Income Dynamics and Consumption Insurance*

Dmytro Hryshko†
University of Alberta

Iourii Manovskii‡
University of Pennsylvania

Abstract

An accurate quantitative analysis using lifecycle incomplete markets models requires that they match both the total amount of insurance available to households in the data and the empirical importance of the sources of insurance such as, e.g., household saving and borrowing or the tax and transfer system. The prominent empirical benchmark estimates of the extent and sources of insurance provided in Blundell, Pistaferri and Preston (American Economic Review, 2008) imply that the currently used models poorly fit these data. We show that this is because the income process used as an input in the incomplete markets models and when constructing the empirical benchmark estimates of insurance abstracts from the irregular nature of income observations at the start and end of family income spells. When ignored, this feature of the income data induces a large bias in the empirical measures of the degree and sources of insurance and results in a misleading assessment of the models' performance. The empirical measures of insurance accounting for the bias imply a limited role of assets and taxes and transfers in insuring permanent shocks to household budgets.

*We benefited from comments of participants at numerous conferences and seminars. Support from the National Science Foundation Grants No. SES-1357903 and SES-1824520, the Danish Social Science Research Council (FSE) Grant No. 300279, and SSHRC IG Grant No. 435-2018-1275 is gratefully acknowledged.

†Department of Economics, University of Alberta, 8–14 HM Tory Building, Edmonton, AB, T6G2H4, 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

Models with incomplete insurance markets are at the heart of modern heterogeneous-agent macroeconomics. To ensure accurate quantitative analysis using this class of models and to properly assess implications of various economic disturbances, it is essential that these models replicate the correct amount of insurance available to agents. For an informative policy analysis, it is also necessary that these models accurately characterize the sources of insurance, e.g., the role of the tax and transfer system in mitigating the impact of shocks to family earnings on consumption.[1] In a seminal contribution, Blundell et al. (2008) – BPP hereafter – developed a measurement methodology and provided estimates of the insurance against permanent and transitory income shocks available to families in the data where insurance is defined by the fraction of shocks that does not translate into movements in consumption. They also showed how to assess in the data the contribution of taxes and transfers to buffering the impact of shocks to household earnings on consumption. These estimates have become the key benchmarks for assessing the performance of incomplete markets models.[2]

The key empirical finding in BPP is that households are able to almost fully insure transitory shocks to their budgets and that over a third of permanent shocks does not induce changes in consumption. As emphasized by Kaplan and Violante (2010), this magnitude of measured insurance against permanent income shocks is too large relative to quantitative predictions of the benchmark incomplete markets model. The insurance role of the tax and transfer system implied by BPP estimates is also surprisingly large. BPP inferred the role of the tax and transfer system by comparing the estimates of insurance against the shocks to family earnings and the shocks to net family income. Because the consumption measure is held constant, the transmission of shocks to both income measures equally reflects the insurance achieved through saving and borrowing, so that the difference between the two measures of pass-through reflects the role of taxes and transfers.[3] BPP also mentioned an alternative

[1]E.g., Blundell et al. (2016), Boar and Midrigan (2020), Heathcote et al. (2014), Heathcote et al. (2017), and Wu and Krueger (2020) use estimates of the importance of the tax and transfer system as an input into their quantitative models.

[2]See, e.g., Kaplan and Violante (2010), Guvenen et al. (2016), and De Nardi et al. (2020).

[3]This difference includes the contribution of both private and public transfers across households, but the

method to infer the role of the tax and transfer system from the extent of reduction in the estimated variance of net family income relative to the variance of family earnings. Although these two methods should yield identical estimates in theory, BPP's results imply that they differ drastically in practice. While the first method implies that 63% of permanent shocks to household earnings are insured by the tax and transfer system, the second method implies that only 42% of them are. The discrepancy is particularly unfortunate because, as we show in the Appendix, both of these methods of measuring the insurance role of taxes and transfers are robust to measurement error in earnings which is prevalent in the data, while the alternative methods available in the literature are not. However, the implausibly high degree of insurance due to taxes and transfers that BPP-based methods imply and the large discrepancy between the estimates based on them prevent these methods from becoming the primary benchmarks in the literature.

In this paper, we show that the surprisingly high overall degree of insurance against permanent income shocks, the unexpectedly important role of the tax and transfer system, and the large difference in the estimates of its role using the two methods trace to the same root cause – the bias in the measured dynamics of income and earnings.

Specifically, BPP estimated consumption insurance and the risk to family income and earnings using data from the Panel Study of Income Dynamics (PSID) which contain families with incomplete income spells. In a recent paper Daly et al. (2016) showed, using survey and administrative data on male earnings, that log earnings observations at the start or end of contiguous earnings spells (i.e., preceded or followed by a missing observation) are lower on average and substantially more volatile than the typical draws from the interior of individual earnings spells. This can be due to tenure effects on wages, incomplete working years, extra volatility of earnings at the start or end of marriages, etc. The presence of these large but transitory earnings deviations at the start and end of contiguous earnings spells leads to an upward bias in the estimate of the variance of permanent shocks when targeting the moments of income in growth rates, as is done in BPP's methodology. In this paper we further

role of private transfers was found to be minor in Attanasio et al. (2015).

show theoretically that this upward bias in the estimate of the permanent-shock variance also induces an upward bias in the estimate of consumption insurance against permanent shocks.

Empirically, we show that the pattern of the low mean and, even more importantly, of the high variance of observations next to the missing ones extends to incomplete family earnings and net family income spells in PSID data.[4] As net family income was used in BPP to measure the extent of self-insurance achieved through saving and borrowing, the estimated degree of insurance is biased upward. Quantitatively, this bias is quite large: while standard BPP estimates imply that as much as 37% of permanent shocks to net family income are insured, the corrected estimates imply that insurance is much closer to zero – the result that appears more consistent with the low wealth balances held by families in these data. While these corrected estimates are based on the modified income process that explicitly models the deviations of income following or preceding the missing observations, a qualitatively similar result is obtained by simply excluding the observations next to the missing ones from BPP's estimation sample: the estimated insurance falls from 37% to 12%.

Considering the insurance of the shocks to family earnings rather than net family income, the bias is even larger: the degree of insurance falls from 77% to 36% once the income process is modified to reflect the properties of observations surrounding the missing ones. This is relevant for measuring the role of the tax and transfer system using the difference in the estimated consumption pass-through coefficients for the shocks to family earnings and net family income. As the downward bias in the transmission of family earnings shocks to consumption, in proportion to its true value, is larger than the bias in the transmission of shocks to net family income, the proportionate difference in transmission is significantly overestimated. This leads to a large upward bias in the estimated contribution of the tax and transfer system to mitigating the risk of fluctuations in family earnings. Specifically, the

[4] A note on our terminology. We consider the PSID sample spanning 1979-1993 survey years. Incomplete spell occurs if a family first enters the sample after 1979, and/or exits the sample before 1993, and/or a very small fraction of cases of families having missing income observations after first entering the sample but before leaving it. When a family first enters the sample, by definition it was not part of the sample in the preceding year and we refer to its income in the preceding year as missing. Thus, the first income observation for a family is necessarily preceded by missing income observations. Applying similar logic, the last income observation observed for a family in the sample is followed by a missing income observation.

estimated share of shocks to family earnings insured by the tax and transfer system falls from 63% to 35% when this bias is corrected.[5]

We next assess the alternative approach of measuring the role of the tax and transfer system through the reduction in the estimated variance of net family income relative to the variance of family earnings. Ignoring the properties of observations surrounding the missing ones leads to a larger upward bias in the estimated variance of permanent shocks to family earnings than to family incomes. The resulting upward bias in the estimated importance of the tax and transfer system is much smaller, however, than the bias obtained by comparing the estimated transmission of these shocks to consumption. The reason for this result is revealed by our theoretical analysis: the upward bias in the estimated variance of permanent shocks when the properties of observations surrounding the missing ones are not taken into account is significantly amplified when these estimates are used to measure the extent of consumption insurance.

The rest of the paper is organized as follows. In Section 2 we review the BPP's methodology, the biases in estimated variances of permanent shocks induced by the lower mean and/or higher variance of observations surrounding the missing ones, and provide a characterization of how these biases distort the estimates of consumption insurance. In Section 3 we describe PSID data, the sample used in estimation, and the special properties of family earnings and net family income observations surrounding the missing ones. Section 4 contains the results of our empirical analysis. We show that the estimated variances of permanent shocks to family earnings and net family income are reduced when the corresponding stochastic process is modified to reflect the special nature of observations surrounding the missing ones or when these observations are simply not used in estimation. The refined estimates for the variance of permanent shocks provide a good fit to the variances of net family income and family earnings in both levels and differences, which is not possible using the canonical income process in BPP that does not account for the properties of observations around missing income records (Krueger et al., 2010; Daly et al., 2016). Estimation with corrected estimates of the

[5]Incidentally, the 35% estimate is close to the optimal insurance in the settings of Doligalski et al. (2020) and Wu and Krueger (2020).

variances of permanent shocks yields much lower estimates of consumption insurance against these shocks. This result suggests that the "excess insurance" puzzle implied by BPP estimates was induced by the bias in the measures of income dynamics. Correcting for this bias also results in a much lower estimate of the importance of the tax and transfer system in insuring permanent shocks to family earnings. Finally, we combine our theoretical analysis and empirical findings to explain how the bias in measured income dynamics induces the diverging estimates of the insurance role of the tax and transfer system. Section 5 concludes.

