A Risky Venture: Income Dynamics Among Pass-Through Business Owners

Jason DeBacker, Vasia Panousi, Shanthi Ramnath

Online Appendix

A-1 Defining business income: A comparison of sources

Panel Survey of Income Dynamics (PSID)

In section G of the PSID Family file, respondents are asked to report wage and salary income for both the head and the wife: "How much did you (head/wife) earn altogether from wages or salaries in year t, that is, before anything was deducted for taxes or other things?" Wages and salaries are top coded at $9,999,997$. By contrast, there is no income top-coding in our tax data, regardless of income category.

The construction of the business income series is very complicated. First, net business income is imputed. Next, it is split between labor and capital income, and allocated to heads/wives/other family members. First, net business income is split equally among head/wife and other family member owners. Next, income is divided between labor and asset income for heads and wives only according to the following rules: 1) If head or wife responds that they work in the business, then their share of income is split evenly between labor and asset income. 2) If head or wife does not work in the business, then their share of income is assigned to asset income. 3) If the business reports a net loss, then 100 percent of the loss is assigned to asset income. 4) For other family members, net business income is split evenly between labor and asset income, unless the business realized a loss, in which case the loss is allocated to asset income.

In general, differences between our administrative data and the PSID could be due to any of a large number of factors. For example, potentially significant disadvantages of the PSID data include measurement error, attrition, consistency issues over time in survey questions and in the definition of income variables, the top coding of income variables, and a shift in the early 1990s from personal interviews to telephone-based interviews. On the other hand, the PSID data might do better at capturing the lower end of the income distribution. Nonetheless a lot of households file taxes even if they do not have to, so as to receive various tax credits.

For additional details, see *psidonline.isr.umich.edu/Publications/Papers/*.

Social Security Administration (SSA)

Confidential data from the SSA have been used in papers co-authored with Jay Song. While not subject to top-coding, SSA data consist only of information from individual W-2 forms, as their purpose is the collection of Social Security taxes. This means annual wage income for each person, no income sources other than wages, and no ability to connect individuals to households. By contrast, our data connect individuals to their partners and children, and include all income sources, not just wages, and at the household level for all. Hence, whereas SSA data can only be used to examine questions related to wage-income dynamics of individuals, usually prime-age men, our data can be used for analysis of comprehensive income measures and income dynamics at the household level, thus providing extremely valuable information for analyses of household self-insurance, consumption, and welfare. For additional details, see *www.irs.gov/pub/irs-pdf/i1040sse.pdf*

Survey of Consumer Finances (SCF)

The SCF is conducted by economists and analysts at the Federal Reserve Board. The SCF asks questions about privately-held (closely-held) businesses. According to the survey, "The forms of business in this category are sole proprietorships, limited partnerships, other types of partnership, subchapter S corporations and other types of corporation that are not publicly traded, limited liability companies, and other types of private business". With the exception of closely-held subchapter C corporations, these are the same businesses which generate the household business income in our data. These businesses, or a subset of them, have been termed "entrepreneurs" by authors like Cagetti and De Nardi, who use these data extensively.

The survey asks about the family's income, before taxes, for the full calendar year preceding the survey. The components of income in the SCF are wages; self-employment and business income; taxable and tax-exempt interest; dividends; realized capital gains; food stamps and other, related support programs provided by government; pensions and withdrawals from retirement accounts; Social Security; alimony and other support payments; and miscellaneous sources of income for all members of the primary economic unit in the household. These are exactly the same income categories as in our tax data, from IRS Form 1040.

In fact, the SCF high-income over-sample is derived from IRS data exactly like the ones we use in our analysis. First, the SCF analysts ask the IRS for a sample of social-security numbers (SSN) drawn from personal-income tax returns of filers with income above a certain threshold. Second, using detailed income information from Forms 1040 provided by the IRS for these households, the SCF analysts perform an asset imputation. This determines whether these high-income households are also high-asset households. Third, the SCF analysts call the identified high-asset households and ask permission to interview them. About 70% of the households accept to be interviewed. They are then asked questions about their income and assets, based on which the SCF analysts have reached two conclusions. First, that their asset imputation is relatively accurate, in that it matches self-reported wealth by households. Second, that households report totals about a broader version of "financial" income, which includes business income, capital gains, dividends, and rental income, that are consistent across the survey and the tax returns. Nonetheless, households sometimes give different responses to the survey about the components of the total financial income, although they are advised to look at their tax returns when they respond to survey questions. However, this discrepancy has been declining over time. This part of the SCF design is highly classified and confidential. Zip codes are not released to the general public. Only American citizens are allowed to participate in the survey design and administration.

In sum, the SCF identifies income from privately-held businesses exactly as we do. It also uses Form 1040 data exactly like ours to identify and interview the households in their over-sample. Clearly, the SCF is mostly a cross-sectional study, whereas our data have a large longitudinal dimension.

The SCF also conducted a small 2007-2018 panel, to assess households' experience with the Great Recession. Barnett and Panousi (2018) provide an extensive analysis using the SCF panel.

National Income and Product Accounts (NIPA)

The aggregate NIPA data derive from IRS tax-return data, and hence are of similar origin as our data, but they are divided into different groupings. In particular, the NIPAs separate aggregate income from US businesses into two categories. First, income from C-corporations and S-corporations together. Second, income from sole proprietors and partners together. We do not have information from C-corporations in our data, but our household-business-income is generated by the same sole proprietorships, partnerships, and S-corporations as in the tax data.

To adjust for potential tax evasion of (smaller) businesses, the NIPAs add an adjustment amount to the total from above. This adjustment has fluctuated a lot over time, although it is less volatile of recent decades In the past, it was supposed to be informed by IRS controls about tax evasion. However, these have been discontinued since the 1990s, so now the adjustment is likely, to a large degree, arbitrary. The time-series of the unadjusted and of the adjusted income business income are almost parallel during our sample period. We then conclude that the IRS adjustment, though potentially relevant for income levels and for analysis of business returns, may not be as relevant for an analysis of variances. In any case, the adjustment is largely arbitrary and, furthermore, as noted in Section I, there is no consensus about the distribution of tax evasion at the micro level.

National Longitudinal Survey of Youth 1979 (NLSY79)

This survey includes cohorts born between 1957 and 1964. At the time of their first interview, respondents' ages ranged from 14 to 22. The respondents were 47 to 56 at the time of their 2012 interviews. Respondents have been asked questions about net business income since 1983. About 12,686 individuals were initially interviewed in 1979, in equal parts male and female.

Longitudinal Business Data Base (LBD)

The LDB is at the establishment level, whereas our data are at the household level. The LBD contains annual observations on employment and payroll for all establishments and firms in the private (non-government) sector. The sources are periodic business surveys conducted by the Census Bureau and federal government administrative records, i.e. confidential tax returns for business entities.

The Longitudinal Business Database (LBD) is constructed from the Census Bureau's Register of US businesses with paid employees and enhanced with survey data collections. It covers all sectors of the economy and all geographic areas. It employs establishment and firm records and has information on employment, payroll, 4-digit SIC, EIN, business name and location. The Compustat database has been used to supplement LBD with info on whether firms are publicly traded. As of 2000, the LBD has almost 5 million firms with positive employment in the non-farm non-government sector, of which about 7,000 are publicly traded. The average LBD firm size is about 18 employees, compared to the average of 4,000 employees of publicly traded firms. Publicly traded firms account for a trivial fraction of all firms and less than 30% of non-farm business employment. The sample period starts in 1976. The unit of observation in the LBD is the establishment, defined as a single physical location where business is conducted. Each establishment-year record has a firm identifier associated with it, so it is possible to track the ownership structure of firms in any given year, as well as changes over time. Firm size is constructed by aggregating employment across all establishments belonging to the firm.

A-2 Probit model estimates, full set of parameter estimates

Table A.1: Exit from business: Probit results

	(1)	(2)
log(positive business income$_{t-1}$)	-0.045	-0.041
	(0.000)	(0.000)
log(negative business income$_{t-1}$)	0.022	0.041
	(0.001)	(0.000)
log(labor income$_{t-1}$)		0.012
		(0.000)
Number of consecutive years with biz income	-0.002	-0.003
	(0.000)	(0.000)
Number of years w/ biz income	-0.012	-0.013
	(0.000)	(0.000)
Age 35-39		0.005
		(0.003)
Age 40-44		0.006
		(0.003)
Age 45-49		0.012
		(0.003)
Age 50-54		0.020
		(0.003)
Age 55-60		0.024
		(0.003)
Has children		0.011
		(0.002)
Single male		0.056
		(0.003)
Single female		0.047
		(0.003)
Married female		-0.027
		(0.006)
t_{1989}	-0.003	-0.003
	(0.007)	(0.007)
t_{1990}	0.009	0.009
	(0.007)	(0.007)

	(1)	(2)
t_{1991}	0.024	0.023
	(0.007)	(0.007)
t_{1992}	0.018	0.017
	(0.007)	(0.007)
t_{1993}	0.010	0.008
	(0.007)	(0.007)
t_{1994}	0.019	0.018
	(0.007)	(0.007)
t_{1995}	0.020	0.019
	(0.007)	(0.007)
t_{1996}	0.020	0.018
	(0.007)	(0.007)
t_{1997}	0.023	0.020
	(0.007)	(0.007)
t_{1998}	0.024	0.021
	(0.007)	(0.007)
t_{1999}	0.034	0.030
	(0.007)	(0.007)
t_{2000}	0.040	0.036
	(0.007)	(0.007)
t_{2001}	0.037	0.033
	(0.007)	(0.007)
t_{2002}	0.027	0.023
	(0.007)	(0.007)
t_{2003}	0.012	0.009
	(0.007)	(0.007)
t_{2004}	0.019	0.015
	(0.007)	(0.007)
t_{2005}	0.025	0.021
	(0.007)	(0.007)
t_{2006}	0.024	0.019
	(0.007)	(0.007)
t_{2007}	0.026	0.022
	(0.007)	(0.007)
t_{2008}	0.028	0.023
	(0.007)	(0.007)

	(1)	(2)
t_{2009}	0.043	0.038
	(0.007)	(0.007)
t_{2010}	0.027	0.024
	(0.007)	(0.007)
t_{2011}	0.040	0.036
	(0.007)	(0.007)
t_{2012}	0.026	0.023
	(0.007)	(0.007)
t_{2013}	0.027	0.022
	(0.007)	(0.007)
t_{2014}	0.027	0.022
	(0.007)	(0.007)
t_{2015}	0.020	0.014
	(0.007)	(0.007)
t_{2016}	0.016	0.011
	(0.007)	(0.007)
t_{2017}	0.005	-0.001
	(0.007)	(0.007)
t_{2018}	0.008	0.007
	(0.007)	(0.007)
Observations	155,151	155,151

[a] Standard errors in parentheses below parameter estimates.

