distributions of reciprocity and amounts received

Table 4 is a cross-tabulation of unweighted counts at the national level, split by 1099-R availability (i.e., whether a 1099-R record is matched) and receipt of retirement income as reported in the CPS ASEC. The first column of Table 4 shows that virtually none of the CPS ASEC respondents that are out-of-universe for the retirement income items have any matching 1099-R records. This suggests both that the Census Bureau has correctly defined the universe for retirement income and that there are relatively few "false matches".

The second and third columns of Table 4 can be interpreted to show both that the CPS ASEC measures reciprocity well, and that the PVS matching process worked well. Nearly 91 percent of CPS ASEC respondents who report receiving retirement income are matched to a Form 1099-R record. 88 percent of respondents who claim they did not receive retirement income also do not have a matching 1099-R. Overall, 89 percent of respondents report reciprocity "correctly." A note of caution is warranted regarding the results shown in Table 2. The PVS processing will tend to assign PIKs to CPS ASEC respondents who correctly report personally identifying data like name and date of birth. It seems likely that there is a positive correlation between a respondent providing reliable personal data and also correctly reporting their retirement income reciprocity; if so, the results in Table 2 overstate the quality of the CPS ASEC reciprocity data.

An alternative interpretation of Table 4 is as follows. Two-thirds of respondents with a matched 1099-R are listed in the CPS ASEC as not receiving retirement income. While 13 percent of CPS ASEC respondents have matched 1099-R, only 5 percent are reported as receiving retirement income. These suggest that reciprocity is actually not measured well. The 1099-R is required for more types of income than are covered by the ASEC concept, which may explain a portion of the discrepancy, and it is difficult to judge the extent to which mismatching may play a role. Yet, taking the existence of a matched 1099-R as "truth," one can still reconcile these seemingly opposite conclusions in that the CPS ASEC has few false positives but potentially substantial false negatives.

Table 5 shows that 1099-R values correlate most closely with three sources of income, as defined in the CPS ASEC: survivor's income, disability benefits, and above all retirement income. Among the 8,076 respondents giving a positive value for retirement income, 21.9 percent of these values are within 5 percent of the matched 1099-R value.

Tables 6 and 7 are cross-tabulations of unweighted counts at the national level, split by the "true" 1099-R amount and the difference between the amount reported on the CPS ASEC and the 1099-R amount. The sample definition is the only dimension that changes across these four tables. Cells with fewer than five cases are suppressed to protect respondent confidentiality.

Table 6 uses the full sample of all 184,883 CPS ASEC cases that are successfully matched to a PIK. The first column represents respondents who are not matched to any Form 1099-R. Although there are 196 respondents in this group who reported receiving income of four or five figures (and thus unlikely to be truly misreporting), this is a small number when compared to the more than 160,000 in this group who did not report that they received any retirement income. This reflects the quality of reciprocity reporting and the PVS match as documented in the top row of Table 2.

The other columns of Table 7 show the discrepancies for respondents who are successfully matched to at least one Form 1099-R. Most of these respondents have 1099-R income of four or five digits (85 percent). In each column, a large majority of cases are off by a negative amount of the same magnitude as the 1099-R amount. Comparison of Table 6 to Table 7, which only includes cases with positive amounts of reported CPS ASEC retirement income, reveals that nearly all of these are due to respondents who report no income although they are matched to a Form 1099-R. The first value in each column of Table 7, as opposed to Table 6, reflects the exclusion of people who did not report retirement income in the CPS ASEC.

People receiving at least \$1,000 (94 percent of people who report a positive amount) sometimes substantially underreport the amounts they receive. Among those receiving income of four figures, 27 percent underreport by a four-figure amount. Also, among five-figure recipients, 27 percent underreport by five figures. Of those reporting a positive amount, only 1 percent receive an amount in six figures, but 35 percent of those underreport by at least \$100,000.

Still, these underreporters are a minority of the sample, and the correlation between reported income and actual income remains strong. Among the 5,880 cases with a matching 1099-R and reporting a positive amount on the CPS ASEC, the correlation coefficient between the log reported amount and the log actual amount is 0.769. The R^2 of a regression of log reported amounts on log actual amounts is the square of the correlation coefficient, 0.591. The mean

(signed) difference is -\$3,983, however, suggesting that underreporting among those receiving large amounts could lead to large differences in means or aggregates.⁶

Figure 4 is a scatterplot of the retirement income amounts reported in ASEC against total amounts in the Form 1099-R extract. It reveals a strong linear relationship between the two amounts, along the 45-degree line of equality. Some horizontal streaks appear in the plot, indicating that the heaping of amounts in the survey data is absent from the administrative 1099-R data. There are 5,874 points in the plot (some outliers are removed for disclosure avoidance), and the correlation coefficient is 0.769.

Figure 5 is an analogous scatterplot, but for survivor income instead of retirement income. This is intended to gauge whether the Form 1099-R extract contains amounts that respondents report as survivor income. This plot has many fewer points than the retirement income plot in Figure 4, because receiving survivor income is relatively rarer. The plot also has more variation, suggesting that many respondents report amounts that are not captured by Form 1099-R. Still, a clear line emerges along the 45-degree line of equality, suggesting that at least some of the underreporting suggested by Table 4 is due to the Form 1099-R having a somewhat larger scope than the retirement income question on the CPS ASEC. Like Figure 4, some horizontal segments appear near the 45-degree line, suggesting heaping of values. These streaks are to the right of the line, suggesting that heaping is usually due to rounding down. There are 851 points in the plot (some outliers are removed for disclosure avoidance), and the correlation coefficient is 0.504.

Figure 6 is yet another scatterplot, similar to Figures 4 and 5, but for disability income. The sparseness of this plot reflects that few people report receiving any disability income, and that there is little correlation between amounts. Yet, again a significant fraction of responses match 1099-R amounts, suggesting that at least some of the Form 1099-R amounts represent payments that respondents consider disability income. There are 173 points in the plot (some outliers are removed for disclosure avoidance), and the correlation coefficient is 0.292.

Figure 7 combines those three ASEC income sources: retirement income, survivor income, and disability income, and compares the sum to the 1099-R total amounts. It looks very similar to Figure 4, reflecting that incorporating survivor and disability makes little qualitative difference, simply because so few people report receiving those two types of income, relative to retirement income. There are 6,692 points in the plot (some outliers are removed for disclosure avoidance), and the correlation coefficient is 0.735.

Splitting by sources of ASEC retirement income

⁶ The correlation coefficient, the R^2 , and the mean signed difference are all statistically significantly different from zero at the 90 percent confidence level. The mean signed difference is the mean difference between the reported amount and the actual amount, averaged across all respondents with a matching 1099-R and reporting a positive amount on the CPS ASEC.

When respondents report that a household member received retirement income, a follow-up question asks the source of the retirement income. Table 8 summarizes the responses to that follow-up question. Table 8 reveals that the top two sources of retirement income in 2009 were company or union pensions and state or local governments, suggesting that the shift from defined benefit plans to defined contribution has yet to appear in the ASEC. Table 9 condenses these results further, categorizing the first five options as “defined benefit” sources and the others as “defined contribution,” then splitting respondents by the proportions of their retirement incomes coming from each of those two groups of sources. This further clarifies the main point from Table 8, that the retirement income reported in the CPS ASEC is overwhelmingly from defined benefit plans.

Figures 8 and 9 are the analogs of Figure 4, split by the groups identified in Table 9. Of the 5,830 people whose income was entirely from defined benefit sources, 5,495 had at least one matching Form 1099-R, and of that group 5,475 are plotted in Figure 8 (20 observations are removed for disclosure avoidance). The correlation coefficient among the group of 5,495 is 0.770, suggesting that defined benefit amounts tend to be accurately reported. As in Figure 5, horizontal streaks appear in Figure 8 just to the right of the line of equality, suggesting that heaping in the survey data is generally due to respondents rounding down.

Figure 9 repeats this exercise for the relatively few individuals who report receiving all of their retirement income was received from defined contribution accounts. Perhaps surprisingly, many of the 317 observations on this plot (6 are omitted for disclosure avoidance) lie on the 45-degree line, suggesting that defined contribution distributions also tend to be generally accurately reported. The correlation coefficient for all 323 individuals with a matching Form 1099-R is 0.741.

Splitting by Form 1099-R distribution codes

Box 7 on Form 1099-R contains distribution codes categorizing type of income being paid, as well as a checkbox for whether the distribution is from an IRA/SEP/SIMPLE plan. Data from Box 7 are not provided directly to the Census Bureau, but instead they are recoded by the IRS into a variable DTYPE, which they do include on the Form 1099-R extract. The recoding scheme is summarized in Table 2. Forms with distribution codes besides the ones included in Table 2 are entirely excluded from the extract. The recoding scheme is intended to approximately group the distribution codes into “defined contribution” and “defined benefit” groups.

This Form 1099-R distribution code grouping is explored in Figure 10. Form 1099-R amounts are summed separately for defined benefit and defined contribution values of DTYPE, then compared to CPS ASEC retirement income which has also been summed separately for the two approximate groupings of income sources. The upper-left panel of Figure 10 shows essentially the same image as Figures 4, 7, and 8.

