# How Much Consumption Insurance in the U.S.?*

**Dmytro Hryshko**[†]
University of Alberta

**Iourii Manovskii**[‡]
University of Pennsylvania

**Abstract**

Most of what the profession knows about joint income and consumption dynamics at the household level in the U.S. is based on the data from the Panel Study of Income Dynamics (PSID). We find that there are two sets of households in the PSID that differ dramatically in the dynamics of their income and consumption. Households headed by the original PSID males and their sons have a highly persistent income process, and permanent shocks to their incomes almost fully pass through to consumption. Households headed by males who marry daughters of the original PSID members have a much less persistent income process and a dramatically higher degree of consumption insurance. These differences are surprising but highly robust. Conditional on income dynamics, the degree of insurance in each subsample is consistent with the prediction of the standard incomplete-markets model. This result stands in contrast to the well known puzzle in Blundell, Pistaferri, and Preston (2008) of excess insurance of permanent income shocks for the combined sample.

KEYWORDS: Consumption inequality, income processes, heterogeneity, labor income risk, insurance, incomplete markets models.

**JEL Classification:** D12, D14, D31, D91, E21

*We are very grateful to David Johnson and Frank Stafford from the PSID for thorough reading of the manuscript and detailed comments. We also benefited from the comments of participants at the CEPR ESSIM Conference, PSID Annual User Conference, SED in Edinburgh, 2017 Asian Meeting of the Econometric Society in Hong Kong, Workshop on Inequality at the Federal Reserve Bank of St. Louis, and XI REDg Workshop at Universidad Carlos III de Madrid and seminar participants at the Universities of Alberta, Delaware, Saskatchewan, and Copenhagen Business School. Support from the National Science Foundation Grant No. SES–1357903 is gratefully acknowledged.

[†]Department of Economics, University of Alberta, 8–14 HM Tory Building, T6G 2H4, Edmonton, AB, Canada. E-mail: `dhryshko@ualberta.ca`.

[‡]Department of Economics, University of Pennsylvania, 160 McNeil Building, 3718 Locust Walk, Philadelphia, PA, 19104-6297 USA. E-mail: `manovski@econ.upenn.edu`.

# 1 Introduction

Much of our knowledge of the joint income and consumption dynamics at the household level in the U.S. is based on the data from the Panel Study of Income Dynamics (PSID). In a seminal contribution, Blundell, Pistaferri, and Preston (2008) (BPP hereafter), use these data to estimate consumption insurance coefficients for permanent and transitory idiosyncratic income shocks, i.e. the fraction of those shocks that does not translate into movements in consumption. This direct evidence on the degree of insurance provides an essential empirical benchmark for assessing the performance of workhorse quantitative models of household consumption and saving choices. BPP find that household consumption is excessively insured against permanent shocks to net household incomes relative to the prediction of the standard incomplete-markets model. In contrast, we provide evidence that the degree of insurance and income dynamics vary quite dramatically and systematically across two sets of households in the PSID. Conditional on income dynamics, the estimated insurance against permanent shocks for both types of households is in line with the prediction of the standard incomplete-markets model.

To understand the distinction between these two types of households, it is necessary to briefly describe the PSID data. The PSID started in 1968 with a representative cross-section of US households. These households, as well as their children, grandchildren, etc. form the current PSID sample. Ignoring the issues of potential sample attrition and post-1968 immigration, this sample continues to represent the US population over time. In other words, by following this branch of the US family tree, we can learn about the population at large. Note that individuals who become married to the core or "sample" PSID members are not considered to be part of the branch, and are labeled as "non-sample" individuals by the PSID. The information on these individuals is collected while they are attached to a core PSID member, but they are not followed either before or after this period of attachment. An analogy might be helpful in highlighting the distinction. Imagine all individuals who were originally interviewed by the PSID in 1968 were endowed with the PSID "gene." All individuals born to or adopted by somebody with the PSID "gene" acquire the gene themselves and are followed by the PSID. The "gene" is not passed to the spouse. Thus, "sample" PSID members are the ones with the "gene" and "non-sample" PSID members are the ones without the "gene."[1]

We find that families headed by sample males (who have the "gene") have drastically different insurance against permanent income shocks to net family incomes relative to the families headed by non-sample males (who do not have the "gene"). If we restrict the BPP

[1] We are grateful to Mariacristina De Nardi for alerting us to the existence of this "gene" terminology.

data to households headed by sample males, we find a virtually complete pass-through of permanent income shocks to consumption. In contrast, the households headed by non-sample males show a dramatically higher degree of insurance against permanent shocks.

The large discrepancy in the degree of insurance remains robust to all our efforts to identify and to control for observable differences among sample and non-sample households. (We refer to households headed by sample males as "sample households" and households headed by non-sample males as "non-sample households.") In the most comparable grouping, we consider the sample males and females (who all have the PSID gene) who marry after 1968. One can roughly describe the two groups as consisting of sons and daughters of the original PSID sample, with their spouses being non-sample females and non-sample males, respectively. As can be expected, these groups are virtually identical with respect to all observables. Yet, over 90% of permanent income shocks are passed through to consumption of households headed by PSID sons, while only 40% of permanent income shocks are passed through to consumption of households headed by non-sample PSID sons-in-law, married to PSID sample daughters.

While our finding on the dramatic difference in the degree of insurance is novel in the literature, the finding that there is little cross-sectional difference among comparable sample and non-sample individuals in the PSID is consistent with early studies by Becketti, Gould, Lillard, and Welch (1988) and Lillard (1989). However, to the best of our knowledge, the literature has never compared the dynamic properties of income or earnings among sample and non-sample PSID individuals or households. This is an important omission as the dynamic properties of incomes are the crucial ingredients in BPP analysis and, indeed, in any model with incomplete insurance markets. We present evidence of substantial differences. Specifically, while the permanent component of the income process among sample-male-headed households is well described by a random-walk model, the families headed by non-sample males have a far less persistent permanent component of income. Although the literature traditionally considers the pooled sample, we argue that it might be essential to recognize the heterogeneity in income dynamics between the two groups.

We show that assuming a common income process, in particular assuming that the persistent income component follows a random walk, leads to a significant misspecification and induces much of the well-known discrepancy between the estimates of the household income process using the moments in levels and differences.[2] In particular, using a random-walk process common to the two groups results in inflated estimates of the variance of permanent shocks when estimation targets the income moments in differences (the procedure followed

[2]In addition, it is important to model the low mean and high variance of observations at the beginning and at the end of income spells, as suggested by Daly, Hryshko, and Manovskii (2016).

by BPP). As these shocks are not truly permanent, consumption responds relatively little to them, as predicted by the standard theory. This results in some (but not very large) overestimation of the degree of insurance of permanent shocks.

Perhaps even more importantly, as pointed out by Kaplan and Violante (2010) and Blundell (2014), correctly measuring the persistence of the income innovation is key for interpretation of the resulting insurance coefficients. For example, the findings of BPP, who considered only the combined sample and assumed a random walk process, suggest a considerably higher degree of insurance against permanent income shocks relative to the predictions of the standard models of imperfect consumption risk-sharing via self-insurance through saving and borrowing. Our estimates, based separately on sample and non-sample households, point to a different conclusion. The amount of insurance achieved by non-sample households is roughly in line with the prediction of the standard model given that "permanent" shocks to their incomes have only limited persistence. On the other hand, the point estimate of no insurance against truly permanent income shocks achieved by sample households is more puzzling in light of the theory. It may point to lack of precautionary motives for accumulating wealth among sample households, but it is unclear why one should expect significant preference heterogeneity across sample and non-sample households.[3] A more plausible explanation is that this point estimate for the insurance against random-walk shocks comes with a fairly sizable standard error so that the point estimate in the data is not statistically different from the prediction of the standard model.

The rest of the paper is structured as follows. Section 2 describes the data used; Section 3 documents differences in consumption insurance among sample and non-sample families in the PSID; Section 4 models and estimates income processes for sample and non-sample households; and Section 5 concludes.

# 2 Data

At the core of our study is the data set used and made publicly available by Blundell, Pistaferri, and Preston (2008). We augment these data with additional variables extracted from the PSID, most importantly, the ones that indicate whether a particular individual is a sample or non-sample PSID member. As summarized above, the PSID started in 1968 interviewing about 4,800 families; 2,930 of them were nationally representative (SRC sample), while the rest belonged to income-poor households (SEO sample). Members of these

[3]Following the approach in Hryshko, Luengo-Prado, and Sørensen (2011), for the comparable samples of the families formed by sons and daughters of original PSID sample members, we find no differences in risk attitudes as revealed by their choices of hypothetical risky gambles in the 1996 wave of the PSID.

original households, as well as their descendants (children, grandchildren etc.) are referred to as sample members by the PSID, whereas individuals entering the PSID due to marriage or living arrangements with the original sample members are labeled non-sample (e.g., a male marrying a sample female after 1968 will become a head of household and will be treated as a non-sample PSID member). The major distinction of non-sample persons is that the PSID makes no attempts to contact these individuals once they separate from a sample person. While the PSID provides weights for sample individuals, which makes it possible to achieve nationally representative results using individual data, the non-sample members have zero (longitudinal, and cross-sectional up to 1997) weights in the PSID.

Unless specifically stated otherwise, we maintain all of the sample restrictions made by BPP and we refer the reader to that paper for the detailed discussion of the motivation behind those restrictions. Briefly, the main objective was to focus on a sample of continuously married couples headed by a male (with or without children). BPP aimed to restrict the sample to households with male heads of ages 30–65 who do not change their marital status and are continuously married to the same spouse during 1978–1992. The focus on continuously married couples is to eliminate the potential effects of dramatic family composition change, such as divorce. As we discuss below, the actual implementation of data construction allows for sample females (but not sample males) to marry and divorce inside the sample. We assess the consequences of this feature of the data below.

Our initial sample is the same as in BPP, and contains 1,765 households, among them 965 families headed by sample males, and 800 families headed by non-sample males. Various modifications to this sample will be considered and explained below.

# 3 Documenting Differences in Insurance among Sample and Non-Sample Households

In this section we document large and robust differences in the measured insurance against permanent income shocks among sample and non-sample households. We begin by briefly summarizing the empirical measures of insurance. These measures are the same as the ones proposed and implemented by BPP.

## 3.1 Methodology

BPP assume that household $i$'s idiosyncratic net family income, $y_{it}$, is composed of a fixed effect, $\alpha_i$, a random-walk permanent component, $p_{it} = p_{it-1} + \xi_{it}$, and a transitory component modeled as a moving average process of order one, $\tau_{it} = \epsilon_{it} + \theta\epsilon_{it-1}$. Idiosyncratic income

and idiosyncratic consumption are residuals from panel regressions of the logs of net family income, and (imputed) nondurable consumption on a number of observables.

BPP consider the following equation for residual consumption growth:

$$\Delta c_{it} = \phi \xi_{it} + \psi \epsilon_{it} + \zeta_{it} + \Delta u_{it}, \tag{1}$$

where $\Delta c_{it}$ is individual $i$'s consumption growth at time $t$, $\xi_{it}$ is the permanent shock to household $i$'s disposable income, $\epsilon_{it}$ is the transitory shock, $\zeta_{it}$ is an innovation to consumption growth independent of the two income components, and $u_{it}$ is an i.i.d. measurement (and imputation) error in nondurable consumption. All of the shocks are assumed to be independent of each other. Coefficients $\phi$ and $\psi$ measure the transmission of permanent and transitory shocks to consumption. Conversely, $1-\phi$ and $1-\psi$ measure the extent of household consumption self-insurance against permanent and transitory shocks to net income due to accumulated assets. For other measures of income, $1-\phi$ and $1-\psi$ will have different interpretations.[4]

Following BPP, we estimate $\phi$ and $\psi$, the parameters of the income process (the moving-average parameter and the time-varying variances of permanent and transitory shocks), the variance of random growth in consumption, $\sigma_\zeta^2$, and time-varying variances of measurement (and imputation) error in consumption using the minimum-distance method. The parameters are recovered by minimizing the weighted distance between the full set of autocovariances of income and consumption growth, the full set of their cross-covariances, and their model counterparts. The weights are obtained from the diagonal weighting matrix constructed from the diagonal of the variance-covariance matrix of the data moments.

## 3.2 Benchmark Insurance Estimates for the Combined Sample and for Households Headed by Sample and Non-sample Males

In column (1) of Table 1 we tabulate the results based on the full sample of 1,765 PSID families. As reported by BPP, consumption is almost perfectly insulated from transitory shocks ($\hat{\psi}$ is close to zero) while about 36% of permanent shocks are insured ($\hat{\phi} = 0.64$).

Next, we consider separately the households headed by sample and non-sample males. The results are in columns (2) and (3): sample families insure only about 6% of permanent shocks while non-sample families insure up to 57% of permanent shocks; the difference in

[4]For instance, Blundell, Pistaferri, and Saporta-Eksten (2016) measure the extent of consumption insurance against permanent and transitory shocks to husband's wages due to changes in own and spousal labor supply, accumulated assets, and the tax and transfer system, whereas Arellano, Blundell, and Bonhomme (2017) study consumption insurance against persistent and transitory shocks to household earnings due to assets, and the tax and transfer system.