Table A.1 displays the average marginal effects from probit regressions where and outcome of 1 denotes continued business activity and 0 denotes exit from business activity. The panel used estimate the probit models is our benchmark 1987-2018 panel, which drop filers who never report business income outside the $\pm\$5,000$ threshold at any point in the sample. We restrict ages to 30-60, and exclude farmers.

A-3 Detailed Descriptive Statistics of the Distributions of Percentage Changes

Table A.2: Descriptive Statistics for Percentage Changes

	Business Income	Labor Income
Mean	-7,905,016	10
Std Dev	2,452	256
Skewness	-277	15
Kurtosis	76,463	1,304
Percentiles		
p10	-100	-25
p25	-58	-6
p50	0	2
p75	70	12
p90	188	38
p95	445	77
p99	2,452	256
Observations	76,466	336,319

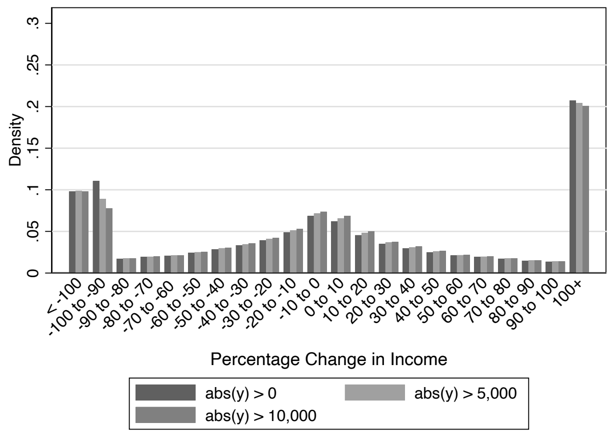
Table A.2 presents the descriptive statistics for the distributions of the percentage changes in income using our benchmark 1987-2018 business income and labor income panels. In the business income panel, we drop households who never report business income outside the $\pm\$5,000$ interval, restrict ages to 30-60, and exclude farmers. In the labor income panel we drop observations with labor income less than $2,575, restrict ages to 30-60, and exclude farmer. Percentage changes are computed as in Equation 2 and "smeared" by averaging 11 observations to protect the confidentiality of the data.

A-4 Percent changes in income: Robustness to alternative samples and time intervals

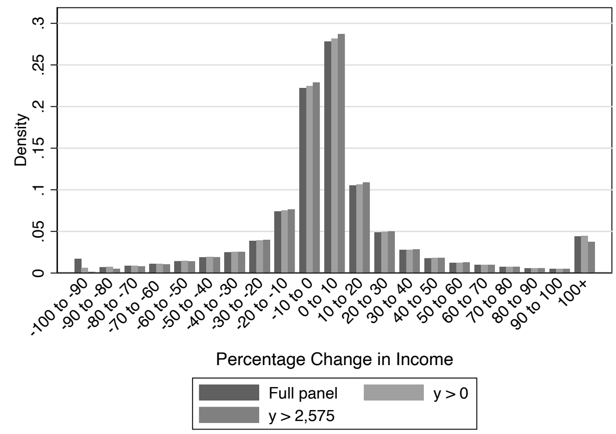
In order to examine the robustness of the percent changes in income results, we perform the following tests.

First, we conduct the analysis using alternative samples. In particular, we compute percentage changes in business income with our benchmark sample, which drops taxpayers who never have business income outside of the (-$5,000, $5,0000) interval, restricts to taxpayers where the primary filers is aged to 30-60, and excludes farmers and then for two other samples where we change the interval of business income to +/-$10,000 and +/-$0. We compute the percentage change in labor income for three samples as well. In addition to our baseline labor income sample that drops observations with labor income less than $2,575, we also consider samples where we only drop observations with labor income that equals zero, and where we keep all observations, regardless of whether labor income is positive. Figure A.1 relates these results and shows that the qualitative results from our benchmark samples obtain.

Figure A.1: **Percentage Changes in Business and Labor Income, various sample selection criteria**



(a) Business Income

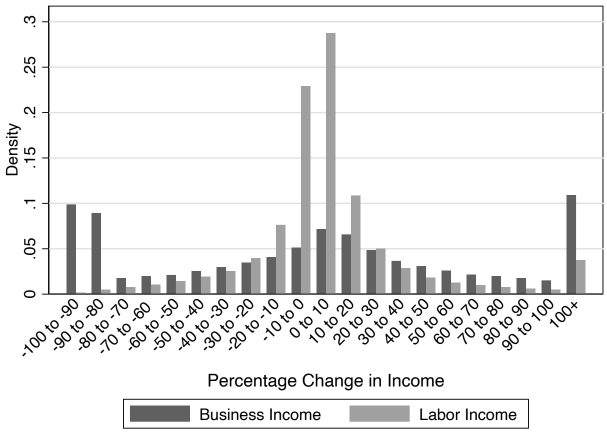


(b) Labor Income

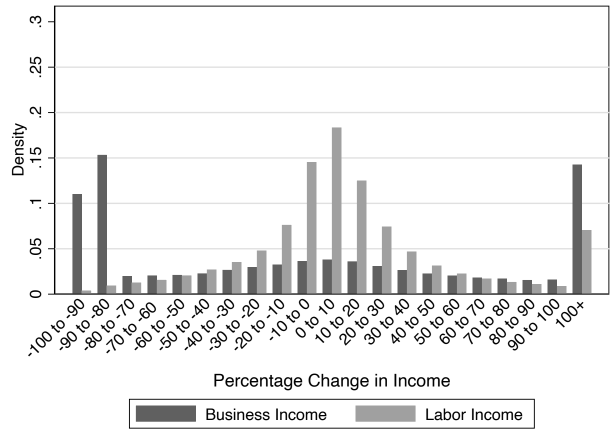
Figure *A.1* presents the distribution of percentages changes in income over one year, for business and labor income over various samples. The panels used to construct percent changes in business income income are our benchmark sample, which drops taxpayers who never have business income outside of the (-$5,000, $5,0000) interval, restricts to taxpayers where the primary filers is aged to 30-60, and excludes farmers and then two other samples where we change the interval of business income to +/-$10,000 and +/-$0. The labor income panels include our benchmark panel, which drops taxpayers who have labor income less than $2,575, restricts to taxpayers where the primary filers is aged to 30-60, and excludes farmers and two other samples, where we include all observations with primary files aged 30-60 without farm income, and where we include all observations with non-zero labor income, who also are non-farmers and aged 30-60. The horizontal axis shows the size of the percent change. All bins have a size of 10 percentage points, except the last bin on the right and the last bin on the left. The last bin on the right groups together all observations for which income increased by more than 100%. The last bin on the left groups together all observations for which income decreased by more than 100%. The vertical axis shows the fraction of all business or labor income observations of percent changes in each size-of-percent-change bin.

Second, we consider different windows over which we compute the percentage changes in income. A.2 related the percentage changes in business income over three, five, and 10 year windows. As the length of the windows extend, larger percentage changes become more likely. But business income always shows much more frequent large percentage changes, both positive and negative.