2 The Income Process and the Measurement of Consumption Insurance

In this section we briefly review the framework developed by BPP for measuring consumption insurance against shocks to family income.[6] The key elements of the framework, outlined below, are the income process and an equation for idiosyncratic consumption growth. Subsequently, we characterize the bias induced on estimates of insurance by the failure to account for the high variance and low mean of observations surrounding the missing ones when measuring income dynamics.

2.1 The Canonical Income Process and Consumption Insurance

BPP consider the following canonical process for characterizing the dynamics of family income:

$$\begin{aligned} y_{it} &= \alpha_i + p_{it} + \tau_{it} \\ p_{it} &= p_{it-1} + \xi_{it} \\ \tau_{it} &= \epsilon_{it} + \theta\epsilon_{it-1}, \end{aligned} \tag{1}$$

where log earnings y_{it} of family i at time t consists of the permanent component, p_{it}, and the transitory component, τ_{it}; α_i is the fixed effect; ξ_{it} is the permanent shock; ϵ_{it} is the transitory

[6] Our discussion of the income process and the insurance against shocks to income in this section equally applies to the earnings prosess and the insurance of shocks to earnings.

shock with persistence θ.

In the standard consumption-savings model with households facing permanent and transitory shocks to earnings and incomplete insurance markets, if the Euler equation holds at equality, consumption growth, Δc_{it}, can be approximated as

$$\Delta c_{it} = \phi_{it}\xi_{it} + \psi_{it}\epsilon_{it} + \zeta_{it}, \tag{2}$$

where $1-\phi_{it}$ is the amount of insurance of permanent shocks available to household i at time t, $1-\psi_{it}$ is the corresponding amount of insurance against transitory shocks, and the random term ζ_{it} represents innovations in consumption independent of the two income components. BPP show that the (average) insurance coefficients for permanent and transitory shocks, in the case of a serially uncorrelated transitory component, can be recovered using the following identifying moments:

Permanent insurance:

$$1-\phi_t = 1 - \frac{E[\Delta c_{it}\Delta y_{it-1}] + E[\Delta c_{it}\Delta y_{it}] + E[\Delta c_{it}\Delta y_{it+1}]}{E[\Delta y_{it}\Delta y_{it-1}] + E[\Delta y_{it}\Delta y_{it}] + E[\Delta y_{it}\Delta y_{it+1}]}, \tag{3}$$

Transitory insurance:

$$1-\psi_t = 1 - \frac{E\left[\Delta c_{it}\Delta y_{it+1}\right]}{E\left[\Delta y_{it}\Delta y_{it+1}\right]}, \tag{4}$$

where each expectation (averaging) is taken over all observations used for estimation of that particular covariance moment. Since available sample sizes are typically small leading to potentially imprecise estimates of these indentifying moments, the literature relies on a minimum-distance procedure for estimating the parameters of interest – including ϕ_t and ψ_t – that utilizes all of the available autocovariance moments in the data.

Note that the parameters ϕ_t and ψ_t reflect the total amount of insurance of permanent and transitory shocks without directly revealing the individual sources and mechanisms of insurance. For example, if y_{it} stands for net family income, measured insurance coefficients will reflect insurance due to accumulated assets.[7] If y_{it} is measured as gross family earnings,

[7]They may also reflect households' advance information about income innovations not available to the

as in Arellano et al. (2017), the insurance coefficients would reflect consumption insurance due to taxes and transfers in addition to the insurance due to assets. If y_{it} is measured as household head's earnings, the insurance coefficients would also reflect the insurance due to spousal labor supply; see Blundell et al. (2016).

2.2 Augmented Income Process and Biases in the Estimated Insurance Coefficients

To understand the connection between measured income dynamics and consumption insurance, note that the denominators of Eqs. (3) and (4) – which identify the extent of insurance – are precisely the moments that measure the variance of permanent and transitory shocks, respectively, using the income moments in growth rates. Thus, any bias arising in measuring the variances of shocks will also affect the measurement of the insurance against those shocks. Daly et al. (2016) identified an important potential source of such a bias. Specifically, they showed that estimation of the income process in Eq. (1) returns a substantially biased estimate of the variance of permanent shocks if one relies on a minimum-distance estimation based on matching the income moments in growth rates in incomplete panels where the mean and/or the variance of observations surrounding missing records in incomplete income spells are systematically different from typical income observations. We establish below that this is the case in PSID family earnings and income data, where the beginning or the end of incomplete spells often coincide with the start or end of marriages.

2.2.1 Augmented Income Process

Daly et al. (2016) proposed two ways of remedying the problem in estimating the variance of permanent income shocks arising from the use of incomplete household panels. First, one can simply drop observations surrounding missing records. Dropping observations at the start and end family spells in the data is consistent with the identification strategy of Blundell et al. (2008) who propose to focus on income risk by selecting a sample of continuously married

econometrician; see Kaufmann and Pistaferri (2009) for a discussion.

couples who do not experience marriage and divorce in the estimation sample, obviating the need for modeling "divorce, widowhood, or other household breaking-up factors" (see p. 1892).

Second, one can explicitly augment the income process in Eq. (1) with an extra transitory shock (with a lower mean and higher variance) next to missing records to account for the irregular nature of those observations and to recover the variance of permanent shocks without biases.[8] Specifically, let t_0 be the first and T the last sample year in the dataset (for example, while PSID data spans the period from 1968 to today, BPP only include observations between $t_0 = 1979$ and $T = 1993$ in their estimation sample). The extended income process that, relative to Eq. (1), introduces an extra transitory component χ_{it}, then becomes:

$$
\begin{aligned}
y_{it} &= \alpha_i + p_{it} + \tau_{it} + \chi_{it}, \quad t = t_0, \ldots, T \\
p_{it} &= p_{it-1} + \xi_{it} \\
\tau_{it} &= \epsilon_{it} + \theta \epsilon_{it-1} \\
\chi_{it+j} &= \begin{cases} \nu_{it} & \text{if } y_{it-1} \text{ or } y_{it+1} \text{ is missing and } t-1 \geq t_0,\, t+1 \leq T, j = 0 \\ \theta \nu_{it} & j = 1 \\ 0 & \text{otherwise,} \end{cases}
\end{aligned} \tag{5}
$$

where we assume the same persistence of ϵ and ν shocks.[9] Note that the records prior to t_0 and past T are missing from the estimation sample by construction, and so χ_{it} is assumed zero for $t = t_0$ and $t = T$ because t_0 and T are the researcher's choice that does not indicate genuine start or end of family income spells but instead simply truncates ongoing spells.[10] Shocks ν_{it} can be drawn from distributions with different means and variances depending on

[8]In Section 3.2 we present evidence in favor of this modeling choice. Specifically, we show that it is only the observations close to the start or end of incomplete income or earnings spells that have systematically different means of variances. Conceptually, a unit of analysis in BPP and in this paper is a continuous marriage over time. The approach in BPP that is based on a sample of continuously married couples yields estimates that are representative of the population at large if marriage and divorce are orthogonal to income shocks. Thus, relative to the income history of a given family, shocks that systematically induce unusual mean and variance of incomes only at the very start or the end of income history must be transitory.

[9]This choice is motivated by similar estimates of θ on the full sample and on the sample with observations surrounding the missing ones removed.

[10]There are some missing observations in the interior income or earnings spells but they are extremely rare, as we will document below.

whether they disturb incomes prior to or following a missing record.[11]

With the extended income process defined in Eq. (5), one can also modify Eq. (2):

$$\Delta c_{it} = \phi_t \xi_{it} + \psi_t \epsilon_{it} + \psi_t^\nu \nu_{it} + \zeta_{it}, \tag{6}$$

where $1 - \psi_t^\nu$ is the amount of insurance against the shock ν_{it}. To the extent that ν-shocks are larger in magnitude and are thus harder to insure against, one can expect that $\psi_t^\nu > \psi_t$.

2.2.2 From the Bias in Measured Income Dynamics to the Bias in Estimated Insurance Coefficients

We now describe the bias associated with ignoring the extra transitory component χ_{it} when measuring the insurance against permanent income shocks. The bias arises if one estimates the model in Eqs. (1)–(2) when the true income process and consumption equation are given instead by Eqs. (5)–(6). Since the denominator in Eq. (3), used for identification of the permanent insurance for the model in Eqs. (1)–(2), utilizes information on income data only, we can use the results in Daly et al. (2016) to characterize the bias in the estimated insurance coefficients.