The second panel of Figure 10, in the top right, compares ASEC DB amounts to 1099-R DC amounts, and it shows essentially no correlation, just as one would expect. The bottom-left panel has a similar mismatch of sources, and yet a linear pattern emerges. This suggests that some of the payment sources included in the DTYPE=2 “defined benefit” group in the Form 1099-R extract may be reported by respondents as defined contribution income. Further investigation is required here to more specifically identify the reason for the surprisingly high correspondence in the bottom left panel. The bottom right panel compares defined contribution amounts, and like Figure 9, this panel suggests that defined contribution amounts are often accurately reported.

Comparison of the age distributions of reciprocity

Results of comparisons of age distributions of the reciprocity of retirement income are illustrated in Figures 11, 12, and 13.

Figure 11 plots proportion receiving retirement income by age. Three series are plotted: ASEC unedited values, ASEC edited values, and 1099-R values. Each series is weighted using ASEC final sample weights. However, this weighting is not intended to estimate national-level parameters, and is done only to describe the weighted sample. Some age cells are collapsed, such that every non-empty cell has at least approximately 10 observations (and never less than 6 observations).

Figure 11 shows that the three data sources all agree that individuals through their early twenties have no retirement income. This is partly by construction in the ASEC, as the universe for retirement income consists of those age 15 and up. Among people aged roughly 25 to 50, Figure 2 shows that 1099-R records indicate that about 10 percent of people of this age group receive retirement income, even though both ASEC sources are at or near zero. The three sources begin to diverge further for the elderly, with the edited ASEC apparently underreporting reciprocity by about half.

Figure 12 is a simple transformation of the previous Figure 11, as Figure 12 illustrates the difference between reciprocity rates in the 1099-R and reciprocity rates in the edited ASEC, across all ages. This figure shows three age ranges of sharp divergence: in one’s 20s, one’s early 60s, and possibly in one’s early 70s, though increased variation in the later years makes the last of these transitions less clear.

Some of these transitions may be due to age cutoffs in the regulations governing individual retirement accounts (IRAs) and 401(k)s. At age 55 the early withdrawal penalty for 401(k)s ends, at age 59½ the early withdrawal penalty on IRAs ends, and at age 70½ account holders are required to begin taking distributions from any 401(k)s and IRAs.

Table 10 shows results from a linear probability regression model for 1099-R retirement reciprocity with a quintic control for age and discontinuities at the relevant cutoffs. Table 11 shows the results from the corresponding tests of differences between the coefficients in Table 5.

These tests in Table 11 show a relatively large and statistically significant discontinuity at age 59½, of 14.0 percentage points. This is the age at which withdrawals from IRAs are allowed without penalty, which suggests that such IRA withdrawals may represent an important source for measurement error in the CPS ASEC. Some previous literature, however, has indicated that IRA withdrawals are often taken as lump-sum distributions, which would fall outside the scope of the ASEC income definition.

Figure 13 plots fitted values from the previous regression model, with reference lines at the specified ages. This figure illustrates that the relatively large jump in 1099-R reciprocity at age 59½, of 14.0 percentage points, represents an increase in the conditional expectation from 19.2 percent at age 59 to 35.7 percent at age 60. This is a proportional increase in the probability of receiving retirement income of 79.5 percent.

CONCLUSION

In this paper I evaluate the quality of the retirement income data in the 2010 CPS ASEC by matching it to individual microdata from 1099-R forms filed with tax returns in the tax year 2009. The main outcome is that the CPS ASEC measures retirement income amounts better than might have been expected, even for amounts from defined contribution sources. Retirement income reciprocity may be less well reported, but differences in the scope of the two data sources and the potential for mismatches implies that the results of this paper set an upper bound to the misreporting of retirement income reciprocity. Many correctly matched Form 1099-R recipients might actually be correctly reporting that they did not receive retirement income, either because it is more appropriately categorized as another income source or because it is not received on a regular basis.

I also analyze the differences in the joint distribution of age and retirement income reciprocity. I find that, while 1099-R microdata confirms the ASEC for the young, the two sources diverge somewhat for individuals aged approximately 25 to 50, leading to a sharp additional divergence at age 59½, which is the age at which IRA withdrawals are allowed without penalty. Although this seems to provide suggestive evidence that IRA withdrawals may be an important source of unmeasured income, previous literature has suggested that many IRA withdrawals are taken as lump-sum distribution rather than as a regular, ongoing payment. Such lump-sum payments are excluded from the CPS ASEC income definition, and thus represent a difference in the respective scopes of the two data sources rather than measurement error *per se*.

This paper should be considered a description of an ongoing research project rather than a final report. Several avenues for future research immediately recommend themselves. First, the extent and nature of the misreporting has not yet been fully explored. An analysis that identifies which demographic characteristics are most predictive of misreporting would be helpful. The age distribution analysis in this paper is a first step in this direction. Disaggregating retirement

income by its constituent sources (e.g., pensions vs. IRAs) may shed light on the contention that the shift in recent years to defined contribution plans has diminished the accuracy of the CPS ASEC.

Second, the opportunities for methodological improvement can be more fully explored in the near future. For example, the 1099-R data may be useful for generating more accurate imputed values. An assessment of this possibility is well within the scope of this project.

Third, the findings of this project will have several implications for the measurement of poverty, income inequality, and the well-being of the elderly, which will need to be more explicitly characterized. For example, it would be useful to know whether misreporting increases the measured official poverty rate or decrease it, in an analysis similar to that of Hokayem, Ziliak, and Bollinger (2012), who gauge the impact of wage misreporting and imputation on poverty rates. Such calculations are important goals of the research agenda initiated with this paper.

