A Risky Venture: Income Dynamics within the Non-Corporate Private Business Sector

Jason DeBacker, Vasia Panousi, Shanthi Ramnath*

November 3, 2014

Abstract

We employ a new, large, and confidential panel of US income tax returns for the period 1987-2009 to extensively characterize and quantify business income risks and to compare to labor income or earnings risks. Our findings show business income to be much riskier than labor income, even conditional on a household's continued participation in private business endeavors. Compared to labor income, business income is less persistent, and it is characterized by higher probabilities of extreme upward or downward mobility. Our data also suggest that high-income households are more likely to face the tail risks on both ends of the business income distribution. Furthermore, we find that, as compared to labor income, the cross-sectional variation in business income is less explained by heterogeneity across households than within household income variation. Finally, we use our models of income dynamics models to calibrate a life-cycle model with endogenous business ownership to explain an apparent business ownership puzzle.

*Jason DeBacker, Middle Tennessee State University, jason.debacker@mtsu.edu; Vasia Panousi, Federal Reserve Board, vasia.panousi@frb.gov; Shanthi Ramnath, US Department of the Treasury, shanthi.ramnath@treasury.gov. We are grateful to Joe Altonji, George-Marios Angeletos, Chris Carroll, Janice Eberly, Glenn Follette, John Haltiwanger, Marco Cagetti, Mariacristina De Nardi, Kevin Moore, Ricardo Reis, Jose-Victor Rios-Rull, John Sabelhaus, Louise Sheiner, Paul Smith, Gianluca Violante, Annette Vissing-Jørgensen, and the seminar participants at the Federal Reserve Board, the 6th annual meeting of the Portuguese Economic Journal, the Summer 2012 NBER Capital Markets Group (EFEL), the Summer 2012 NBER Conference on Research in Income and Wealth (CRIW), Indiana University, the Census Bureau, the Brookings Institute, the Bank of Portugal, and the New York Fed for helpful comments and discussions. We are thankful to our discussants, Nikolai Roussanov (EFEL) and Ron S. Jarmin (CRIW), for extremely insightful suggestions. The views in this paper are solely those of the authors and do not necessarily represent those of the Treasury Department, the Board of Governors of the Federal Reserve System, or members of their staffs.

1 Introduction

Privately held businesses account for nearly half of aggregate production and employment in the United States (Davis, Haltiwanger, Jarmin and Miranda (2007)), yet very little is known about the risks that private business owners face. This gap in the literature stands in contrast to the large body of work that examines the theoretical and empirical importance of risks associated with earning labor income. Instead, most empirical work to date focuses on heterogeneity in business outcomes rather than risks faced by owners. This is in part due to data limitations of past studies that often use cross section data for their analyses and are therefore unable to follow businesses over time–a crucial element for distinguishing true risk from heterogeneity within the population.

Our paper is one of the first to empirically document properties of income risk faced by privately held businesses using a panel of US individual income tax return data over the period 1987-2009.[1] We focus our analysis specifically on the non-corporate sector, which accounts for roughly 58% of aggregate net business income.[2] As Figure 1 shows, the non-corporate sector is becoming an increasingly important part of the economy. Non-corporate business income is a source of income for over 25 million households each year, and thus an important topic of study. Administrative tax data are uniquely suited to study properties of business income risk, since unlike many surveys, the income variables are not subject to top coding. In addition, these data allow one to above both positive and negative income realizations. Finally, the panel nature of the data allows us to decompose the total variance of business income realizations into its components that are due to heterogeneity among business owners in (time invariant) ability and to true risk faced by business owners.

[Figure 1 about here.]

We employ several different methods to examine the properties of business income risk as well as provide a comparison of business income risk to that of labor income. First, using one-year transition matrices, we find that business income has lower persistence than labor income with an immobility ratio of roughly 0.4 for business income and roughly 0.6 for labor income. Conditional on leaving their initial decile, a household faces a higher probability of moving further in the business income distribution than in the labor income distribution. Households starting at the lowest decile of the business income distribution face a 9% probability of transitioning to one of the three topmost deciles, whereas the corresponding probability is essentially zero for labor income. Likewise, households starting at the highest decile of the business income distribution face

[1] Notably, Heaton and Lucas (2000) use a publicly available panel of 1979-1990 tax returns to test the hypothesis that owners of private businesses are more likely than wage earners to hold the rest of their portfolios in riskless assets. However, their analysis uses a measure of business income that includes many other income categories from Form 1040, in addition to "true" business income, and it furthermore excludes negative income observations.

[2] This statistic comes from the authors' calculations using the Statistics of Income Tax Stats Integrated Business Data tabulations of business tax returns.

a 4% probability of transitioning to the bottom decile within a year, whereas the corresponding probability is close to zero for labor income. This picture is preserved qualitatively over longer time horizons, indicating that at least some business income risk persists over time.

Second, we analyze the factors that affect exit from business endeavors. We find that conditional on business participation in the previous year, the probability of business exit in the current year is 10%. Our evidence also indicates that exit is associated with poor business outcomes. In particular, years immediately before exit are more likely to be years with business income losses and more likely to have been preceded by a longer streak of previous losses. In addition, most households realize lower levels of capital gains in exit years, compared to other years. Furthermore, we find that the probability of exit decreases as proxies for business age and household business experience increase.

Third, we calculate households' exposure to risk derived from business participation over time as captured by year-to-year income changes. Looking at year-to-year changes in income, we find that the risk distribution is more dispersed for business than for labor income. Importantly, the distribution of business income risk has less mass in the middle and more mass in the tails compared to the distribution of labor income risk.

Overall, we find that business income appears to be riskier than labor income, not only because of its higher probability of exit, but also because of its more pronounced fluctuations conditional on not exiting. Most importantly, the risk involved in earning business income appears to be, to a large extent, tail risk, namely risk associated with large positive and/or negative income changes. Furthermore, our evidence suggests that both ends of the tail risk are actually borne by high-income households. In other words, the common assumption that business income variance does not depend on income or wealth appears to be at odds with the data.