Table 1: Minimum-Distance Partial Insurance Estimates for the Combined Sample and Households Headed by Sample and Non-sample Males.

| | Combined (1) | Sample (2) | Non-sample (3) |
|---|---|---|---|
| $\phi$, transmission of perm. shock | 0.6436 | 0.9430 | 0.4303 |
| | (0.0858) | (0.1508) | (0.0950) |
| $\psi$, transmission of trans. shock | 0.0291 | –0.0108 | 0.1014 |
| | (0.0436) | (0.0469) | (0.1009) |

*Notes*: Standard errors in parentheses. p-value for test of equal $\phi$ ($\psi$) between sample and non-sample families equals 0.4% (31%).

the permanent insurance between sample and non-sample families is significant at the 1% level whereas the difference in the transitory insurance is not statistically significant at any conventional level. The degree of insurance estimated on the combined sample in column (1) appears to reflect a "weighted average" of insurance achieved by the two types of families.

## 3.3 The Effects of Marriage and Divorce

As we mentioned above, the dataset construction by BPP treats sample and non-sample households asymmetrically by allowing households headed by non-sample males to be formed through marriage or end in divorce inside the 1978–1992 sample window while this is not allowed for the households headed by sample males. The reason for this asymmetry is technical but very simple. BPP select the sample based on the following criteria: head's marital status does not change and the head remains married to the same wife. Husbands of PSID sample females are only tracked by the PSID while they are married to PSID sample females. Thus, even if a PSID sample female marries or divorces within the sample window, her non-sample husband (defined by the PSID as the household head) will only be tracked while he is married to her, so his marital status as recorded by the PSID cannot change. In contrast, sample males are continuously tracked before and after the marriage so all their marital status changes are recorded and, if present, are used to deselect them form the sample. The quantitative importance of this asymmetry is highlighted by Table 2 which indicates that only 8% of sample households and more than 50% of non-sample households are formed inside the sample window (the shares of households with code values for the family composition change variable 2 through 6).

Table 2: Family composition change in the year first entered the sample 1978–1992.

| | Code value | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 4 | 5 | 6 | Total |
| Sample | 710 | 175 | 23 | 0 | 57 | 0 | 965 |
| | (73.58) | (18.13) | (2.38) | (0.00) | (5.91) | (0.00) | (100) |
| Non-sample | 290 | 73 | 0 | 316 | 1 | 120 | 800 |
| | (36.25) | (9.13) | (0.00) | (39.50) | (0.13) | (15.00) | (100) |

*Notes*: 0="No change," 1="Change in members other than head or wife," 2="Head same but wife left/died and/or head has new wife," 4="Female head from previous year got married, husband (non-sample member) now head," 5="Some sample member other than head or wife has become head of this family unit," 6="Some female in family unit other than the previous-year head got married and non-sample member is now head." Numbers in parentheses are percentages of the "Total."

Clearly, newlywed or divorcing couples may experience substantial changes in spending or labor supply behavior at the start or end of their marriages which may lead to atypical income and/or consumption dynamics, and may potentially affect the consumption insurance estimates.[5] Thus, the key question addressed in this section is whether the differences in insurance between sample and non-sample households documented above are induced by the asymmetric treatment of marriage and divorce between them.

To evaluate the importance of marriage, we group households into those who got married before the start of the sample in 1978 (the code for the family composition change variable in 1978 equals 0 or 1 in Table 2), and those who got married in 1978 or after (the code for the family composition change variable in 1978 equals 2, 4, 5 or 6). As reported in Panels A and B of Table 3, the estimated insurance of permanent shocks for sample households is close to none while non-sample households appear to insure a substantial fraction of permanent shocks regardless of whether we look at the couples who enter the BPP sample being married (columns (2) and (3)), or marry into the sample (the results in column (5) are based only on 80 sample families formed in 1978 or later, and therefore are somewhat imprecise).

To assess the impact of keeping non-sample households whose marriage ends in divorce in the estimation sample, we restrict the data to only households that had been surveyed in 1992 – the last sample year. The results reported in Panel C of Table 3 continue to exhibit

[5] Mazzocco, Ruiz, and Yamaguchi (2007), using the PSID, document substantial differences in the labor supply behavior for the members of newlywed (separated) couples around the time of marriage (divorce). Daly, Hryshko, and Manovskii (2016), using Danish and German administrative data, and survey data from the PSID show that systematic differences in the earnings dynamics at the start or end of earnings spells for a subsample of households may result in substantial biases in the estimated insurance of permanent shocks to earnings.

Table 3: Minimum-Distance Partial Insurance Estimates: The Effects of Marriage and Divorce.

| | Combined (1) | Sample (2) | Non-sample (3) | Combined (4) | Sample (5) | Non-sample (6) |
|---|---|---|---|---|---|---|
| | Panel A. Married before 1978 | | | Panel B. Married in/after 1978 | | |
| $\phi$, transmission | 0.7138 | 0.9265 | 0.5310 | 0.5937 | 1.4665 | 0.3283 |
| perm. shock | (0.0981) | (0.1313) | (0.1073) | (0.1815) | (0.5491) | (0.1153) |
| $\psi$, transmission | 0.0213 | –0.0369 | 0.0554 | 0.0205 | 0.0762 | 0.1570 |
| trans. shock | (0.0458) | (0.0428) | (0.1174) | (0.1309) | (0.1281) | (0.1277) |
| | Panel C. Surveyed in 1992 | | | Panel D. Balanced | | |
| $\phi$, transmission | 0.6293 | 0.8920 | 0.4802 | 0.6844 | 0.8986 | 0.3665 |
| perm. shock | (0.0914) | (0.1584) | (0.1044) | (0.1787) | (0.1602) | (0.1032) |
| $\psi$, transmission | 0.0249 | 0.0269 | 0.0445 | 0.1077 | 0.0231 | 0.0886 |
| trans. shock | (0.0463) | (0.0481) | (0.0990) | (0.0399) | (0.0481) | (0.1378) |

*Notes*: In Panel A, p-value for test of equal $\phi$ ($\psi$) between sample and non-sample families equals 3% (46%); in Panel B, the respective p-values are 4% and 65%; in Panel C, 3% and 87%; in Panel D, 1% and 65%.

the same pattern: families headed by sample males appear to be substantially less insured against permanent shocks than families headed by non-sample males.

Finally, we consider the set of households that had been first surveyed in 1978 and last surveyed in 1992, the first and last years in the BPP sample, respectively. We label those samples "Balanced."[6] This set of households lacks both newlywed couples and the couples divorcing inside the 1978–1992 period. The results are in Panel D. They are based on much smaller samples than the full sample which is reflected in the precision of the estimates. Remarkably, the point estimates for the transmission coefficient for permanent shocks for sample and non-sample families are fairly similar to those obtained for the full sets of those households – non-sample households achieve a substantial insurance of permanent shocks while sample families achieve virtually none.

In summary, all of the panels point to the statistically significant difference in the estimated insurance of permanent shocks. We do not find an important role of the differential selection of non-sample families with respect to marriage and divorce in inducing the differences in the extent of permanent insurance achieved by those families relative to their sample counterparts.

## 3.4 Consumption Insurance among Sample and Non-Sample Households Matched on Observables

Sample and non-sample households differ in some observable characteristics which may, in turn, lead to different levels of consumption insurance. We, therefore, next, form pairs of sample and non-sample households of exactly the same age, schooling, and race.[7] We have 501 of such pairs. These pairs are not unique – for example, there are sixteen households headed by white sample males born in 1920 with no college education, and only one such household headed by a non-sample male; these households will create one pair which shares the same year of birth, schooling and race but there are 16 different pairs we could possibly form. We therefore estimate insurance coefficients for a random set of 501 pairs of sample and non-sample households, repeat this exercise 1,000 times, and average the results. It is worth highlighting that, at each iteration, sample and non-sample households are exactly matched on the distribution of age, education, and race. The results are reported in Table 4. The results for matched pairs are somewhat less precise as they are based on a smaller set of

[6]Note that the label doesn't mean that they contain all 15 observations on family income and nondurable consumption as there are at most 13 consumption observations during the 1978–1992 period; it simply means that the families in the "Balanced" samples are observed both in 1978 and 1992.

[7]Households are grouped into two education groups – with (some) college and no college education, labeled "College" and "No college," respectively.

Table 4: Minimum-Distance Partial Insurance Estimates. Sample and Non-sample Households Matched on education, race, and year of birth of Head.

| | Combined (1) | Sample (2) | Non-sample (3) |
|---|---|---|---|
| $\phi$, transmission of perm. shock | 0.6741 | 1.0810 | 0.4279 |
| | (0.1643) | (0.3434) | (0.1675) |
| | [0.4467,0.9602] | [0.6153,1.6930] | [0.1723,0.7228] |
| $\psi$, transmission of trans. shock | 0.0371 | –0.0086 | 0.0920 |
| | (0.0611) | (0.0821) | (0.1272) |
| | [–0.0692,0.1334] | [–0.1445,0.1193] | [–0.0922,0.3173] |

*Notes*: Standard errors calculated as the standard deviations of the estimates across 1,000 random samples in parentheses; 95 percent confidence interval reported in square parentheses. p-value for test of equal $\phi$ ($\psi$) between sample and non-sample families equals 9% (51%).

households; they are similar, however, in terms of point estimates to the results in Table 1 – nondurable consumption of sample households absorbs permanent shocks almost fully while non-sample households insure more than 50 percent of permanent shocks.

## 3.5 Consumption Insurance among Households Formed by PSID Sons and Daughters

Although our previous experiment matches families on head's year of birth, education, and race, sample and non-sample families may still potentially differ on a variety of other characteristics. By construction, non-sample families do not include couples formed in 1968 but will contain females marrying non-sample males in 1969 and later, keeping their families intact until they are last observed during 1978–1992, or (re-)marrying during 1978–1992. To put selection of sample and non-sample families on an equal footing, we allow PSID sample males to (re-)marry and divorce during 1978–1992, keeping data for each newly-formed couple with the same sample male head in the final dataset.[8] We further split the resulting dataset of sample families into those who had been married in 1968 and stayed married until they are last seen in 1978–1992, and those who, similarly to non-sample families, married or re-married in 1969 or later. We label them "Sample orig." and "Sample sons," respectively, because the latter sample is dominated by the sons of original PSID households in addition to a few original sample members who married after 1969. In total, we have 569 original sample families, 1,057 families formed and headed by sample "sons," and 804 families formed

[8]This selection is also recently used in Blundell, Pistaferri, and Saporta-Eksten (2016).

by sample "daughters" and headed by their non-sample husbands (the latter group is the same as the set of non-sample families).[9]

In Table A-1 we tabulate means for various characteristics for the resulting three subsamples. Original sample families are naturally older and thus different from the other two subsamples with respect to many cross-sectional characteristics. In contrast, households formed by sample sons and sample daughters are very similar with respect to age, average nondurable consumption, net family income, head's earnings, assets, head's and wife's hours worked, incidence of unemployment, disability and displacement, occupation and industry switching, precision of food and income measurement, immigrant status of the head, incidence of owning a business and homeownership rates, among many other things. We further used a LASSO regression for predicting if a family (among the set of households formed by sample sons and daughters) belongs to the group of non-sample families using a wide variety of variables.[10] The results are depicted in Figure A-1 – although some regressors are picked by LASSO as having nonzero predictive power for the non-sample status, their predictive strength is minimal as the average prediction and the range of predicted values for sample and non-sample families are similar, and substantially deviate from their true respective values of zero and one. The analysis corroborates the conclusion based on a simple comparison of means in Table A-1 that the families formed by sample sons and daughters do not significantly differ on a wide variety of observable characteristics.

Despite the sets of households formed by sample sons and sample daughters being nearly identical with respect to their cross-sectional characteristics, the results in columns (1) and (2) of Table 5 indicate that they differ dramatically in the degree of consumption insurance against permanent income shocks, with non-sample households (i.e., the ones formed by PSID sample daughters) being significantly better insured.

In contrast, despite being different in many observable dimensions, original sample families and younger sample families formed mostly by their sons have very similar insurance against permanent income shocks – columns (3) and (1), respectively. In column (4), therefore, we group them together obtaining similar in magnitude but a more precise estimate of the insurance coefficient for permanent income shocks. In the following, we will concentrate on this larger sample, and will compare the estimates of insurance coefficients for this larger sample with the estimates for non-sample families.

[9]Relative to the original BPP data, additional four non-sample families are added as non-sample males from those families changed their marital status during 1978–1992 and were followed by the PSID after the change (some non-sample individuals were designated as followable since 1990).

[10]We are grateful to Max Sties for proposing and helping us implement this experiment.