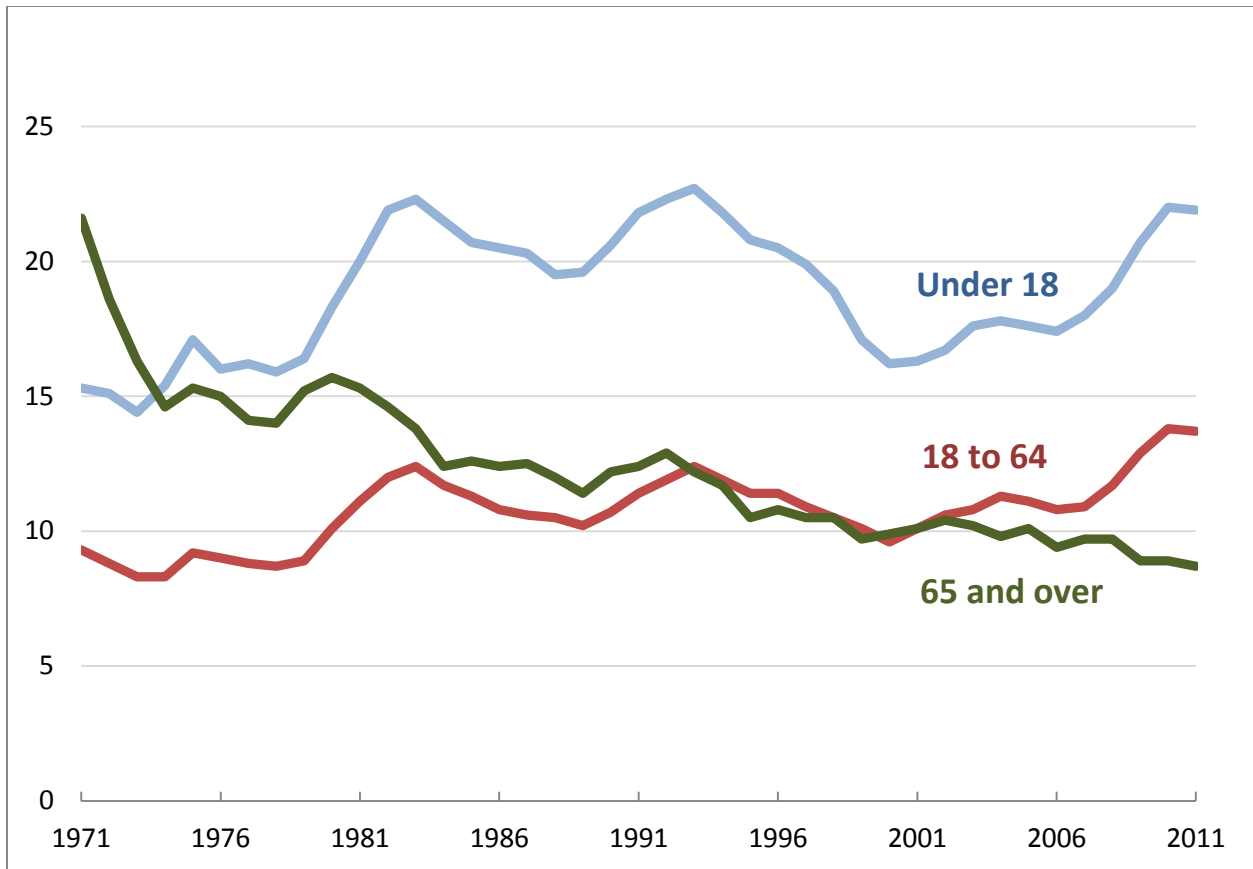
DRAFT

REFERENCES

- Bryant, Victoria L., Sarah Holden, and John Sabelhaus. 2011. "Qualified Retirement Plans: Analysis of Distribution and Rollover Activity." Pension Research Council Working Paper No. 2011-01. Philadelphia, PA: University of Pennsylvania. Available at <http://www.pensionresearchcouncil.org/publications/document.php?file=935>
- Coder, John F. 1990. "Exploring Nonsampling Errors in the Wage and Salary Income Data from the March Current Population Survey." Paper presented at the Allied Social Sciences Association/Society of Government Economists Meetings, Washington, D.C., December 28-30.
- Czajka, John L. and Gabrielle Denmead. 2012. "Getting More from Survey Income Measures: Empirically-based Recommendations for Improving Accuracy and Efficiency." FCSM paper.
- Czajka, John L. and Gabrielle Denmead. 2012. "Income Measurement for the 21st Century: Updating the Current Population Survey." Mathematica Policy Research final report. August 8.
- David, Martin, Roderick J.A. Little, Michael E. Samuhel, and Robert K. Triest. 1986. "Alternative Methods for CPS Income Imputation." *Journal of the American Statistical Association* 81(393): 29-41.
- DeNavas-Walt, Carmen, Bernadette D. Proctor, and Jessica C. Smith, 2013. U.S. Census Bureau, Current Population Reports, P60-245, *Income, Poverty, and Health Insurance Coverage in the United States: 2012*, U.S. Government Printing Office, Washington, DC.
- Gustman, Alan L., Thomas L. Steinmeier, and Nahid Tabatabai. 2012. "Mismeasurement of Pensions Before and After Retirement: The Mystery of the Disappearing Pensions with Implications for the Importance of Social Security as a Source of Retirement Support." NBER Working Paper No. 18542. November.
- Hokayem, Charles, James Ziliak, and Chris Bollinger. 2012. "CPS Non-Response and the Measurement of Poverty." Working paper, presented at the 2012 Joint Statistical Meetings.
- Iams, Howard M. and Patrick J. Purcell. 2013. "The Impact of Retirement Account Distributions on Measures of Family Income." *Social Security Bulletin* 73(2): 77-84.
- Kilss, Beth, and Frederick J. Scheuren. 1978. "The 1973 CPS-IRS-SSA Exact Match Study." Reprinted in *Social Security Bulletin*, July 1988, 51(7): 23-31.
- Meyer, Bruce D., and James X. Sullivan. 2010. "Consumption and Income of the Poor Elderly Since 1960." Working paper.

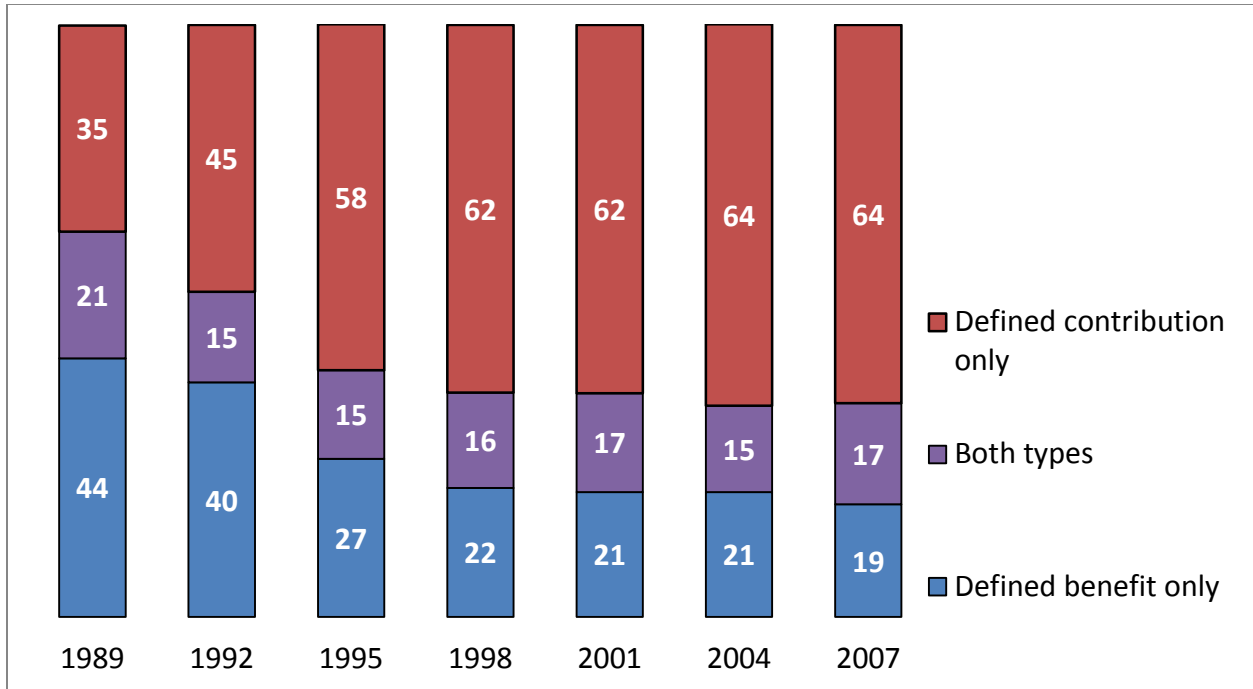
- Moore, Jeffrey C., Linda L. Stinson, and Edward J. Welniak, Jr. 2000. "Income Measurement Error in Surveys: A Review." *Journal of Official Statistics* 16(4): 331-361.
- Nelson, Charles T. 1983. "A Comparison of CPS and IRS Interest Amounts from the March 1983 CPS-1982 IMF Exact Match File," unpublished Census Bureau memorandum.
- Erard, Brian, Mark Payne, and Alan Plumley. 2012. "Advances in Nonfiling Measures." Paper presented at the 2012 IRS-TPC Research Conference, June 21.
- NORC at the University of Chicago. 2011. "Assessment of the U.S. Census Bureau's Person Identification Validation System." Final report presented to U.S. Census Bureau, March 31.
- Roemer, Marc. 2002. "Using Administrative Earnings Records to Assess Wage Data Quality in the March Current Population Survey and the Survey of Income and Program Participation." Technical Paper No. TP-2002-22, Longitudinal Employer-Household Dynamics, US Census Bureau, Washington, DC. Available at <ftp://ftp2.census.gov/ces/tp/tp-2002-22.pdf>
- Sabelhaus, John, and Daniel Schrass. 2009. "The Evolving Role of IRAs in U.S. Retirement Planning." *Investment Company Institute Research Perspective* 15(3): 1-32. November.

Figure 1: Poverty rate by age



Source: DeNavas-Walt, Proctor, and Smith (2013)

Figure 2: Percentage of covered, full-time workers with a DB or DC plan at their current job



Source: Sabelhaus and Schrass (2009) tabulations of the Federal Reserve Board Survey of Consumer Finances, 1989-2007.

Figure 3: Sample copy of the 2009 Form 1099-R

9898 VOID CORRECTED

PAYER'S name, street address, city, state, and ZIP code		1 Gross distribution \$		OMB No. 1545-0119 <div style="font-size: 2em; font-weight: bold; text-align: center;">2009</div> Form 1099-R		Distributions From Pensions, Annuities, Retirement or Profit-Sharing Plans, IRAs, Insurance Contracts, etc. Copy A For Internal Revenue Service Center File with Form 1096. For Privacy Act and Paperwork Reduction Act Notice, see the 2009 General Instructions for Forms 1099, 1098, 3921, 3922, 5498, and W-2G.	
		2a Taxable amount \$		Total distribution <input type="checkbox"/>			
		2b Taxable amount not determined <input type="checkbox"/>					
PAYER'S federal identification number	RECIPIENT'S identification number	3 Capital gain (included in box 2a) \$		4 Federal income tax withheld \$			
RECIPIENT'S name		5 Employee contributions /Designated Roth contributions or insurance premiums \$		6 Net unrealized appreciation in employer's securities \$			
Street address (including apt. no.)		7 Distribution code(s)	IRA/SEP/SIMPLE <input type="checkbox"/>	8 Other \$			
City, state, and ZIP code		9a Your percentage of total distribution %		9b Total employee contributions \$			
1st year of desig. Roth contrib.		10 State tax withheld \$		11 State/Payer's state no.		12 State distribution \$	
Account number (see instructions)		13 Local tax withheld \$		14 Name of locality		15 Local distribution \$	