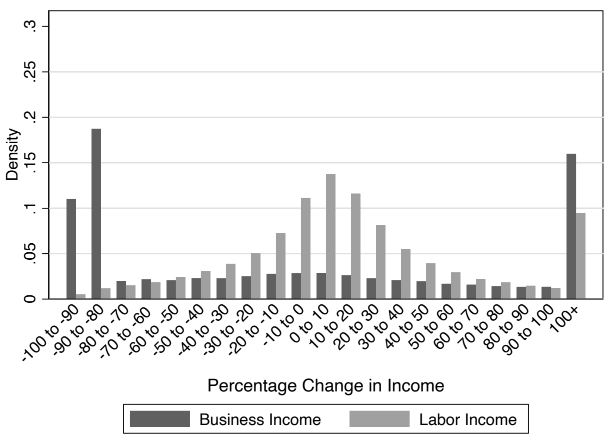
Figure A.2: **Percentage Changes in Business and Labor Income, variation in window over which percentage changes are computed**



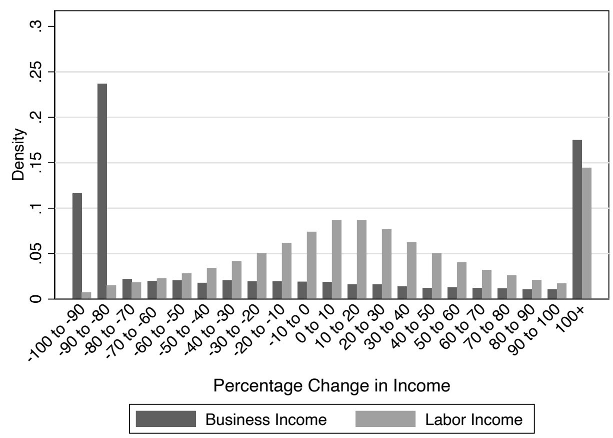
(a) One Year Percentage Changes



(b) Three Year Percentage Changes



(c) Five Year Percentage Changes



(d) Ten Year Percentage Changes

Figure A.2 presents the distribution of percentages changes in income over one, three, five and ten years, for business and labor income. The panels used to construct percent changes in income for each household are our benchmark 1987-2018 panels. The business income panel drops taxpayers who never have business income outside of the (-$5,000, $5,0000) interval, restricts to taxpayers where the primary filers is aged to 30-60, and excludes farmers. The labor income panel drops taxpayers who have labor income less than $2,575, restricts to taxpayers where the primary filers is aged to 30-60, and excludes farmers. The horizontal axis shows the size of the percent change. All bins have a size of 10 percentage points, except the last bin on the right and the last bin on the left. The last bin on the right groups together all observations for which income increased by more than 100%. The last bin on the left groups together all observations for which income decreased by more than 100%. The vertical axis shows the fraction of all business or labor income observations of percent changes in each size-of-percent-change bin.

A-5 Additional Mobility Estimates: Alternative transition windows

In order to examine the robustness of the income mobility results, we create transition matrices for 3, 5, and 10 year windows. These are presented in Tables A.3 to A.8.

Table A.3: Three-year Transition Matrix Across Deciles of Business Income Distribution

From/To	0	1	2	3	4	5	6	7	8	9	10
0	0.742	0.032	0.036	0.034	0.034	0.031	0.027	0.022	0.018	0.014	0.010
1	0.228	0.296	0.126	0.066	0.045	0.036	0.031	0.033	0.039	0.044	0.056
2	0.278	0.123	0.203	0.135	0.071	0.047	0.035	0.032	0.032	0.029	0.016
3	0.324	0.057	0.109	0.191	0.104	0.065	0.039	0.036	0.027	0.029	0.019
4	0.314	0.040	0.069	0.114	0.161	0.103	0.065	0.050	0.038	0.027	0.019
5	0.272	0.037	0.055	0.072	0.111	0.157	0.108	0.074	0.054	0.038	0.020
6	0.236	0.030	0.040	0.053	0.079	0.122	0.170	0.130	0.081	0.042	0.019
7	0.174	0.032	0.032	0.042	0.050	0.073	0.133	0.204	0.155	0.078	0.027
8	0.124	0.035	0.027	0.028	0.037	0.050	0.072	0.139	0.258	0.179	0.050
9	0.084	0.043	0.021	0.021	0.026	0.033	0.034	0.059	0.152	0.364	0.164
10	0.044	0.054	0.013	0.012	0.012	0.014	0.014	0.021	0.037	0.129	0.650

This table shows the three-year transition matrix for deciles of business income. The zero state denotes no business income. The numbers in the table denote probabilities, and are calculated as the number of household-year observations for which there is a transition from decile x to decile y over the period, divided by the number of household-year observations of any transition over that same period. The calculations include households that are in the panel at both ends of the transition.

Table A.4: Five-year Transition Matrix Across Deciles of Business Income Distribution

From/To	0	1	2	3	4	5	6	7	8	9	10
0	0.678	0.040	0.044	0.041	0.039	0.035	0.032	0.028	0.024	0.021	0.017
1	0.275	0.233	0.111	0.069	0.049	0.040	0.033	0.032	0.038	0.053	0.067
2	0.343	0.105	0.159	0.119	0.065	0.044	0.040	0.031	0.034	0.036	0.023
3	0.358	0.056	0.095	0.157	0.090	0.060	0.046	0.041	0.033	0.036	0.027
4	0.359	0.043	0.066	0.104	0.128	0.087	0.054	0.048	0.049	0.036	0.026
5	0.326	0.040	0.049	0.070	0.105	0.120	0.083	0.071	0.058	0.046	0.030
6	0.284	0.033	0.041	0.061	0.073	0.108	0.123	0.114	0.083	0.052	0.028
7	0.226	0.036	0.040	0.044	0.052	0.075	0.115	0.154	0.146	0.077	0.035
8	0.172	0.034	0.033	0.035	0.042	0.060	0.075	0.123	0.198	0.167	0.061
9	0.120	0.045	0.025	0.025	0.032	0.039	0.041	0.064	0.136	0.302	0.172
10	0.061	0.066	0.017	0.017	0.015	0.014	0.015	0.024	0.047	0.129	0.596

This table shows the five-year transition matrix for deciles of business income. The zero state denotes no business income. The numbers in the table denote probabilities, and are calculated as the number of household-year observations for which there is a transition from decile x to decile y over the period, divided by the number of household-year observations of any transition over that same period. The calculations include households that are in the panel at both ends of the transition.

Table A.5: Ten-year Transition Matrix Across Deciles of Business Income Distribution

From/To	0	1	2	3	4	5	6	7	8	9	10
0	0.595	0.050	0.053	0.050	0.046	0.040	0.036	0.033	0.033	0.033	0.031
1	0.336	0.155	0.087	0.063	0.042	0.042	0.029	0.037	0.050	0.067	0.091
2	0.387	0.092	0.112	0.099	0.059	0.044	0.033	0.037	0.044	0.043	0.047
3	0.393	0.054	0.085	0.130	0.087	0.055	0.041	0.034	0.039	0.038	0.043
4	0.394	0.054	0.058	0.089	0.086	0.075	0.057	0.051	0.043	0.045	0.048
5	0.395	0.044	0.055	0.066	0.079	0.080	0.063	0.063	0.056	0.055	0.045
6	0.338	0.035	0.050	0.056	0.075	0.095	0.091	0.079	0.084	0.055	0.041
7	0.301	0.041	0.044	0.048	0.055	0.073	0.086	0.100	0.112	0.083	0.056
8	0.250	0.046	0.035	0.051	0.045	0.054	0.057	0.093	0.150	0.142	0.078
9	0.189	0.048	0.035	0.036	0.036	0.035	0.041	0.058	0.108	0.234	0.180
10	0.087	0.073	0.021	0.024	0.019	0.018	0.025	0.029	0.054	0.135	0.516

This table shows the ten-year transition matrix for deciles of business income. The zero state denotes no business income. The numbers in the table denote probabilities, and are calculated as the number of household-year observations for which there is a transition from decile x to decile y over the period, divided by the number of household-year observations of any transition over that same period. The calculations include households that are in the panel at both ends of the transition.

Table A.6: Three-year Transition Matrix Across Deciles of Labor Income Distribution

From/To	1	2	3	4	5	6	7	8	9	10
1	0.421	0.262	0.132	0.073	0.044	0.030	0.018	0.010	0.005	0.004
2	0.183	0.371	0.228	0.099	0.051	0.030	0.018	0.010	0.006	0.004
3	0.093	0.149	0.351	0.226	0.087	0.046	0.025	0.013	0.006	0.003
4	0.057	0.068	0.129	0.345	0.230	0.088	0.042	0.024	0.012	0.005
5	0.040	0.040	0.057	0.127	0.348	0.228	0.089	0.041	0.022	0.008
6	0.025	0.024	0.035	0.053	0.132	0.358	0.236	0.086	0.037	0.014
7	0.019	0.017	0.021	0.031	0.054	0.137	0.378	0.245	0.077	0.022
8	0.013	0.011	0.011	0.019	0.033	0.057	0.149	0.420	0.240	0.048
9	0.009	0.007	0.008	0.011	0.016	0.031	0.053	0.151	0.514	0.200
10	0.007	0.006	0.005	0.006	0.009	0.012	0.020	0.035	0.133	0.768

This table shows the one-year transition matrix for deciles of labor income. The zero state denotes no labor income. The numbers in the table denote probabilities, and are calculated as the number of household-year observations for which there is a transition from decile x to decile y over the period, divided by the number of household-year observations of any transition over that same period. The calculations include households that are in the panel at both ends of the transition.

Table A.7: Five-year Transition Matrix Across Deciles of Labor Income Distribution

From/To	1	2	3	4	5	6	7	8	9	10
1	0.342	0.258	0.149	0.093	0.058	0.040	0.025	0.016	0.010	0.007
2	0.167	0.315	0.234	0.117	0.067	0.042	0.027	0.016	0.009	0.006
3	0.096	0.144	0.286	0.226	0.110	0.062	0.038	0.021	0.011	0.006
4	0.060	0.070	0.124	0.283	0.233	0.111	0.058	0.035	0.019	0.008
5	0.042	0.045	0.064	0.119	0.278	0.229	0.114	0.063	0.032	0.014
6	0.030	0.029	0.040	0.060	0.125	0.284	0.235	0.116	0.058	0.023
7	0.021	0.021	0.024	0.036	0.059	0.134	0.309	0.254	0.108	0.034
8	0.016	0.015	0.015	0.023	0.040	0.068	0.147	0.341	0.263	0.071
9	0.012	0.010	0.011	0.014	0.023	0.041	0.067	0.153	0.433	0.235
10	0.009	0.007	0.007	0.008	0.013	0.017	0.026	0.046	0.143	0.724

This table shows five-year transition matrix for deciles of labor income. The zero state denotes no labor income. The numbers in the table denote probabilities, and are calculated as the number of household-year observations for which there is a transition from decile x to decile y over the period, divided by the number of household-year observations of any transition over that same period. The calculations include households that are in the panel at both ends of the transition.