We next use error components models of income dynamics to further analyze the observed income fluctuations. We find that a positive one standard deviation model-estimated shock leads to a percentage increase in the level of business income that is about 3-4 times larger than the corresponding increase for labor income. Thus, we again find that business income risk is substantially higher than labor income risk given that much of the variation of labor income can be explained by individual-specific differences. We are then able to decompose the total variation in business income into the part that is attributed to heterogeneity in individual business owners and the remainder that is due to risk faced by those individual business owners. The question of the importance of heterogeneity in business income earning ability was first raised by Lucas (1978). Others, such as Evans and Jovanovic (1989), further consider the importance of heterogeneity in the skill of business owners. Our paper is the first to answer this question empirically. We find that individual heterogeneity accounts for about 45% of the cross-sectional variance in labor income and about 25% of the cross-sectional variance in business income. This is relatively stable over our sample period. The high fraction of income variance that is due to within household variation underscores the risks business owners face.

Given the uncertainty in income realizations facing business owners relative to labor providers. and the lower mean income of business owners, a natural question that arises is: why do individuals choose business ownership as a source of income? Using our parameter estimates from the error components models, we calibrate a life-cycle model with endogenous business participation to answer this question. With the life-cycle model, we are able to account for heterogeneity in income processes and show that selection into business ownership is done by those most likely to realize a better return from business ownership. That is, when considering the appropriate counterfactual of what business owners would earn in the labor market, the apparent business ownership puzzle disappears.

2 Previous literature and motivation

Both theoretical macro and finance literatures are increasingly examining topics related to entrepreneurs and privately held businesses. For example, past studies have explored the implications of (potentially uninsurable) business income or investment risk on saving, investment, capital accumulation, global imbalances, fiscal policy, asset returns, and entrepreneurial choices. A non-exhaustive list includes Polkovnichenko (2003), Bitler, Moskowitz and Vissing-Jørgensen (2005), Angeletos (2007), Angeletos and Calvet (2006), Albanesi (2012), Angeletos and Panousi (2009), Angeletos and Panousi (2011), Roussanov (2010), Panousi (2012), Benhabib and Zhu (2008), Zhu (2010), Parker and Vissing-Jørgensen (2009), and Wang, Wang and Yang (2011). Most of the empirical evidence on entrepreneurs and private businesses derives from cross sectional data, such as the Survey of Consumer Finances (SCF). For example, Quadrini (2000), Gentry and Hubbard (2000), Carroll (2000), Moskowitz and Vissing-Jørgensen (2002), Moore (2004), Cagetti and De Nardi (2006), and De Nardi, Doctor and Krane (2007) document significant cross sectional heterogeneity in private business outcomes. Some aspects of entrepreneurship, especially those related to business entry and exit, are examined in Holmes and Schmitz (1990), Holmes and Schmitz (1995), Holmes and Schmitz (1996), Evans and Leighton (1989), Evans and Jovanovic (1989), Holtz-Eakin, Joulfaian and Rosen (1994), and Blanchflower and Oswald (1998). We use large, annual cross-sections of income tax returns from the IRS to replicate certain cross-sectional findings from these studies before turning to our analysis of business income risk using panel data. Using cross-sections of tax data, we confirm that the large and increasing variance in business income found in the SCF also exists in the administrative data we use.

Given the large variance found in cross sections of business income, a natural question that arises is to what extent is heterogeneity a contributing factor (see, for example, Lucas (1978) and Evans and Jovanovic (1989)). This question is widely studied with respect to labor income including studies such as Meghir and Pistaferri (2004), Guvenen (2009), and Schulhofer-Wohl and Norets (2009). All find an important role for heterogeneity in explaining labor income risks and

Meghir and Pistaferri (2004) further highlights the importance for both observable and unobservable heterogeneity. Unfortunately, a comparable measure of business income risk has been difficult to obtain due to data limitations. Moskowitz and Vissing-Jørgensen (2002) suggest that the standard deviation of returns in privately held business is approximately 50% per annum, though due to a lack of time variation in their data, this measure of risk is unable to separate the component of variance that is driven by heterogeneity. Nonetheless, the estimate from Moskowitz and Vissing-Jørgensen (2002) has been widely used to calibrate many of the macro and finance models mentioned above.

Of particular interest to the questions we explore in this paper includes analysis from Moskowitz and Vissing-Jørgensen (2002). Moskowitz and Vissing-Jørgensen (2002) document the puzzling result that some households invest heavily in a single privately held firm where the risk-return trade-off appears to be worse than had they invested instead in large public firms. Similarly, we address the seemingly puzzling choice of business ownership despite the lower mean income and higher variance for business income compared to labor earnings. Unlike Moskowitz and Vissing-Jørgensen (2002), our focus is on the choice to pursue labor income or business income rather than the investment of capital. Others, such as Carrington, McCue and Pierce (1996), Hamilton (2000), and Kawaguchi (2002) have also documented this puzzle. Although we acknowledge that other factors such as differences in preferences, expectations, and non-pecuniary benefits highlighted by Moskowitz and Vissing-Jørgensen (2002), Hamilton (2000), and Hurst and Pugsley (2011) may also drive the business ownership puzzle, we focus attention on heterogeneity in income processes and selection effects, which we show may be important for the business owner's decision. In what follows, we show that such income processes, with both a persistent and transitory income component, when used in models of endogenous business participation, can account for the apparent business ownership puzzle, even with rational, forward looking agents.

3 Individual Tax Return Data

For our analysis, we use data from the Statistics of Income (SOI) division of the Internal Revenue Service (IRS). The data consist of 23 years of individual tax returns, spanning the period 1987-2009.[3] Each year, the IRS draws a stratified random sample of tax returns, consisting of two subsamples. The first subsample includes all tax units where the primary filer's social security number ends in one of a set of four-digit combinations. From 1987 through 1997, two four-digit endings were sampled. This number was increased to five combinations between 1998-2005, and then to ten combinations from 2006 onwards. The second subsample, known as the high income oversample, consists of returns sampled at progressively higher rates at higher income levels, where the highest income returns are selected with certainty. The overall sample has about 120,000-200,000 observations per year, and is weighted to represent the universe of tax filers (which includes

[3]The data is kept at the US Treasury Department and is subject to strict confidentiality rules.

about 140 million returns per year). Over the 23 years, these repeated cross sections in total are comprised of over three million observations. Each cross section contains information from taxpayers' Form 1040 and from a number of other supplementary forms and schedules. In addition, the data include information on age and gender of the primary filer from matched Social Security Administration (SSA) records.