Form 1099-R

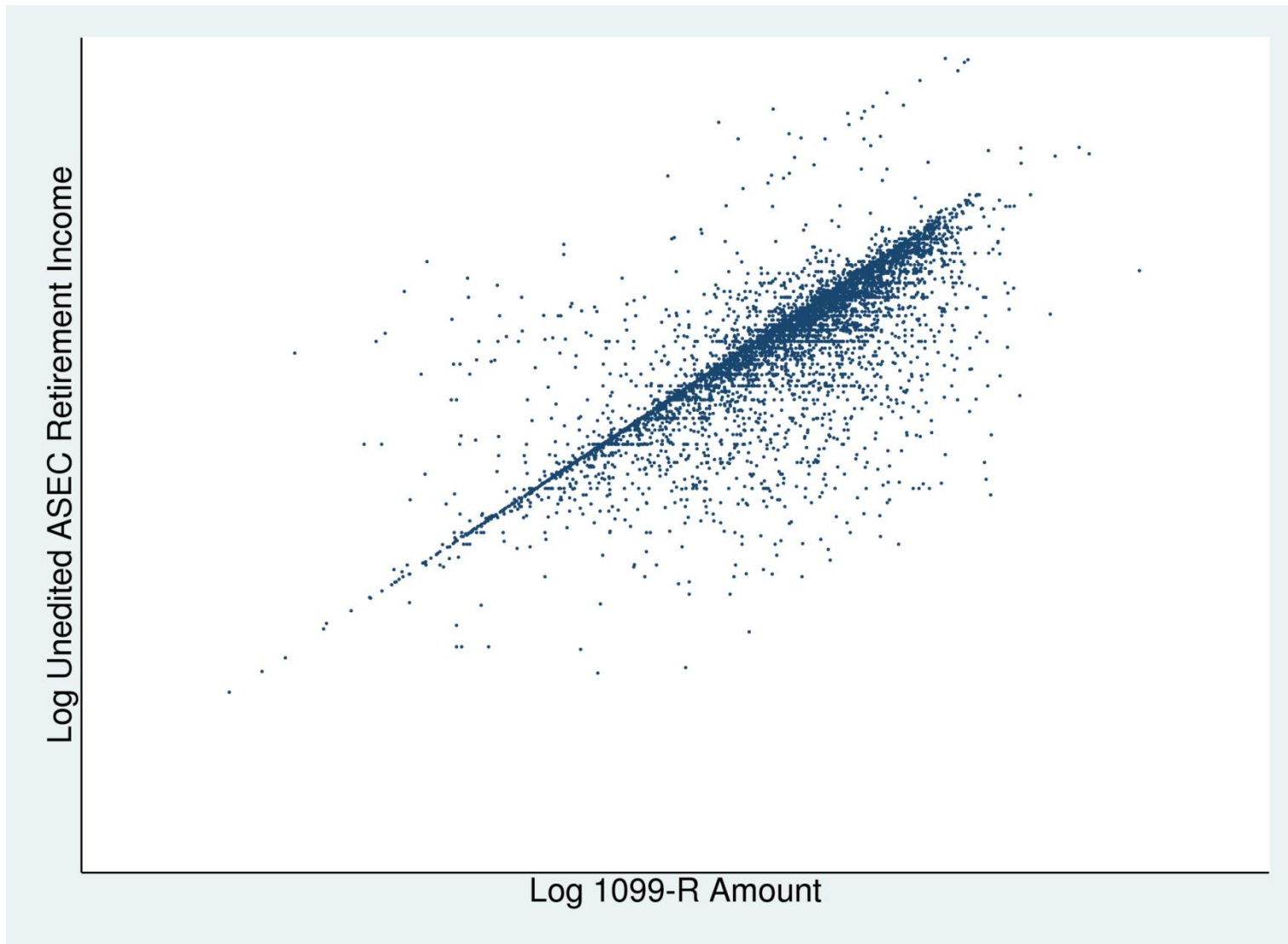
Cat. No. 14436Q

Department of the Treasury - Internal Revenue Service

Do Not Cut or Separate Forms on This Page — Do Not Cut or Separate Forms on This Page

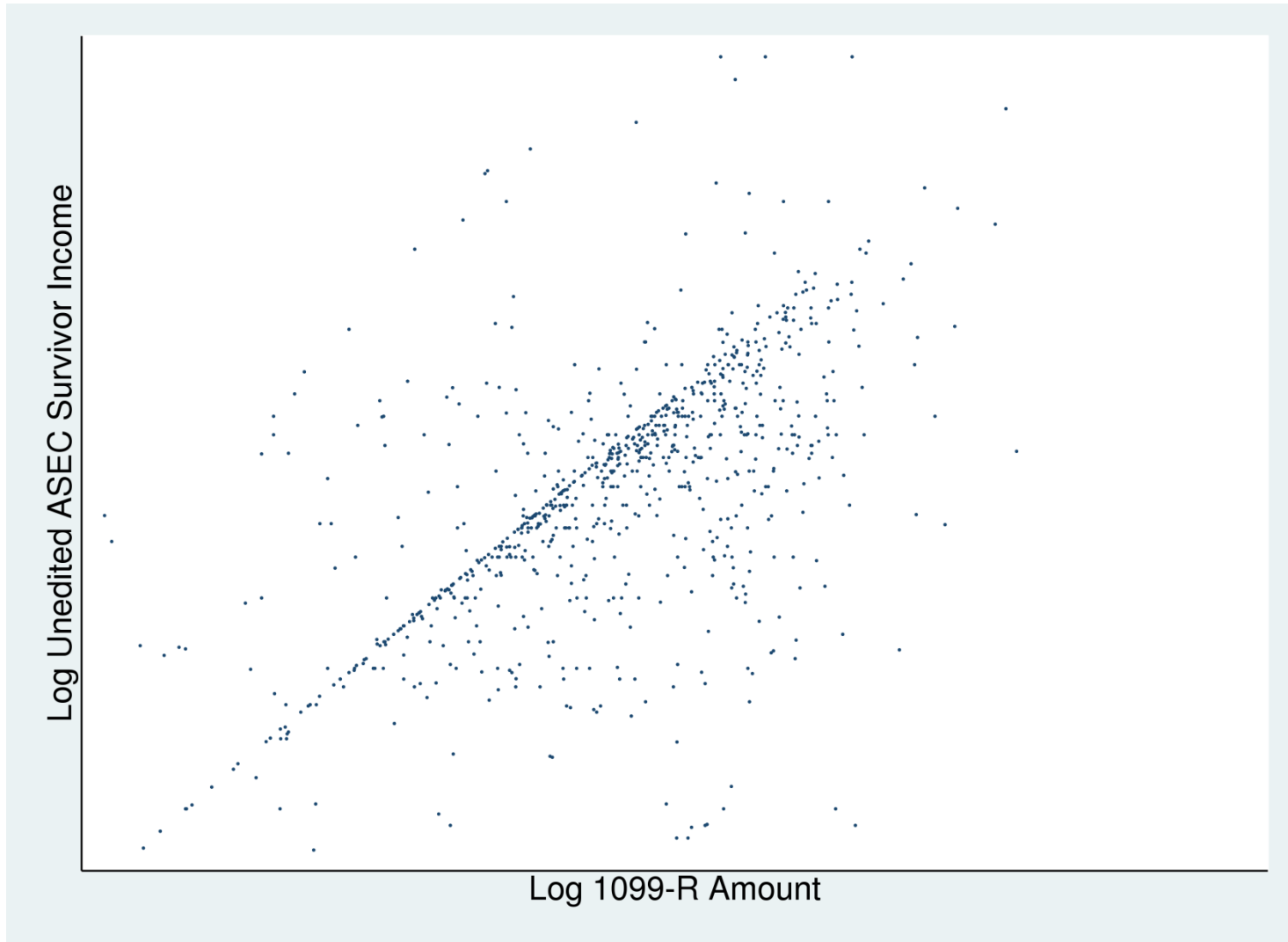


Figure 4: Scatterplot of individual log unedited ASEC retirement income by log 1099-R amount



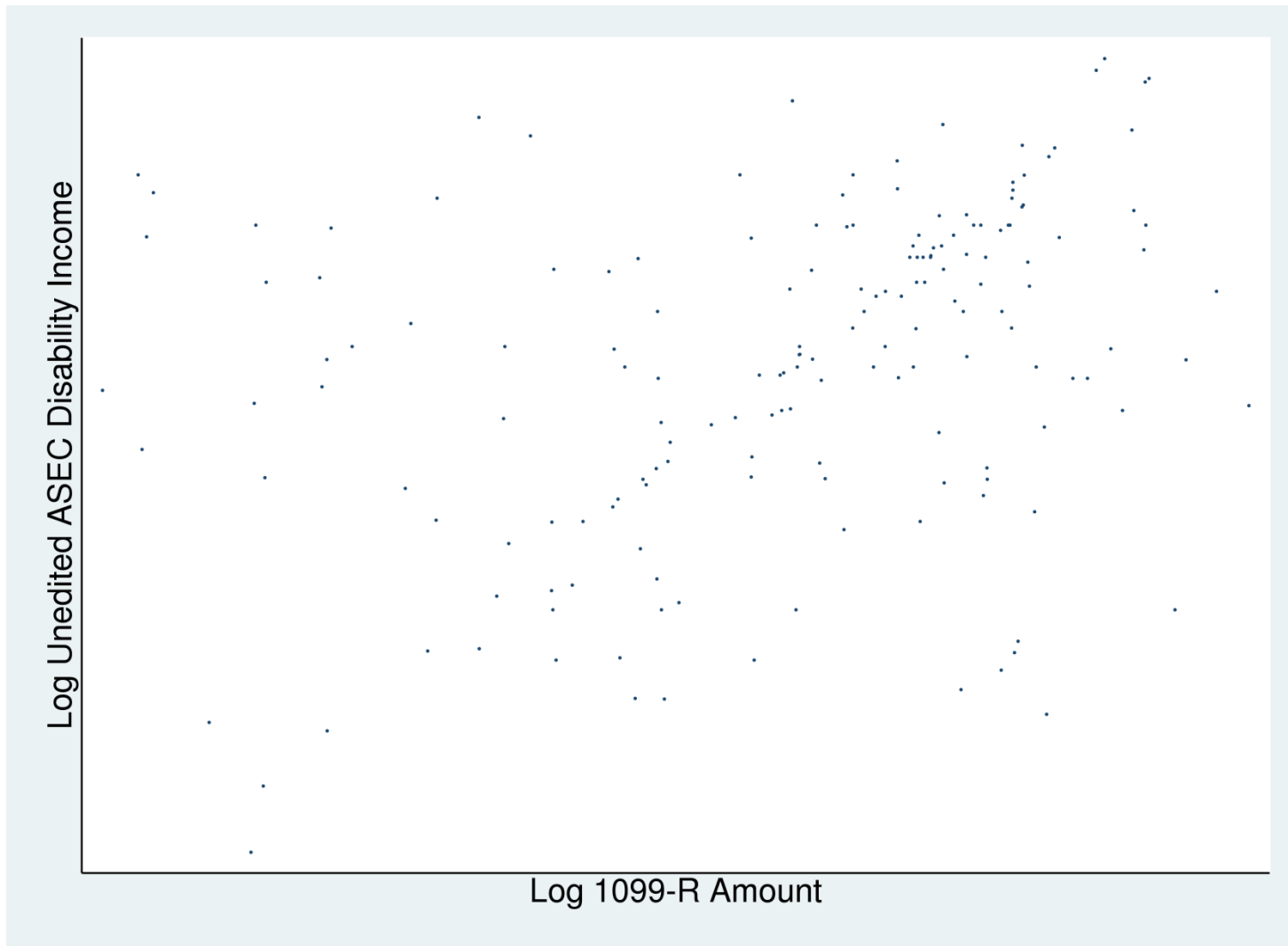
Sources: 2010 Current Population Survey Annual Social and Economic Supplement, IRS Tax Year 2009 Form 1099-R microdata.