Consider first an unbalanced sample with consecutive income observations such that part of the sample is comprised of individuals who start their incomplete income spells at $t > t_0$ while the rest of individuals have nonmissing income and consumption data throughout the whole sample period. The denominator of Eq. (3) is equal to the identifying moment for the variance of permanent shocks and will result in an upward-biased estimate of $\sigma^2_{\xi,t+1} + s_{t,t+1}(\mu^2_\nu + \sigma^2_\nu)$, where μ_ν and σ^2_ν are the mean and variance of the shock ν.[12] If consumption reacts to the

[11] As we will document in Section 3.2, it is not only the first or last observations of earnings spells are characterized by lower mean and higher variance, but up to three observations at the extremes of the earnings spells. Daly et al. (2016) show that to obtain unbiased estimates of the variance of permanent shocks it is sufficient to account only for the special properties of observations right next to the missing ones. The remaining extra variation will be subsumed in the estimated variance of transitory shocks, which we are not interested in in this paper. In contrast, to recover the true variance of permanent shocks when dropping observations preceding or following the missing ones, it is not enough to drop only the first and last earnings records of incomplete spells, instead, one must drop all observations at the ends of earnings spells that are systematically different.

[12] $s_{t,t'}$ is the share of individuals with incomplete income spells in the total sample of individuals observed at times t and t'.

current shocks only (which will be the case when an intertemporal shift of resources is allowed to the extent desired by a household), none of the moments in the numerator of Eq. (3) will be affected, so that the bias in the estimated permanent insurance at $t+1$ will equal

$$\left(1-\frac{\phi_{t+1}\sigma^2_{\xi,t+1}}{\sigma^2_{\xi,t+1}+s_{t,t+1}(\mu^2_\nu+\sigma^2_\nu)}\right)-(1-\phi_{t+1})=\lambda_{t+1}\phi_{t+1},\ \text{where } \lambda_{t+1}=\frac{s_{t,t+1}(\mu^2_\nu+\sigma^2_\nu)}{s_{t,t+1}(\mu^2_\nu+\sigma^2_\nu)+\sigma^2_{\xi,t+1}}.$$

Consider next an unbalanced sample with consecutive income observations such that part of the sample consists of individuals who end their incomplete income spells at $t<T$, while the other individuals have nonmissing income and consumption data throughout the whole sample period. In this case, the denominator of Eq. (3) will equal $\sigma^2_{\xi,t}+s_{t-1,t}(\mu^2_\nu+\sigma^2_\nu)$, an upward-biased estimate of the variance of permanent shocks. Since the shock ν is assumed to occur at t and consumption reacts to the current shocks only, the moment $E[\Delta c_{it}\Delta y_{it-1}]$ equals zero, while the moment $E[\Delta c_{it}\Delta y_{it+1}]$ will be identified by averaging over the sample of individuals who have complete income spells, and will equal $-\psi_t\sigma^2_{\epsilon_t}$. The moment $E[\Delta c_{it}\Delta y_{it}]$ will, however, be affected by incomplete income spells. Averaging over all individuals observed at times $t-1$ and t, the moment will be estimated as $\phi_t\sigma^2_{\xi,t}+\psi_t\sigma^2_{\epsilon,t}+s_{t-1,t}\psi^\nu_t(\mu^2_\nu+\sigma^2_\mu)$. Summing up, the bias in the estimated permanent insurance in this case will equal

$$\left(1-\frac{\phi_t\sigma^2_{\xi,t}+s_{t-1,t}\psi^\nu_t(\mu^2_\nu+\sigma^2_\mu)}{\sigma^2_{\xi,t}+s_{t-1,t}(\mu^2_\nu+\sigma^2_\nu)}\right)-(1-\phi_t)=(\phi_t-\psi^\nu_t)\lambda_t,\ \text{where } \lambda_t=\frac{s_{t-1,t}(\mu^2_\nu+\sigma^2_\nu)}{s_{t-1,t}(\mu^2_\nu+\sigma^2_\nu)+\sigma^2_{\xi,t}}.$$

The bias is unambiguously positive and potentially large if $\phi \gg \psi^\nu$ (which is likely to hold because permanent shocks are harder to self-insure against), and λ is large (i.e., the mean and/or the variance of the shock ν are large relative to the variance of permanent shocks).

Consider now a nonconsecutive unbalanced sample that consists of individuals with missing income records at time t, in the interior of the sample period, and individuals with nonmissing income and consumption records throughout the sample period. Individuals with missing income records at time t will bias upward the estimated permanent variance and permanent insurance at times $t+2$ and $t-1$. The biases for permanent insurance, respectively, are $\phi_{t+2}\lambda_{t+2}$ and $(\phi_{t-1}-\psi^\nu_{t-1})\lambda_{t-1}$ with properly defined λ's.

If the transitory component τ and the extra transitory component χ are both moving average processes of order one, the identifying moment (3) should be modified, adding second-order income autocovariances to the denominator, and adding the cross-covariances of consumption growth at time t and income growth at times $t+2$ and $t-2$ to the numerator. It is straightforward to show that the biases outlined above will change little if θ is close to zero (as is typically found in the data).[13]

3 Data

In this section we describe PSID data, the sample used in estimation, and the special properties of family earnings and net family income observations surrounding the missing ones.

3.1 Estimation Sample

As in BPP, our data contain married couples with heads of ages 30–65 observed within the 1979–1993 period in the PSID. As explained in Hryshko and Manovskii (2019), BPP included some families that marry or divorce between 1979 and 1993, but not all of them. We consider the sample that includes all such families, a selection also used in Blundell et al. (2016).[14] The income measures we consider are net family income and family earnings that combine earnings of the head and wife. Our sample comprises 2,430 families: among them, 534 (493) families have their first income (earnings) observation in 1979 and last observation in 1993, with nearly 80% of the sample having their first and/or last income observation after 1979 or before 1993.[15] Among 1332 families entering the sample after 1979, 64% were married after 1979 including 42% who were married in their first spell year. For comparison, only 4% of the families whose first spell year is 1979 were married in that year. It is more challenging to estimate what fraction of 959 families who left the sample prior to 1992 did so due to the end of marriage because some families simply stop responding to the PSID (attrit) and we do not

[13]The identifying moment (4) should also be modified in this case to $1 - \frac{E[\Delta c_{it}\Delta y_{it+2}]}{\frac{1}{\theta}E[\Delta y_{it}\Delta y_{it+2}]}$.

[14]The results on a smaller BPP sample are qualitatively similar.

[15]In addition, there are 123 missing earnings and 46 missing income records in the interior of earnings (income) spells.

know whether they divorced or not. Among families which exit the sample before 1993 but their members do not permanently attrit at the point of exit, 77% leave the sample because their marriage ended. We can obtain the lower bound of 53% of sample exits due to the end of marriage if we assume that none of the attriting families ended their marriage when leaving the sample. Using the same specification and the set of controls as in BPP, we extract earnings, income and (imputed) nondurable consumption residuals from regressions of earnings, income and consumption on a number of family characteristics such as the head's age, family size, number of kids, head's education, employment status dummies, etc. These residuals are used to construct idiosyncratic earnings, income and consumption growth utilized in estimating the models in Eqs. (1)–(2) and Eqs. (5)–(6).

3.2 Properties of Income and Earnings Observations Surrounding the Missing Ones

In Figure 1 and Table 1 we provide evidence that earnings and income observations in the beginning and end of incomplete spells are systematically different.

Consider first the properties of family earnings. The top two graphs in panel (a) of Figure 1 plot means and variances of residual family earnings for the spells that start in 1979 (solid lines) and spells that start later than 1979 (dashed lines). Earnings are somewhat lower on average and have higher variance in the first few years of the spells that start late. Those who enter the estimation sample later than 1979 are about ten years younger on average than those who enter in 1979 – since earnings dispersion increases with age, one would expect, absent any earnings effects in the beginning of marriages, that the variance of earnings for those who start later is *lower* than for those who start in 1979, in contrast to what we observe in the data. By year four of the spells, these effects dissipate, highlighting their transitory nature.

The bottom two graphs in panel (a) of Figure 1 plot means and variances of residual family earnings for the spells that end in 1993 (solid lines) and spells that end earlier than 1993 (dashed lines). Many of the spells that end earlier than 1993 represent the marriages that end in divorce or widowhood while spells that end in 1993 are ongoing but are mechanically truncated by the choice of the sample window. Earnings are somewhat lower and become more unstable a few years prior to the end of these spells. Since the average age of heads of families that exit the sample early and those that do not is nearly the same, the volatility pattern is consistent with the same earnings process for the two sets of families, with an extra transitory volatility induced by the marriage dissolution process in the last few years for those spells that exit the sample early.

We summarize this graphical evidence using regression anaysis reported in Table 1. The table contains the results of four regressions of family earnings' residuals and their variances on the dummies indicating the first, second or third observation preceding or following a missing one. Each regression is summarized in two columns. The odd-numbered columns contain coefficients on the dummies created for observations preceding or following a missing observation between 1979 and 1993. For comparison, in even-numbered columns, the coefficients are on the dummies created for the first observation of the sample window (1979) and two observations following it, and on the dummies created for the last observation in the sample window (1993) and two observations preceding it.[16]

The results in column (1) indicate that earnings residuals are about 0.15 log points lower in the first and last periods around the missing observation. In contrast, earnings residuals are not different from the unconditional mean of zero in the few first and last periods of the sample window – column (2). In the regression of columns (3) and (4), we first net out the mean effects of outlying observations on the residuals, and then regress squared (net) residuals on the same dummies as in columns (1) and (2), respectively. Squared residuals are lower in the few first sample years and higher in the last sample years due to the increase in inequality over the life cycle – column (4). The volatility of earnings, however, is much higher in the first and last sample years if individuals' first earnings records are not in the first sample year and last earnings records are not in the last sample year, as can be seen by comparing the first six regressors in columns (3) and (4).[17]

[16] By construction, all observations before 1979 and after 1993 are missing. However, they are merely missing due to the choice to consider the sample spanning the period 1979–1993; they are not missing from most actual individual income histories that overlap with the sample window.