Table 5: Minimum-Distance Partial Insurance Estimates. Updated Sample.

| | Sample sons (1) | Non-sample (2) | Sample orig. (3) | Sample all (4) |
|---|---|---|---|---|
| $\phi$, transmission perm. shock | 0.9280 | 0.4143 | 0.8991 | 0.8832 |
| | (0.2289) | (0.0943) | (0.1577) | (0.1405) |
| $\psi$, transmission trans. shock | 0.0742 | 0.1184 | –0.0002 | 0.0245 |
| | (0.0650) | (0.1020) | (0.0472) | (0.0408) |

*Notes*: p-value for test of equal $\phi$ ($\psi$) in columns (1) and (2) equals 4% (71%); in columns (2) and (3) equals 1% (29%); in columns (2) and (4) equals 1% (39%).

## 3.6 Differences in Consumption Insurance among Sample and Non-Sample Households by Gender of the Respondent

The characteristic in Table A-1 that differs the most between sample and non-sample households is that non-sample families have a higher incidence of wives responding to the survey. In Table 6, we consider whether consumption insurance differs for samples who have either male or female heads permanently responding to the survey.[11]

Table 6: Minimum-Distance Partial Insurance Estimates by Gender of Respondent.

| | Head respondent | | | Wife respondent | | |
|---|---|---|---|---|---|---|
| | Sample orig. (1) | Sample sons (2) | Non-samp. (3) | Sample orig. (4) | Sample sons (5) | Non-samp. (6) |
| $\phi$, transmission perm. shock | 0.8818 | 0.7474 | 0.2553 | 1.7467 | 1.4051 | 0.3598 |
| | (0.1951) | (0.1821) | (0.1187) | (0.7361) | (0.4109) | (0.0930) |
| $\psi$, transmission trans. shock | –0.0449 | –0.1189 | 0.3037 | 0.2955 | 0.2687 | 0.2804 |
| | (0.0602) | (0.0932) | (0.2104) | (0.1843) | (0.1314) | (0.2663) |

*Notes*: p-value for test of equal $\phi$ ($\psi$) in columns (1) and (3) equals 1% (11%); in columns (4) and (6) equals 6% (96%); in columns (2) and (3) equals 2% (6%); in columns (5) and (6) equals 1% (97%).

Similar to the patterns documented above, conditional on the respondent's gender, the consumption insurance of permanent income shocks is higher for the families headed by non-sample males. The transmission coefficient for permanent income shocks found for the sample families in which the wife always responds is somewhat high but imprecisely estimated

[11] Permanent responding status is less likely to be endogenous to income shocks and insurance.

because of a small number of observations for those families (38 and 76 families headed by the original PSID males and their sons, respectively). Importantly, the transmission coefficients are similar for non-sample families with male or female respondents, who have about equal representation in each of the groups based on the respondent's gender.[12]

## 3.7 Consumption Insurance among Sample and Non-Sample Households by Educational Attainment and Birth Cohort

In Table 7 we consider additional splits of sample and non-sample households across observable characteristics. As expected, consumption insurance is higher in families with college-educated heads, and in families whose heads are closer to retirement. Across all the splits, however, the estimated insurance of permanent shocks is substantially higher for non-sample families.

Table 7: Minimum-Distance Partial Insurance Estimates.

| | No college | | College | | Born 1940/50s | | Born 1920/1930s | |
|---|---|---|---|---|---|---|---|---|
| | Sample (1) | Non-samp. (2) | Sample (3) | Non-samp. (4) | Sample (5) | Non-samp. (6) | Sample (7) | Non-samp. (8) |
| $\phi$, transmission | 1.0449 | 0.8685 | 0.6210 | 0.4087 | 0.9953 | 0.4886 | 0.7726 | 0.3696 |
| perm. shock | (0.2307) | (0.2053) | (0.1424) | (0.0981) | (0.2031) | (0.1062) | (0.1552) | (0.1570) |
| $\psi$, transmission | 0.0987 | 0.1045 | –0.0522 | –0.1149 | 0.0825 | 0.0898 | –0.0239 | 0.0005 |
| trans. shock | (0.0564) | (0.1410) | (0.0585) | (0.0894) | (0.0752) | (0.1038) | (0.0474) | (0.1305) |

*Notes*: p-value for test of equal $\phi$ ($\psi$) in columns (1) and (2) equals 56% (96%); in columns (3) and (4) equals 21% (55%); columns (5) and (6) equals 2% (95%); columns (7) and (8) equals 6% (86%).

## 3.8 Consumption Insurance among Sample and Non-Sample Households for Other Income and Consumption Measures

In Table 8 we report results based on different concepts of income and consumption. To ensure that the results are not driven by potential miscalculation of household taxes or poor measurement of non-labor income, we consider earnings as a source of risk to household budgets. Specifically, we report results based on the combined head's and wife's earnings in columns (1) and (2), and male earnings in columns (3) and (4). As was the case with net

[12]The remaining group consists of the families with switching respondents; those families have a higher transmission coefficient of permanent shocks which may be due to endogeneity of switching to the shocks that are hard to insure against such as, e.g., the incidence of disability.

Table 8: Minimum-Distance Partial Insurance Estimates. Robustness to Different Income and Consumption Measures.

| | Sample (1) | Non-sample (2) | Sample (3) | Non-sample (4) |
|---|---|---|---|---|
| | Panel A. Nondur. cons., total earnings | | Panel B. Nondur. cons., male earninigs | |
| $\phi$, transmission perm. shock | 0.3428 | 0.1626 | 0.3776 | 0.1578 |
| | (0.0636) | (0.0573) | (0.0704) | (0.0469) |
| $\psi$, transmission trans. shock | 0.0470 | 0.1504 | 0.0560 | 0.0223 |
| | (0.0301) | (0.0724) | (0.0327) | (0.0615) |
| | Panel C. Food, total earnings | | Panel D. Food, male earnings | |
| $\phi$, transmission perm. shock | 0.2839 | 0.1229 | 0.2809 | 0.0925 |
| | (0.0537) | (0.0482) | (0.0547) | (0.0358) |
| $\psi$, transmission trans. shock | 0.0228 | 0.1085 | 0.0319 | 0.0352 |
| | (0.0269) | (0.0544) | (0.0266) | (0.0497) |

*Notes*: In Panel A, p-value for test of equal $\phi$ ($\psi$) between sample and non-sample families equals 3% (18%); in Panel B, the respective p-values are 1% and 62%; in Panel C, 2% and 15%; in Panel D, 0.4% and 95%.

income, non-sample households are found to be better insured against permanent shocks to earnings than sample households; see Panels A and B.

In Panels C and D, we use food as a measure of household consumption, because food is free of potential imputation biases which may result in divergent estimates of insurance among sample and non-sample households. Not surprisingly, food is better insured than nondurable consumption but non-sample households are once again found to be substantially more insured against permanent income and earnings shocks than their sample counterparts. The transmission coefficient for permanent earnings shocks is at least twice as high for the sample relative to non-sample families across all specifications.

## 3.9 Effects of Consumption Imputation

In Table 9 we consider alternative imputation procedures for nondurable consumption to further evaluate whether the procedure adopted in BPP induces the systematic difference in insurance between sample and non-sample households. BPP used the food demand equation

Table 9: Minimum-Distance Partial Insurance Estimates. Modified Imputation Procedures.

| Data: | Original, BPP | | | | Updated BPP sample | | | |
|---|---|---|---|---|---|---|---|---|
| Instruments: | BPP, real | | Alt. IV, real | | BPP, real | | Alt. IV, real | |
| | Sample (1) | Non-samp. (2) | Sample (3) | Non-samp. (4) | Sample (5) | Non-samp. (6) | Sample (7) | Non-samp. (8) |
| $\phi$, transmission | 0.8798 | 0.3919 | 1.0117 | 0.4681 | 0.8245 | 0.3763 | 1.0058 | 0.4470 |
| perm. shock | (0.1410) | (0.0862) | (0.1648) | (0.1060) | (0.1301) | (0.0854) | (0.1644) | (0.1052) |
| $\psi$, transmission | –0.0101 | 0.0824 | –0.0045 | 0.1150 | 0.0197 | 0.0988 | 0.0331 | 0.1369 |
| trans. shock | (0.0436) | (0.0917) | (0.0533) | (0.1148) | (0.0382) | (0.0928) | (0.0466) | (0.1160) |

*Notes*: p-value for test of equal $\phi$ ($\psi$) in columns (1) and (2) equals 0.31% (36%); in columns (3) and (4) equals 0.55% (34%); in columns (5) and (6) equals 0.40% (43%); in columns (7) and (8) equals 0.42% (40%). "BPP, real" estimation uses the original BPP instruments for imputation – the average, by cohort, year and education, head's and wife's hourly wages but in real terms. "Alt. IV, real" estimation uses the alternative instruments for imputation – the average, by cohort and year, real hourly head's and wife's wages.

estimated on CEX data to impute nondurable consumption to the PSID households. Since nominal nondurable expenditures, a right-hand-side variable of the equation, are potentially measured with error, BPP instrumented nominal nondurable consumption with the average – by cohort, year and education – hourly nominal head's and wife's wages. Campos and Reggio (2014) recently suggested that the instruments might be correlated with measurement error, which would result in inconsistent estimates of the regression coefficients in the equation used for imputation. Instead, they proposed to use real nondurable expenditures as a right-hand-side variable in the imputation equation and real hourly head's and wife's wages as instruments; they also suggested alternative instruments – real hourly head's and wife's wages averaged by cohort and year ("Alt. IV, real" in Table 9) rather than by cohort, year, and education as in BPP ("BPP, real" in Table 9).

The results in the first four columns of Table 9 are based on the original BPP data whereas columns (5)–(8) contain the results for an updated BPP data that, in addition, contains sample families who divorce, and divorce and remarry inside the period 1978–1992. The coefficients are somewhat smaller if we use original BPP instruments in real terms. The relative difference among sample and non-sample households in the insurance of permanent income shocks is, however, preserved across various imputation procedures.

Table 10: Minimum-Distance Partial Insurance Estimates. Data: 1999–2009.

| | Food | | Nondur. consumption | | Food away | |
|---|---|---|---|---|---|---|
| | Sample sons (1) | Non-samp. (2) | Sample sons (3) | Non-samp. (4) | Sample sons (5) | Non-samp. (6) |
| $\phi$, transmission perm. shock | 0.4307 | 0.1087 | 0.2815 | 0.1111 | 0.8469 | 0.4734 |
| | (0.1010) | (0.0618) | (0.0996) | (0.0680) | (0.2164) | (0.1262) |
| $\psi$, transmission trans. shock | –0.0397 | 0.0174 | 0.0904 | 0.0109 | –0.1268 | –0.0330 |
| | (0.0526) | (0.0614) | (0.0498) | (0.0656) | (0.0893) | (0.1121) |

*Notes*: p-value for test of equal $\phi$ ($\psi$) in columns (1) and (2) equals 0.7% (48%); in columns (3) and (4) equals 16% (33%); columns (5) and (6) equals 14% (51%).

## 3.10 Recent Data

Our objective in this section is twofold. First, we consider whether the differences in insurance between sample and non-sample households are also prominent in more recent PSID data. Second, we address the concerns about the effects of consumption imputation in yet another way, by exploiting the fact that in recent years a measure of nondurable consumption can be directly constructed using PSID data.

Specifically, since 1997 the PSID switched to biennial collection of the data and included a set of questions on expenditure categories other than food. We use income and consumption data for 1999–2009 from Blundell, Pistaferri, and Saporta-Eksten (2016) and accommodate the biennial nature of the data in the minimum-distance procedure. As the PSID does not report taxes since 1992, our income measure is gross family money income. For consumption, we use food and a measure of nondurable consumption which, for comparability with the measure utilized in BPP and used by us above, is the sum of food, utilities, and transportation expenditures and does not include housing-related expenses, child care, and education expenditures. We keep households whose head is 25 to 57 years of age and utilize all other sample selection criteria as in Blundell, Pistaferri, and Saporta-Eksten (2016). Our final data contains 1,528 households formed by sample sons and 1,612 households formed by sample daughters.[13]

Table 10 contains the results.[14] Columns (1) and (2) report the results for food that sums food at home, delivered food, food out, and the value of food stamps. As before, relative to the families of sample sons, food consumption of non-sample families appears to

[13]We dropped much older original sample families.

[14]We assume that the moving average parameter is zero because it is not identified if the biennial data are used for estimation. The results are similar if we assume other reasonable values for this parameter.

be better insured against permanent shocks to family incomes. Columns (3) and (4) contain the results for nondurable consumption. Again, non-sample families appear to have more insurance against permanent shocks but, somewhat surprisingly, nondurable consumption is more inelastic to permanent income shocks than food consumption. This is perhaps due to committed expenditures forming the bulk of transportation and utility costs. A relatively higher sensitivity of nondurable consumption to permanent income shocks reported in earlier sections of the paper is likely due to omission of more elastic components of nondurable expenditures in the recent waves of the PSID such as clothing, alcohol, tobacco, and expenditures on recreation. To verify this conjecture, in columns (5) and (6) we find that expenditures on food away from home, arguably a more elastic component of family consumption, is much more sensitive to permanent income shocks. As for the other measures of household consumption, non-sample families show a lower transmission of permanent income shocks to expenditures on food away from home than do sample families.