Table A.8: Ten-year Transition Matrix Across Deciles of Labor Income Distribution

From/To	1	2	3	4	5	6	7	8	9	10
1	0.241	0.245	0.163	0.113	0.079	0.055	0.041	0.032	0.018	0.014
2	0.137	0.246	0.224	0.139	0.089	0.062	0.044	0.030	0.018	0.011
3	0.078	0.118	0.217	0.220	0.146	0.088	0.061	0.038	0.022	0.011
4	0.058	0.071	0.110	0.205	0.210	0.141	0.098	0.056	0.033	0.017
5	0.043	0.045	0.064	0.104	0.198	0.212	0.147	0.099	0.062	0.026
6	0.034	0.030	0.041	0.064	0.108	0.200	0.222	0.160	0.094	0.045
7	0.024	0.023	0.032	0.043	0.066	0.120	0.225	0.241	0.160	0.065
8	0.020	0.019	0.022	0.030	0.050	0.079	0.137	0.249	0.272	0.122
9	0.017	0.014	0.014	0.020	0.034	0.053	0.084	0.151	0.326	0.288
10	0.013	0.012	0.011	0.013	0.017	0.026	0.034	0.060	0.158	0.655

This table shows ten-year transition matrix for deciles of labor income. The zero state denotes no labor income. The numbers in the table denote probabilities, and are calculated as the number of household-year observations for which there is a transition from decile x to decile y over the period, divided by the number of household-year observations of any transition over that same period. The calculations include households that are in the panel at both ends of the transition.

A-6 Accounting for negative income

One transformation that addresses negative values is the inverse hyperbolic sine (IHS). Letting y denote income, the IHS transformation of the y-observations is given by:

$$y_{ihs}(\theta) = \frac{1}{\theta} sinh^{-1}(\theta y) = \frac{1}{\theta} log(\theta y + \sqrt{1 + \theta^2 y^2}), \ , \tag{A.6.1}$$

where θ is a location parameter. As shown in Figure $A.3$, for a given θ, the IHS is very similar to the log when y is positive: It limits to the same slope as the log as y increases and is simply vertically shifted by a constant. For y negative, the IHS function is the mirror image of its shape in the positive quadrant. For y positive, the difference between the IHS and the log essentially lies in the way each function treats small observations. In particular, the log goes to minus infinity as y goes to zero, whereas the IHS is approximately linear in a symmetric interval around the origin. If y is large relative to $\frac{1}{\theta}$, the IHS function approximates the $log(y)$ function for positive values and $-log(abs(y))$ for negative values. We note that an attractive property of the IHS for our purposes is that it treats large absolute values of income symmetrically. This is important for our analysis, because, as we have seen, households experiencing either positive or negative percent changes are rich households, and therefore they should be treated in a symmetric fashion in model estimation. We use a location parameter of $\theta = 1$, which means that differences in income with the IHS transformation are almost exactly the same as differences in income with log transformation for values of income in our sample (e.g., $ln(4000) - ln(2000) = 0.693147$ and $ihs(4000) - ihs(2000) = 0.693147$). Thus, the results below can be interpreted the same way (and generally directly compared to) results using a log transformation, with the caveat that our results incorporate the effects of negative income realizations.

Figure A.3: Inverse Hyperbolic Sine (IHS) vs. Log

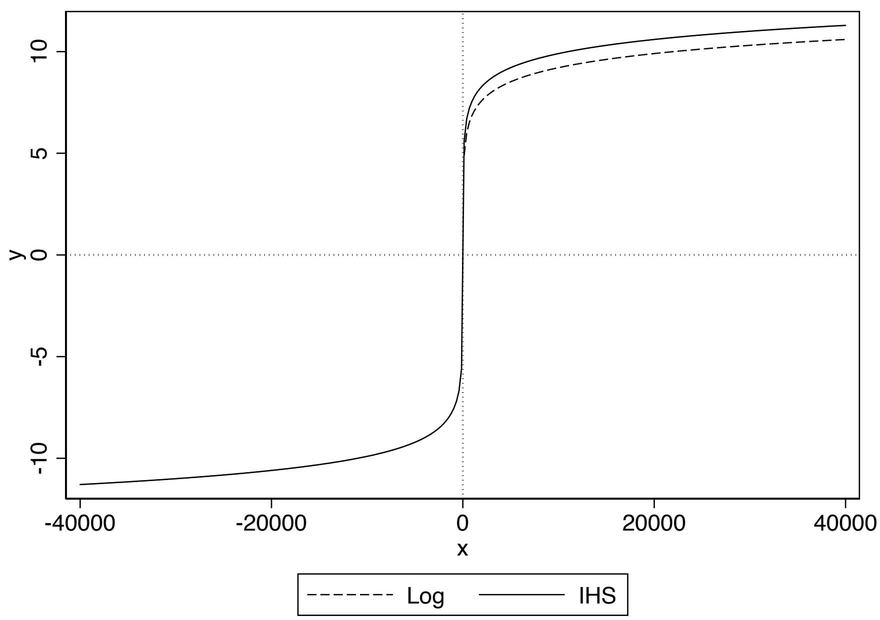
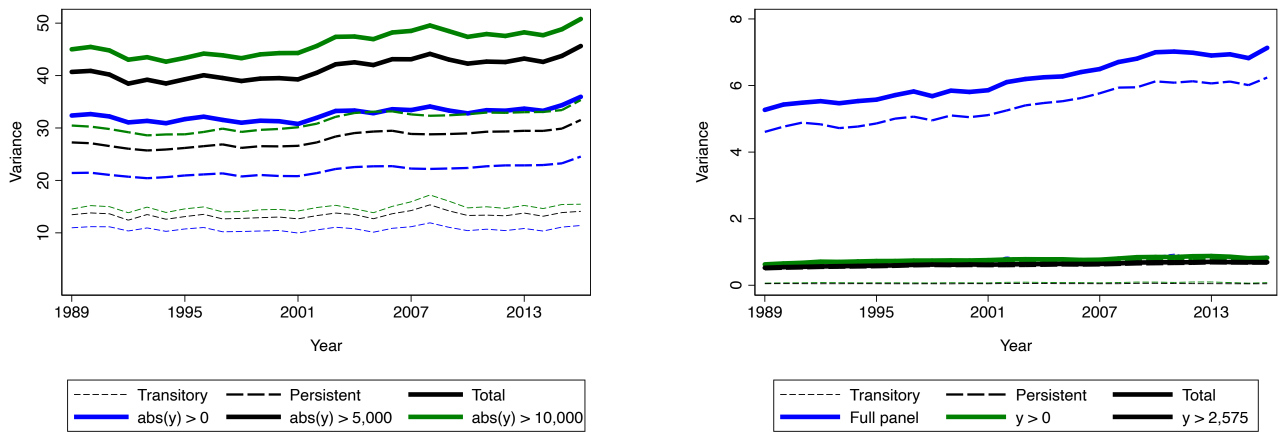


Figure A.3 plots the Inverse Hyperbolic Sine (IHS), for θ=1, versus the log function. Income, y, is on the horizontal axis. The log of income (dashed line) or the IHS of income (solid line) are on the vertical axis. The IHS of y is given in (A.6.1).

A-7 Non-parametric Variance Decomposition: Robustness to alternative samples

Figure A.4: **Persistent vs. Transitory Risk, KSS (2010), various sample selection criteria**



(a) Business Income (b) Labor Income

Figure A.4 decomposes the total cross sectional variance into its persistent and transitory components using the decomposition of Kopczuk et al. (2010). The decomposition using a window of 5 years (P=5). See text for more details. The panels used to construct the variances in business income income are our benchmark sample, which drops taxpayers who never have business income outside of the (-$5,000, $5,0000) interval, restricts to taxpayers where the primary filers is aged to 30-60, and excludes farmers and then two other samples where we change the interval of business income to +/-$10,000 and +/-$0. The labor income panels include our benchmark panel, which drops taxpayers who have labor income less than $2,575, restricts to taxpayers where the primary filers is aged to 30-60, and excludes farmers and two other samples, where we include all observations with primary files aged 30-60 without farm income, and where we include all observations with non-zero labor income, who also are non-farmers and aged 30-60.

A-8 Estimation of error-components model

We employ a method of moments estimator for our error components models. Specifically, the estimation minimizes the distance between the model's theoretical variances and autocovariances and those calculated in our panel data. For each combination of normalized age, a, calendar year, t, and lead, k, the error-components models described by equations 9-12 imply the parametric form for each autocovariance of residual, transformed income such as $cov(\xi_{a,t}, \xi_{a+k,t+k})$. As an example, if $a = 5$, $t = 2006$, $k = 0$, the counterpart to the theoretical moment is the variance (since $k = 0$)

of income for 35 year olds (since normalized age is 5) in year 2006. The theoretical variances and autocovariances ($cov(a,t,k)$) are functions of the model parameters: $\sigma^2_\alpha, \sigma^2_r, \sigma^2_z, \rho, \lambda_t, \pi_t$, for $t = 1987, ..., 2018$. The empirical variance and autocovariances are computed from our panel of tax returns for $a = 1, ..., 31$, $t = 1987, ..., 2018$, and $k = 0, ..., 22$. There are 6,532 of these variances and autocovariances use in the estimation of each model. We weight each moments using the inverse of the number of observations used to compute each empirical statistical moment.[36]

Note that our long panel allows use to distinguish between permanent and very persistent shocks. The half life of an AR(1) process with a persistence parameter of 0.9 is about 7 years. With our 32 year panel, we observe a large number of business owners over more than a 7 year window (the average is 13.9 years in the business income panel), allowing us to separately identify permanent and highly persistent shocks to income.