[17] The coefficients on the dummies measure the variances in the respective periods relative to the average

We perform the corresponding analysis of family income spells in Panel (b) of Figure 1 and Columns (5)–(8) of Table 1. The results indicate that, on average, family income residuals in the few first and last periods (if they differ from the first and last sample years) are not different from zero. However, net family incomes are substantially more volatile in the first and last years of incomplete income spells.

4 Income Dynamics, Consumption Insurance, and the Role of Taxes and Transfers

4.1 The Transmission of Permanent Shocks to Family Earnings and Incomes to Consumption

In Table 2 we report the estimated transmission coefficients for permanent and transitory shocks to incomes, $\hat{\phi}$ and $\hat{\psi}$ respectively, and the averages of estimated variances of permanent and transitory shocks.[18]

Consider first the results based on estimating the baseline model in Eqs. (1)–(2). Using the full sample in column (1), the estimated transmission coefficient for permanent shocks to family earnings is 0.23, implying insurance of 77%. The value is similar to that reported in BPP for a somewhat different sample. This standard estimation approach does not attempt to account for the contribution of the systematically different observations surrounding the missing ones. Thus, the variance of permanent shocks and the insurance against permanent shocks is likely to be upward-biased. In columns (2)–(4) we report the results of three experiments designed to eliminate this bias.

In column (2), we restrict the sample to families that are continuously present in the sample from 1979 through 1993 and have no missing observations. The objective of this experiment

variance in the sample overall measured by a constant. For instance, the estimated constant in the regression of columns (3) and (4) is 0.25, so that the variance of residual earnings in the first year for the earnings spells that start in the first sample year equals 0.19 (=0.25 − 0.06), while the volatility of earnings in the first year for the earnings spells that start later than the first sample year equals 0.54 (=0.25+0.29).

[18] In principle, ϕ and ψ can vary by age and/or time but since our sample is small we estimate the transmission coefficients for the sample overall (in-sample averages) as in BPP.

is to get a sense of the bias caused by the presence of low mean or high variance observations at the start and end of contiguous family income spells by limiting the sample to families that do not have such observations. Relative to the full sample, the transmission coefficient for permanent shocks to family earnings rises dramatically to 0.68, implying insurance of only 32%, while the variance of permanent shocks drops to one third of the value estimated for the full sample. Of course, this is a selected sample so that a lower estimated insurance against permanent shocks might be due to these families being younger and/or less wealthy. The data do not support this interpretation as households in the balanced sample are somewhat older on average (average age of 46 vs. 43 in the full sample) and hold more wealth (average (median) net worth is about 10% (20%) higher in the balanced sample).[19]

To minimize the potential impact of sample selection, in column (3), we once again use the full original sample of column (1) but now drop 3 income observations around the missing records in incomplete income spells. Similarly to the results for the balanced sample and in line with the theoretical bias outlined above, both the estimated insurance of permanent shocks and the variance of permanent shocks are substantially lower than the corresponding estimates for the full sample. In this sample, the average age and average wealth – the key determinants of consumption insurance in lifecycle models with incomplete insurance markets – are nearly the same as in the full sample and, yet, insurance of permanent shocks is substantially lower.[20]

Finally, in column (4) we retain all families and all observations from the original sample but estimate the extended model in Eqs. (5)–(6) which accounts for the potentially lower mean and higher variance of observations adjacent to the missing ones.[21] The estimated

[19]Following BPP, net worth is calculated as the sum of dividend and interest income divided by the interest rate, and housing equity.

[20]To further verify that it is the specific nature of the first and last observations of incomplete spells that is the reason for the result, we conducted a Monte Carlo experiment, in which we dropped random observations on family earnings residuals such that the total number of retained observations is the same as in our estimation in column (3). The average transmission coefficient for permanent shocks from such an experiment across 1,000 replications is 0.24, virtually the same as the transmission coefficient for the full sample. The (average) estimated variances of permanent and transitory shocks to family earnings are also the same as for the full sample. A similar Monte Carlo experiment conducted for the estimation using net family income recovers, on average, the estimates reported in column (5).

[21]Specifically, in addition to all of the moments in the original BPP estimation we target the regression coefficients in four regressions similar to the ones reported in Table 1, with income residuals (consumption growth) and net squared income residuals (net squared consumption growth) on the left-hand side, and seven regressors on the right-hand side: three "One year after miss." dummies, one defined for the first record of

transmission coefficient for permanent shocks to family earnings is 0.64, close its value in columns (2) and (3) which restricted the sample to eliminate the influence of observations surrounding the missing ones, but almost three times higher than the estimate obtained on the full sample. This change in the estimated extent of insurance against permanent shocks is induced by a substantial reduction in the estimated variance of permanent shocks which leads to a significant improvement in the measurement of earnings dynamics. This can be seen in Figure 2 which plots the fit of the models estimated in column (1) (long dashed line), and column (4) (short dashed line) to the moments of log family earnings in levels and differences (solid line), in panels (a) and (b), respectively.[22] Since the growth moments are targeted in the estimation, both models fit them similarly well, as can be seen in panel (b). The variances of log family earnings residuals, which are not explicitly targeted in estimation, are fit substantially better by the extended model in column (4).[23]

In columns (5)–(8), we repeat the same experiments as in columns (1)–(4) but with net family income instead of earnings. The results follow the same pattern in that the estimated insurance of permanent shocks and the variance of permanent shocks fall substantially once the properties of observations surrounding the missing ones are modeled or the sample is restricted to remove such observations. Figure 3 illustrates the associated improvement in the fit to family income data moments. These results indicate that the "excess insurance"

income spells that start in 1979, another created for the first record of income spells starting later than in 1979, and the third one defined for income records after missing income observations in the interior of income spells; three "One year before miss." dummies, one defined for the last record of income spells that end in 1993, another created for the last record of income spells ending earlier than in 1993, and the third one created for income records before missing income observations in the interior of income spells; and a constant. We estimated the model by the method of simulated minimum distance, assuming that permanent, transitory, rare transitory shocks, are drawn from normal distributions, and using the diagonal weighting matrix calculated by block-bootstrap. We verified that the simulated minimum distance with the assumption of normal permanent and transitory shocks delivers virtually the same parameter estimates as the standard BPP estimation, which allows for any distributions of permanent and transitory shocks.

[22] We use the estimated variance for permanent shocks prior to 1979 from the model in Table 1 column (4) to calculate the variance of log family earnings in 1979 for the model in column (1) of Table 2.

[23] Interestingly, the superior fit to the moments in levels is not due to the extra moments in levels we utilized in the estimation of the extended income process in addition to the standard BPP moments. To confirm this, we performed an estimation where we targeted instead the relative variances of incomes in 1979 and 1993 to the overall cross-sectional variance as well as the relative variance of incomes for the first and last years of the spells that started after 1979 and/or ended before 1993 to the overall cross-sectional variance of incomes. This estimation does not target any levels moments. Yet, it delivered nearly the same transmission coefficients and a similarly good fit to both the moments in levels and growth rates as our baseline estimation of the extended income process.

puzzle implied by the BPP's estimates disappears once the bias induced by the properties of observations surrounding the missing ones is removed.[24]

4.2 Measuring the Role of the Tax and Transfer System

BPP have proposed to infer the insurance role of the tax and transfer system through a comparison of estimated insurance of permanent shocks to family incomes and estimated insurance of permanent shocks to family earnings. The logic of this experiment can be formalized as follows. Let net family income be represented, following Heathcote et al. (2014) and Blundell et al. (2016), as

$$\text{Net Family Income}_{it} = (1-\kappa)(\text{Family Earnings}_{it})^{1-\gamma}, \tag{7}$$

where γ measures progressivity of the tax and transfer system and also the share of the shocks to family earnings that does not pass through to net family incomes, i.e., insurance. The permanent shock to log family earnings ξ^e_{it} will be mediated by the tax and transfer system, mapping into the $\xi_{it} = (1-\gamma)\xi^e_{it}$ shock to log net family income. Recall that using shocks to net family income to estimate ϕ in Eq. (2) reveals the extent of insurance achieved through saving and borrowing. When the shocks to family earnings are used instead, the estimated coefficient for permanent shocks to family earnings, $\widehat{\phi(1-\gamma)}$, reveals the insurance achieved both through taxes and transfers and through saving and borrowing. Thus, the ratio of these estimates reveals the insurance provided by the tax and transfer system.