# 4 Income Processes of Sample and Non-Sample Households

The body of evidence presented so far points to the substantial differences in insurance against permanent shocks to net family incomes for sample versus non-sample families. Underlying these findings was the standard maintained assumption that the income prosess is the same across sample and non-sample households, assumed to consist of a random walk permanent component and an MA(1) component as in BPP. The assumption of common income process is consistent with the evidence of cross-sectional similarity of sample and non-sample households across observable characteristics. However, the dynamic properties of incomes of sample and non-sample households have never been examined in the literature, to our knowledge. If they differ, the estimates of insurance might be biased. Moreover, the interpretation of insurance coefficients depends on the dynamic properties of shocks to household budgets. For example, the insurance of about 60% of permanent shocks, found for non-sample families, appears excessive for consumption models with incomplete markets when the permanent component is a random walk process, but the value is reasonable for the income process with low persistence of shocks to the permanent component. In this section, we provide evidence that income processes indeed differ systematically across the two types of households and reinterpret the differences in consumption insurance in light of the differences in income processes.

## 4.1 Differences in the Moments Targeted in the Minimum-Distance Estimation across Sample and Non-sample Families

The minimum-distance estimation used above is based on the moments constructed from residual consumption and income growth. We therefore begin by examining the key moments used in the estimation procedure.

Figure A-2(a) plots the trends for the cross-sectional variance of income growth, consumption growth, and cross-covariance of income and consumption growth for sample and non-sample PSID households. There are some noticeable differences in the trends. In particular, for non-sample families, the variance of consumption growth rates had undergone a steeper rise in mid-1980s, the cross-covariance of income and consumption growth rates had reached its peak later, and the variance of income growth rates had not experienced any clear trend. The latter fact manifests itself in the correlation of just 18% between the variance of income growth rates for the two subsamples.[15] This may suggest that the two subsamples are different in terms of their income processes.

In Figure A-2(b) we therefore plot additional income moments – the first-, second-, and third-order autocovariances – for the two samples. While there are some differences in the trends for the first- and second-order autocovariances, the most important is the plot for the third-order autocovariance. Under the null of the income process in BPP – net family income is the sum of a random walk and an MA(1) component – the autocovariances beyond second order should not differ from zero. While the average third-order autocovariance is not significantly different from zero for the sample families, it is statistically different from zero, at the 2% level, for the families headed by non-sample males. As the minimum-distance estimation targets not only the third-order but all of the higher-order autocovariances of income growth, we next test if all higher-order autocovariances above the second order are jointly equal to zero, as in Abowd and Card (1989). The p-value of the test for sample families is nearly 28%, but only about 2% for non-sample families.[16]

These results suggest that the random-walk plus an MA(1) component is an adequate description of the income process for the sample families. However, the permanent component of non-sample households' incomes appears to be less persistent than a random walk. In the following, we generalize it to an AR(1) process.[17]

[15] In a regression of the cross-sectional variances in income growth rates on a constant and trend, the estimated coefficient on trend is not significantly different from zero for non-sample families, but significant at about 2% level for sample families.

[16] These results also hold for vaious partitions of the data, e.g. the p-values of the test for college, no college, born 1920/30s, and born 1940/50s partitions of sample families equal 15%, 19%, 35%, and 22%, respectively, while they equal 0.4%, 0%, 0%, and 1% for the same partitions of non-sample families.

[17] This evidence is also consistent with a possibility that an MA(1) process does not fully capture dynamics

## 4.2 Difference in the Persistence of Permanent Shocks

In this section, we use an alternative way to evaluate the persistence of income shocks that does not rely on the use of the autocovariance function of incomes. Specifically, we use GMM for estimation of the persistence of income shocks for sample and non-sample households and confirm our conjecture of the relatively lower persistence of the shocks to net family incomes for non-sample families.

Recall that the income process in BPP is $y_{it} = \alpha_i + p_{it} + \tau_{it}$, where $y_{it}$ is income residual for household $i$ in year $t$ from the first-stage regression that controls for year-of-birth effects, year effects, education dummies, family size dummies, etc. Instead of imposing random walk, we generalize the permanent component to an autoregressive process, $p_{it} = \rho p_{it-1} + \xi_{it}$. Following BPP, we assume that permanent and transitory shocks to income are independent,[18] and that the transitory component is an MA(1) process ($\tau_{it} = \epsilon_{it} + \theta\epsilon_{it-1}$). There is a large literature with an objective of estimating $\rho$ in a GMM setting that commonly restricts $\alpha_i$ to be an i.i.d. component; see, e.g., Arellano and Honoré (2001) for a review. To form orthogonality conditions, we further need to make restrictions on the initial conditions for the permanent component, $p_{i0}$. We assume that the permanent component at the start of an individual's working career is zero for all individuals as is done, e.g., in Guvenen (2009).

For convenience, the income process can be written as $y_{it} = (1-\rho)\alpha_i + \rho y_{it-1} + \xi_{it} + \epsilon_{it} - (\rho - \theta)\epsilon_{it-1} - \rho\theta\epsilon_{it-2}$. Given our assumptions, the time-$t$ quasi-difference $y_{it} - \rho y_{it-1}$ will be uncorrelated with income growth measured at times $t - j$, $j \geq 3$. In particular, we can use the following set of orthogonality conditions to identify $\rho$: $E\left[(y_{it} - \rho y_{it-1})\Delta y_{it-j}\right]$, $t = 1982, \ldots, 1992$, $j \geq 3$. This is the GMM estimator in levels that, under our formulation of the income process, satisfies the constant-correlated effects assumption required for its validity; see Bun and Sarafidis (2015) for more details.

The GMM estimates of the persistence of permanent shocks are reported in Table 11. We consider two notions of income – the residual income used in BPP and "raw" net family incomes measured as the income residuals from a regression that controls for year dummies only. For both notions of income, the estimated persistence for non-sample families is lower than for sample families.[19] Therefore, the differential persistence is not spuriously induced by the controls of the first-stage regression. We also performed the GMM estimations for

of the transitory component of income for non-sample families. We have found a significantly better fit to the data for the parsimonious AR(1) plus MA(1) process than an alternative of maintaining the random walk assumption for permanent shocks and relaxing the assumption of an MA(1) transitory component instead.

[18] The assumption is standard in the literature. See Ejrnæs and Browning (2014) and Hryshko (2014) for exceptions.

[19] Adding first-differenced moments to the levels moments – the estimation known as system-GMM; see Blundell and Bond (1998) – delivers qualitatively similar results in that the estimated persistences for the sample families are higher than the corresponding values obtained for the non-sample families.

Table 11: GMM Estimates of Persistence.

| | Residuals | | | Raw data | | |
|---|---|---|---|---|---|---|
| | Sample sons (1) | Non-samp. (2) | Sample all (3) | Sample sons (4) | Non-samp. (5) | Sample all (6) |
| $\rho$, persistence | 0.94 | 0.78 | 0.93 | 0.94 | 0.87 | 0.96 |
| perm. shock | (0.02) | (0.04) | (0.01) | (0.02) | (0.02) | (0.01) |

*Notes*: Standard errors in parentheses.

college and no college families separately and the results were similar in that, conditional on education, non-sample families have a much lower persistence of income shocks.[20]

## 4.3 The Effect of Misspecification of the Income Process on the Estimated Variances of Income Shocks

### 4.3.1 Evidence of Misspecification

If the misspecified income process is the main culprit of the different estimated insurances of permanent shocks for sample versus non-sample families, we may expect that the random walk-MA(1) decomposition will perform relatively worse in terms of the fit to the income moments for non-sample families. Following Heathcote, Perri, and Violante (2010), we therefore examine the fit of the estimated BPP models to the income moments in levels and growth rates.

First, in Figure A-3(a) we consider non-sample families and examine the fit of the BPP model (short-dashed line), which assumes that the permanent component is a random walk and targets the moments for income and consumption growth rates, to the moments of income in levels (solid line). In the data, the variance of log residual incomes rises from about 0.12 to 0.18, while the model predicts a rise to about 0.45. Thus, remarkably, the variance of incomes in levels is overestimated by about 150% in the last sample year.

Figure A-4(a) similarly plots the fit of the BPP model to the income moments in levels for sample families. The variances of log incomes are overestimated in 1992 by about 40%, which, however, is substantially lower relative to overestimation for the non-sample families.

[20] We also estimated the income process using the methodology proposed by Browning, Ejrnæs, and Alvarez (2010) that relies neither on the autocovariance function of income levels and/or growth rates nor on the orthogonality conditions exploited in this section. The results once again revealed lower persistence of income shocks for non-sample households than for their sample counterparts.

In Figures A-3(b) and A-4(b), we plot the fit (short-dashed line) to the key income moments in growth rates (solid line).[21] The fit of the BPP model, which specifically targets these moments, is fairly good for the variances, and first two autocovariances.

Figure A-5 provides further evidence that the income dynamics in levels is different for the families headed by sample versus non-sample males. Specifically, we plot the autocorrelation function of net family incomes for households headed by sample sons, all families headed by sample males ("sample sons" and "original sample" members), and households headed by non-sample males.[22] As can be seen, income dynamics differs markedly between sample and non-sample families. Below, we will provide the fit of various estimated models to these autocorrelation functions (that are not targeted in the minimum-distance estimation).

### 4.3.2 Sources of Misspecification

***1. Persistence of Shocks.*** The identifying moments used by Heathcote, Perri, and Violante (2010) can be used to better understand these findings. Consider the income process which is the sum of the permanent random-walk component and an i.i.d. transitory shock.[23] The following moments identify the variances of permanent and transitory shocks, $\sigma^2_{\xi,t}$ and $\sigma^2_{\epsilon,t}$ when targeting the moments of incomes in levels and growth rates, respectively:[24]

Differences :

$$\sigma^2_{\xi,t,\text{diffs}} = E[\Delta y_{it}\Delta y_{it-1}] + E[\Delta y_{it}\Delta y_{it}] + E[\Delta y_{it}\Delta y_{it+1}] \tag{2}$$

$$\sigma^2_{\epsilon,t,\text{diffs}} = -E[\Delta y_{it}\Delta y_{it+1}]. \tag{3}$$

Levels :

$$\sigma^2_{\xi,t,\text{levs}} = E[y_{it}y_{it+1}] - E[y_{it}y_{it-1}] \tag{4}$$

$$\sigma^2_{\epsilon,t,\text{levs}} = E[y_{it}y_{it}] - E[y_{it}y_{it+1}]. \tag{5}$$

The estimated variance of permanent and transitory shocks using these identifying moments in levels and differences should be identical if the true and estimated income processes are identical. If, however, the permanent component is an AR(1) process, $p_{it} = \rho p_{it-1} + \xi_{it}$,

[21] Since autocovariances of income growth rates beyond the second order are typically of smaller absolute magnitude and less precise (and therefore receive a smaller weight in estimations), it is reasonable to expect a relatively worse fit for them.

[22] Autocorrelation of order $j$ in year $t$ is calculated as $\frac{E[y_{it}y_{it+j}]}{\sqrt{E[y_{it}y_{it}]}\sqrt{E[y_{it+j}y_{it+j}]}}$. In the figure, for each $j$, we plot autocorrelations averaged over all $t$'s.

[23] The assumption of an i.i.d. transitory component is consistent with a small and insignificant estimate of the moving average parameter in the BPP estimation for non-sample families.

[24] Heathcote, Perri, and Violante (2010) used the moments identifying the variances of the shocks from biennial data but these moments are analogous if one instead relies on the annual data for identification of the variances in each year.

we show in the Appendix that the difference in the estimated moments (2) and (4), and (5) and (3) will equal, respectively:

$$\sigma^2_{\xi,t,\text{diffs}} - \sigma^2_{\xi,t,\text{levs}} = (1-\rho)(\rho+\rho^3)\text{var}(p_{it-2}) > 0 \tag{6}$$

$$\sigma^2_{\epsilon,t,\text{levs}} - \sigma^2_{\epsilon,t,\text{diffs}} = \rho(1-\rho)\text{var}(p_{it-1}) > 0. \tag{7}$$

Clearly, misspecification of the permanent component may lead to inflated estimates of the variances of permanent (transitory) shocks when targeting the moments in growth rates (levels), to the extent the minimum-distance estimation relies on the identifying moments (2)–(5). Misspecification will lead to negligible biases if the persistence, $\rho$, is close to one; the biases, however, are expected to be larger for smaller values of $\rho$. As the results above point to a value of $\rho$ substantially lower than one for non-sample families, the variance of permanent shocks would be significantly overestimated when moments in growth rates are targeted, leading to the dramatic overestimation of the variance of income levels observed above. We also show in the Appendix that one may expect a larger downward bias in the estimated transmission coefficient for permanent shocks using the random-walk assumption for smaller values of the true persistence $\rho$.

***2. Income Observations at the Start or End of Contiguous Income Spells.*** Daly, Hryshko, and Manovskii (2016) have pointed another potential source of bias in estimating the variances of shocks. The bias arises if income records in the beginning or end of incomplete income spells, or around missing records are systematically different in their means or variances. This can occur, for example, at the beginning of marriages for the newly-formed couples, or in the end of marriages for the couples which dissolve during 1978–1992.