A-8.A First stage estimates

Please find parameter estimates from the first-stage regressions from which the residual IHS-transformed income is determined in the attached file:
`FirstStageResults_Formatted_ForOnlineAppendix.xlsx`

A-8.B Moment conditions

Let a be "normalized age", defined as a = age-30+1, or years since age 30. Then, the theoretical moments implied by our baseline error-components model in equations 9-12 are as follows:

$$cov(\xi^i_{a,t}, \xi^i_{a+k,t+k}) = \lambda_t \cdot \lambda_{t+k} \cdot \left(\sigma^2_\alpha + (\sigma^2_r \cdot a)\right) + \rho^k var(z^i_t) \tag{A.8.2}$$

For $t = 1987$, $2 \leq a \leq 36$:

$$var(z^i_{a,1987}) = \sigma^2_z \frac{1-\rho^{2a}}{1-\rho^2} \tag{A.8.3}$$

For $1987 \leq t \leq 2018$, $a = 1$:

$$var(z^i_{1,t}) = \pi^2_t \sigma^2_z \tag{A.8.4}$$

For $1988 \leq t \leq 2018$, $2 \leq a \leq 36$:

$$var(z^i_{a,t}) = \rho^2 var(z^i_{a-1,t-1}) + \pi^2_t \sigma^2_z \tag{A.8.5}$$

To obtain identification, we impose the normalization $\lambda_t = \pi_t = 1$ for all calendar years $t \leq 1987$,

[36] We do not use an optimal weighting matrix for reasons discussed in Altonji and Segal (1996)

where 1987 is the first year in the sample. We also impose $\pi_{2018} = \pi_{2017}$, since 2018 is the last year in the sample.

A-8.C Error component model parameter estimates, full set of parameter estimates

Table A.9: Error Component Model Parameter Estimates, Business Income

	FE_AR1_RW	FE_AR1	FE_RW_WN	RW_WN
ρ	0.615	0.762	0.000	0.000
	(0.004)	(0.003)	-	-
σ^2_α	4.616	6.087	5.440	0.000
	(0.271)	(0.468)	(0.322)	-
σ^2_r	0.483	0.000	0.593	1.185
	(0.029)	-	(0.036)	(0.072)
σ^2_z	19.172	14.123	28.805	28.943
	(0.835)	(0.588)	(1.603)	(1.245)
π_{1988}	1.002	1.010	0.956	0.955
	(0.042)	(0.056)	(0.037)	(0.028)
π_{1989}	0.977	0.976	0.952	0.949
	(0.039)	(0.051)	(0.040)	(0.030)
π_{1990}	0.992	1.035	0.948	0.944
	(0.036)	(0.046)	(0.038)	(0.028)
π_{1991}	0.970	0.986	0.940	0.938
	(0.033)	(0.045)	(0.036)	(0.027)
π_{1992}	0.961	0.972	-0.932	0.934
	(0.033)	(0.043)	(0.034)	(0.027)
π_{1993}	0.947	0.972	0.940	0.941
	(0.033)	(0.046)	(0.033)	(0.025)
π_{1994}	0.891	0.902	-0.901	-0.917
	(0.032)	(0.047)	(0.032)	(0.026)
π_{1995}	0.941	1.017	-0.903	-0.914
	(0.032)	(0.046)	(0.032)	(0.025)
π_{1996}	0.904	0.927	-0.917	-0.926
	(0.032)	(0.050)	(0.033)	(0.025)
π_{1997}	0.931	0.958	-0.912	-0.926
	(0.031)	(0.047)	(0.031)	(0.024)
π_{1998}	0.918	0.930	-0.920	0.936
	(0.032)	(0.048)	(0.032)	(0.024)
π_{1999}	0.924	0.947	-0.926	0.938
	(0.032)	(0.048)	(0.032)	(0.025)
π_{2000}	0.924	0.942	-0.922	0.934
	(0.032)	(0.048)	(0.032)	(0.026)
π_{2001}	0.938	0.953	-0.915	-0.930
	(0.030)	(0.046)	(0.030)	(0.026)
π_{2002}	0.988	1.012	0.946	0.958
	(0.029)	(0.044)	(0.030)	(0.026)
π_{2003}	0.984	0.970	0.975	0.985
	(0.029)	(0.046)	(0.031)	(0.026)
π_{2004}	0.976	0.962	0.955	0.972
	(0.029)	(0.046)	(0.031)	(0.028)

	FE_AR1_RW	FE_AR1	FE_RW_WN	RW_WN
π_{2005}	0.956	0.949	0.945	0.958
	(0.028)	(0.044)	(0.030)	(0.027)
π_{2006}	0.980	0.989	0.957	0.969
	(0.028)	(0.042)	(0.030)	(0.028)
π_{2007}	0.973	0.957	0.966	0.978
	(0.030)	(0.044)	(0.031)	(0.029)
π_{2008}	1.000	1.014	0.972	0.984
	(0.030)	(0.041)	(0.031)	(0.029)
π_{2009}	1.016	1.040	0.961	0.972
	(0.030)	(0.039)	(0.031)	(0.030)
π_{2010}	1.000	1.004	0.945	0.957
	(0.030)	(0.039)	(0.031)	(0.031)
π_{2011}	0.969	0.967	0.933	0.945
	(0.031)	(0.039)	(0.031)	(0.032)
π_{2012}	0.993	1.000	-0.924	0.935
	(0.029)	(0.037)	(0.031)	(0.032)
π_{2013}	0.972	0.955	0.925	0.936
	(0.030)	(0.038)	(0.032)	(0.034)
π_{2014}	1.003	0.989	0.923	0.938
	(0.032)	(0.037)	(0.034)	(0.037)
π_{2015}	1.034	1.022	0.951	0.963
	(0.033)	(0.039)	(0.035)	(0.037)
π_{2016}	1.068	1.061	0.982	0.991
	(0.037)	(0.042)	(0.037)	(0.040)
π_{2017}	1.119	1.176	1.059	1.066
	(0.032)	(0.031)	(0.036)	(0.034)
π_{2018}	#N/A	#N/A	#N/A	#N/A
	#N/A	#N/A	#N/A	#N/A
λ_{1988}	1.047	1.031	1.107	1.100
	(0.039)	(0.053)	(0.041)	(0.043)
λ_{1989}	1.025	0.981	1.102	1.088
	(0.039)	(0.053)	(0.043)	(0.044)
λ_{1990}	0.978	0.858	1.078	1.055
	(0.038)	(0.052)	(0.042)	(0.041)
λ_{1991}	0.967	0.838	1.064	1.026
	(0.037)	(0.050)	(0.040)	(0.039)
λ_{1992}	0.938	0.816	1.038	0.984
	(0.036)	(0.051)	(0.039)	(0.038)
λ_{1993}	0.965	0.836	1.045	0.986
	(0.036)	(0.051)	(0.038)	(0.037)
λ_{1994}	1.045	1.042	1.092	0.994
	(0.036)	(0.053)	(0.038)	(0.036)

	FE_AR1_RW	FE_AR1	FE_RW_WN	RW_WN
λ_{1995}	1.089	1.049	1.130	1.030
	(0.038)	(0.056)	(0.039)	(0.036)
λ_{1996}	1.057	1.015	1.088	0.988
	(0.036)	(0.055)	(0.037)	(0.035)
λ_{1997}	1.081	1.094	1.108	0.990
	(0.037)	(0.056)	(0.037)	(0.035)
λ_{1998}	1.083	1.142	1.100	0.973
	(0.036)	(0.059)	(0.037)	(0.034)
λ_{1999}	1.047	1.079	1.067	0.944
	(0.035)	(0.058)	(0.036)	(0.033)
λ_{2000}	1.034	1.081	1.056	0.932
	(0.035)	(0.058)	(0.035)	(0.032)
λ_{2001}	1.052	1.136	1.075	0.944
	(0.035)	(0.059)	(0.035)	(0.032)
λ_{2002}	1.022	1.081	1.063	0.934
	(0.034)	(0.057)	(0.035)	(0.032)
λ_{2003}	1.027	1.118	1.069	0.938
	(0.034)	(0.058)	(0.036)	(0.033)
λ_{2004}	1.112	1.301	1.144	0.996
	(0.036)	(0.065)	(0.038)	(0.034)
λ_{2005}	1.106	1.283	1.138	0.994
	(0.036)	(0.064)	(0.037)	(0.034)
λ_{2006}	1.095	1.255	1.133	0.989
	(0.035)	(0.062)	(0.037)	(0.034)
λ_{2007}	1.077	1.246	1.116	0.971
	(0.035)	(0.061)	(0.036)	(0.033)
λ_{2008}	1.039	1.172	1.097	0.953
	(0.034)	(0.058)	(0.035)	(0.032)
λ_{2009}	1.013	1.106	1.096	0.952
	(0.034)	(0.055)	(0.036)	(0.033)
λ_{2010}	1.034	1.142	1.118	0.969
	(0.035)	(0.056)	(0.037)	(0.033)