Using the full estimation sample, the estimate of ϕ in column (5) of Table 2 is 0.63 whereas the estimate of $\phi(1-\gamma)$ in column (1) is 0.23, which yields $1-\hat{\gamma} = 0.37$. This implies that 63% of permanent shocks to household earnings are insured by the tax and transfer system. Controlling for the bias induced on the measurement of permanent insurance by the high variance

[24] Of course, an assessment of whether the measured degree of insurance is "excessive" is relative to a model. We do not clutter this paper with an explicit model but base our conclusion on the quantitative analysis in Hryshko and Manovskii (2019) and Kaplan and Violante (2010). Those papers also highlight that misspecified persistence of the permanent shock (e.g., assuming that it follows a random walk while its true persistence is lower) imparts only a small bias on the estimated insurance coefficients but is crucial for interpreting their meaning. Because the potential bias induced by misspecified persistence is small, it has a minor impact on the estimates in this paper, and we ignore this issue to keep the analysis focused and transparent.

and low mean of observations surrounding the missing ones reduces the estimated insurance role of the tax and transfer system dramatically. Simply dropping these observations, as is done in columns (3) and (7), yields $\hat{\phi} = 0.88$ and $\widehat{\phi(1-\gamma)} = 0.58$, implying $1 - \hat{\gamma} = 0.66$, and an estimate of insurance due to the tax and transfer system of 34%, nearly 50% lower than the biased estimate calculated for the full sample. A comparison of the estimates in columns (4) and (8) which explicitly account for the properties of observations adjacent to the missing ones implies $1 - \hat{\gamma} = 0.65$ so that 35% of permanent family earnings shocks are insured by the tax and transfer system. Thus, the tax and transfer system plays an important role in insuring shocks to household earnings but not nearly as large a role as predicted by the biased insurance estimates that do not control for the properties of observations surrounding the missing ones.

BPP also suggest that the role of the tax and transfer system in insuring permanent shocks to family earnings can be inferred through comparison of the estimated variances of permanent shocks to family earnings and net family income.[25] Indeed, the minimum-distance estimation yields an alternative estimate of $1 - \hat{\gamma} = \sqrt{\frac{\sigma^2_\xi}{\sigma^2_{\xi,e}}}$, where σ^2_ξ and $\sigma^2_{\xi,e}$ are the variances of permanent shocks to net family income and family earnings, respectively.[26] Using this approach, and comparing the estimated variances of permanent shocks for the full sample

[25] Note that one may also use estimated transmission coefficients for transitory shocks and variances of transitory shocks to earnings and incomes for measuring the insurance role of taxes and transfers, γ. In Appendix II we show that since the variance of transitory shocks cannot be separately identified from the variance of measurement error in incomes, an estimate of γ inferred through comparison of the transitory shock variances for earnings and incomes might be biased. Likewise, the transmission coefficients for transitory shocks will be biased if income is measured with error and so will be an estimate of γ based on comparison of the transmission coefficients for transitory shocks to earnings and family incomes. For this reason, we focus on the measures of γ based on the estimates of permanent risk and its transmission to consumption which are robust to measurement error.

[26] Heathcote et al. (2014) and Blundell et al. (2016) estimate $1 - \gamma$ by regressing log net income on log earnings and obtain 0.815 and 0.92, respectively, whereas the corresponding estimate on our sample is 0.63. These estimates differ for a variety of reasons: e.g., Heathcote et al. (2014) include singles in their samples while Blundell et al. (2016) do not include unemployment insurance in their measure of disposable incomes. They also use PSID data for the 2000's whereas we use an earlier period that stretches to early 1990's. Using the TAXSIM that enables us to calculate comparable measures of household taxes for an earlier and later periods, we found the regression-based estimates of γ that are comparable to our estimate for an earlier period and to Heathcote et al. (2014)'s estimate for the later period, which indicates that the U.S. tax and transfer system became less progressive with time. One may also expect the regression-based estimate of γ to be biased due to endogeneity and measurement error in pretax earnings; in Appendix II we explores the role of measurement error for alternative approaches to measuring γ using the data generated from an incomplete-markets lifecycle model.

in columns (1) and (5) yields $1-\hat{\gamma}=0.58$, i.e., a substantially smaller role of the tax and transfer system relative to the one implied by relating the estimated transmission coefficients for permanent shocks to earnings and income. The fact that two alternative estimates of the role of taxes and transfers based on the same model yield strikingly different implications points to model misspecification. The root of the problem can once again be traced to the failure to account for the unique properties of observations next to the missing ones. If these properties are not accounted for, both methods for recovering the role of taxes and transfers will produce biased estimates of $1-\gamma$, but the size of the bias will differ between them. The biases will depend on the true value of γ, the variance of permanent shocks, and the mean and variance of income observations surrounding the missing ones.

Specifically, an estimate of $1-\gamma$ using estimated variances of permanent shocks will be

$$1-\hat{\gamma}=\left(\frac{(1-\gamma)^2\sigma^2_{\xi,e}+s_{ni}(\mu^2_{\nu,ni}+\sigma^2_{\nu,ni})}{\sigma^2_{\xi,e}+s_e(\mu^2_{\nu,e}+\sigma^2_{\nu,e})}\right)^{\frac{1}{2}}, \tag{8}$$

where s_{ni} (s_e) is the share of individuals with incomplete family income (family earnings) spells in a typical year, $\mu_{\nu,ni}$ ($\mu_{\nu,e}$) is the mean and $\sigma^2_{\nu,ni}$ ($\sigma^2_{\nu,e}$) is the variance of family income (family earnings) records surrounding the missing ones.

If one instead uses the relative estimates of transmission coefficients for permanent shocks to estimate the role of the tax and transfer system neglecting the properties of observations adjacent to the missing ones, an estimate of $1-\gamma$ would be

$$1-\hat{\gamma}=\frac{1}{1-\gamma}\frac{\sigma^2_{\xi,e}(1-\gamma)^2+s_{ni}(\mu^2_{\nu,ni}+\sigma^2_{\nu,ni})}{\sigma^2_{\xi,e}+s_e(\mu^2_{\nu,e}+\sigma^2_{\nu,e})}. \tag{9}$$

One estimate is not necessarily more biased than another but we can use, e.g., our (unbiased) estimate of $\gamma=0.30$ implied by the estimated variances of permanent earnings and income shocks in columns (4) and (8) and the estimated parameters of the transitory component χ in Eq. (5) to evaluate the relative bias in estimated γ's using two different methods. An estimate of $1-\gamma$ given by Eq. (9) using the relative transmission coefficients for the full sample equals 0.39, close to the data estimate of 0.37 obtained using the transmission coefficients for permanent shocks in columns (1) and (5). Eq. (8), utilizing relative permanent variances, produces an estimate of $1-\gamma$ of 0.53, which is also fairly close to the data estimate of 0.58. This calculation confirms that the large difference in the estimated insurance role of the tax and transfer system estimated using the two methods is largely induced by neglecting the properties of income and earnings observations surrounding the missing ones.

5 Conclusion

The informativeness of the quantitative analysis based on lifecycle models with idiosyncratic income risk and incomplete insurance markets has been recently put into doubt because of the apparent inability of this class of models to match the extent of consumption insurance of permanent income shocks in the data. Direct empirical measurement of the extent of insurance in the seminal contribution by BPP raised the "excess insurance" puzzle because families in the data were found to be much better insured than the standard incomplete markets model predicts. Moreover, an accurate policy analysis using such a model requires that it is consistent with the empirical evidence on the relative importance of various sources of insurance, such as the public tax and transfer system or household saving and borrowing. Yet, this empirical evidence is quite contradictory. The main measurement methodology proposed by BPP and based on comparing the extent of transmission of permanent shocks to family income and earnings to consumption implies a very important role of the tax and transfer system. An alternative measurement approach based on the same model and data implies a 50% smaller role.

In this paper we provided evidence that the standard empirical estimates of the extent and sources of insurance are significantly biased. The bias arises because family income (earnings) at the start and end of contiguous income (earnings) spells tend to have a lower mean and higher variance than interior observations. This feature of the data induces an upward bias in the estimate of the variance of permanent shocks when targeting the moments of income in growth rates, as is done in BPP's methodology. The upward bias in the estimate of the permanent-shock variance in turn induces an upward bias in the estimate of consumption insurance against permanent shocks.

The income process used as an input in incomplete markets models and when constructing the empirical benchmark estimates of consumption insurance abstracts from the irregular nature of income and earnings observations adjacent to the missing ones. There is therefore a disconnect between the commonly used models and the empirical benchmark estimates of consumption insurance, as the latter are and the former are not affected by this feature of the data. One could incorporate this feature of the data into the income process specified in the model. In this case, the BPP methodology applied to the model and to the data will yield biased estimates but the bias will be the same in the model and in the data. This approach, however, adds significant complexity to the model without clear substantive payoff. An alternative approach is to leave the income process in the model unchanged but to measure its components in the data without bias.

When we perform such measurement, the excess insurance puzzle disappears and the data imply the extent of insurance largely consistent with the incomplete markets model's prediction of only a limited role of assets in insuring consumption against permanent income shocks (Kaplan and Violante, 2010; Carroll, 2009; Hryshko and Manovskii, 2019). Moreover, we find a smaller role of the tax and transfer system in insuring the shocks to family earnings and align the estimates of this role obtained using two measurement approaches. These corrected estimates of the extent and sources of insurance thus serve as a better empirical benchmark for the standard quantitative incomplete markets model.

