On the Heterogeneity in Family Earnings and Income Dynamics in the PSID

By Dmytro Hryshko and Iourii Manovskii*

The Panel Study of Income Dynamics (PSID) was launched in 1968 interviewing about 5,000 families. It has continuously followed those families, their children, grandchildren, etc. – collectively labeled as "sample" individuals – since then. In addition, the PSID collects information on individuals who get married to the sample individuals – referred to as "nonsample" individuals – but only while they remain married to the PSID sample members.

Becketti et al. (1988) compared the estimates of cross-sectional earnings equations between PSID sample and nonsample heads, and between sample and nonsample wives and found no major differences. No research known to us, however, compared the dynamics of family incomes for sample and nonsample PSID members. Filling this gap in the literature, Hryshko and Manovskii (2016) found that the PSID families headed by sample males are dramatically different from the families headed by nonsample males in their consumption and income dynamics. In particular, families headed by sample males have more persistent income shocks, and their consumption is more sensitive to longer-lasting shocks.

In this article we supplement the analysis on income dynamics of families headed by sample and nonsample males in Hryshko and Manovskii (2016) by using several additional methods for identification of persistence of income shocks. Specifically, we evaluate persistence of longer-lasting shocks, first, assuming it is homogeneous in each subsample of PSID families (as in, e.g., Blundell, Pistaferri and Preston 2008), second, assuming it is different across all families (as in Browning, Ejrnæs and Alvarez 2010), and, third, assuming that it is different across earnings histories (as in Arellano, Blundell and Bonhomme 2017). The results of the analyses based on all three methods confirm that families headed by sample males have a significantly more persistent income process.

I. Data and Samples

Ignoring Latino and Immigrant samples added to the PSID in the 1990's, the universe of PSID couples in any survey year consists of the 1968 couples who stay married until that survey year, original sample males and females who divorce after 1968 and remarry, and sample males and females who form their families after 1968 (roughly, sample sons and daughters). As original stable families, by definition, do not contain any nonsample individuals, we focus on the families formed after 1968 headed by sample or nonsample males. Since most of the information in the PSID is gathered for heads, and males are typically designated heads in PSID families, in the following, we label the former group as families of "sample sons" and the latter group as "nonsample" families.

As in Hryshko and Manovskii (2016), we add families of sample sons who divorced and remarried inside the 1978–1992 period to the data from Blundell, Pistaferri and Preston (2008) whose focus was on that period. As we need relatively long income histories for recovering heterogeneity in persistence, following Browning, Ejrnæs and Alvarez (2010), we select the data for families

* Hryshko: University of Alberta, Department of Economics, 8–14 HM Tory Building, T6G 2H4, Edmonton, AB, Canada, dhryshko@ualberta.ca. Manovskii: University of Pennsylvania, Department of Economics, 160 McNeil Building, 3718 Locust Walk, Philadelphia, PA, 19104-6297, USA, manovski@econ.upenn.edu. Acknowledgements: We are grateful to David Johnson for detailed comments. This research received financial support from the Killam Research Fund and EFF-SAS Fund to Hryshko, and from NSF Grant No. SES-1357903 to Manovskii.

with at least nine consecutive observations on incomes during the 1978–1992 period. For comparability with Browning, Ejrnæs and Alvarez (2010) and Arellano, Blundell and Bonhomme (2017), we further select families with heads of ages 25–60. In total, we have 514 families of sample sons and 468 nonsample families.

In Online Appendix we compare those families across many demographic characteristics such as age, education, and race, and across variables that are related to income dynamics such as, e.g., occupation switching and job displacement. As might be expected, these families are similar: among about 40 variables we considered the only significant differences, at the 10 percent level, are in wife's age, the incidence of business ownership, the incidence of members other than head and wife in having income, and the average number of observations in an income spell. Quantitatively, these differences are quite small.

II. Empirical Results

We consider two measures of family income in our analysis – net family income as in Blundell, Pistaferri and Preston (2008) and combined earnings of the head and wife as in Arellano, Blundell and Bonhomme (2017). Both income measures are important for a broad spectrum of family decisions. As is typical in the literature, we consider residuals from the first-stage regressions of those income measures on the full set of controls in Blundell, Pistaferri and Preston (2008), including the head's education, year-of-birth dummies, race, etc.