Figure 5: Scatterplot of individual log unedited ASEC survivor income by log 1099-R amount



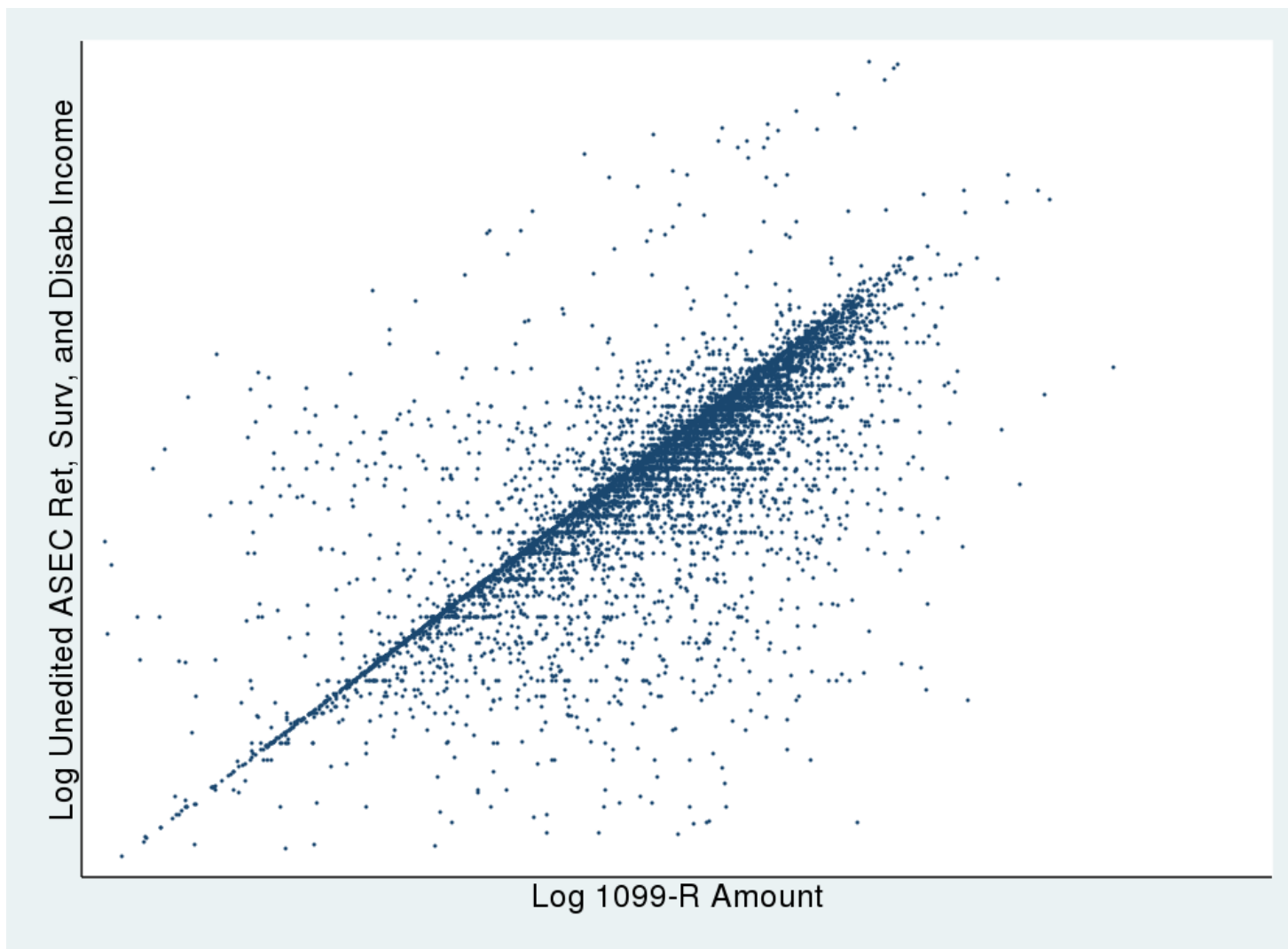
Sources: 2010 Current Population Survey Annual Social and Economic Supplement, IRS Tax Year 2009 Form 1099-R microdata.

Figure 6: Scatterplot of individual log unedited ASEC disability income by log 1099-R amount



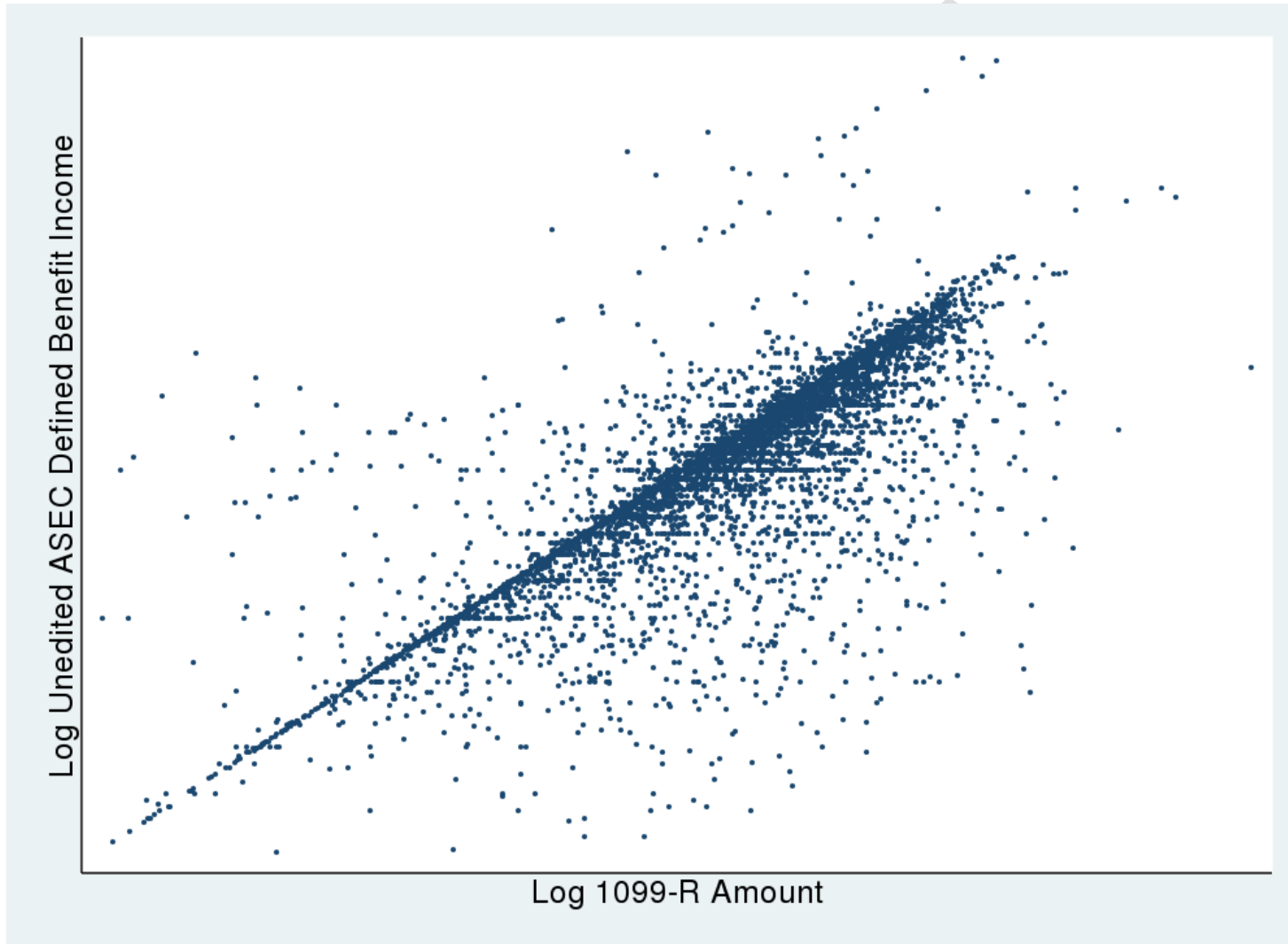
Sources: 2010 Current Population Survey Annual Social and Economic Supplement, IRS Tax Year 2009 Form 1099-R microdata.