References

Arellano, M., R. Blundell, and S. Bonhomme (2017): "Earnings and Consumption Dynamics: A Nonlinear Panel Data Framework," *Econometrica*, 85, 693–734.

Attanasio, O., C. Meghir, and C. Mommaerts (2015): "Insurance in Extended Family Networks," NBER Working Paper 21059.

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

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

Boar, C. and V. Midrigan (2020): "Efficient Redistribution," NBER Working Paper 27622.

Bound, J., C. Brown, G. J. Duncan, and W. L. Rodgers (1994): "Evidence on the Validity of Cross-Sectional and Longitudinal Labor Market Data," *Journal of Labor Economics*, 12, 345–368.

Bound, J. and A. B. Krueger (1991): "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?" *Journal of Labor Economics*, 9, 1–24.

Carroll, C. D. (2009): "Precautionary Saving and the Marginal Propensity to Consume Out of Permanent Income," *Journal of Monetary Economics*, 56, 780–790.

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

De Nardi, M., G. Fella, and G. P. Pardo (2020): "Nonlinear Household Earnings Dynamics, Self-Insurance, and Welfare," *Journal of the European Economic Association*, 18, 890–926.

Doligalski, P., A. Ndiaye, and N. Werquin (2020): "Redistribution with Performance Pay," Mimeo.

Guvenen, F., F. Karahan, S. Ozkan, and J. Song (2016): "What Do Data on Millions of U.S. Workers Reveal about Life-Cycle Earnings Dynamics?" mimeo, University of Minnesota.

Heathcote, J., K. Storesletten, and G. L. Violante (2014): "Consumption and Labor Supply with Partial Insurance: An Analytical Framework," *American Economic Review*, 104, 2075–2126.

——— (2017): "Optimal Tax Progressivity: An Analytical Framework," *Quarterly Journal of Economics*, 132, 1693–1754.

Hryshko, D. and I. Manovskii (2019): "How Much Consumption Insurance in the U.S.?" mimeo, University of Alberta and University of Pennsylvania.

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

Kaufmann, K. and L. Pistaferri (2009): "Disentangling Insurance and Information in Intertemporal Consumption Choices," *American Economic Review: Papers and Proceedings*, 99, 387–392.

Krueger, D., F. Perri, L. Pistaferri, and G. L. Violante (2010): "Cross-sectional Facts for Macroeconomists," *Review of Economic Dynamics*, 13, 1–14.

Wu, C. and D. Krueger (2020): "Consumption Insurance Against Wage Risk: Family Labor Supply and Optimal Progressive Income Taxation," *American Economic Journal: Macroeconomics*, forthcoming.

Table 1: Net Family Income and Earnings Residuals.

Income measure:	Panel A: Family earnings				Panel B: Net family income			
Dependent variable:	Means		Variances		Means		Variances	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
One year after miss.	–0.15***	–0.03	0.29***	–0.06***	–0.02	–0.00	0.04**	–0.06***
	(–6.18)	(–1.61)	(4.40)	(–2.65)	(–1.12)	(–0.14)	(2.49)	(–6.15)
Two years after miss.	–0.09***	–0.02	0.11***	–0.09***	–0.00	–0.00	0.02	–0.05***
	(–4.59)	(–1.13)	(2.89)	(–5.19)	(–0.18)	(–0.38)	(1.21)	(–4.66)
Three years after miss.	–0.07***	–0.02	0.09**	–0.07***	0.01	–0.01	0.03	–0.04***
	(–3.12)	(–1.33)	(2.39)	(–2.76)	(0.77)	(–0.76)	(1.37)	(–3.18)
Three years before miss.	–0.03	–0.01	0.24***	0.03	0.00	–0.00	0.05***	0.03***
	(–1.12)	(–0.48)	(4.34)	(1.64)	(0.25)	(–0.50)	(2.91)	(2.80)
Two years before miss.	–0.03	–0.01	0.28***	0.07***	–0.02	0.00	0.08***	0.04***
	(–1.37)	(–0.40)	(6.59)	(3.35)	(–1.10)	(0.02)	(4.14)	(3.91)
One year before miss.	–0.14***	–0.03	0.64***	0.18***	–0.02	–0.00	0.08***	0.05***
	(–4.44)	(–1.56)	(7.22)	(4.79)	(–1.47)	(–0.31)	(4.54)	(3.70)
Constant	0.04***		0.25***		0.00		0.19***	
	(2.66)		(16.64)		(0.34)		(22.76)	
No. obs.	20,465		20,465		21,076		21,076	
No. indiv.	2,420		2,420		2,429		2,429	

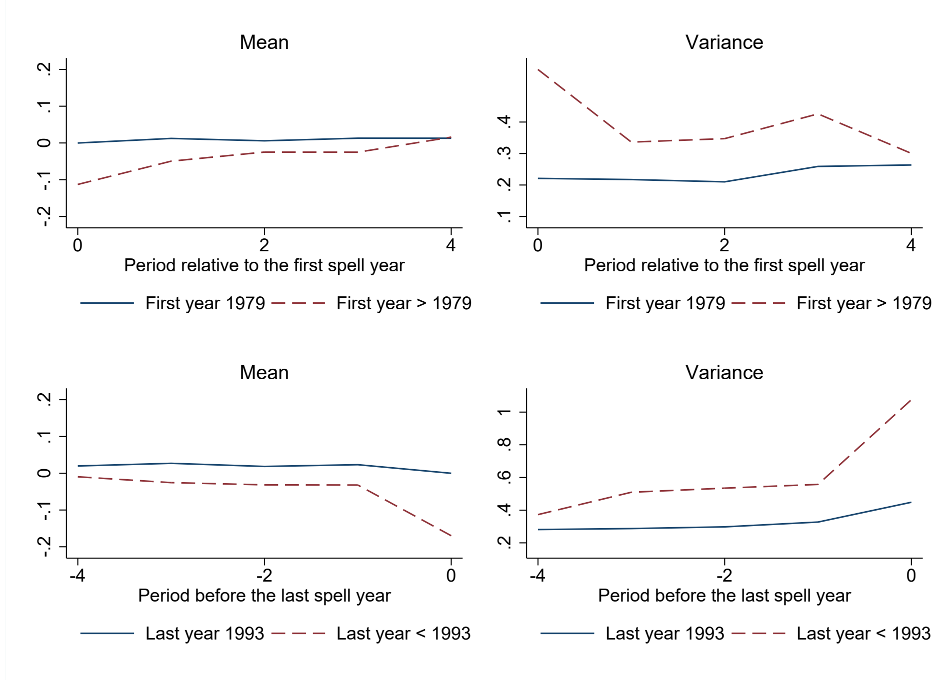
Notes: We use PSID data spanning the period 1979–1993. In odd-numbered columns, the dummies "One year after miss."–"Three years after miss." are equal to one for individual income observations after missing ones if the missings occur later than in 1979, and are zero otherwise; "Three years before miss."–"One year before miss." are equal to one for individual income observations before missing ones if the missings occur earlier than in 1993, and are zero otherwise. In even-numbered columns, the dummies "One year after miss."–"Three years after miss." are equal to one if an individual's first earnings record is in 1979, and are zero otherwise; "Three years before miss."–"One year before miss." are equal to one if an individual's last earnings record is in 1993, and are 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.

Table 2: Minimum-Distance Partial Insurance and Variance Estimates.

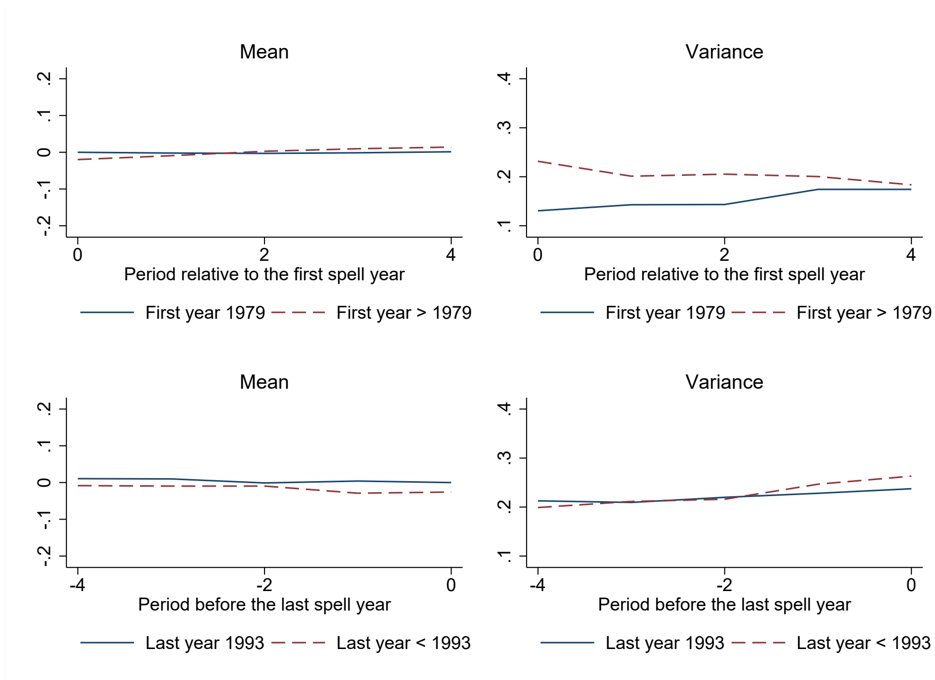
	Panel A: Family earnings				Panel B: Net family income			
	Full samp.	Balanc.	Drop first & last 3 obs.	Extended model	Full samp.	Balanc.	Drop first & last 3 obs.	Extended model
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ϕ, transm. of	0.23	0.68	0.58	0.64	0.63	0.83	0.88	0.99
perm. shock ξ	(0.04)	(0.17)	(0.12)	(0.17)	(0.08)	(0.18)	(0.14)	(0.13)
ψ, transm. of	0.09	0.09	0.08	0.08	0.06	0.08	0.06	0.07
trans. shock ϵ	(0.03)	(0.05)	(0.04)	(0.03)	(0.04)	(0.06)	(0.04)	(0.04)
ψ^{ν}, transm. of	—	—	—	0.10	—	—	—	0.07
trans. shock ν	—	—	—	(0.06)	—	—	—	(0.22)
σ_{ξ}^{2}, var. perm.	0.06	0.02	0.03	0.02	0.02	0.01	0.01	0.01
shock (avg.)	(0.006)	(0.004)	(0.005)	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)
σ_{ϵ}^{2}, var. trans.	0.07	0.06	0.06	0.07	0.03	0.04	0.04	0.04
shock (avg.)	(0.005)	(0.005)	(0.005)	(0.006)	(0.002)	(0.003)	(0.002)	(0.002)
Age (avg.)	43	46	43	43	43	47	43	43
Wealth (avg.)	104,837	116,498	104,524	104,837	105,313	114,227	105,664	105,313
Wealth (median)	50,200	60,222	50,679	50,200	50,288	61,626	51,231	50,288
No. househ.	2,430	478	2,430	2,430	2,430	516	2,429	2,430