	FE_AR1_RW	FE_AR1	FE_RW_WN	RW_WN
λ_{2011}	1.015	1.115	1.099	0.952
	(0.034)	(0.055)	(0.036)	(0.033)
λ_{2012}	1.000	1.080	1.099	0.953
	(0.034)	(0.054)	(0.036)	(0.033)
λ_{2013}	1.032	1.162	1.118	0.968
	(0.035)	(0.056)	(0.037)	(0.034)
λ_{2014}	1.029	1.196	1.130	0.974
	(0.035)	(0.057)	(0.037)	(0.034)
λ_{2015}	1.022	1.205	1.134	0.979
	(0.036)	(0.057)	(0.039)	(0.036)
λ_{2016}	0.989	1.171	1.116	0.966
	(0.035)	(0.056)	(0.039)	(0.036)
λ_{2017}	0.985	1.170	1.088	0.939
	(0.037)	(0.055)	(0.040)	(0.038)
λ_{2018}	1.036	1.221	1.081	0.938
	(0.038)	(0.057)	(0.039)	(0.038)
Observations	10,548	10,548	10,548	10,548
Root MSE	4.329	5.324	6.171	7.024

Table A.9 presents the results of model estimation using our benchmark 1987-2018 business income panel where we drop households who never report income outside the $\pm\$5,000$ interval, restrict ages to 30-60, and exclude farmers. In the models, FE indicates fixed effects, AR indicates an AR(1) component, RW a random walk component, and WN a white noise component, see text for details.

Table A.10: Error Component Model Parameter Estimates, Labor Income

	FE_AR1_RW	FE_AR1	FE_RW_WN	RW_WN
ρ	0.699	0.848	0.000	0.000
	(0.005)	(0.002)	-	-
σ_α^2	0.182	0.197	0.194	0.000
	(0.005)	(0.007)	(0.005)	-
σ_r^2	0.004	0.000	0.005	0.025
	(0.000)	-	(0.000)	(0.001)
σ_z^2	0.160	0.086	0.277	0.290
	(0.006)	(0.003)	(0.012)	(0.029)
π_{1987}	1.000	1.000	1.000	1.000
	-	-	-	-
π_{1988}	1.040	1.178	1.001	1.009
	(0.043)	(0.075)	(0.031)	(0.073)
π_{1989}	1.047	1.165	0.995	1.011
	(0.041)	(0.073)	(0.031)	(0.077)
π_{1990}	1.007	1.168	0.935	0.966
	(0.043)	(0.069)	(0.032)	(0.082)
π_{1991}	1.007	1.131	0.937	0.966
	(0.038)	(0.066)	(0.032)	(0.076)
π_{1992}	1.011	1.067	0.969	1.009
	(0.044)	(0.074)	(0.038)	(0.077)
π_{1993}	0.949	1.066	0.928	0.969
	(0.041)	(0.070)	(0.032)	(0.075)
π_{1994}	0.954	1.075	0.923	0.968
	(0.043)	(0.071)	(0.032)	(0.074)
π_{1995}	0.950	1.086	0.904	0.952
	(0.042)	(0.069)	(0.032)	(0.074)
π_{1996}	0.934	1.071	0.909	0.959
	(0.039)	(0.067)	(0.029)	(0.073)
π_{1997}	0.868	0.872	0.882	0.931
	(0.043)	(0.084)	(0.030)	(0.072)
π_{1998}	0.923	1.084	0.897	0.953
	(0.043)	(0.069)	(0.029)	(0.073)
π_{1999}	0.854	0.823	0.902	0.957
	(0.041)	(0.086)	(0.026)	(0.069)
π_{2000}	0.892	1.021	0.910	0.966
	(0.044)	(0.071)	(0.030)	(0.071)
π_{2001}	0.883	1.020	0.913	0.969
	(0.042)	(0.067)	(0.029)	(0.073)
π_{2002}	0.905	1.030	0.918	0.966
	(0.036)	(0.064)	(0.026)	(0.069)
π_{2003}	0.982	1.186	0.955	1.004
	(0.038)	(0.060)	(0.029)	(0.072)
π_{2004}	0.887	0.941	0.926	0.974
	(0.045)	(0.077)	(0.032)	(0.074)

	FE_AR1_RW	FE_AR1	FE_RW_WN	RW_WN
π_{2005}	0.878	0.961	0.916	0.966
	(0.050)	(0.086)	(0.033)	(0.074)
π_{2006}	0.959	1.121	0.935	0.987
	(0.048)	(0.072)	(0.033)	(0.076)
π_{2007}	0.931	1.075	0.926	0.983
	(0.046)	(0.067)	(0.035)	(0.080)
π_{2008}	0.946	1.103	0.901	0.957
	(0.039)	(0.064)	(0.029)	(0.076)
π_{2009}	1.004	1.115	0.930	0.989
	(0.040)	(0.064)	(0.033)	(0.080)
π_{2010}	0.957	1.015	0.934	0.991
	(0.046)	(0.079)	(0.033)	(0.082)
π_{2011}	0.993	1.130	0.907	0.973
	(0.051)	(0.076)	(0.039)	(0.089)
π_{2012}	1.000	1.084	0.912	0.981
	(0.045)	(0.072)	(0.036)	(0.088)
π_{2013}	0.984	1.079	0.895	0.963
	(0.042)	(0.067)	(0.035)	(0.085)
π_{2014}	1.014	1.107	0.917	0.982
	(0.041)	(0.066)	(0.036)	(0.085)
π_{2015}	0.992	1.092	0.915	0.980
	(0.042)	(0.067)	(0.037)	(0.087)
π_{2016}	1.007	1.126	0.953	1.007
	(0.039)	(0.066)	(0.034)	(0.083)
π_{2017}	1.029	1.260	0.975	1.009
	(0.029)	(0.042)	(0.028)	(0.069)
π_{2018}	1.029	1.260	0.975	1.009
	-	-	-	-
λ_{1988}	1.053	1.027	1.079	1.054
	(0.018)	(0.025)	(0.019)	(0.034)
λ_{1989}	1.055	1.005	1.103	1.054
	(0.018)	(0.024)	(0.020)	(0.033)
λ_{1990}	1.061	0.981	1.122	1.045
	(0.017)	(0.023)	(0.020)	(0.033)
λ_{1991}	1.083	0.989	1.141	1.050
	(0.018)	(0.024)	(0.020)	(0.032)
λ_{1992}	1.097	1.015	1.150	1.029
	(0.019)	(0.025)	(0.020)	(0.032)
λ_{1993}	1.101	1.007	1.149	1.017
	(0.018)	(0.024)	(0.019)	(0.030)
λ_{1994}	1.111	1.013	1.153	1.004
	(0.018)	(0.024)	(0.020)	(0.029)

	FE_AR1_RW	FE_AR1	FE_RW_WN	RW_WN
λ_{1995}	1.141	1.041	1.177	1.011
	(0.018)	(0.023)	(0.019)	(0.028)
λ_{1996}	1.162	1.061	1.188	1.007
	(0.018)	(0.024)	(0.019)	(0.028)
λ_{1997}	1.198	1.129	1.208	1.015
	(0.019)	(0.026)	(0.019)	(0.028)
λ_{1998}	1.207	1.133	1.212	1.003
	(0.018)	(0.026)	(0.019)	(0.027)
λ_{1999}	1.215	1.176	1.205	0.987
	(0.018)	(0.027)	(0.019)	(0.027)
λ_{2000}	1.206	1.163	1.194	0.972
	(0.018)	(0.026)	(0.019)	(0.026)
λ_{2001}	1.227	1.185	1.208	0.980
	(0.018)	(0.026)	(0.018)	(0.026)
λ_{2002}	1.235	1.193	1.216	0.990
	(0.018)	(0.026)	(0.018)	(0.026)
λ_{2003}	1.212	1.151	1.206	0.972
	(0.018)	(0.026)	(0.018)	(0.026)
λ_{2004}	1.221	1.174	1.211	0.977
	(0.018)	(0.027)	(0.018)	(0.026)
λ_{2005}	1.237	1.200	1.220	0.983
	(0.019)	(0.028)	(0.019)	(0.027)
λ_{2006}	1.240	1.198	1.231	0.987
	(0.018)	(0.026)	(0.019)	(0.027)
λ_{2007}	1.247	1.203	1.240	0.989
	(0.018)	(0.026)	(0.019)	(0.027)
λ_{2008}	1.246	1.193	1.248	0.998
	(0.018)	(0.026)	(0.019)	(0.027)
λ_{2009}	1.242	1.195	1.255	0.999
	(0.018)	(0.026)	(0.019)	(0.027)
λ_{2010}	1.282	1.252	1.289	1.030
	(0.019)	(0.028)	(0.020)	(0.028)

	FE_AR1_RW	FE_AR1	FE_RW_WN	RW_WN
λ_{2011}	1.278	1.244	1.299	1.035
	(0.020)	(0.027)	(0.020)	(0.028)
λ_{2012}	1.269	1.241	1.298	1.032
	(0.019)	(0.026)	(0.020)	(0.029)
λ_{2013}	1.247	1.217	1.285	1.022
	(0.019)	(0.025)	(0.020)	(0.029)
λ_{2014}	1.239	1.214	1.282	1.020
	(0.019)	(0.025)	(0.021)	(0.029)
λ_{2015}	1.255	1.237	1.294	1.032
	(0.020)	(0.025)	(0.021)	(0.030)
λ_{2016}	1.199	1.175	1.237	0.986
	(0.018)	(0.024)	(0.021)	(0.030)
λ_{2017}	1.211	1.171	1.234	0.998
	(0.018)	(0.023)	(0.020)	(0.030)
λ_{2018}	1.212	1.162	1.209	0.980
	(0.018)	(0.023)	(0.020)	(0.031)
Observations	10,548	10,548	10,548	10,548
Root MSE	0.056	0.0623	0.0736	0.167

Table A.10 presents the results of model estimation using our benchmark 1987-2018 labor income panel where we drop observations below $2,575, restrict ages to 30-60, and exclude farmers. In the models, FE indicates fixed effects, AR indicates an AR(1) component, RW a random walk component, and WN a white noise component, see text for details.