Figure 7: Scatterplot of individual log unedited ASEC sum of retirement, survivor, and disability income by log 1099-R amount



Sources: 2010 Current Population Survey Annual Social and Economic Supplement, IRS Tax Year 2009 Form 1099-R microdata.

Figure 8: Scatterplot of individual log unedited ASEC retirement income by log 1099-R amount, including only respondents whose retirement sources are all defined-benefit



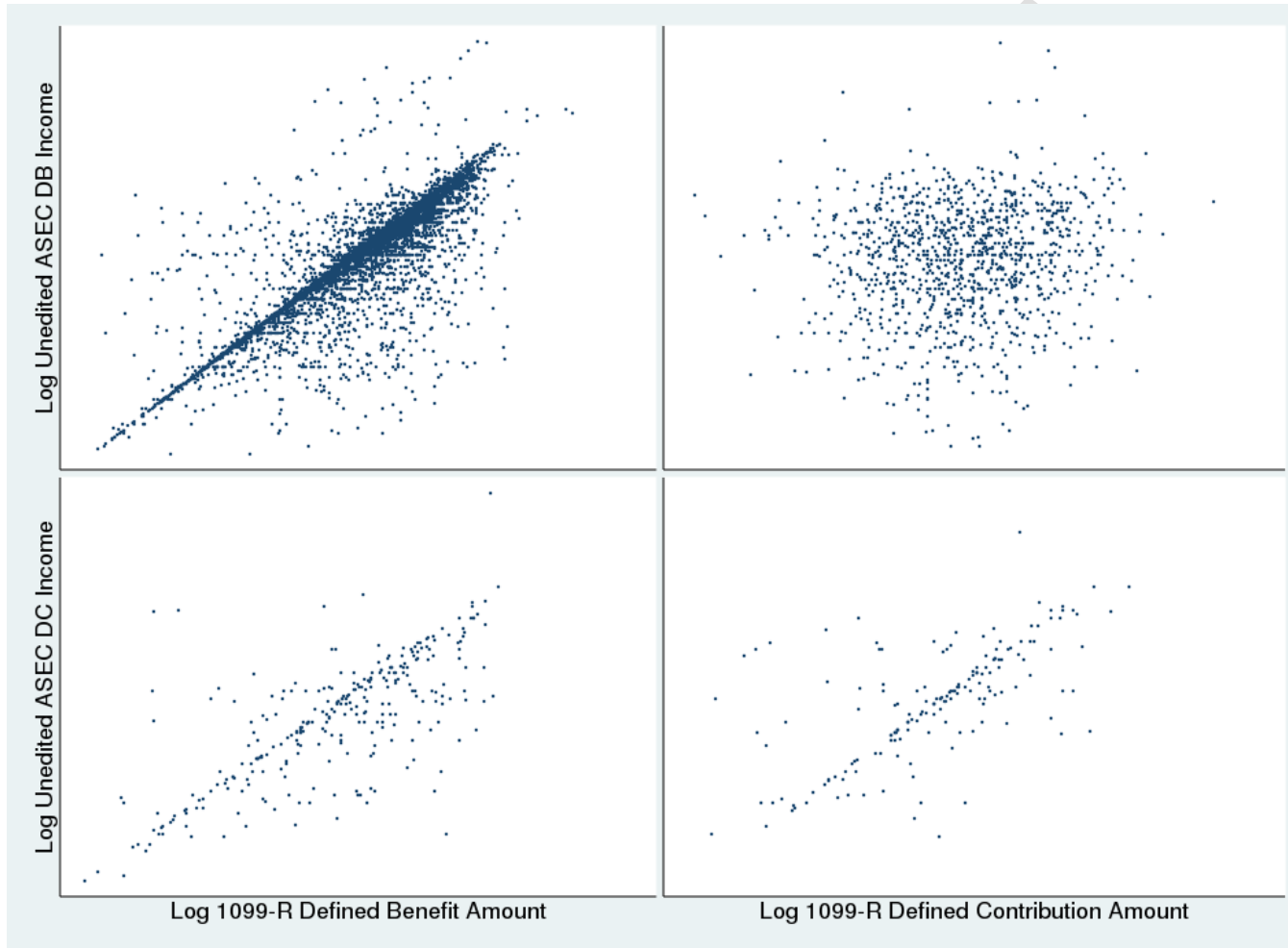
Sources: 2010 Current Population Survey Annual Social and Economic Supplement, IRS Tax Year 2009 Form 1099-R microdata.

Figure 9: Scatterplot of individual log unedited ASEC retirement income by log 1099-R amount, including only respondents whose retirement sources are all defined-contribution



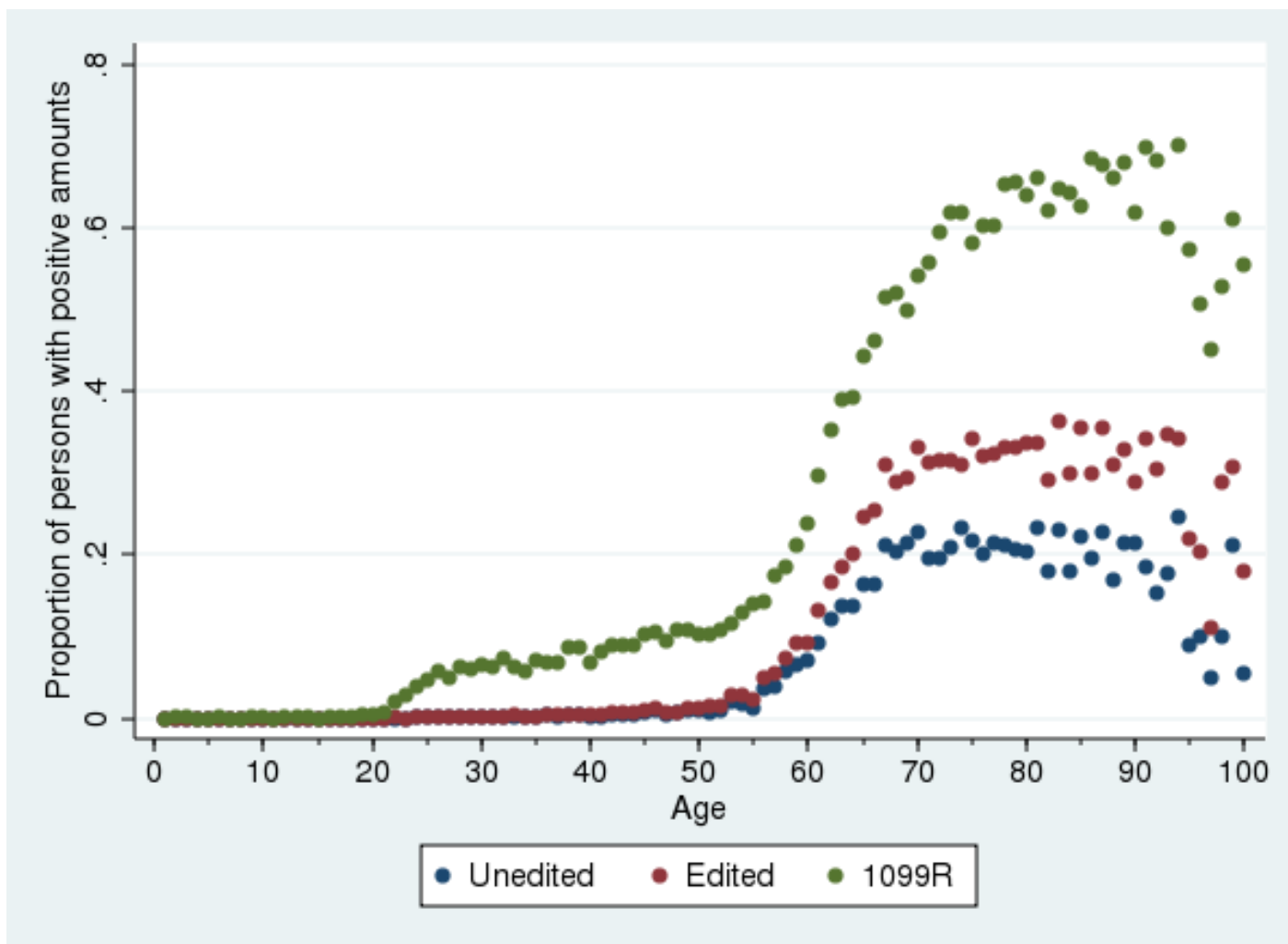
Sources: 2010 Current Population Survey Annual Social and Economic Supplement, IRS Tax Year 2009 Form 1099-R microdata.