Notes: Standard errors in parentheses. Full estimation contains results for ϕ and ψ; σ_{ξ}^{2} and σ_{ϵ}^{2}, by year; θ; variances of measurement error in consumption, and σ_{ζ}^{2}.

Figure 1: Family Earnings and Net Family Income Spells. PSID data.



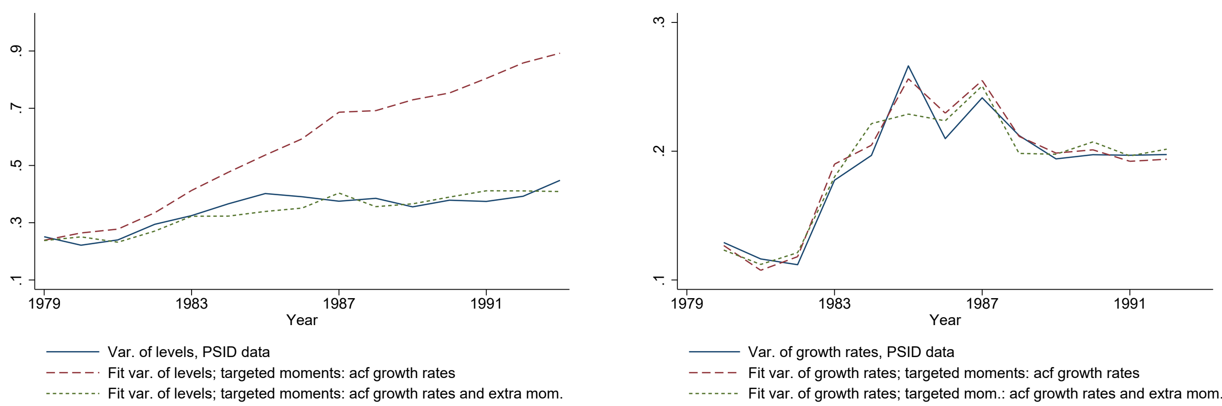
(a) Family earnings spells



(b) Net family income spells

Notes: The top two graphs in panel (a) plot means and variances of residual family earnings for the spells that start in 1979 (solid lines) and spells that start later than 1979 (dashed lines). The bottom two graphs in panel (a) plot means and variances of residual family earnings for the spells that end in 1993 (solid lines) and spells that end earlier than 1993 (dashed lines). Panel (b) plots the same information as panel (a) but for residual net family incomes.

Figure 2: Fit to the Moments of Log Family Earnings in Levels and Differences.

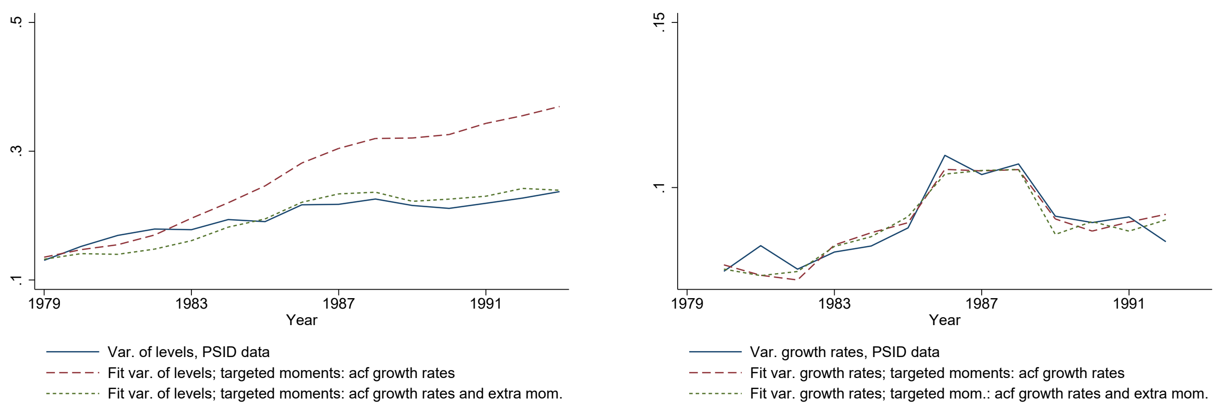


(a) Variances of log family earnings

(b) Variances of family earnings growth rates

Notes: In panel (a) solid line depicts the variance of log family earnings in PSID data for the period 1979–1993; long dash line depicts, for the same period, the variance of log family earnings implied by the estimates of the model targeting the standard BPP moments; short dash line depicts, for the same period, the variance of log family earnings implied by the model that targets the standard BPP moments and, in addition, the mean and the variance of observations surrounding missing earnings records. In panel (b) solid line depicts the variance of family earnings growth in PSID data for the period 1979–1993; long dash line depicts, for the same period, the variance of family earnings growth implied by the estimates of the model targeting the standard BPP moments; short dash line depicts, for the same period, the variance of family earnings growth implied by the model that targets the standard BPP moments and, in addition, the mean and the variance of observations surrounding missing earnings records.

Figure 3: Fit to the Moments of Log Net Family Income in Levels and Differences.



(a) Variances of log net family income

(b) Variances of net family income growth rates

Notes: In panel (a) solid line depicts the variance of log net family income in PSID data for the period 1979–1993; long dash line depicts, for the same period, the variance of log net family income implied by the estimates of the model targeting the standard BPP moments; short dash line depicts, for the same period, the variance of log net family income implied by the model that targets the standard BPP moments and, in addition, the mean and the variance of observations surrounding missing net family income records. In panel (b) solid line depicts the variance of net family income growth in PSID data for the period 1979–1993; long dash line depicts, for the same period, the variance of net family income growth implied by the estimates of the model targeting the standard BPP moments; short dash line depicts, for the same period, the variance of net family income growth implied by the model that targets the standard BPP moments and, in addition, the mean and the variance of observations surrounding missing net family income records.

Appendix

The objective of this appendix is to assess the performance of several methods for measuring the insurance role of tax and transfer system available in the literature in presence of measurement error in earnings, which is widely accepted to be a prominent feature of the data. First, we provide evidence on the properties of measurement error in survey data. We then introduce such measurement error in a fully specified lifecycle model and assess quantitatively the biases it induces on different approaches to measuring the insurance role of taxes and transfers.

I Measurement Error in HRS Earnings Data

To get an idea about the properties of measurement error in family earnings in survey data, we use nonimputed earnings data from the Health and Retirement Study (HRS) and administrative data from the Social Security Administration (SSA) matched to the HRS survey years 1992–2012.

We select males and their spouses present in the 2010 HRS wave, born before 1960, and use the social security weights to correct for selective matching.[27] Since sampling in the HRS is biennial, we retain administrative earnings (including deferred compensation) records for the corresponding years. We select earnings records for couples whose male is aged 50–65, and drop the data if there are earnings growth outliers in the survey earnings reports. Our data contain 2,207 couples with at least two observations on family earnings both in the HRS and administrative data.

Following Bound and Krueger (1991) and Bound et al. (1994), we define (log) measurement error as the difference between log family earnings reported in the HRS and log family earnings from SSA data. We drop observations above the ninety-ninth and below the first percentiles of the distribution of log measurement error in the data. The median log error is zero, the mean is 0.05, and the standard deviation is 0.47.

[27] Many of these households have both survey and administrative data available for 2012, which we also utilize.