A-9 Robustness of Findings: Alternative Age Restrictions

Our benchmark panels restrict the sample to households with a primary filer aged 30-60. We make this restriction to focus on household with the most attachment to the labor force and to make our results more comparable to past work on labor earnings, which typically focuses on prime-aged workers (for example, Kopczuk et al. (2010) restrict their sample to those aged 25-60). This section provides some insights regarding the effects of these sample restrictions.

One of the largest differences across the lifecycle are exit rates from business ownership. Figure A.5 plots exit rates by age (denoting with vertical dotted line our sample age restrictions). Business exit rates increase with age, but the gradient becomes much steeper after age 60. Therefore, if we were to relax our age restrictions and include older primary filers, we would find higher rates of exit and more frequent transitions to the zero state in our business income transition matrices.

Figure A.5: Rates of Business Exit by Age

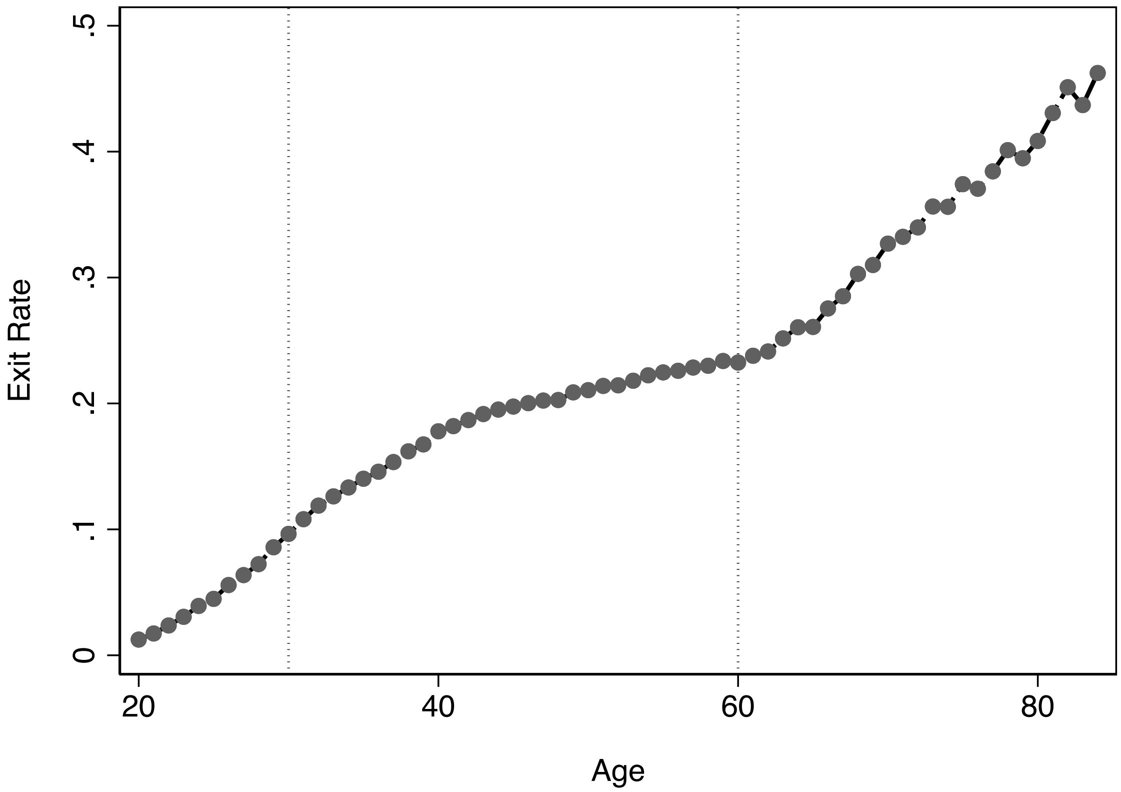
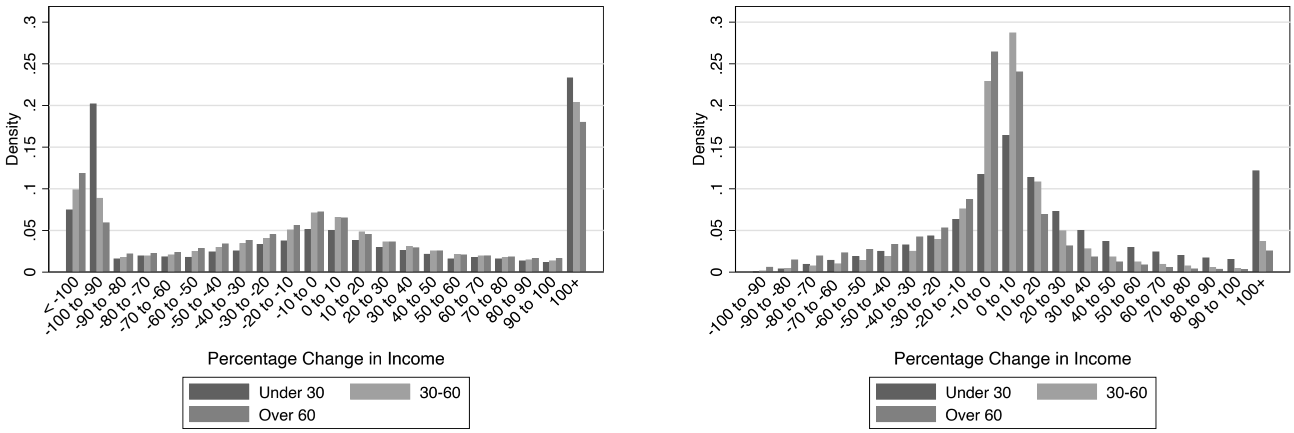


Figure A.5 plots the lifecycle profile of business exit rates using a panel that drops farmers households who never never earn business income outside of the (-$5,000, $5,0000) interval. Business exit is defined as earning non-zero business income last year and zero business income in the current year.

Figure A.6 show the percentage changes in business and labor income using our benchmark panels and then alternative samples that restrict the primary filer age to be less than 30 and over 60, respectively. Those outside of our benchmark ages 30 to 60 have fairly similar distributions of percentage changes. The notable exception being that household with younger primary filers see more large, positive percentage changes than other age groups. This is true for both business income and labor income.

Figure A.6: **Percentage Changes in Business and Labor Income, various age restrictions**

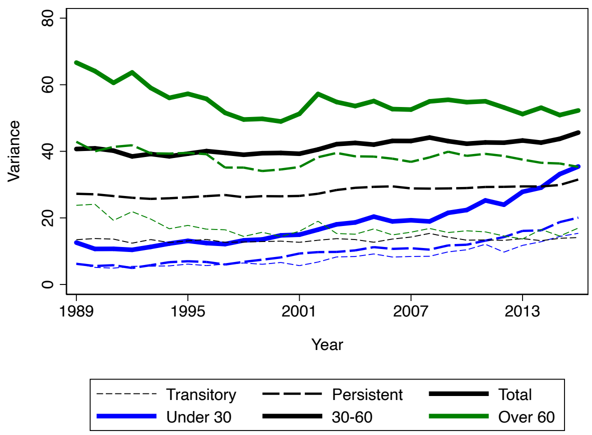


(a) Business Income (b) Labor Income

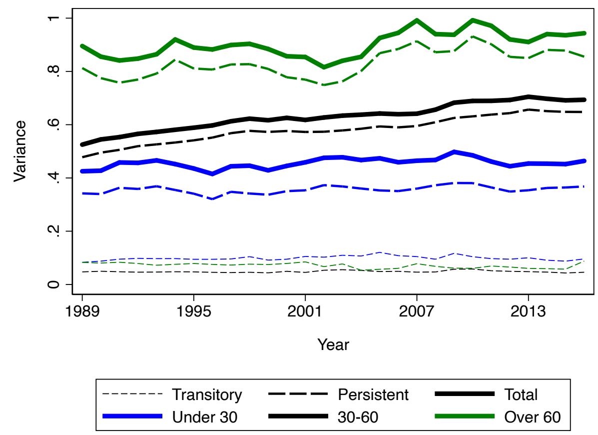
Figure *A.6* presents the distribution of percentages changes in income over one year, for business and labor income over various samples. The panels used to construct percent changes in business income income are our benchmark sample, which drops taxpayers who never have business income outside of the (-$5,000, $5,0000) interval, restricts to taxpayers where the primary filers is aged to 30-60, and excludes farmers and then two other samples where we restrict ages to those under 30 and over 60, respectively. The labor income panels include our benchmark panel, which drops taxpayers who have labor income less than $2,575, restricts to taxpayers where the primary filers is aged to 30-60, and excludes farmers and two other samples, where we include all observations with primary files aged 30-60 without farm income, then two other samples where we restrict ages to those under 30 and over 60, respectively. The horizontal axis shows the size of the percent change. All bins have a size of 10 percentage points, except the last bin on the right and the last bin on the left. The last bin on the right groups together all observations for which income increased by more than 100%. The last bin on the left groups together all observations for which income decreased by more than 100%. The vertical axis shows the fraction of all business or labor income observations of percent changes in each size-of-percent-change bin.