We restrict the data to households where the primary filer has either some nonzero labor income or some nonzero business income (or both). We also restrict the ages of the primary filer to 30-60, primarily to draw a better comparison between labor and business income, as filers over 60 are likely to have a lower attachment to the labor market or may take on a business as a hobby.[4] Finally, we exclude primary filers that report farm income (Schedule F), since their business income is likely to include at least some amount of government subsidies.

3.1 Income Definitions

Observations in the data are at the tax unit level. A tax unit consists of a primary filer (as well as any dependents) if the taxpayer is single, married but filing separately, a head of household, or a widow(er). A tax unit consists of a primary and secondary filer (as well as any dependents) if the taxpayer is married filing jointly. As a notational clarification, in this paper we use the terms "tax unit" and "household" interchangeably. In addition, all of our income measures are in real 2005 dollars.

Our measure of business income includes the sum of income generated from sole proprietorships, partnerships, and S corporations, each of which are reported on Form 1040 of an individual's income tax return. This income is the owner's share of net profit or loss from business operations after all expenses, costs, and deductions have been subtracted. For tax purposes, sole proprietorships, partnerships, and S corporations are passthrough entities and are not subject to double taxation. This definition of business income does not include rental income, dividend income, or capital gains.

A sole proprietor is an individual owner of an unincorporated business. Sole proprietors report income and expenses to the IRS on Schedule C of Form 1040. A partnership consists of a group of people who form a business where each person contributes money, property, labor or skill and shares in the profits and losses. A partnership files a Form 1065 with the IRS and each partner additionally receives a Schedule K-1, which reports his or her share of the partnership's profits or losses. This amount is then included on Schedule E and Form 1040.[5] Finally, S corporations are a type of corporation that satisfy a number of criteria and elect to pass their income and losses

[4] We have looked into the impact of these age restrictions on our analysis and found that they are minimal for business owners. In particular, older filers do not differ significantly from those in the 30-60 range. There are differences between those in the 30-60 range and those in the 20-29 range. Notably, those 20-29 have much lower mean business income and a significantly lower variance in business income than those prime age filers.

[5] We note that information returns, such as Schedule K-1 or 1099 forms, are not included in our data. Our measure of net income from partnerships includes income and losses from both passive and active participation partnerships. Active participation is defined as substantial involvement in the day-to-day business activities.

through to their shareholders for federal tax purposes.[6] These shareholders then report income or losses from the S corporation on Schedule E and Form 1040.[7]

All three types of businesses can carry back or carry forward net operating losses to other tax years, as long as these losses come from active owners or partners. Before 1998, losses could be carried back three years and carried forward 15 years.[8] Starting in 1998, carry-backs and carry-forwards were changed to two and 20 years, respectively. Net operating loss carry-backs and carry-forwards do not affect our measure of business income, since they are claimed on a separate area of the tax form.[9]

Our measure of labor income comes from the line item for "Wages, salaries, tips, etc." found on Form 1040. This measure does not include income that is excludable for tax purposes including, for example, contributions to employer-sponsored retirement accounts or health insurance premiums. To create a more comprehensive measure for income, which we call "Adjusted Total Income" (ATI), we define a new variable comprised of the tax unit's total amount of income (e.g. labor, business, capital) in addition to some excluded income items (e.g. employee contributions to retirement income, tax-exempt interest, nontaxable IRA and pension distributions, and nontaxable Social Security benefits).[10] Our measure for capital gains comes from Form 1040, which includes gains reported on Schedule D and Form 4797.[11]

In a given year, a household may report positive business income (business profits), negative business income (business losses), or zero business income in a given year. Due to IRS coding, a value of zero could indicate that either exactly zero income was earned from the operation of an existing business, or that the household did not submit a return with business income in that year. For sole proprietors, we have additional information to help distinguish the two sources of a zero income report. We find that less than 1% of the zeros reported on a sole proprietor return were due to exactly zero profits from operating a business. Hence, a zero value for business income in any

[6]In 2009, to qualify for S corporation status, the corporation must be a domestic corporation, have only allowable shareholders (including individuals, certain trusts, and estates) and may not include partnerships, corporations or non-resident alien shareholders, have no more than 100 shareholders, have one class of stock, and not be an ineligible corporation, i.e. certain financial institutions, insurance companies, and domestic international sales corporations. See IRS Publication 542.

[7]It is important to note that the attractiveness of these business forms, compared to organizing as a C corporation subject to the corporation income tax, changed after the Tax Reform Act of 1986, partly guiding our sample period choice. For a discussion of these issues, see Slemrod (1992).

[8]For example, a loss in 1997 could be used to offset income in tax years 1994-1996, or held to offset income in tax years 1998-2012.

[9]Carry-forwards may, however, affect our measure of adjusted total income, defined below, since they are included in the "other income" line on Form 1040, and the data do not contain sufficient information in a number of years to net these out. However, since losses of sole proprietorships, partnerships, and S corporations can be used to offset any income, at least 85% of net operating losses is generally used in the year of the loss. See Cooper and Knittel (2010).

[10]We do not have information on excludable tax-deferred retirement contributions made by employers or employer sponsored health insurance premiums, and so those amounts are omitted from our adjusted total income measure.

[11]Because the tax treatment of capital gains changed a number of times over the sample period, our data unfortunately do not contain sufficient detail to make our measure more consistent over time.

given year likely indicates that the household did not operate a business that year. Although we do not have corresponding data for partnerships and S corporations, we will nonetheless proceed under the assumption that a value of zero for business income represents non-participation. In other words, the business may have been operating during that year, but the household did not participate in these operations, and hence received zero compensation from the business. We will be referring to this process as "exit".[12] As long as business income is non-zero, then we will assume that the household is participating in business endeavors and is making profits or losses. We will be referring to this process as "conditional on no exit".[13]

3.2 Data Limitations

When dealing with tax data, and with reported business income in particular, it is important to note that income misreporting could emerge as a potential problem. This issue, however, is not unique to tax data in that survey data can also suffer from similar misreporting. For example, Hurst, Li and Pugsley (2010) estimate that self-employment income underreporting in household surveys is about 30%. We note that while there is no consensus about the distribution of tax non-compliance either in the cross section or over time, there is nonetheless evidence that tax evasion exists, especially among sole proprietors.[14] In general, the existing evidence suggests that smaller businesses are more likely to misreport income, though estimates vary depending on the year, methodology, and type of small business. Roughly 30% of sole proprietors underreport income according to the 1992 IRS tax gap study and work by Feldman and Slemrod (2007). The IRS estimates that underreporting for partnerships and S corporations was about 7-8% in 1992 and 18% in 2001, indicating that these entities likely report their income more accurately than sole proprietors. In addition to size, Andreoni, Erard and Feinstein (1998) argue that misreporting also varies by occupation and industry, with lower percentages found in finance, real estate, insurance, agriculture, and wholesale trade.