Table A-1: Autocovariance structure of measurement error in family earnings in HRS

	Autocovariance of order:			
	0	1	2	3
Estimate	0.22	–0.08	0.001	–0.01
(s.e.)	(0.01)	(0.008)	(0.005)	(0.009)

Notes: Standard errors, clustered by household, in parentheses.

In Table A-1, we show the autocovariance function for the growth rate in measurement error. The autocovariance function is significant up to the first order – since our data are biennial, this is consistent with measurement error being no more persistent than a moving average process of order one.

There is also a possibility, however, that the stochastic component of measurement error contains a random walk component along with a short-lived transitory component. To test for the presence of permanent component in measurement error, we construct, for each year, a statistic $E[\Delta\epsilon_{it}^m(\Delta\epsilon_{it-2}^m + \Delta\epsilon_{it}^m + \Delta\epsilon_{it+2}^m)]$, where ϵ_{it}^m is measurement error in earnings for family i at time t. The statistic should be statistically different from zero under the null of a random-walk component in measurement error. Table A-2 contains the estimates for each year – they are quite noisy, with a p-value of the test that all of them cannot be distinguished from zero equal to 21%. Thus, using the combined evidence in Tables A-1–A-2, we may conclude that measurement error in family earnings does not contain a permanent component and is short-lived.

The implication of these findings is that the variance of permanent shocks and the consumption transmission coefficient for permanent shocks will be estimated without biases using the BPP methodology even in presence of realistic measurement error. In contrast, the variance of transitory shocks is not separately identified from the measurement error in income or earnings using this methodology, and the resulting transmission coefficients will be biased. This will be relevant when assessing various approaches to measuring the role of tax and

Table A-2: Testing for the presence of a permanent component in measurement error in family earnings. HRS data

Year	1995	1997	1999	2001	2003	2005	2007	2009
Stat.	0.009	–0.002	0.03	0.014	0.022	0.082	0.016	0.005
S.e.	(0.017)	(0.016)	(0.019)	(0.015)	(0.020)	(0.033)	(0.014)	(0.012)
p-value, all zero				21%				

Notes: Standard errors, clustered by household, in parentheses.

transfer system in the following section.

II The Effect of Measurement Error in Model-Generated Data

In this section, we assess several available approaches to measuring the insurance role of taxes and transfers in presence of realistic measurement error in earnings. While all these approaches require data on gross earnings and net (after taxes and transfers) family incomes, the approach that relies on comparing the transmission of permanent and transitory shocks to family gross earnings and net incomes to consumption also requires consumption data. We therefore need a model that features transitory and permanent risk to earnings and net family incomes, a mapping from earnings to net incomes that is similar to Eq. (7), and a model of consumption choices with incomplete insurance markets where assets, and taxes and transfers provide consumption insurance. We utilize simulated data from a calibrated model of Wu and Krueger (2020) that contains all of the required ingredients.

In the absence of measurement error in incomes, we can recover γ using the following three approaches.

1. Running a regression:

$$\log(\text{Net fam. inc.})_{it} = \text{const} + (1-\gamma)\cdot\log(\text{Fam. earnings})_{it} + \text{error}_{it}.$$

2. Using estimated transmission coefficients:

$$\hat{\gamma}_\phi = 1 - \frac{\hat{\phi}_{\text{[using earn.]}}}{\hat{\phi}_{\text{[using net fam. inc.]}}} \qquad \hat{\gamma}_\psi = 1 - \frac{\hat{\psi}_{\text{[using earn.]}}}{\hat{\psi}_{\text{[using net fam. inc.]}}}$$

3. Using estimated variances of shocks:

$$\hat{\gamma}_{\sigma_\xi} = 1 - \frac{\hat{\sigma}_{\xi,\text{[using net fam. inc.]}}}{\hat{\sigma}_{\xi,\text{[using earn.]}}} \qquad \hat{\gamma}_{\sigma_\epsilon} = 1 - \frac{\hat{\sigma}_{\epsilon,\text{[using net fam. inc.]}}}{\hat{\sigma}_{\epsilon,\text{[using earn.]}}}$$

The first approach should recover the same estimate if the regression is run in levels, first differences, or controls for individual fixed effects provided there is no unobserved heterogeneity across families (the condition that will be satisfied in our simulated data).[28] The second and third approaches rely on the estimates of permanent and transitory insurance, and permanent and transitory income and earnings risk, respectively, from the model in Eqs. (1)–(2).[29] This can be done by matching the model and data moments using the minimum distance method.

We simulate model data on earnings and net family incomes and consumption using the programs available as an online supplementary material to Wu and Krueger (2020). Their model is an incomplete-markets lifecycle model of two-earner couples facing permanent and transitory risk to each earner's wages, valuing household consumption and each spouse's leisure.[30] In Wu and Krueger (2020), the model true γ equals 0.1327, that is, slightly more than 13% of the shocks to family earnings – permanent and transitory – do not pass through to net family incomes. We simulate data for 50,000 families for the benchmark calibration

[28]Interestingly, in our PSID sample the estimated insurance role of taxes and transfers increases by over 30% when the regression is run in first differences or includes individual fixed effects relative to the regression in levels. We will observe a similar pattern in the model that includes measurement error in earnings.

[29]To focus on the role of measurement error, in this section we abstract form the issues related to low mean and high variance of earnings and incomes at the start and end of incomplete spells studied above.

[30]Wu and Krueger (2020) follow Blundell et al. (2016) in assuming that the exogenous source of risk is to hourly wages, but we verified that the combined spousal earnings can be modeled, similarly to wages, as the sum of a random walk component and a purely transitory shock. In particular, we cannot reject the null hypothesis that the persistence of an autoregressive component in spousal earnings equals one, and find that the autocovariances of spousal earnings growth above order one are small in magnitude and statistically not different from zero.

Table A-3: Implied insurance due to taxes & transfers, $\hat{\gamma}$

	No meas. error (1)	Meas. error (2)
From regression in levels	0.1327	0.1775
From regression in levels, FE	0.1327	0.2444
From regression in diffs.	0.1327	0.4304
Using trans. cons. insur., $\hat{\gamma}_{\psi}$	0.1327	0.4873
Using var. trans. shocks, $\hat{\gamma}_{\sigma_{\epsilon}}$	0.1327	0.3312
Using perm. cons. insur., $\hat{\gamma}_{\phi}$	0.1327	0.1319
Using var. perm. shocks, $\hat{\gamma}_{\sigma_{\xi}}$	0.1327	0.1323

Notes: Based on the data for 50,000 families from the benchmark calibration of Wu and Krueger (2020).

of Wu and Krueger (2020) and use ages 30–65 as in our empirical analysis to evaluate the performance of the three approaches to estimating γ when earnings do and do not contain measurement error.

In column (1) of Table A-3 we show the estimated insurance role of taxes and transfers, γ, using the three approaches from the model data with no measurement error in incomes. Not surprisingly, all approaches yield the estimates that recover the true value of γ.

We now assess the impact of measurement error in earnings on the estimated value of γ. To this end, we add measurement error to family earnings which is consistent with the measurement error properties we observe in the HRS data matched to the administrative tax records. We find that measurement error in family earnings is large, not very persistent, as shown above, and its variance increases with age. The second property is important because it would enable robust estimates of γ if we use the estimates of permanent income risk and consumption insurance against such risk, as will be demonstrated shortly. The third property is important if one is interested in the estimate of γ by age as, e.g., in Heathcote et al. (2017). If the variance of measurement error does not vary with age, it will be relatively unimportant at older ages when the variance of true incomes is relatively high due to accumulation of persistent shocks. If, however, the variance of measurement error gets larger with age, it will potentially produce a nontrivial bias in γ using the first approach outlined above even at older ages. We assume that measurement error is i.i.d., its variance at age 30 amounts to 10% of the variance of true family earnings,[31] and the variance increases by 2.7% each year, the growth we recover from the matched survey and administrative data. Column (2) of Table A-3 shows the estimated insurance role of taxes and transfers using the three approaches outlined above. The first three rows display the regression coefficients. The regression in levels produces the least whereas the regression in differences produces the most biased estimate of γ. The latter is not surprising as differencing amplifies the bias of the regression coefficients for variables measured with i.i.d. measurement errors. Interestingly, the bias of the fixed-effects regression estimate is also larger than the bias in the regression in levels despite our simulated data having no permanent heterogeneity. The fourth and fifth rows and column (2) show that the estimates of γ using estimated transmission coefficients for transitory shocks and the relative variances of transitory shocks to family earnings and disposable income are substantially biased upward. In contrast, estimating γ using the transmission coefficients for permanent shocks or the variances of permanent shocks to earnings and disposable income results in a near perfect recovery of the true insurance role of taxes and transfers, rows (6) and (7).

The upshot of this exercise is that estimation of the insurance role of taxes and transfers based on the transmission coefficients for permanent shocks or using the relative variances of permanent shocks is robust to measurement error in earnings (as long as the measurement error does not contain a random walk itself), unlike the other methods which might be biased.

[31] Our analysis of measurement error in the HRS data is for ages 50–65. Extrapolating to age 30 would result in an estimate of the variance of measurement error of about 50% of the true variance of family earnings in the model data. We use a more modest value for the variance of measurement error at age 30 to see if there are biases in γ under this milder scenario.