To assess the presence of these effects in our net family income data, in Table 12 we compare first and last income records within incomplete income spells, as well as the records around missing observations with other income observations. On average, income residuals in those periods are not different from zero both for sample and non-sample families – columns (1) and (3), respectively (with observations after missing income records for non-sample families deviating from this pattern). This is in contrast to the results of similar regressions based on the male earnings data, both in administrative and PSID data, as documented in Daly, Hryshko, and Manovskii (2016). This is likely the result of smoothing of shocks to male earnings provided by family labor supply and the tax and transfer system. In columns (2) and (4) we estimate similar regressions with squared residual incomes as dependent variables. For the families headed by sample males, incomes are more volatile in

Table 12: Net Family Income Residuals.

| | Sample, All | | Non-sample | |
|---|---|---|---|---|
| | Means | Vars. | Means | Vars. |
| | (1) | (2) | (3) | (4) |
| Year observed: first | –0.00 | 0.08*** | –0.03 | 0.09*** |
| | (0.01) | (22.44) | (0.75) | (17.43) |
| Year observed: last | –0.00 | –0.01 | –0.06* | 0.05** |
| | (0.05) | (0.14) | (3.18) | (4.36) |
| 1 year before inc. miss. | 0.03 | 0.44* | 0.24 | 0.13 |
| | (0.04) | (3.66) | (0.76) | (0.32) |
| 1 year after inc. miss. | –0.13 | 0.75** | 0.43** | –0.13 |
| | (0.52) | (4.65) | (4.83) | (0.90) |
| No. obs. | 14,726 | 14,726 | 6,350 | 6,350 |
| No. indiv. | 1,625 | 1,625 | 804 | 804 |

*Notes:* This table summarizes the results of Table A-2. Income data span the period 1979–1993. Income recorded in year $t$ reflects income received in year $t-1$. "Year observed: first" is equal to the difference between the estimated coefficients for the dummies "Year observed: first, year $\neq$ 1978" (equal to one if an individual's first income record is after 1979, and zero otherwise) and "Year observed: first, year = 1978" (equal to one if an individual's first income record is in 1979, and zero otherwise). "Year observed: last" is the difference between the estimated coefficients for the dummies "Year observed: last, year $\neq$ 1992" (equal to one if an individual's last income record is before 1993, and is equal to zero otherwise) and "Year observed: last, year = 1992" (equal to one if an individual's last income record is in 1993, and is equal to zero otherwise). "1 year before inc. miss." is the difference between the estimated coefficient for the dummy "1 year before inc. miss." (equal one if next period's income is missing, and zero otherwise) and a constant, and "1 year after inc. miss." (equal one if previous period's income is missing, and zero otherwise) is the difference between the estimated coefficient for the dummy "1 year after inc. miss." and a constant. Standard errors are clustered by individual; F-statistics are in parentheses. *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level.

the first year of incomplete income spells, and are substantially more volatile than typical income records in the periods around missing records. For non-sample families, incomes are more volatile in the first and last years of incomplete income spells. Incomes are also more volatile than typical observations right before missing income records but this effect is not statistically significant.

### 4.3.3 Addressing the Two Sources of Misspecification

Motivated by these findings, we introduce two modifications to the specification of the income process. First, we relax the assumption of a random walk in incomes, and model the permanent component as a persistent AR(1) process, $p_{it} = \rho p_{it-1} + \xi_{it}$, estimating, in addition, persistence $\rho$. Second, we follow Daly, Hryshko, and Manovskii (2016) and augment the estimating consumption equation with an additional shock to household incomes to which consumption may react: $\Delta c_{it} = \zeta_{it} + \phi\xi_{it} + \psi\epsilon_{it} + \psi_{\text{rare},t}\nu_{it} + \Delta u_{it}$, where $\nu_{it}$ is an i.i.d. shock (with mean and variance estimated from the data), which appears only in the first and last periods of incomplete income spells, or in the periods right before and after a missing income record.[25] The resulting income process is as follows:

$$
\begin{aligned}
y_{it} &= \alpha_i + p_{it} + \tau_{it} + \chi_{it}, \quad t = t_0, \ldots, T \\
p_{it} &= \rho p_{it-1} + \xi_{it} \\
\tau_{it} &= \epsilon_{it} + \theta\epsilon_{it-1} \\
\chi_{it+j} &= \begin{cases} \nu_{it} & \text{if } y_{it-k} \text{ or } y_{it+k} \text{ is missing and } t-k \geq t_0,\ t+k \leq T, j = 0 \\ \theta\nu_{it} & j = 1 \\ 0 & \text{otherwise,} \end{cases}
\end{aligned}
\tag{8}
$$

where $\alpha_i$ is individual $i$'s fixed effect, $t_0$ is the first sample year (1978), and $T$ is the last sample year (1992). We make restriction $k = 1$ to isolate only the first and last observation of an income spell (if it is different from the first or last year of the sample window), and observations around missing income records inside of an incomplete income spell.

To recover additional parameters, in addition to all of the moments in the original BPP estimation, we also target all the regression coefficients reported in Table A-2. We estimate the model by the method of simulated minimum distance, assuming that persistent, transitory, and rare transitory shocks are drawn from normal distributions, and using the diagonal weighting matrix calculated by block-bootstrap.[26] In estimations, we assumed that the fixed

[25] Daly, Hryshko, and Manovskii (2016) showed that it is sufficient to account for the mean and variance of the first records around missing ones in order to eliminate the biases in the estimated permanent-transitory decomposition of earnings.

[26] We verified that the simulated method of moments with the assumption of normal permanent and

Table 13: Minimum-Distance Partial Insurance Estimates.

| | Sample (1) | Non-sample (2) |
|---|---|---|
| $\rho$, AR coeff. | 0.9908 | 0.8961 |
| | (0.0099) | (0.0338) |
| $\theta$, MA coeff. | 0.1234 | 0.0618 |
| | (0.0251) | (0.0598) |
| $\phi$, transmission perm. shock | 1.0 | 0.6013 |
| | (0.2052) | (0.1251) |
| $\psi$, transmission trans. shock | 0.0465 | 0.0917 |
| | (0.0417) | (0.1114) |

effect in family incomes is independent of the shocks.

### 4.3.4 Estimation Results and Model Fit

Table 13 contains estimation results.[27] For the families headed by sample males, the persistence of longer-lasting shocks is estimated to be close to one, while the transmission coefficient for permanent shocks binds at one. In contrast, for non-sample families, the AR(1) coefficient is estimated to be about 0.90, and the insurance of "permanent" shocks of about 40%.[28,29] In Figures A-3–A-4 we show the fit of the models (long-dashed lines) in Table 13 to the data moments. The models produce an overall good fit both to the moments in levels and growth rates both for sample and non-sample families.

transitory shocks delivers virtually the same parameter estimates as the estimations which assume that the shocks are drawn from a fat-tailed Student t-distribution, the degrees of freedom of which are estimated by matching kurtosis of residual consumption and income growth observed in the data. We also experimented with updating the BPP estimation to allow for fitting the third moments of income and consumption growth, but we did not find any substantial differences in the estimated insurance against permanent income shocks when fitting both the second and third moments. Importantly, the differences in the estimated permanent insurance for sample versus non-sample families remain a robust feature of the data.

[27] As the transmission coefficient for rare shocks was estimated with a large standard error both for sample and non-sample families, we restricted it to equal the transmission coefficient for transitory shocks. The estimated persistence of permanent shocks is invariant to this assumption. The full results of estimations are available upon request.

[28] The insurance of permanent shocks is higher than the insurance estimated under the assumption of a random-walk permanent component, which is consistent with the biases we outline in Appendix.

[29] Note that the values of the estimated persistence for sample and non-sample families are higher than the values reported in Table 11. This is not surprising as the methods use different information; in particular, Table 13 uses, in addition, consumption information to identify the parameters of the income process. Han and Phillips (2010) show that system-GMM estimates of the persistence may be downward-biased when the true persistence is close to unity.

In Figure A-6, we show the fit of the estimated models to the autocorrelation function of income levels. Lines with circles reproduce the autocorrelation function in the data, short-dashed lines produce the autocorrelation function implied by the estimates of the BPP model assuming that incomes contain a random-walk permanent component, and solid lines produce the autocorrelation function implied by the estimates in Table 13. Panels (a) and (b) contain the plots for non-sample and sample families, respectively.[30] Up to order six, the autocorrelation functions produced by the original BPP estimation, and estimation with a modified income process show a similar fit to the autocorrelation function in the data. After order six, however, the estimation with a modified income process shows a much tighter fit to the data moments. For non-sample families, the tighter fit is mainly achieved by a lower estimate of the persistence of longer-lasting shocks, whereas for sample families it is achieved by allowing for additional income shocks at the extremes of contiguous income spells, which appear to be relatively more important for sample families as we documented in Table A-2. Importantly, none of the moments in Figure A-6 had been targeted in estimation.

The main substantive takeaway from these findings is on the relationship between the degree of insurance and the dynamic properties of income shocks. Specifically, we do not observe evidence of excess insurance. In particular, the estimated insurance of permanent shocks of 40% is consistent with the persistence of permanent shocks of 0.90 under reasonable parameterizations of standard self-insurance models. The point estimates for sample households suggest essentially a random walk component in income and a complete lack of insurance against these shocks. The point estimate of the transmission coefficient comes with a substantial standard error wich does not allow us to exclude the possibility that consumption insurance of sample households is also consistent with the model.

# 5 Conclusion

In macro and labor economics, there is a long history of using data from the PSID to address various important issues such as consumption and income inequality, consumption smoothing and income dynamics, and completeness of insurance markets, to name a few. In this paper, we use the PSID to examine how much consumption insurance against the shocks to net family incomes is achieved by U.S. households. We find consistent evidence of drastically different insurance patterns in two distinct PSID subsamples. The PSID comprises the original sample members interviewed in 1968 and their offspring, and non-sample members, whose presence in the data grows over time as they marry PSID sample males or females.

[30] For a better reading, we normalized the plot for the BPP, random-walk estimation, so that the autocorrelation function at order one equals to that observed in the data.

Using data from an influential contribution by Blundell, Pistaferri, and Preston (2008), we find a nearly complete pass-through of permanent income shocks to consumption for households headed by PSID sample males. In contrast, families headed by non-sample males show a dramatically higher degree of insurance against permanent income shocks. We explore the reasons for this discrepancy and find evidence that the dynamics of income is very different among households headed by PSID sample and non-sample males. In particular, income shocks of households headed by non-sample males are considerably less persistent. Allowing for this differential persistence, aligns the estimate of insurance among households headed by non-sample males with the prediction from the standard incomplete markets model. In contrast to the recent findings, we do not find excess consumption insurance beyond that provided by self-insurance due to accumulated household wealth.

While the patterns documented in the paper appear highly robust, the existence of large differences in the stochastic properties of income and consumption between households headed by sample and non-sample males is somewhat surprising. Our extensive exploration of the data yielded no reason to suspect that these differences are not genuine but are, e.g., an artifact of the PSID procedures. However, the PSID is the foundation of our knowledge of household income and consumption dynamics. Virtually all quantitative incomplete markets models in the literature are either estimated using the PSID data or take PSID estimates of income processes as the key inputs. Given the central role played by these data in addressing key economic questions, we hope our findings will stimulate additional research on better understanding the relationship between sample/non-sample status of the respondent on the one hand and the dynamics of income and consumption of sample/non-sample respondents on the other.

# References


Abowd, J., and D. Card (1989): "On the Covariance Structure of Earnings and Hours Changes," *Econometrica*, 57(2), 411–445.

Arellano, M., R. Blundell, and S. Bonhomme (2017): "Earnings and Consumption Dynamics: A Nonlinear Panel Data Framework," *Econometrica, forthcoming*.

Arellano, M., and B. Honoré (2001): "Panel Data Models: Some Recent Developments," in *Handbook of Econometrics, Vol. 5*, ed. by J. J. Heckman, and E. E. Leamer, chap. 53, pp. 3229–3296. Elsevier, Amsterdam.

Becketti, S., W. Gould, L. Lillard, and F. Welch (1988): "The Panel Study of Income Dynamics after Fourteen Years: An Evaluation," *Journal of Labor Economics*, 6(4), 472–492.

Blundell, R. (2014): "Income Dynamics and Life-Cycle Inequality: Mechanisms and Controversies," *Economic Journal*, 124, 289–318.

Blundell, R., and S. Bond (1998): "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, 87, 115–143.

Blundell, R., L. Pistaferri, and I. Preston (2008): "Consumption Inequality and Partial Insurance," *American Economic Review*, 98(5), 1887–1921.

Blundell, R., L. Pistaferri, and I. Saporta-Eksten (2016): "Consumption Inequality and Family Labor Supply," *American Economic Review*, 106(2), 387–435.

Browning, M., M. Ejrnæs, and J. Alvarez (2010): "Modelling Income Processes with Lots of Heterogeneity," *Review of Economic Studies*, 77, 1353–1381.