Although income misreporting raises some concern when estimating the level of business returns, it may pose less of a problem for estimating risk or income changes over time, as long as misreported income does not systematically vary with time or with the business cycle. Indeed, Plumley (1996) finds that economic variables, such as the unemployment rate, are not a statistically significant

[12] Exit does not necessarily indicate a business failure or closure. For example, an owner who continues the operation of a business, but who reorganizes their business into a C-corporation will also show up in the data as having zero business income (i.e., an exit). It is not feasible to track such changes through the tax data, but we do note that these are likely to represent a very small fraction of our exits.

[13] Because our data is at the tax unit level, we do not know which business the household's income comes from over time. For example, if a household owns two sole proprietorships, we are unable to separate whether both survive over the years, or one survives and the other does not. As a result, we use the terminology "conditional on no exit" rather than "conditional on business survival" and we discuss participation in business endeavors or having business income, rather than having a particular business.

[14] The literature on tax evasion and income misreporting is extensive, and we will not attempt to review it here. For a more detailed review see Slemrod (2007).

determinant of income under-reporting.

Tax data poses additional limitations with respect to calculating income for sole proprietors and partnerships. We define business income as the residual or net income accruing to owners after all expenses have been paid, including any wage payments to the owners themselves. However, neither sole proprietors nor partnerships issue W2 forms to their owners. This means our business income measure for both of these types of businesses could also include any labor income the individual(s) earned from the business. In other words, for sole proprietors and partners, all returns to labor and capital are included in our measure of business income (net profits/losses). For S corporations, on the other hand, W2 forms are issued for employees. In fact, if an owner of an S corporation is also an employee of the firm, he is required to report "reasonable compensation" in the form of wages and salaries. Therefore, his labor income from the firm will not be included in our business income measure. However, S corporation owners have an incentive to take their compensation in forms other than wages in order to avoid payroll taxes on wage income.[15] We leave disentangling labor and capital income for sole proprietors and partnerships for future work and will treat all income that is reported as business income equally.

3.3 Panel of Individual Tax Returns

The main contribution of our paper is providing an analysis of business income risk that exploits the panel dimension of tax return data. To create this panel, we restrict each cross section of returns to include only observations from the random subsample beginning in 1987, where the primary filer had a social security number (SSN) ending in one of the two original four-digit combinations. As a result, we are able to track those tax units over the 23 years of our sample.[16] The panel is unbalanced, with some tax units exiting the sample due to death, emigration, or falling below the filing threshold, and others added due to immigration or becoming filers. Also, unlike the repeated cross-sections, the panel does not oversample high income taxpayers.

We impose additional income limitations to the panel data in order to exclude observations with little attachment to either the business or labor sector. The business income sample excludes

[15] Individual partners may also receive "guaranteed payments" from the partnership (line 10, Form 1065). These payments could be in lieu of wages, but they could also be used for payments to capital. Because guaranteed payments are the only form of partnership income subject to payroll taxes (for limited partners), partners may also want to minimize the amount they receive through such payments. However, these guaranteed payments do show up in our measure of partnership income.

[16] Note that changes in family circumstances can result in taxpayers being dropped from or added to the sample. For example, if a woman who has a sampled SSN four-digit ending marries a man who does not, and he is listed as the primary filer on the couple's joint return, then the woman will be dropped from the sample. In addition, if a couple divorces, only the primary filer with the four-digit SSN ending will be followed after the divorce. Conversely, if a single man with a SSN four-digit ending gets married, and if that man is listed as the primary filer on the couple's joint return, his wife will be added to the sample. For married filing jointly returns, which account for over 80% of total business income, the primary filer is overwhelmingly the husband. However, in our panel, we have taken care to track households whose composition did not change but the primary filer did, as well as households who misreported the primary filer's SSN.

households who never report business income outside of the (-$5,000,$5,000) interval for at least one year during the sample period. This restriction is similar to methodology used in Knittel, Nelson, DeBacker, Kitchen, Pearce and Prisinzano (2011). Our labor income sample imposes an analogous restriction, by excluding filers whose labor income always fell below one-quarter of the amount of income one would receive working full time at the minimum wage ($2,575). With these restrictions, the resulting panel spanning 1987-2009 has roughly 90,000 filer-year observations in the business income sample and roughly 287,000 filer-year observations in the labor income sample.

4 Cross Section Descriptive Statistics

Before moving to our panel analysis, we start by providing some descriptive statistics of how business income in the cross section evolves over time. The percentage of total tax returns claiming income from a business remains relatively stable over our sample period, increasing from about 23% in 1987 to 25% in 2009. Hence, about a quarter of all returns filed included some form of business income. The relative importance of business income in the composition of aggregate income, however, increases over our sample period. Business income as fraction of adjusted total income (ATI) increased from 5% in 1987 to 10% in 2005, though it has since dropped to 8% in the aftermath of the Great Recession. Figure 2 presents the time series for mean business income and mean labor income. In 1987, average business income per business return filed was about 27% of average labor income per labor return filed, but this ratio increased to 49% in 2005 and, despite a drop from 2005 to 2009, is still higher in 2009 than in 1987. This suggests that households receive non-negligible amounts of income from their participation in private businesses.

[Figure 2 about here.]