A. Homegeneous persistence

Our analysis starts with the canonical model of the income process: family i's idiosyncratic income at time t, y_{it}, is assumed to consist of a permanent component p_{it}, a transitory component ϵ_{it}, and a fixed effect, α_i. That is, $y_{it} = \alpha_i + p_{it} + \epsilon_{it}$. The permanent component is modeled as an autoregressive process of order one, $p_{it} = \rho p_{it-1} + \xi_{it}$, and the transitory component is assumed to be mean-zero shock, independent over time and across families.[1] In this section, our interest is in identification of ρ, the persistence of the permanent shock to income, ξ_{it}.

We assume that permanent and transitory shocks to income are independent,[2] α_i is i.i.d., and the permanent component at the start of an individual's working career, p_{i0}, equals $m + \xi_{i0}$ such that $y_{i0} = m + \alpha_i + \xi_{i0} + \epsilon_{i0}$.

For convenience, the income process for $t > 0$ can be written as

$$(1) \quad y_{it} = (1-\rho)\alpha_i + \rho y_{it-1} + \xi_{it} + \epsilon_{it} - \rho\epsilon_{it-1}.$$

Given our assumptions, the time-t quasi-difference $y_{it} - \rho y_{it-1}$ will be uncorrelated with income growth measured at times $t-j$, $j \geq 2$. In particular, we can use the following set of orthogonality conditions in levels to identify ρ: $E\left[(y_{it} - \rho y_{it-1})\Delta y_{it-j}\right] = 0$, $t = 1981, \ldots, 1992$, $j \geq 2$. We can also use the following set of moments in differences to identify ρ: $E\left[(\Delta y_{it} - \rho \Delta y_{it-1}) y_{it-j}\right] = 0$, $t = 1981, \ldots, 1992$, $j \geq 3$. The two sets of moments combined form the so-called system GMM estimator; see, e.g., Blundell and Bond (1998).

In Table 1, Panel A, columns (1) and (2), and Panel B, columns (5) and (6) we tabulate system GMM estimates of persistence of permanent shocks ξ_{it} for net family income and couples' earnings residuals, respectively. Clearly, incomes are more persistent for the families of sample sons than for nonsample families despite cross-sectional similarity of the two samples established above. In columns (3)–(4) and (7)–(8) we tabulate the results for "raw" net family incomes and earnings. There, we take out annual means from net family incomes and couples' earnings but do not control for any other first-stage regressors. We do so to safeguard our conclusions against any possible differences in the esti-

[1]Our GMM results are similar if we assume an MA(1) transitory component instead. See also Hryshko and Manovskii (2016) for the results with an MA(1) component for somewhat different samples.

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

Table 1—GMM estimates of persistence.

	Panel A: Net family income				Panel B: Couples' earnings			
	Resid.		Raw data		Resid.		Raw data	
	Samp.	Nons.	Samp.	Nons.	Samp.	Nons.	Samp.	Nons.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ρ, persistence perm. shock	0.90	0.78	0.93	0.81	0.89	0.73	0.89	0.78
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
	Panel C: Net family inc., res.				Panel D: Couples' earnings, res.			
	College		No college		College		No college	
	Samp.	Nons.	Samp.	Nons.	Samp.	Nons.	Samp.	Nons.
ρ, persistence perm. shock	0.85	0.79	0.86	0.69	0.84	0.74	0.78	0.57
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)

Note: System GMM estimates. Asymptotic standard errors in parentheses.

mated persistence induced by the first-stage regressions. Raw income measures are, to no surprise, more persistent than residual income measures but, importantly, incomes of nonsample families are far less persistent.