Bun, M. J., and V. Sarafidis (2015): "Dynamic Panel Data Models," in *The Oxford Handbook of Panel Data*, ed. by B. H. Baltagi, chap. 3, pp. 76–110. Oxford University Press, Oxford.

Campos, R. G., and I. Reggio (2014): "Measurement Error in Imputation Procedures," *Economics Letters*, 122, 197–202.

Daly, M., D. Hryshko, and I. Manovskii (2016): "Improving the Measurement of Earnings Dynamics," NBER Working Paper # 22938.

Ejrnæs, M., and M. Browning (2014): "The Persistent-Transitory Representation for Earnings Processes," *Quantitative Economics*, 5, 555–581.

Guvenen, F. (2009): "An Empirical Investigation of Labor Income Processes," *Review of Economic Dynamics*, 12(1), 58–79.

Han, C., and P. C. B. Phillips (2010): "GMM Estimation for Dynamic Panels with Fixed Effects and Strong Instruments at Unity," *Econometric Theory*, 26, 119–151.

Heathcote, J., F. Perri, and G. L. Violante (2010): "Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States: 1967–2006," *Review of Economic Dynamics*, 13(1), 15–51.

Hryshko, D. (2014): "Correlated Income Shocks and Excess Smoothness of Consumption," *Journal of Economic Dynamics and Control*, 48, 41–62.

Hryshko, D., M. J. Luengo-Prado, and B. E. Sørensen (2011): "Childhood Determinants of Risk Aversion: The Long Shadow of Compulsory Education," *Quantitative Economics*, 2, 37–72.

Kaplan, G., and G. L. Violante (2010): "How Much Consumption Insurance Beyond Self-Insurance?," *American Economic Journal: Macroeconomics*, 2, 53–87.

Lillard, L. A. (1989): "Sample Dynamics: Some Behavioral Issues," in *Panel Surveys*, ed. by D. Kasprzyk, G. Duncan, G. Kalton, and M. P. Singh, pp. 497–511. Wiley Series in Probability and Statistics.

Mazzocco, M., C. Ruiz, and S. Yamaguchi (2007): "Labor Supply, Wealth Dynamics, and Marriage Decisions," Discussion paper, University of California Los Angeles, Mimeo.

# Appendix

If the permanent component is an AR(1) process with persistence $\rho$, the misspecified variance of permanent shocks using the moments in levels equals

$$\sigma^2_{\xi,t,\text{levs}} = E[y_{it}y_{it+1}] - E[y_{it}y_{it-1}] = (\rho^3 - \rho)\text{var}(p_{t-1}) + \rho\sigma^2_{\xi_t}.$$

Using the moments in differences instead, the variance will equal

$$\sigma^2_{\xi,t,\text{diffs}} = E\left[\Delta y_{it}\sum_{j=-1}^{1}\Delta y_{it+j}\right] = (\rho-1)(\rho^4-\rho)\text{var}(p_{t-2}) + (\rho^3-\rho^2)\sigma^2_{\xi_{t-1}} + \rho\sigma^2_{\xi_t}.$$

Since $\text{var}(p_{t-1}) = \rho^2\text{var}(p_{t-2}) + \sigma^2_{\xi_{t-1}}$,

$$\sigma^2_{\xi,t,\text{diffs}} - \sigma^2_{\xi,t,\text{levs}} = \text{var}(p_{t-2})\left[(\rho-1)(\rho^4-\rho) - \rho^2(\rho^3-\rho)\right] = (1-\rho)(\rho+\rho^3)\text{var}(p_{t-2}),$$

which is greater than zero for $0 < \rho < 1$.
The misspecified variance of transitory shocks using the moments in levels equals

$$\sigma^2_{\epsilon,t,\text{levs}} = E[y_{it}y_{it}] - E[y_{it}y_{it+1}] = \rho^2(1-\rho)\text{var}(p_{t-1}) + (1-\rho)\sigma^2_{\xi_t} + \sigma^2_{\epsilon,t}.$$

Using the moment in differences instead,

$$\sigma^2_{\epsilon,t,\text{diffs}} = -E[\Delta y_{it}\Delta y_{it+1}] = -\rho(1-\rho)^2\text{var}(p_{t-1}) + (1-\rho)\sigma^2_{\xi_t} + \sigma^2_{\epsilon,t}.$$

This implies that

$$\sigma^2_{\epsilon,t,\text{levs}} - \sigma^2_{\epsilon,t,\text{diffs}} = \rho(1-\rho)\text{var}(p_{t-1}),$$

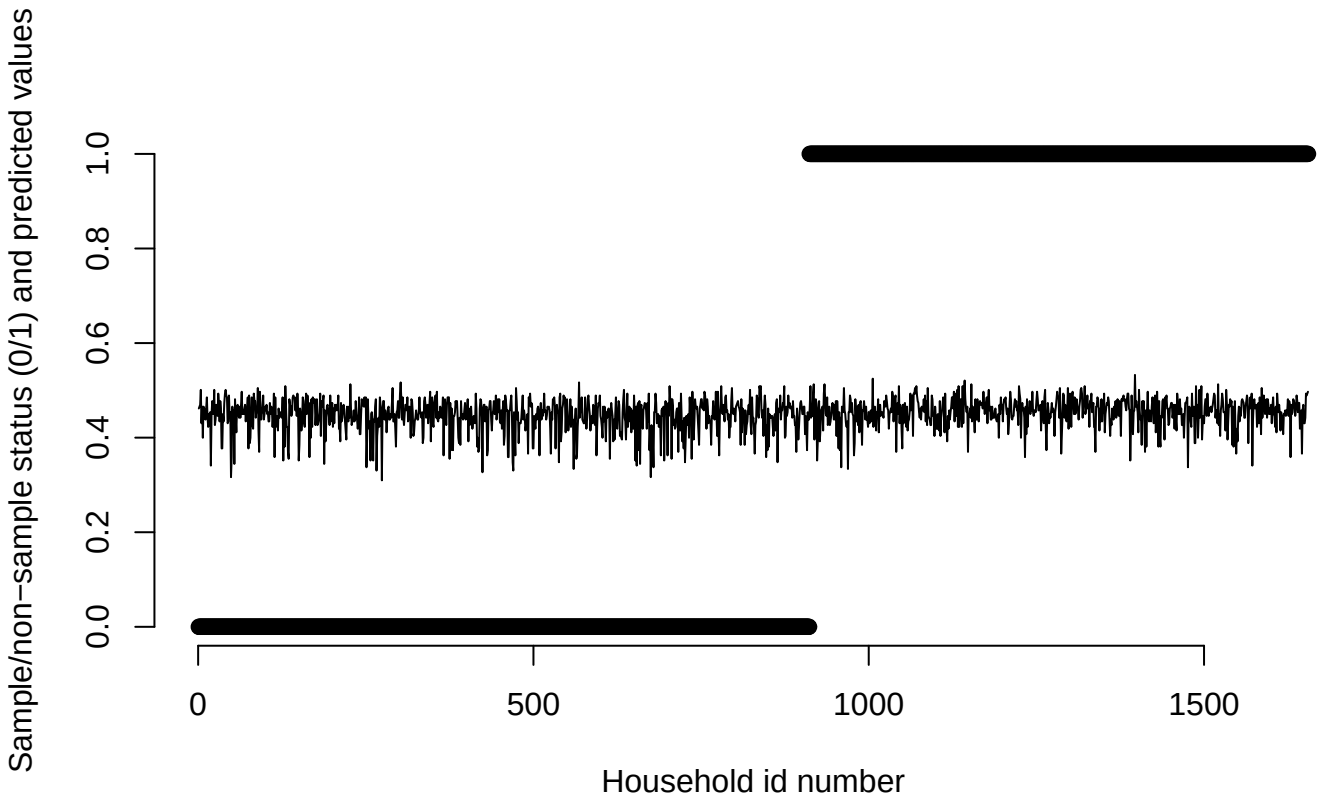
which, again, is greater than zero for $0 < \rho < 1$.
The identifying moment for the insurance of permanent shocks is

$$\hat{\phi}_t = \frac{E[\Delta c_{it}\sum_{j=-1}^{1}\Delta y_{it+j}]}{E[\Delta y_{it}\sum_{j=-1}^{1}\Delta y_{it+j}]} = \frac{\phi_t\rho\sigma^2_{\xi_t}}{(\rho-1)(\rho^4-\rho)\text{var}(p_{t-2}) + (\rho^3-\rho^2)\sigma^2_{\xi_{t-1}} + \rho\sigma^2_{\xi_t}}.$$

Assuming that the variance of persistent shocks does not change over time, $\text{var}(p_{t-2}) = \frac{1-\rho^{2(t-2)}}{1-\rho^2}\sigma^2_\xi$. It then follows that

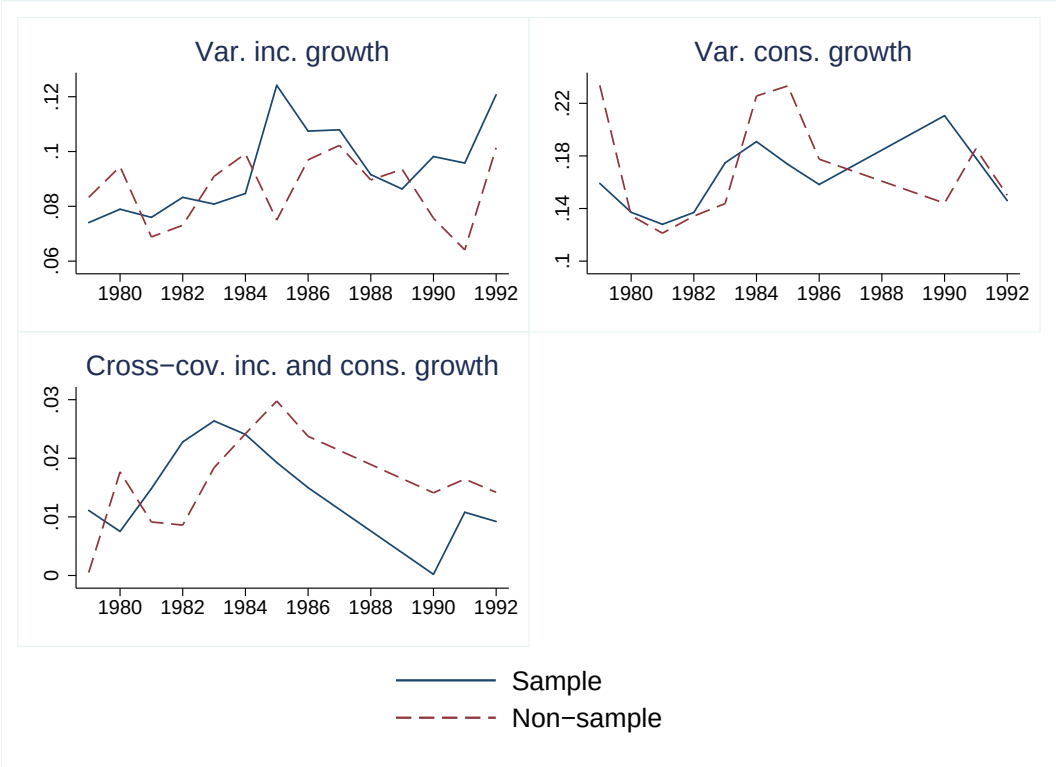
$$\hat{\phi}_t = \frac{E[\Delta c_{it}\sum_{j=-1}^{1}\Delta y_{it+j}]}{E[\Delta y_{it}\sum_{j=-1}^{1}\Delta y_{it+j}]} = \frac{\rho\phi_t}{(\rho-1)(\rho^4-\rho)\frac{1-\rho^{2(t-2)}}{1-\rho^2} + \rho^3 - \rho^2 + \rho} = \frac{\phi_t}{(1-\rho^3)\frac{1-\rho^{2(t-2)}}{1+\rho} - \rho(1-\rho) + 1}.$$

Figure A-1: LASSO predictions of the sample/non-sample status.