In Figure A.7, we using the KSS non-parametric decomposition of the variance in IHS-transformed earnings to show how the important the persistent and transitory shocks are for different age groups. These plots decompose the persistent and transitory components of the total variance for our benchmark panels and then analogous panels that restrict ages to those under 30 and over 60, respectively.

Figure A.7: **Persistent vs. Transitory Risk, KSS (2010), various age restrictions**



(a) Business Income



(b) Labor Income

Figure A.7 decomposes the total cross sectional variance into its persistent and transitory components using the decomposition of Kopczuk et al. (2010). The decomposition using a window of 5 years (P=5). See text for more details. The panels used to construct the variances in business income income are our benchmark sample, which drops taxpayers who never have business income outside of the (-$5,000, $5,0000) interval, restricts to taxpayers where the primary filers is aged to 30-60, and excludes farmers and then two other samples where we restrict ages to those under 30 and over 60, respectively. The labor income panels include our benchmark panel, which drops taxpayers who have labor income less than $2,575, restricts to taxpayers where the primary filers is aged to 30-60, and excludes farmers and then two other samples where we restrict ages to those under 30 and over 60, respectively.

Figure A.7 relates some interesting patterns. First, the total variance in income is largest for those over 60 and smallest for those under 30. This is true for both business and labor income. Second, the relative importance of the persistent component (relative to the transitory component) is similar across age groups. For all ages, the transitory component is more important in explaining the cross-sectional variance in business income, whereas most of the cross-sectional variance is explained by the persistent component of earnings. Finally, the variance in business income has increased substantially for those under 30 over the sample period. In contrast, the variance in labor earnings for this age group has remained steady from 1987-2018. This interesting pattern with young business owners may warrant further study.

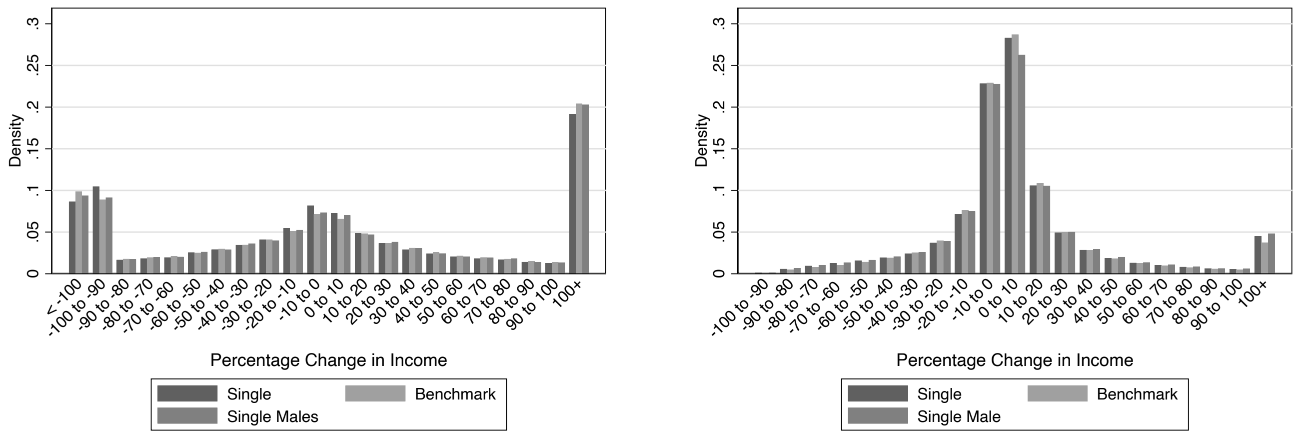
Overall, considering alternative age restrictions would show an increase in business exit and overall variance. However, the general findings of larger tail risk (both positive and negative) for business owners, the large variance of business income relative to labor income, and the importance of transitory shocks in business income and persistent shocks in labor income all obtain regardless of the age restrictions used.

A-10 Robustness of Findings: Alternative Household Construction

In this appendix section we show how restricting our sample to specific household structure affect our findings. Our benchmark panels make no restrictions on household type and therefore includes both married and single filers. As a results, some earning volatility is attributable to changes in household structure. For example, the earnings of a primary filer may decrease if a marriage is terminated and the spouse had non-zero earnings. While changes in household structure are a part of the earnings volatility households face, we may want to understand how much earnings volatility remains when holding household structure constant. To that end, this section presents two sets of results, percentage changes in income and a decomposition of the cross-sectional variance in income, for our benchmark sample and then two other samples. These samples include just single filer and then just single male filers (following the restriction to focus only on mail earners in a large number of papers estimating labor earnings processes).

Figure A.8 relates percentage changes in business and labor income for our benchmark panels, as well as panels that restrict to single households and single male households. The distributions of percentage changes are very similar regardless of the restrictions on household type that are used. Regardless of the sample considered, business income shows a much larger variance of percentage changes and much more action in the tails, than labor income.

Figure A.8: **Percentage Changes in Business and Labor Income, alternative household structure restrictions**



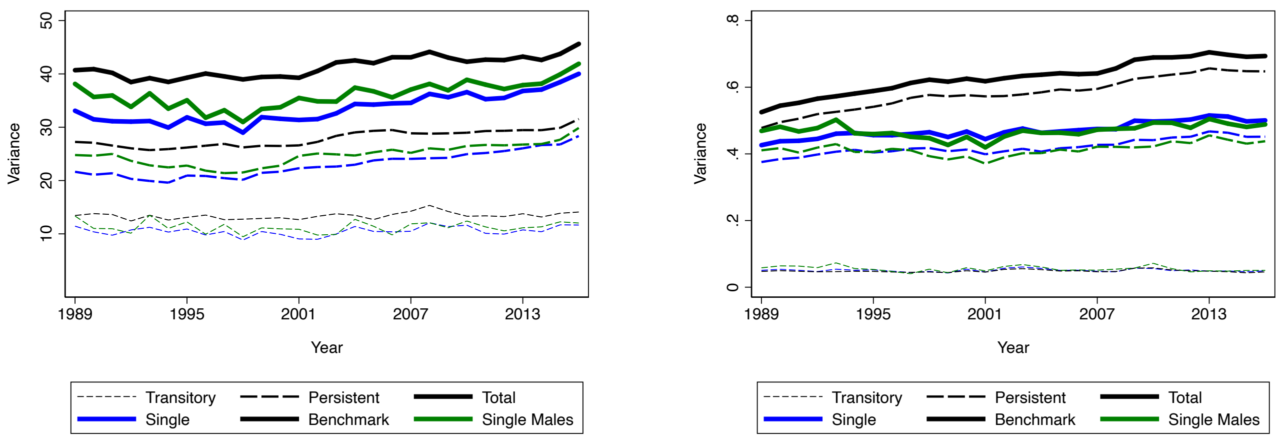
(a) Business Income

(b) Labor Income

Figure A.8 presents the distribution of percentages changes in income over one year, for business and labor income over various samples. The panels used to construct percent changes in business income income are our benchmark sample, which drops taxpayers who never have business income outside of the (-\$5,000, \$5,0000) interval, restricts to taxpayers where the primary filers is aged to 30-60, and excludes farmers and then two other samples where we restrict to single households and single male households, respectively. The labor income panels include our benchmark panel, which drops taxpayers who have labor income less than \$2,575, restricts to taxpayers where the primary filers is aged to 30-60, and excludes farmers and two other samples, where we include all observations with primary files aged 30-60 without farm income, then two other samples where we restrict to single households and single male households, respectively. The horizontal axis shows the size of the percent change. All bins have a size of 10 percentage points, except the last bin on the right and the last bin on the left. The last bin on the right groups together all observations for which income increased by more than 100%. The last bin on the left groups together all observations for which income decreased by more than 100%. The vertical axis shows the fraction of all business or labor income observations of percent changes in each size-of-percent-change bin.

The larger cross-sectional variance in our benchmark panels is clear in Figure A.9, where we plot the KSS decomposition of IHS-transformed business and labor income. Including married households increases the cross-sectional variance of business and labor income by about 1/3 (though this percentage difference grows over time for labor income). The variance in earnings for single households overall and single males are quite similar. Furthermore, regardless of the sample considered, the relative importance of the persistent and transitory components to explaining the overall variance in business and labor income remain. For all samples, persistent difference in earnings explain the vast majority of the cross-sectional variance in labor income, but only about two-thirds of the variance in business income.

Figure A.9: **Persistent vs. Transitory Risk, KSS (2010), alternative household structure restrictions**



(a) Business Income

(b) Labor Income

Figure A.9 decomposes the total cross sectional variance into its persistent and transitory components using the decomposition of Kopczuk et al. (2010). The decomposition using a window of 5 years (P=5). See text for more details. The panels used to construct the variances in business income income are our benchmark sample, which drops taxpayers who never have business income outside of the (-$5,000, $5,0000) interval, restricts to taxpayers where the primary filers is aged to 30-60, and excludes farmers and then two other samples where we restrict to single households and single male households, respectively. The labor income panels include our benchmark panel, which drops taxpayers who have labor income less than $2,575, restricts to taxpayers where the primary filers is aged to 30-60, and excludes farmers and then two other samples where we restrict to single households and single male households, respectively.