In Panel C, we tabulate the estimates of persistence for families whose heads have (at least some) college education, columns (1) and (2), and whose heads have no college education, columns (3) and (4). Families with college-educated heads have more persistent incomes than families with less educated heads, and, conditional on education, families headed by sample sons have more persistent incomes than nonsample families. The same holds for couples' earnings residuals as can be verified in Panel D.[3]

B. Heterogeneous persistence

Although GMM has an advantage of delivering a summary measure of persistence without estimating the income process, it is restrictive, e.g., in assuming homogeneity of the income process parameters. In a recent contribution, Browning, Ejrnæs and Alvarez (2010) use a selected set of moments to estimate an income process based on Eq. (1), which, they note, is equivalent to estimating the income process $y_{it} = (1-\rho)\alpha_i + \rho y_{it-1} + \varepsilon_{it} + \theta\varepsilon_{it-1}$. In particular, they assume that not only α_i and p_{i0} are individual-specific but also are the parameters ρ, θ, and the variance of ε. They also allow for an arbitrary correlation between the income process parameters, and add heterogeneous trends to the model.

With heterogeneity in the persistence, the GMM moments above will deliver inconsistent estimates of average ρ (e.g., Imbs et al. 2005). Following Browning, Ejrnæs and Alvarez (2010) we therefore estimate the following income process by indirect inference: $y_{it} = (1-\rho_i)\alpha_i + \rho_i\gamma_i + \rho_i y_{it-1} + (1-\rho_i)\gamma_i t + \varepsilon_{it} + \theta_i\varepsilon_{it-1}$, where t denotes age. We use the moments in Browning, Ejrnæs and Alvarez (2010) and add the moments informative about the differential persistence between sample and nonsample families. In Hryshko and Manovskii (2016), we found that sample and nonsample families significantly differ in higher-order autocovariances of income growth rates, and in the autocorrelation function of income levels. We, therefore, added the third, fourth and fifth-order autocovariances of growth rates, and the first ten autocorrelations of income

[3]Note that the persistence estimate, e.g., in Panel A, column (1) is lower than the estimate in either column (1) or (2) in Panel C, which is likely due to small-sample issues. Han and Phillips (2010) show that system-GMM estimates of the persistence may be downward-biased when the true persistence is high, which may explain the general tendency of our somewhat low estimates.

Table 2—Indirect inference estimates of heterogeneous persistence.

	Panel A: Net family income				Panel B: Couples' earnings			
	Uncorr.		Corr.		Uncorr.		Corr.	
	Samp.	Nons.	Samp.	Nons.	Samp.	Nons.	Samp.	Nons.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ρ, avg. pers.	0.82	0.73	0.87	0.70	0.81	0.66	0.74	0.67
perm. shock	(0.03)	(0.02)	(0.06)	(0.10)	(0.02)	(0.02)	(0.10)	(0.10)
ρ, first decile	0.79	0.70	0.79	0.57	0.79	0.63	0.60	0.54
ρ, fifth decile	0.82	0.73	0.88	0.71	0.81	0.66	0.76	0.68
ρ, ninth decile	0.86	0.75	0.93	0.82	0.83	0.68	0.87	0.80

Note: Standard deviations of the estimated persistence in parentheses.

levels to the set of moments. For each income measure, we estimated two variants of the model—when α_i, ρ_i, γ_i, θ_i and $\sigma^2_{\varepsilon,i}$ are assumed to be uncorrelated (estimating 12 parameters with 37 moments), and when the correlation between the parameters is arbitrary (estimating 26 parameters with 58 moments). Full details of estimation are provided in online appendix.

Panel A in Table 2 contains the results for net family income residuals. Both models, with independent and codependent heterogeneity, are in agreement that average persistence of longer-lasting shocks is higher for the families of sample sons. Interestingly, for the model with uncorrelated heterogeneity, columns (1) and (2), even the 90th percentile of persistence for nonsample families is below the 10th percentile of the persistence in the families of sample sons. In columns (3) and (4), we fit more moments but the average persistence is similar to columns (1) and (2), although the model with correlated parameters predicts more heterogeneity. The results for couples' earnings residuals presented in Panel B are qualitatively similar.

C. *Nonlinear persistence*

Arellano, Blundell and Bonhomme (2017) develop an estimation technique for the income process of the permanent-transitory type where persistence depends on the past value of income and the size of the current shock.