The importance of business income is even more pronounced for those households at the top of the distribution of adjusted total income (ATI). As Figures 3 and 4 show, the top 1% of households based on ATI account for about 41% of aggregate business income on average over our sample period, and for only 9% of aggregate labor income. This corroborates the evidence of Piketty and Saez (2003) who find that those at the top of the distribution receive a disproportionate share of income from "entrepreneurial" sources, like sole proprietorships, partnerships, and S corporations. Furthermore, the fraction of business income reported by households at the top 1% of ATI has increased much faster than the corresponding fraction of labor income; by 81% versus 31% over our sample period. Similar comparisons hold for the top 5% and the top 10% of high-ATI households.

Overall, households who report having business income also tended to report higher levels of ATI. In particular, conditional on reporting non-zero business income, roughly 52% of those households had ATI above $50,000. Furthermore, households with ATI greater than $100,000 account for about 71% of aggregate business profits and 33% of aggregate business losses.[17] Although me-

[17]These households also account for 39% of aggregate capital losses.

chanically there exists a positive correlation between high profits and high ATI, the interesting point to note here is that high-ATI households also account for roughly one third of all business losses. Similarly, we note that, on average, over 50% of returns reporting business losses larger than $100,000 are returns with positive ATI.

[Figure 3 about here.]

[Figure 4 about here.]

On average, top 1% of the business income distribution accounts for 68% of business income over the period 1987-2009. Thus, business income appears to be skewed towards the top end of the distribution. By contrast, the top 1% of the labor income distribution accounts for only about 11% of labor income. As an alternative measure of skewness, we look at the ratio of different percentiles of business income, conditional on reporting non-zero business income. The $p90/p50$ percentile ratio, which shows how much more unequal the top of the distribution is compared to the middle, is about 13 for business income. The analogous ratio on average for labor income is 3. In addition, conditional on non-zero business income, the ratio $(p50 - p10)/p50$ is 2.6 for business income and 0.74 for non-zero labor income.[18]

Conditional on reporting non-zero business (labor) income, Figure 5 (6) shows the 2009 distribution of business (labor) income. We note that in 2009 about 29% of all business returns filed include losses. This pattern holds true in all years highlighting the necessity of dealing with negative values for business income.

[Figure 5 about here.]

[Figure 6 about here.]

In sum, the cross-sectional data show that business income is of growing importance as source of total income and that it displays much more variance than does business income, especially in the tails of the distribution. We next analyze our panel data, which allow us to better decompose the cross section variation while accounting for idiosyncratic differences among business owners, thereby providing a better understanding of business income risk.

5 Measuring Business Income Risk in the Panel

To gain a better sense for the different risks associated with earning business income, we separately analyze the extensive and intensive margins of change. In particular, using panel data described in Section 3.3, we analyze a household's decision to exit from their business endeavors as well as

[18] The $p10$ is negative for business income, hence we use the ratio $(p50 - p10)/p50$ as an appropriate comparison.

business income fluctuations conditional on participation. For both analyses, we offer corresponding measures for labor income to serve as a comparison. The methodology we use is similar to that found in Carroll (1992) and Gourinchas and Parker (2002), who estimate the earnings (or wage) risk facing households, and in Davis, Haltiwanger, Jarmin, Krizan, Miranda, Nucci and Sandusky (2009) and Moskowitz and Vissing-Jørgensen (2002), who both use similar methodology. In particular, Moskowitz and Vissing-Jørgensen (2002) estimate the probability of going out of business and then characterize heterogeneity within entrepreneurial return conditional on staying in business.

5.1 Mobility

We start by constructing a one year transition matrix for levels of business income, shown in Table 1 where we split business income (measured in real 2005 dollars and conditioning on non-zero observations) into deciles labeled from 1 through 10. The rows (labeled 0 through 10) show the decile in which a household starts in any given year (rows 1-10) and business non-participation (row 0). The columns show the decile the household reaches at the end of the transition period (or the zero state). The numbers in the table denote probabilities that a household transitions into a particular decile (or exits) conditional on their starting in a particular decile (or not participating). The probabilities 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 present in the panel for the duration of the transition period.[19] Thus, for any business activity, the filer will be in deciles 1-10, but will be in the zero state if there were no business activity. For purposes of comparison, we also construct the corresponding one year transition matrix for labor income, shown in Table 2. [20] The zero state in this case denotes observations where labor income is equal to zero.

[Table 1 about here.]

[Table 2 about here.]

From the transition matrices, we see that business income tends to exhibit lower staying probabilities (lower probabilities on the main diagonal) than labor income, and therefore exhibits lower persistence. This information can be summarized by the immobility ratio, which is essentially the average of the diagonal elements of a transition matrix. The immobility ratio for business income

[19] Due to confidentiality, we cannot report any raw numbers coming from actual tax returns, such as minimum or maximum or income decile cutoffs. Here, we are reporting numbers rounded to the closest thousand. Averaging across time, the "blurred" bounds for the business income deciles 1 through 10, respectively, are: lower than -6,000, [-6,000; -1,000], [-1,000; 1000], [1,000; 4,000], [4,000; 8,000], [8,000; 12,000], [12,000; 19,000], [19,000; 32,000], [32,000; 74,000] higher than 74,000.

[20] The time averages for the blurred labor income deciles are: [0; 11,000], [11,000; 20,000], [20,000; 27,000], [27,000; 35,000], [35,000; 42,000], [42,000; 52,000], [52,000; 65,000], [65,000; 82,000], [82,000; 114,000], higher than 114,000.

is 0.38, whereas for labor income it is 0.57. Notably, at the top decile of the business income distribution the staying probability of 74% is higher than the staying probabilities in the rest of the business income distribution. This staying probability is still lower than the corresponding probability from the labor income distribution, which shows an 84% chance of staying in the top decile from one year to the next. We also see that conditional on leaving the starting business income decile, a household faces a 44% probability of moving to either of the two immediately adjacent deciles, whereas the corresponding probability is 69% for labor income. Therefore, movement across deciles that are far apart is more likely to occur in the business income than in the labor income distribution. Finally, households starting in the lowest decile of the business income distribution face a 9% probability of transitioning to decile eight or higher, whereas the corresponding probability is essentially zero for labor income. At the other end of the distribution, a household faces a 4% probability of transitioning from the very top to the very bottom of the business distribution within one year, whereas the corresponding probability is again about zero for labor income. The probability of exit from the business sector is much higher than that from the labor market. In particular, those with above median labor income have a 2% probability of exiting the labor force the next year while, those with above median business income face a 43% chance of exit. Overall, this evidence is consistent with a riskier business income than labor income process.[21]