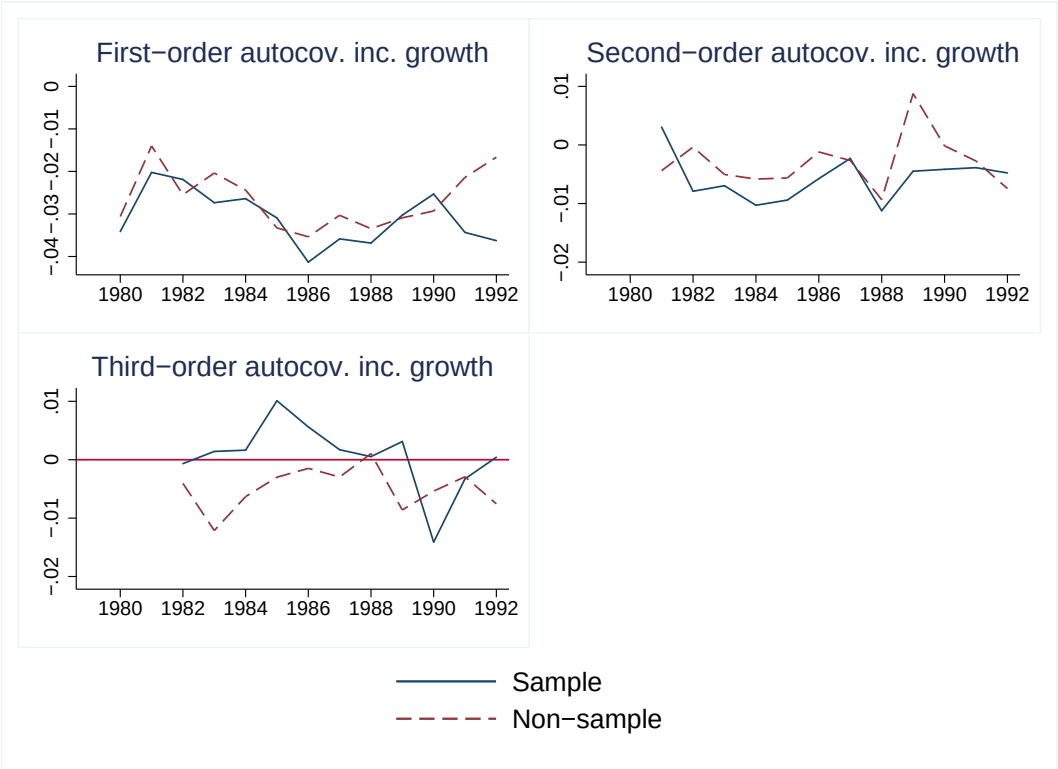



*Notes:* The following regressors and their squared values were used for predicting the sample/non-sample status: annual hours worked of head and wife, labor income of head and wife, net family income, family size, number of children, combined transfer income of head and wife, transfer income of other family members, head's year of birth, wife's year of birth, help from others, net assets; employment, unemployment and temporary layoff dummies, dummies for disability, residence in a large city, outside dependents, presence of extra earners other than head and wife; net family income and nondurable consumption residuals, food at home, and away from home, food stamps; dummies for a working wife, displacement, business ownership, college and non-college degrees, region of growing up, homeownership; number of observations on net family incomes and (imputed) nondurable consumption, race dummies, nondurable consumption, standard deviation of family nondurable consumption, and standard deviation of net family income. All variables that vary over time were averaged.

Figure A-2: Data moments.

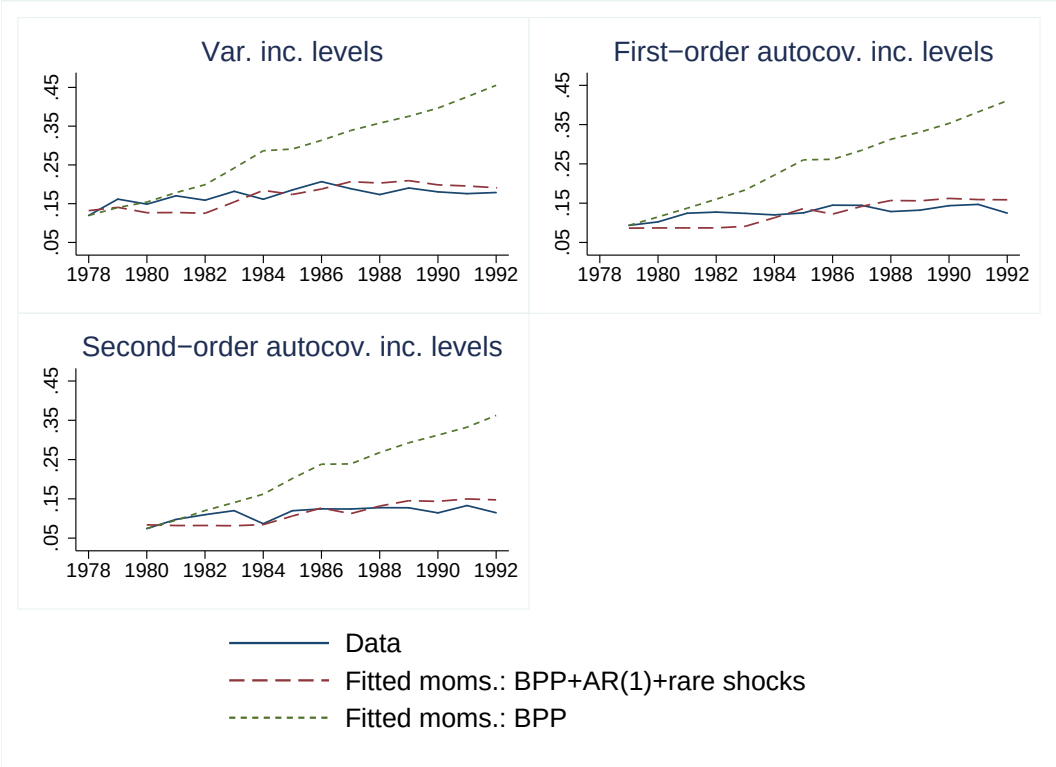



(a) Income growth, consumption growth, and cross-covariance of consumption and income growth.

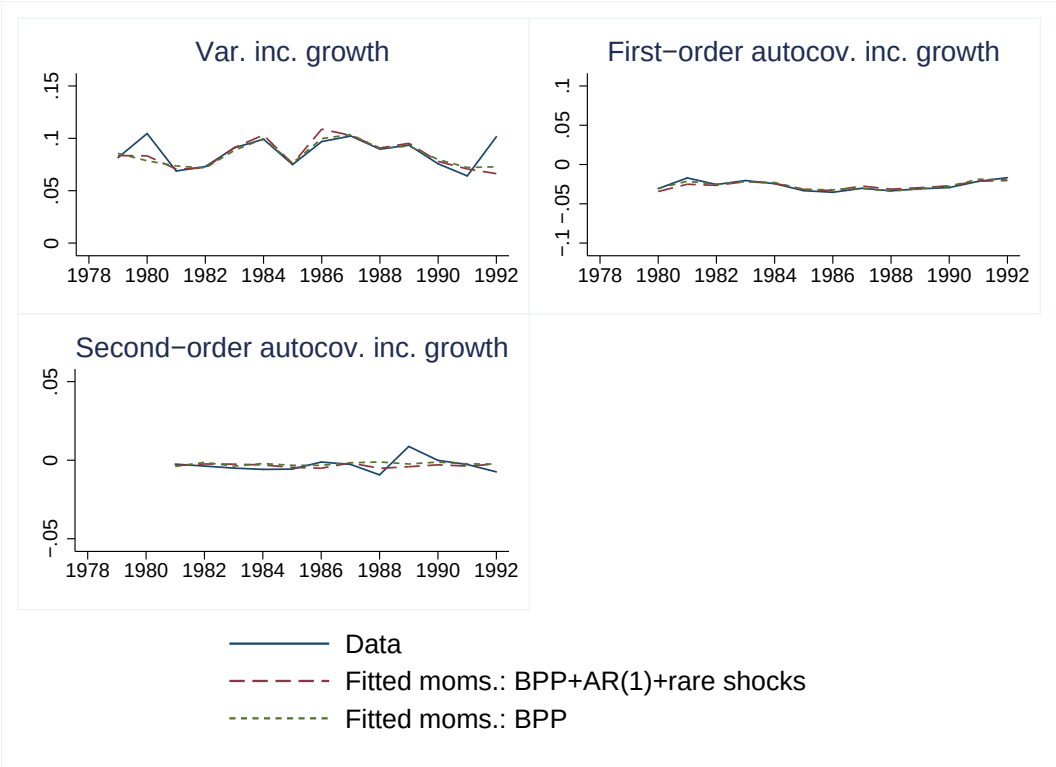



(b) Autocovariances of income growth rates.

Figure A-3: Model Fit. Non-Sample Households.

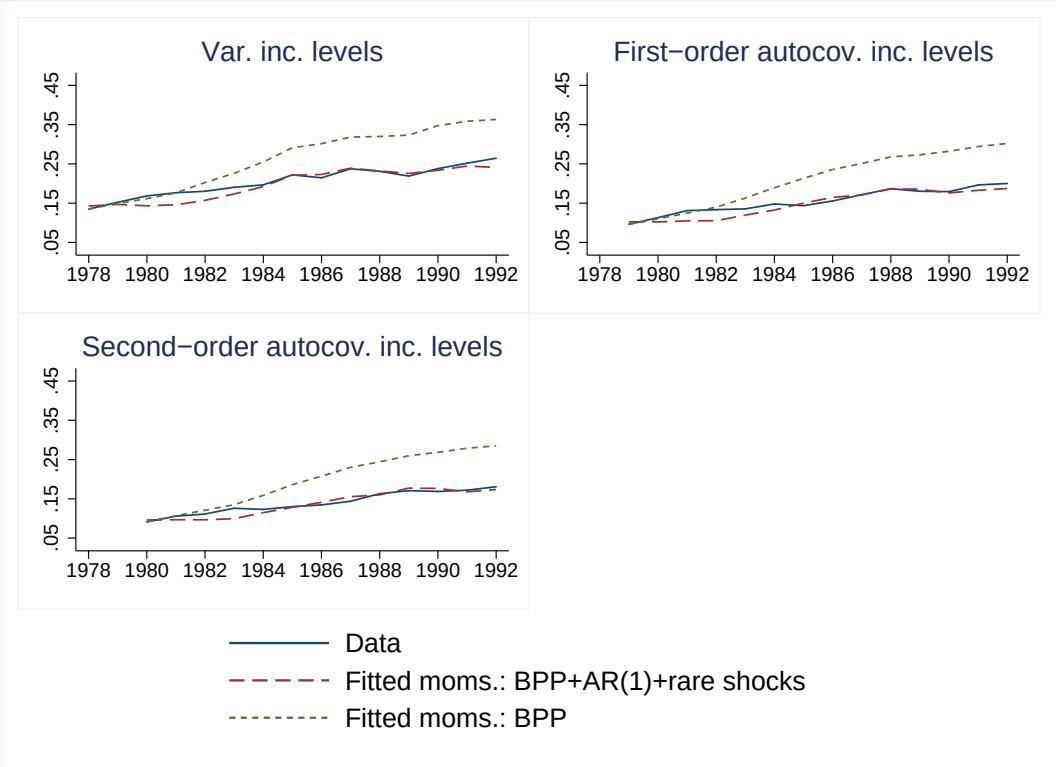



(a) Autocovariances of incomes in levels

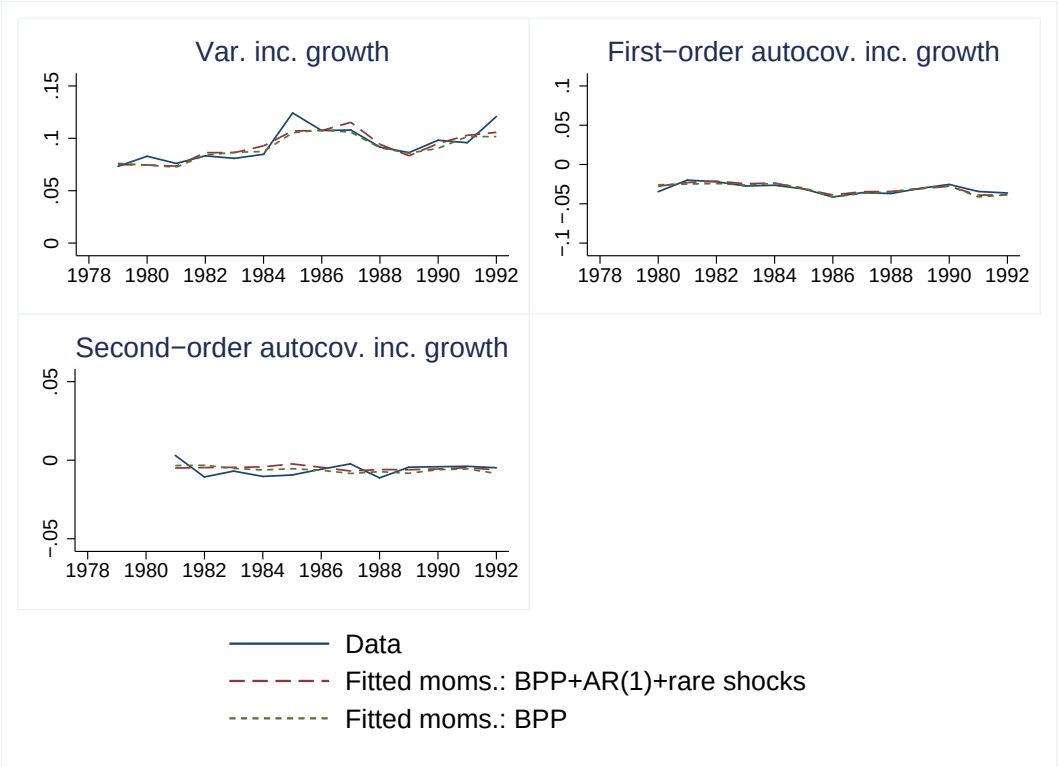



(b) Autocovariances of income growth rates

Figure A-4: Model Fit. Sample Households.