If such an income process is true, the in-sample distribution of past income components and shocks for sample and nonsample families may be such that the (average) persistence we recover from applying GMM is different for the two types of families (despite the income process being the same). We can explore if this is the case by looking at the estimated persistence for sample and nonsample families for given quantiles of past incomes and the current income shock. Specifically, we look at the average estimates of $\rho_t(y_{i,t-1},\tau) = \frac{\partial}{\partial y_{it-1}} Q_t(y_{it-1},\tau)$, where $y_{it} = Q_t(y_{it-1},\tau)$ is the τ-th conditional quantile of y_{it} given y_{it-1}, $\tau \in (0,1)$, and t denotes age.[4]

We follow Arellano, Blundell and Bonhomme (2017) using balanced data from six consecutive PSID waves (we use six consecutive years from 1984 to 1989) but, because our samples are smaller cross-sectionally, we consider only 6 quantiles for the shocks (i.e., 6 values of τ), and 6 quantiles of y_{it-1}, at which we evaluate the persistence of the income history embedded in y_{it-1}, $\rho(y_{it-1},\tau)$, instead of 11 quantiles for both in Arellano, Blundell and Bonhomme (2017).[5] In total, we have 36 estimates of persistence for each quantile of y_{it-1} and

[4] Full estimation of the income process as in Arellano, Blundell and Bonhomme (2017) is beyond the scope of this article; our discussion to follow corresponds to Figure 1(a) in Arellano, Blundell and Bonhomme (2017).

[5] Our conclusions are similar if we make biennial sample selection and model 11 quantiles as in Arellano, Blundell and Bonhomme (2017).

each quantile of the shock to y_{it-1}. For net family income residuals, only 2 out of 36 estimates of persistence are higher for nonsample families; e.g., the persistence equals 0.97 (0.97) for the lowest (highest) income quantile and the lowest (highest) income shock for the families of sample sons vs. 0.89 (0.91) for the same quantiles for nonsample families. The results for earnings residuals are qualitatively similar: 31 out of 36 estimates of persistence are higher for the families of sample sons, and the respective numbers for our example above are 0.93 vs. 0.89 for the lowest income quantile and lowest shock, and 0.94 vs. 0.91 for the highest income quantile and highest income shock for sample and nonsample families, respectively.

III. Conclusion

The dynamic properties of earnings and income play a central role in modern macro and labor economics. As emphasized in Daly, Hryshko and Manovskii (2016), they are key for understanding the nature of inequality, for understanding the variation in consumption, and for the optimal design of tax and transfer policies. Virtually all of the extensive research on these issues is based on the data from the PSID. Typically, these studies pool together the households headed by the PSID sample and nonsample individuals. Indeed, there is no immediate reason for these two sets of households to differ from each other except for random sampling variation. This makes the large and systematic differences we find between them quite surprising. It is not clear whether these differences reflect true differences in the underlying income processes, but, regardless of whether they do, pooling the two samples in estimation may yield quite misleading estimates.

REFERENCES

Arellano, Manuel, Richard Blundell, and Stéphane Bonhomme. 2017. "Earnings and Consumption Dynamics: A Nonlinear Panel Data Framework." *Econometrica*, 85(3): 693–734.

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

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

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

Browning, Martin, Mette Ejrnæs, and Javier Alvarez. 2010. "Modelling Income Processes with Lots of Heterogeneity." *Review of Economic Studies*, 77: 1353–1381.

Daly, Moira, Dmytro Hryshko, and Iourii Manovskii. 2016. "Improving the Measurement of Earnings Dynamics." NBER Working Paper # 22938.

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

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

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

Hryshko, Dmytro, and Iourii Manovskii. 2016. "How Much Consumption Insurance in the U.S.?" University of Pennsylvania mimeo.

Imbs, Jean, Haroon Mumtaz, Morten O. Ravn, and Hélène Rey. 2005. "PPP Strikes Back: Aggregation and the Real Exchange Rate." *Quarterly Journal of Economics*, 80: 1–43.