5.2 Exit

The transition matrices show that the probability of exit from the business sector, defined as business income equal to zero, is higher than the corresponding probability for the labor sector. We next consider factors influencing a household's decision to exit its business endeavors. In general, a household shows up in the panel for an average for 16.5 years and has non-zero business income for about nine of them. Conditional on reporting non-zero business income, the average continuous streak of business losses is 2.9 years.[22] From the data, it appears that exit is likely related to negative business outcomes: years immediately before exit are more likely, on average, to be years with losses and a lack of significant capital gains realizations. In particular, in the year immediately before exit, the probability of losses is higher than in other years (36% versus 14%), and average business income is lower ($4,000 versus $28,000). In the year of exit, average capital gains are lower compared to other years ($1,000 vs. $14,000). One notable exception pertains to households who exit in years with very high business income. Specifically, those with business income above the 99 percentile of the business income distribution the year before exit, realize large amounts of capital gains in exit years.

[21] See Appendix A-4 for a description of results using alternative samples and transition periods. The results of these robustness tests remain qualitatively similar to those presented here.

[22] Recall from above, that the panel data used to generate results on business income are restricted to include only households who have ever reported having business income greater than $5,000 in absolute value. Similarly, results on labor income are generated from panel data that include only households who ever reported having labor income greater than $2,575.

Conditional on reporting non-zero business income in the previous year, the probability of exit in a given year is 10%. Using an indicator to denote years where a business reports zero income, we estimate the probability of exit using a probit model that controls for individual characteristics that may influence the likelihood of exit. Specifically, we run the following probit regression:

$$Pr(Exit)_{i,t} = \alpha + \beta X_{i,t} + \varepsilon_{i,t} \tag{1}$$

Included in the vector $X_{i,t}$ are dummy variables for age and year effects, marital status of the primary filer and number of children. We also control for prior-year income, including controls for the log of positive business income, the log of negative business income, and in certain specifications the log of labor income in the prior year. In addition, we control for a household's general business experience. We use two different proxies for experience since it could be the case that what matters for a business to succeed is not simply the experience acquired while running the current business, but rather the cumulative experience acquired by the household through all of its business endeavors over time, even if this experience was obtained through operating different or eventually unsuccessful businesses. Hence, we estimate the probability of exit as not only a function of the most recent unbroken streak of non-zero business income, but also as a function of the accumulated number of years that a household previously reporting having business income,

[Table 3 about here.]

Table 3 summarizes the average marginal effects from the probit models we estimate.[23]. The probability of exit is decreasing in business age and a household's cumulative business experience. Each additional year in this streak lowers the probability of exit by 1.5%.[24] Demographic controls, where the excluded group is 30-34 year old married filers whose primary filer is male, are all statistically significant and have the expected signs - older filers and those with more attachment to the work force (single filers and males) are more likely to remain in business. Including these controls, however, has essentially no impact on the average marginal effects of the income and business age variables (comparing Columns (1) and (3))

Turning to the effect of income realizations on the probability of exit, we find that those with higher income realizations in the previous year are more likely to remain in business. The average marginal effect of positive income suggests a 1% increase in business income increases the probability of staying in business by four percentage points. Negative income realizations have the opposite effect when included only as logs. A one percent increase in negative income increases the probability of exit by about four percentage points. When we include the log of labor income (Column (2)), we find that increases in labor income increase the probability of exit. A one percent

[23]Here we omit the coefficients on the year effects in the interest of space. We include the full set of results in Appendix $A-2$.

[24]The predicted probabilities are evaluated at the sample means of the control variables.

increase in labor income corresponds to an increase in the probability of exit from business ownership of approximately one percentage point. In sum, exits from businesses are heavily influenced by income realizations, with losses and small income realizations significantly increasing the likelihood of exit.[25]

5.3 Income Fluctuations

We next study general patterns in income fluctuations by first measuring each household's year to year fluctuations in labor or business income. We remove the part of income that is explained by observable characteristics (and thus does not contribute to risk) with the following regression:

$$y_t^i = g(\zeta_t; X_t^i) + \xi_t^i, \tag{2}$$

where y_t^i denotes income (in levels of real 2005 dollars), X_t^i is a vector of observable characteristics, $g(\cdot)$ is the part of income that is common to all individuals conditional on X_t^i, ζ_t is a vector of parameters (possibly including parameters that depend on calendar year t), and ξ_t^i is the unobservable error term. We regress each type of income (business and labor), separately for each year, on a full set of age dummies for the primary filer, the number of children in the household (up to ten), as well as the gender and the marital status of the primary filer. We note that the average R^2 of this regression is about 0.3 for labor income and about 0.03 for business income. Therefore, the predictive power of demographics is higher for labor income than for business income. That demographics have little predictive power for business income is also demonstrated by Moore (2004), who uses the SCF and hence an even richer set of demographics, compared to our tax data. However, DeBacker, Heim, Panousi, Ramnath and Vidangos (2013a) show that predictable variation over the lifecycle is importance in explaining the total variation in labor income, thus, we proceed using residuals derived from Equation 2.

Here we focus on year-to-year fluctuations in residual income, conditional on operating a business and on labor participation. Thus, we drop observations that fall below $5,000 in absolute value for business income and below $2,575 for labor income to include only those who are active participants in their corresponding sector and accordingly, percent changes reflect movement around the income distribution rather than movements between participation and non-participation. We then estimate Equation 2 using the new restricted data to remove the portion of income explained by observable characteristics. Letting e_t denote either residual business or labor income, we define the one year percent change, g_t, between year t and year $t-1$, as follows:

$$g_t = \frac{\xi_t - \xi_{t-1}}{abs(\xi_{t-1})}, \tag{3}$$

[25] Appendix A-3 shows the results of probit models run on an alternative sample, where we include all individuals who ever have non-zero business income. The results are similar.

where $abs(\xi_{t-1})$ is the absolute value of residual income in period $t-1$. We use the absolute value of residual income in the denominator to address the fact that business income can take on negative values. For example, without our modification, a change from -\$5,000 to -\$10,000 would register as a positive percent change even though there was a clear decrease.