Figure 10: Scatterplots of log unedited ASEC defined contribution and defined benefit income, by log 1099-R defined benefit and defined contribution amounts



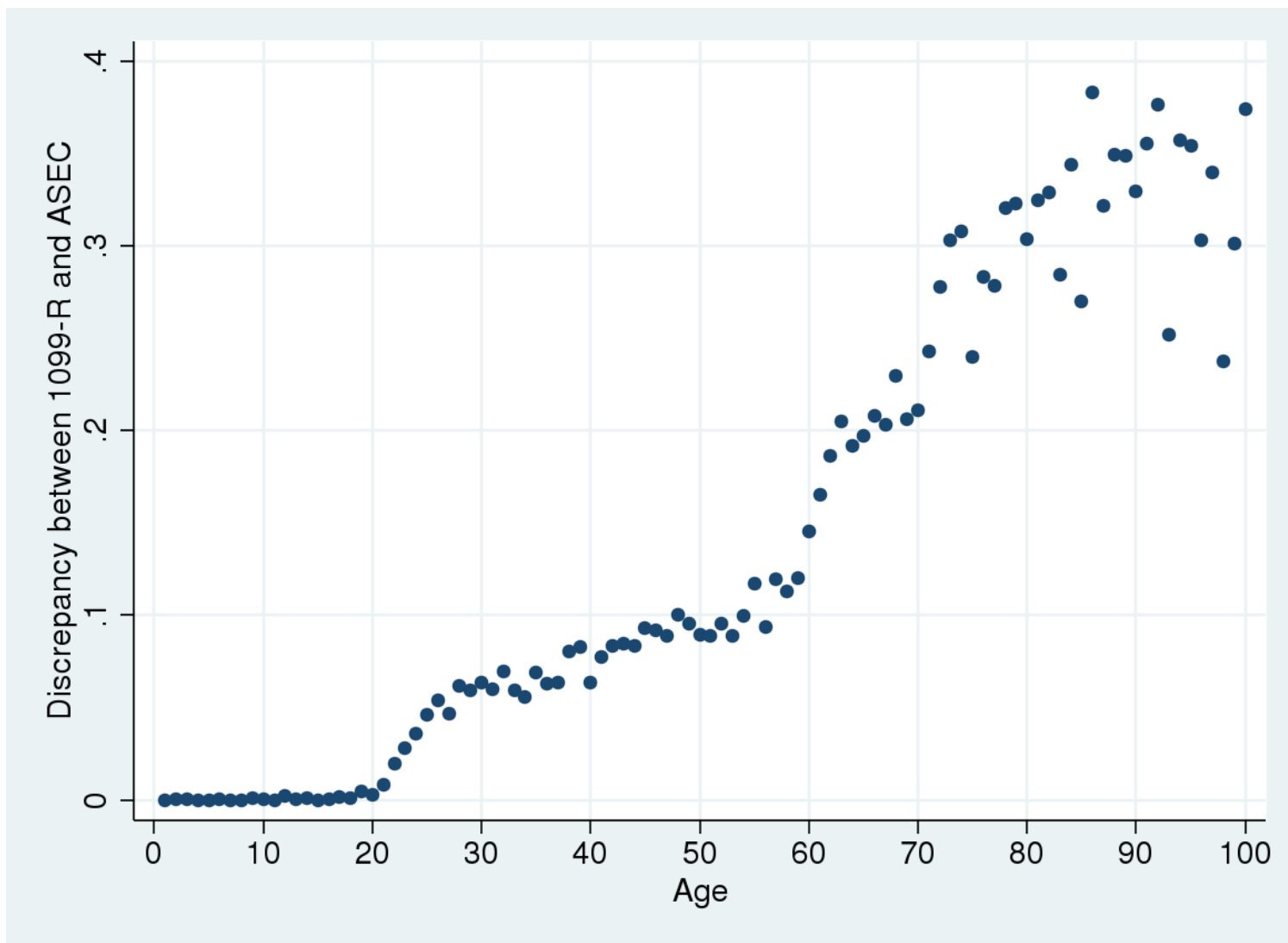
Sources: 2010 Current Population Survey Annual Social and Economic Supplement, IRS Tax Year 2009 Form 1099-R microdata.

Figure 11: Comparison of age distributions across unedited ASEC, edited ASEC, and 1099-R microdata



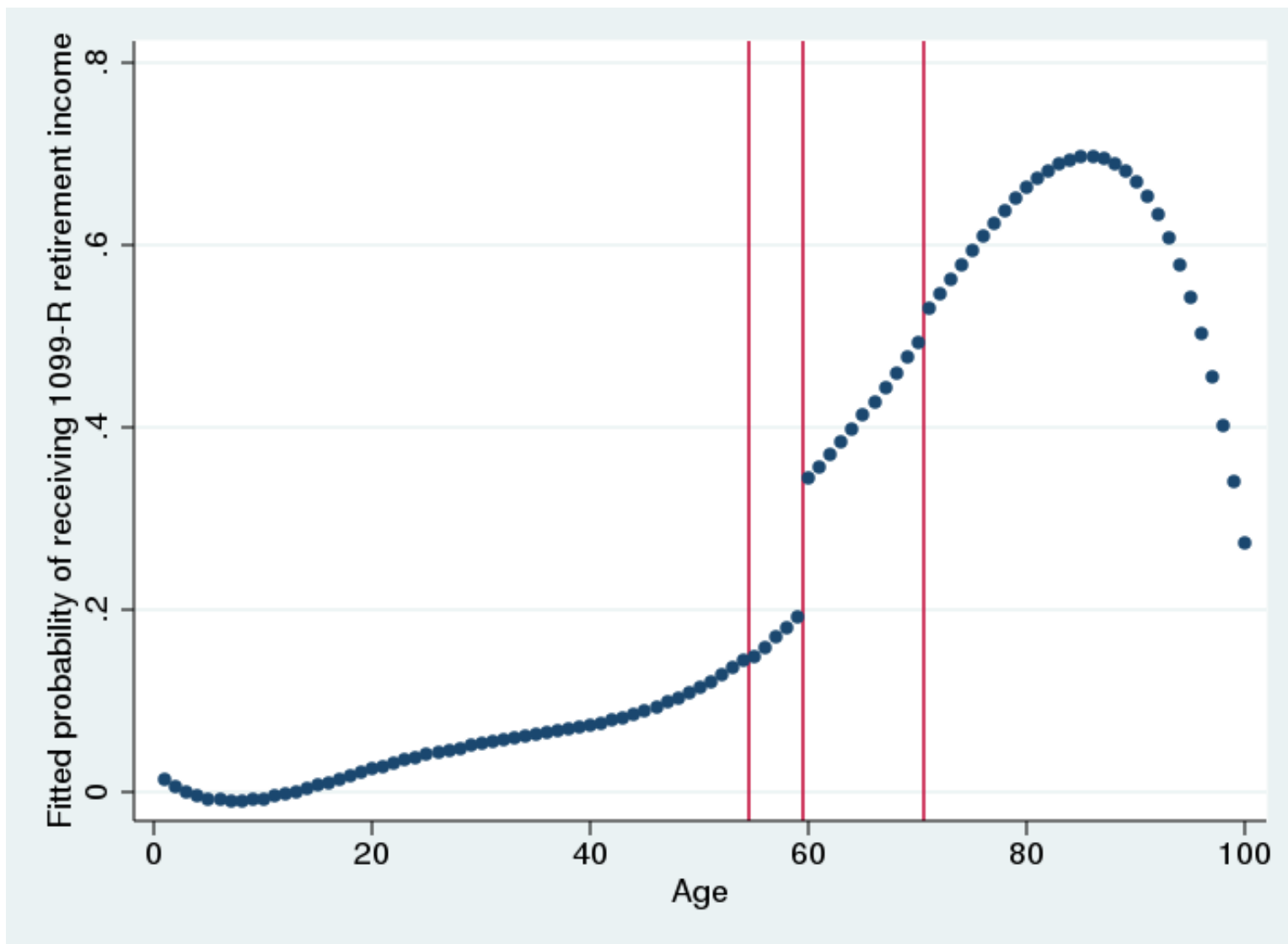
Sources: 2010 Current Population Survey Annual Social and Economic Supplement, IRS Tax Year 2009 Form 1099-R microdata.

Figure 12: Differences in reciprocity rates between edited ASEC and 1099-R microdata



Sources: 2010 Current Population Survey Annual Social and Economic Supplement, IRS Tax Year 2009 Form 1099-R microdata.

Figure 13: Fitted values from a linear probability model with a quintic control for age and discontinuities at specific age cutoffs



Sources: 2010 Current Population Survey Annual Social and Economic Supplement, IRS Tax Year 2009 Form 1099-R microdata.

Table 1: Per-capita income within income deciles for individuals aged 65 or older that had DC plan or IRA withdrawals, tax year 2006

Income decile¹	Income excluding DC plan and IRA withdrawals	Income including DC plan and IRA withdrawals	Percent difference
Lowest	\$5,502	\$8,061	47 %
2nd	11,135	15,848	42
3rd	14,300	17,496	22
4th	17,390	22,087	27
5th	20,909	26,964	29
6th	24,943	30,748	23
7th	30,517	41,174	35
8th	37,255	41,432	11
9th	53,410	60,258	13
Highest	190,169	229,003	20
All	40,574	49,363	22

¹ *Income includes wages and salaries, income from a sole proprietorship or farm, businesses or investments, interest and dividends, Social Security, and other pensions. Income excludes withdrawals from IRAs and DC plans, unemployment or worker's compensation, welfare assistance, child support, and alimony.*

Source: Sabelhaus and Schrass (2009) tabulations of the Federal Reserve Board Survey of Consumer Finances, 2007.