(a) Autocovariances of incomes in levels




(b) Autocovariances of income growth rates

Figure A-5: Autocorrelation function of net family incomes.

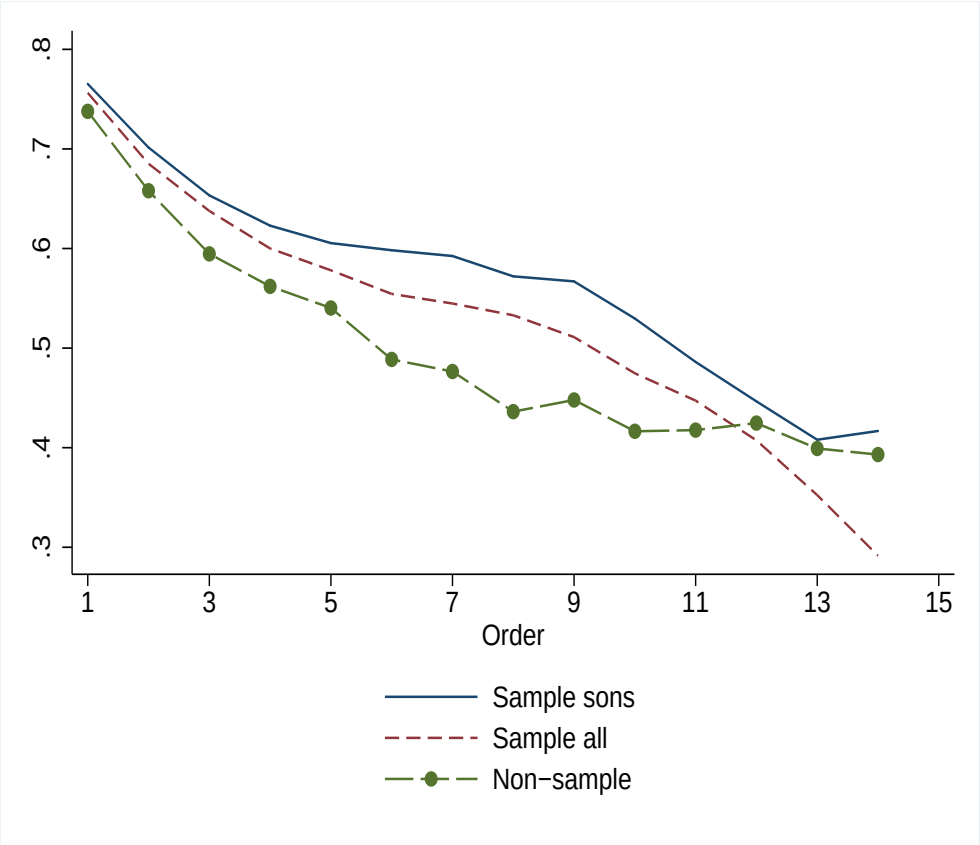
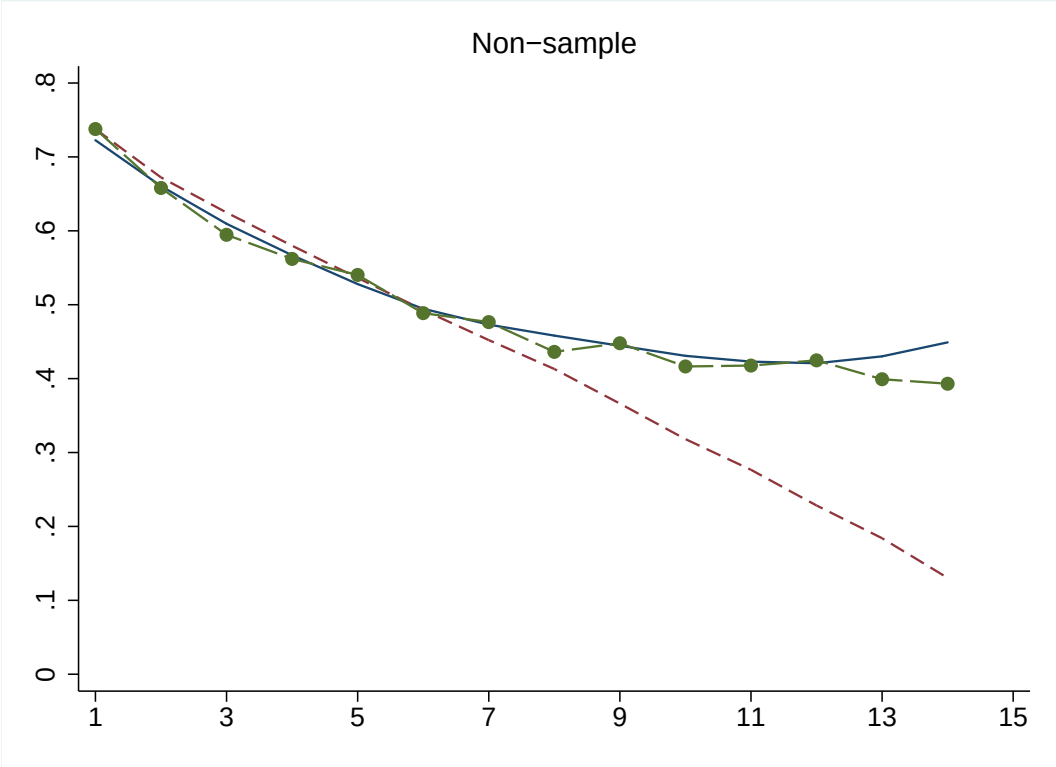
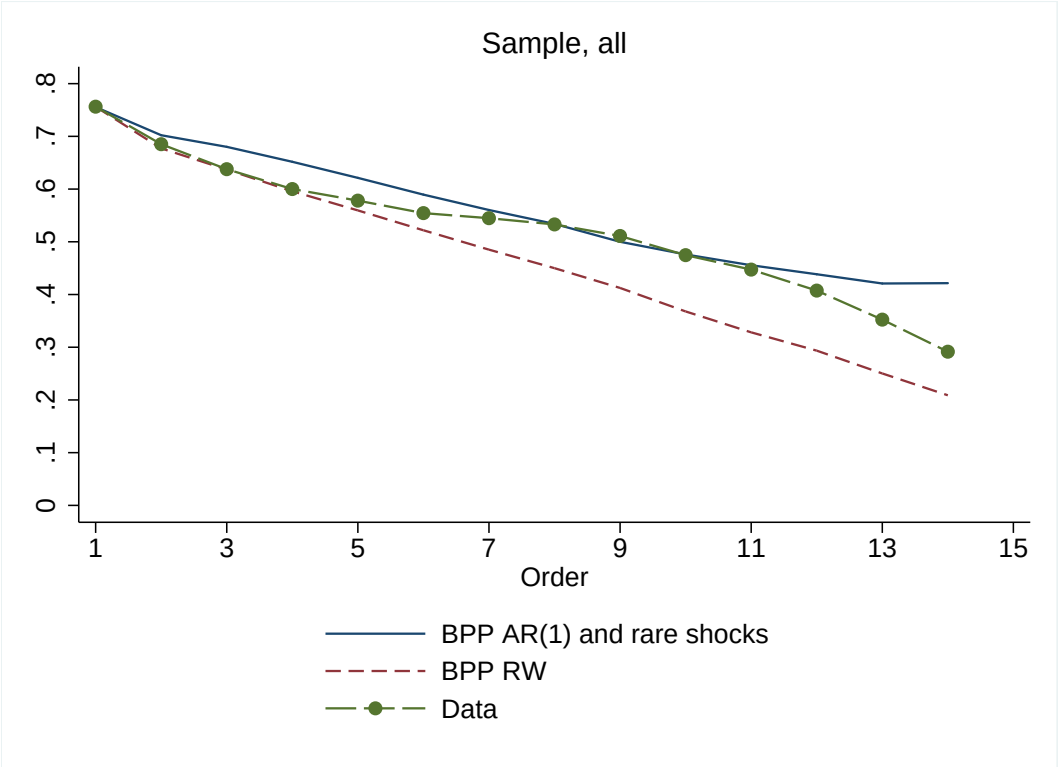


Figure A-6: Model Fit. Autocorrelation function of income levels.




(a) Non-sample




(b) Sample, all

Table A-1: Means of selected variables for various PSID samples.

| | Sample orig. | Sample sons | Non-sample |
|---|---|---|---|
| Head's age | 51.586 | 39.618 | 38.746 |
| Wife's age | 49.156 | 36.862 | 35.313 |
| Nondur. cons. | 43,767 | 28,432 | 25,867 |
| Fam. taxable income | 42,913 | 39,721 | 40,680 |
| Head's earnings | 28,252 | 27,330 | 28,179 |
| Assets | 240,755 | 113,704 | 114,133 |
| Amount of help from relatives | 22.067 | 67.759 | 97.516 |
| If head empl. | 0.819 | 0.904 | 0.918 |
| If head unempl. | 0.009 | 0.011 | 0.014 |
| Head's hours | 1938.697 | 2137.562 | 2186.915 |
| Wife's hours | 959.421 | 1179.178 | 1207.42 |
| If head works | 0.897 | 0.954 | 0.968 |
| If wife works | 0.65 | 0.77 | 0.798 |
| No college | 0.589 | 0.414 | 0.43 |
| College | 0.411 | 0.586 | 0.570 |
| No. children | 0.642 | 1.492 | 1.542 |
| Fam. size | 3.137 | 3.648 | 3.662 |
| White | 0.924 | 0.942 | 0.927 |
| Black | 0.058 | 0.045 | 0.053 |
| North East | 0.243 | 0.192 | 0.206 |
| Midwest | 0.309 | 0.304 | 0.306 |
| South | 0.303 | 0.311 | 0.295 |
| West | 0.145 | 0.193 | 0.193 |
| If provided mon. supp. to others | 0.17 | 0.161 | 0.168 |
| If inc. other members >0 | 0.472 | 0.263 | 0.229 |
| Food at home, minor assign. | 0.004 | 0.003 | 0.003 |
| Food at home, major assign. | 0.006 | 0.003 | 0.003 |
| Food away, minor assign. | 0.004 | 0.002 | 0.002 |
| Food away, major assign. | 0.006 | 0.003 | 0.004 |
| Percent tot. fam. inc., major assign | 5.061 | 2.78 | 2.787 |
| Percent tot. fam. inc., minor assign. | 6.038 | 3.446 | 3.354 |
| Head changed occ. | 0.303 | 0.346 | 0.34 |
| Head changed industry | 0.262 | 0.292 | 0.308 |
| Head disabled | 0.172 | 0.118 | 0.101 |
| Head displaced | 0.035 | 0.052 | 0.056 |
| Fam. owns business | 0.202 | 0.205 | 0.239 |
| Respondent, head | 0.769 | 0.825 | 0.58 |
| Respondent, wife | 0.229 | 0.173 | 0.418 |
| Total tax exemptions, head and wife | 3.01 | 3.674 | 3.678 |
| Region head grew up: foreign country | 0.038 | 0.018 | 0.012 |
| Fam. owns a house | 0.935 | 0.822 | 0.812 |

Table A-2: Net Family Income Residuals.

| | Sample, All | | Non-sample | |
|---|---|---|---|---|
| | Means | Vars. | Means | Vars. |
| | (1) | (2) | (3) | (4) |
| Year observed: first, year = 1978 | 0.00 | –0.08*** | –0.01 | –0.06*** |
| | (0.22) | (–6.65) | (–0.53) | (–3.81) |
| Year observed: first, year ≠ 1978 | 0.00 | 0.01 | –0.04** | 0.03* |
| | (0.03) | (0.41) | (–2.14) | (1.89) |
| Year observed: last, year = 1992 | –0.02 | 0.06*** | 0.02 | 0.00 |
| | (–1.13) | (3.40) | (1.08) | (0.23) |
| Year observed: last, year ≠ 1992 | –0.02 | 0.05** | –0.04 | 0.05*** |
| | (–1.07) | (2.24) | (–1.51) | (2.68) |
| 1 year before inc. miss. | 0.03 | 0.64*** | 0.26 | 0.31 |
| | (0.19) | (2.78) | (0.92) | (1.33) |
| 1 year after inc. miss. | –0.13 | 0.95*** | 0.44** | 0.05 |
| | (–0.73) | (2.72) | (2.27) | (0.38) |
| Constant | –0.00 | 0.20*** | 0.01 | 0.18*** |
| | (–0.18) | (23.49) | (0.88) | (17.99) |
| No. obs. | 14,726 | 14,726 | 6,350 | 6,350 |
| No. indiv. | 1625 | 1625 | 804 | 804 |

*Notes:* Income data span the period 1979–1993. Income recorded in year $t$ reflects income received in year $t-1$. The dummies "Year observed: first, year = 1978" ("Year observed: first, year ≠ 1978") is equal to one if an individual's first income record is in 1979 (after 1979), and is equal to zero otherwise; "Year observed: last, year = 1992" (Year observed: last, year ≠ 1992) is equal to one if an individual's last income record is in 1993 (before 1993), and is equal to zero otherwise. Standard errors are clustered by individual; t-statistics are in parentheses. *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level.