Table 4 shows the ratio of different percentiles of the distribution of percent changes for residual business income, over the corresponding percentile for residual labor income. For both residual business and labor incomes, the 50th percentile (median) of percent changes is around zero. However, all other percentiles are 1.5-2 times more dispersed for residual business income, compared to the corresponding percentiles for residual labor income. For example, on average over our sample, the ratio of the 5th percentile of the change in residual business income to the 5th percentile of the change in residual labor income is 1.7, the ratio of the 40th percentiles is 2.2, the ratio of the 60th percentiles is 1.7, and the ratio of the 95th percentiles is 1.8. This shows that the distribution of percent changes is more dispersed for residual business income than for residual labor income, demonstrating that business income risk appears to be greater than labor income risk.[26]

[Table 4 about here.]

Figure 7 presents the (pooled) distribution of one year percent changes in income residuals, by size of percent change. The horizontal axis shows the size of the percent changes where the bin termed "20% to 30%" indicates that residual income increased between 20% and 30% over the period and the bin termed "-20% to -30%" indicates that residual income fell between 20% and 30% over the period. All bins are 10 percentage points wide except for the last bins on the right and left-hand sides. The last bin on the right groups all observations for which residual income increased by more than 100%, while the last bin on the left groups all observations for which residual income decreased by more than 100%.[27] The blue bars indicate residual business income and the red bars indicate residual labor income. The vertical axis shows, for each type of income, the fraction of all percent change observations in each size-bin.

[Figure 7 about here.]

The overall picture that emerges is striking. First, the distribution of percent changes for residual business income has less mass in the middle. In other words, the unexplained portion of business income is characterized by fewer small increases or small declines, compared to that of labor income. For example, about 11% of all changes in residual business income were increases smaller than 10% or decreases smaller than 10%, compared to about 23% for residual labor income. Second, the distribution of percent changes in residual business income has thicker tails than residual labor income. This is demonstrated by the fact that, away from the origin, the blue bars are taller than

[26]We note that all of our qualitative results remain if we use levels of income rather than residuals of income.

[27]Declines of more than 100% are possible for labor income, because the analysis here is in terms of residuals.

the red bars, and they also decline more slowly. For example, about 33% of all changes in residual business income were increases greater than 100% or decreases greater than 100%, compared to about 20% for labor income.

Analysis of the panel data confirm cross-sectional results regarding the variance of business income relative to labor income. They also provide further evidence showing that the larger variance of business income is not only driven by the probability of exit, but also by the distribution of business income conditional on participation. In addition, the panel data confirm that even when controlling for observables, with-in household fluctuations in business income exceed those from labor income.

6 Estimating Business Income Risk

Thus far, we have utilized the panel dimension of our data to consider the transition of households out of the non-corporate business sector and their income fluctuations within that sector. We find that exit is more likely to occur from the business sector than from the labor sector. Also, year-to-year fluctuations in business income tend to be more volatile than labor income and business income has higher tail risk. While related to the cross-sectional moments presented in Section 4, these panel estimates better capture the true risks facing individual business owners in that they account for heterogeneity. In particular, increases in cross-sectional variance of business income that are driven by heterogeneity do not imply increased risk faced by individual business owners, as owners may have knowledge of their specific skills and thus can forecast and anticipate the income they might derive therefrom. On the other hand, the residual risk (i.e., the income variation not accounted for by these household fixed effects), is unanticipated by the agent and thus represents risk in the true sense. In this section, we use error components models of income dynamics to further decompose income fluctuations into the component that is due to heterogeneity in the productivity of business owners or labor providers and that which is due to the residual risk these owners face .

To estimate the model, we continue to drop observations below our thresholds of business (and labor) income and study the risk of business endeavors conditional on participation. Therefore, our benchmark considers the risk implied by all business spells for a given household over time, even if they are not continuous. By spell, we mean a continuous streak of non-zero business income and by continuous, we mean not interrupted by zeros. Intuitively, our measure of business risk captures risk from a lifetime perspective. For example, a first business may fail, at which point the household exits business endeavors and lets some time elapse before it re-enters with another business that may (or may not) be more successful.

6.1 Modeling income dynamics

We use error-component models to decompose the income fluctuations of business owners and, for comparison, labor providers. The income process is modeled using the following regression equation described in Section 5.3:

$$y_t^{\theta,i} = g(\zeta_t; X_t^i) + \xi_t^i \ , \tag{4}$$

where $y_t^{\theta,i}$ is transformed income (we discuss the income transformations in the following section). We impose additional structure on the error term, $\xi_{a,t}^i$, from equation (10) using the following model:

$$\xi_{a,t}^i \ = \lambda_t \cdot (\alpha^i + r_{a,t}^i) + z_{a,t}^i \tag{5}$$

where $r_{a,t}^i$ and $z_{a,t}^i$ are given by

$$r_{a,t}^i \ = \ r_{a-1,t-1}^i \ + \ \epsilon_{a,t}^i \tag{6}$$

$$z_{a,t}^i \ = \ \rho z_{a-1,t-1}^i \ + \ \pi_t \eta_{a,t}^i \tag{7}$$

$$\alpha^i \sim \ iid \ (0, \ \sigma_\alpha^2), \ \epsilon_{a,t}^i \sim \ iid \ (0, \ \sigma_r^2). \ \eta_{a,t}^i \sim \ iid \ (0, \ \sigma_z^2), \tag{8}$$

This framework is desirable for its flexibility in that the error term comprises both a permanent component and a transitory component. The permanent component consists of an individual-specific, time-invariant fixed effect, α_i, and a random-walk, r_t^i, each of which are pre-multiplied by year-specific factor loadings, λ_t, to allow for variation in the permanent component's relative importance by calendar year. For example, λ_t may reflect changes in the skill premium by year. Importantly, by including a fixed effect, we are able to remove the fraction of the total risk that business owners face that is due to heterogeneity. The transitory component, z_t^i, is specified as an AR(1) process that includes transitory innovations, $\eta_{a,t}^i$, which are multiplied by year-specific factor loadings, π_t. This allows the variance of the innovations, and hence the relative importance of the transitory component to vary by calendar year. Such variation in the transitory component of income may be driven by the business cycle or structural changes in the labor or business income markets. The persistence of the transitory AR(1) component is given by the parameter ρ. When $\rho = 0$, the transitory part becomes essentially white noise (with time-varying variance).