Table 2: Form 1099-R Distribution Codes included and recoded in the IRS extract provided to the Census Bureau

DTYPE = 1 “Defined Contribution”	DTYPE = 2 “Defined Benefit”
IRA/SEP/SIMPLE checkbox marked	1 Early distribution, no exception
B Designated Roth account	2 Early distribution, exception applies
C Designated Roth account, qualified	3 Disability
J Roth IRA	4 Death benefit
S Early distribution from SIMPLE	7 Normal distribution
T Roth IRA, exception applies	A May be eligible for 10-yr
Q Distribution from Roth	W RRB dual (windfall)
	X RRB Tier 1
	Y RRB Tier 2
	Z RRB Supplemental

DRAFT

Table 3: Sample sizes at each stage of the matching process

1099-R		
Person-form records in TY2009 1099-R extract	69,509,404	
MINUS Unprocessed due to blank name	-339,312	-0.5%
MINUS Not found in PVS searches	-282,894	-0.4%
Records available for matching to CPS ASEC	68,887,198	
Persons in 1099-R extract (unique PIK values)	44,868,790	
CPS ASEC (PVS Matching)		
Person records in 2010 CPS ASEC crosswalk file	211,384	
MINUS Not searched due to respondent opt-out	-1,002	-0.5%
MINUS Not searched due to blank name	-2,642	-1.3%
MINUS Not found in PVS searches	-21,440	-10.1%
MINUS Missing household sequence number	-1,292	-0.7%
Records available for matching to 1099-R	185,008	
1099-R persons matched to CPS ASEC persons	23,565	
CPS ASEC (Sample selection)		
Person records in 2010 CPS ASEC crosswalk file	211,384	
MINUS Missing household sequence number	-1,582	-0.7%
Analysis sample size	209,802	

Source: Internal documentation from Census Bureau Center for Administrative Records Research and Applications, "TY2009 PVS MAF Match Results – IRS1099R"

Table 4: Unweighted cross-tabulation of 1099-R availability and receipt of retirement income

	Reported receiving retirement income (RET_YN)			Total
	Not in universe	Yes	No	
Does not have matched 1099-R	42,740 99.96%	840 9.42%	117,748 88.39%	161,328 87.26%
Has matched 1099-R	16 0.04%	8,073 90.58%	15,466 11.61%	23,555 12.74%
	42,756	8,913	133,214	184,883

Sources: 2010 Current Population Survey Annual Social and Economic Supplement, IRS Tax Year 2009 Form 1099-R microdata.

DRAFT

Table 5: Comparisons to 1099-R person-level sums by income source

Edited income source	Correlation coefficient	Percent of responses within 5% of 1099-R	Unweighted count
Wage and salary	0.022	1.3	10,277
Own business self-employment	0.118	2.2	906
Farm self-employment	0.102	0.6	154
Unemployment	0.021	2.2	1,504
Worker's compensation	0.088	1.4	146
Social Security	0.079	3.3	11,778
Supplement Security Income	-0.038	2.5	200
Public assistance	0.035	2.0	49
Veterans' payments	0.065	6.8	893
Survivor's income	0.438	15.3	1,194
Disability	0.214	14.2	267
Retirement	0.611	21.9	8,076
Interest	0.149	1.7	12,341
Dividends	0.074	1.2	4,460
Rent	0.057	1.6	1,544
Educational assistance	-0.054	3.2	284
Child support	0.073	1.6	318
Alimony	0.188	6.0	50
Financial assistance	0.028	1.2	84
Other income	0.157	1.8	274

Sources: 2010 Current Population Survey Annual Social and Economic Supplement, IRS Tax Year 2009 Form 1099-R microdata.

Table 6: Differences between ASEC reported amount and 1099-R amount

<i>Discrepancy</i> (Reported on ASEC - Actual)	Amount of Actual Retirement Income (Recorded on Form 1099-R)						Total
	\$0	\$1 to \$99	\$100 to \$999	\$1,000 to \$9,999	\$10,000 to \$99,999	\$100,000 or more	
-\$100,000 or less	-	-	-	-	-	272	272
-\$99,999 to -\$10,000	-	-	-	-	8,057	34	8,091
-\$9,999 to -\$1,000	-	-	-	7,825	1,605	5	9,435
-\$999 to -\$100	-	-	2,340	376	396	< 5	3,113
-\$99 to -\$10	-	642	17	173	74	0	906
-\$9 to \$9 excluding exact matches	< 5	86	47	255	75	0	466
\$10 to \$99	< 5	0	17	54	36	< 5	110
\$100 to \$999	12	< 5	10	109	151	< 5	284
\$1,000 to \$9,999	93	< 5	17	124	282	< 5	520
\$10,000 to \$99,999	103	< 5	13	54	88	< 5	260
\$100,000 or more	< 5	0	0	< 5	28	< 5	38
Exact match (\$0)	161,111	< 5	27	109	137	< 5	161,388
Total	161,328	733	2,488	9,082	10,929	323	184,883

Sources: 2010 Current Population Survey Annual Social and Economic Supplement, IRS Tax Year 2009 Form 1099-R microdata.

Table 7: Differences between ASEC and 1099-R amounts among respondents reporting a positive amount of retirement income

<i>Discrepancy (Reported on ASEC - Actual)</i>	Amount of Actual Retirement Income (Recorded on Form 1099-R)						Total
	\$0	\$1 to \$99	\$100 to \$999	\$1,000 to \$9,999	\$10,000 to \$99,999	\$100,000 or more	
-\$100,000 or less	-	-	-	-	-	27	27
-\$99,999 to -\$10,000	-	-	-	-	1,049	34	1,083
-\$9,999 to -\$1,000	-	-	-	456	1,605	5	2,066
-\$999 to -\$100	-	-	13	376	396	< 5	786
-\$99 to -\$10	-	0	17	173	74	0	264
-\$9 to \$9 excluding exact matches	< 5	< 5	47	255	75	0	382
\$10 to \$99	< 5	0	17	54	36	< 5	110
\$100 to \$999	12	< 5	10	109	151	< 5	284
\$1,000 to \$9,999	93	< 5	17	124	282	< 5	520
\$10,000 to \$99,999	103	< 5	13	54	88	< 5	260
\$100,000 or more	< 5	0	0	< 5	28	< 5	38
Exact match (\$0)	-	< 5	27	109	137	< 5	277
Total	217	7	161	1,713	3,921	78	6,097

Sources: 2010 Current Population Survey Annual Social and Economic Supplement, IRS Tax Year 2009 Form 1099-R microdata.

Table 8: Descriptive statistics by retirement income source

ASEC Retirement Income Source	Count if Amount > 0	Unedited Mean if Amount > 0	Unedited Mean
Company or union pension	3,341 obs	\$14,747	\$235
Federal government	591	30,615	86
U.S. military	430	23,381	48
State or local government	1,569	25,697	192
U.S. railroad retirement	74	21,121	7
Annuities, paid insurance policies ¹	72	13,339	5
IRA, Keough, 401(k) accounts ¹	222	15,205	16
Other sources	179	21,143	18

¹ Regular payments from these sources.

Source: 2010 Current Population Survey Annual Social and Economic Supplement

Table 9: Counts by ASEC retirement income source category

ASEC Retirement Income Source	Unweighted Count
All defined benefit	5,830
Mostly defined benefit	47
Mostly defined contribution	20
All defined contribution	395

Source: 2010 Current Population Survey Annual Social and Economic Supplement

Table 10: Linear probability model for 1099-R match, with discontinuities at specific age cutoffs

Number of observations	=	184883
R-squared	=	0.2939
Root MSE	=	0.2993

Dependent variable: Indicator for whether person has a matching 1099-R

Independent variable	Coefficient	Robust std. err.	t-stat	P > t
Age 55 to 59	-0.005	0.006	-0.78	0.438
Age 60 to 70	0.136	0.010	13.90	0.000
Age over 70	0.156	0.020	7.86	0.000

Note: The regression specification includes a constant and a 5th-order polynomial control for age. The omitted reference group consists of those age under 55. The regression is weighted using final sample weights, but it does not employ replicate weights in the calculation of standard errors.

Sources: 2010 Current Population Survey Annual Social and Economic Supplement, IRS Tax Year 2009 Form 1099-R microdata.

Table 11: Tests of the discontinuities estimated in the linear probability model for 1099-R match

Age cutoff	Change in reciency rate at cutoff	Robust std. err.	t-stat	P > t
Age 55	-0.005	0.006	-0.78	0.438
Age 59 ½	0.140	0.008	17.66	0.000
Age 70	0.020	0.132	1.56	0.120

Note: Results represent tests of differences between regression coefficients displayed in Table 5 above. The regression specification includes a constant and a 5th-order polynomial control for age. The omitted reference group consists of those age under 55. The regression is weighted using final sample weights, but it does not employ replicate weights in the calculation of standard errors.

Sources: 2010 Current Population Survey Annual Social and Economic Supplement, IRS Tax Year 2009 Form 1099-R microdata.