We estimate the full model above as well as restricted versions. The restrictions we impose are standard in the literature and parsimoniously capture the main features of our business income data. For example, the lifecycle variance profile in our business income data is slightly concave which lends to the restriction that the AR(1) process has persistence $\rho < 1$. On the other hand, a random walk process would imply a linear lifecycle variance profile, but a model that includes both a random walk and an AR(1) component implies a concave variance profile. Below is a list of the models we estimate.

(i) Fixed effect, random walk, AR(1) (FE_RW_AR):

(ii) Fixed effect with AR(1) (FE_AR): $\sigma_r^2 = 0$

(iii) Fixed effect, random walk, white noise (FE_RW_WN): $\rho = 0$

(iv) Random walk with white noise (RW_WN): $\sigma_\alpha^2 = \rho = 0$

The model parameters, namely the persistence of the AR(1) process and the variances of the shocks, are estimated using a minimum distance estimator that matches the model's theoretical variances and autocovariances to their empirical counterparts. In particular, each model above implies a specific parametric form for the variance and autocovariance of residual income, for each calendar year t and each lead k. These theoretical variances and autocovariances, denoted by $cov(t, k)$, are functions of the model parameters, for $t = 1987, ..., 2009$. Hence, we estimate the model parameters by minimizing the distance between the theoretical variances and autocovariances implied by the model, and their empirical counterparts, which we compute from our panel data for $t = 1987, ..., 2009$ and $k = 0, ..., 22$. Our minimum distance estimator uses a diagonal matrix as the weighting matrix, with weights equal to the inverse of the number of observations used to compute each empirical statistical moment.[28,29]

6.1.1 Accounting for negative income

Before proceeding to the model estimation, we must first impose some additional structure to our data. The use of statistical models often requires imposing some structure on the data. For example, in the labor or earnings dynamics literature, such structure is attained by excluding very low income observations and by logging the data to reduce the extreme skewness of the labor income distribution. The distribution of business income is skewed; however, the typical log transformation is not appropriate since business income can take on negative values. One transformation that

[28]We do not use an optimal weighting matrix for reasons discussed in Altonji and Segal (1996).
[29]Appendix $A-6$ provides more detail on the estimation of the error-components models.

addresses negative values is the inverse hyperbolic sine (IHS).[30] Letting y denote income, the IHS transformation of the y-observations is given by:

$$y^{\theta} = \frac{1}{\theta} sinh^{-1}(\theta y) = \frac{1}{\theta} log(\theta y + \sqrt{1+\theta^2 y^2}), \ , \tag{9}$$

where θ is a location parameter. As shown in Figure 8, 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. The location parameter θ may vary across datasets, adjusting the degree to which large values are down-weighted and of appropriately handling the interval of observations around zero.[31] The location parameter θ is chosen so as to best fit each particular data set, in the sense of estimating the degree to which large values are down-weighted and of appropriately handling the interval of observations around zero. Note from (9) that the IHS transformation depends on the units for y.[32]

[Figure 8 about here.]

Hence, the IHS operates like a generalized Box-Cox transformation, both in terms of its location parameter, θ, and in terms of how that parameter is estimated. Here, we present the intuition for the estimation of θ, while the technical details are delegated to the Appendix A-8. The methodology is similar to that in the original Box and Cox paper. Pick any given θ and transform business income, y, based on (9), to get transformed income, y^{θ}. Then, perform our standard first stage regression from (2) as follows:

$$y_t^{\theta,i} = g(\zeta_t; X_t^i) + \xi_t^i \ , \tag{10}$$

This is possible because, for a constant θ, the log-likelihood of (10) is the same, up to a constant, as in the standard OLS estimation in (2), and hence we can proceed with OLS to estimate the regressor coefficients and the variance of the error term in (10). Repeat this process for a grid

[30]The IHS can actually also address zero observations, contrary to the log. This might suggest that there is no need to exclude the zeros from our benchmark analysis, but that we could instead leave them in and have a continuous process of business income, interspersed with zeros. The problem with this approach turns out to be a technical one. Specifically, when the income distribution has a large mass at any given point (here at zero), then neither the log nor the IHS transformation can be successfully applied to the data.

[31]For more discussion on this point see Pence (2006).

[32]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.

of (positive) θ-values and search for the value of θ that maximizes the log likelihood. In general, there is no reason for the estimated θ to be the same for different samples. We transform labor income using the log function and estimate the θ for the IHS transformation of business income. We find the θ for business income to be 0.7. To make comparisons, we untransform the model results back to levels of income. The first stage regression performs two roles. One, it provides the framework for estimating the location parameter, θ. Two, it makes the business income treatment comparable to that for labor income, in that it removes predictable variation in income over the lifecycle. As already mentioned, this variation has been shown to matter much for labor income, but its predictability means that it is not a source of risk.

6.2 Results

Tables 5 and 6 present the parameter estimates for models (i)-(iv) for business income and labor income, respectively.[33] Note that models (i)-(iii), which allow for household fixed effects, fit the data much better than model (iv), as measured by the root-mean-square-error. This suggests that heterogeneity is an important source of the cross-sectional variation in income. Since the FE_RW_AR fits the data best, we use this as our baseline case in the analysis below. Because the transformations of business and labor income are not the same, one cannot directly compare the parameter estimates from both income processes by looking at the parameter estimates here.

[Table 5 about here.]

[Table 6 about here.]

For a more intuitive interpretation of our variance estimate, we calculate the effect of a one standard deviation shock, based on the model, on the levels of business and labor income thereby undoing the IHS and log transformations of the data. We then compare this to a benchmark case where the household's primary filer is a 35 year old married male with two kids.[34] For example, for the FE_RW_AR model, we take a one standard deviation shock to income in year t (σ_t) to be the sum of the shocks to the permanent and the transitory components, weighted by their respective factor loadings, i.e. $\sigma_t = (\sigma_r^2\lambda_t^2 + \sigma_z^2\pi_t^2)^{1/2}$. Then, we find the level of income for our benchmark average household without the shock and with, for example, a positive σ_t-shock.[35] Next, for each year t, we calculate the percent change implied by these two income levels, using Equation (3). We then average across years, and we compare the resulting average percent changes for business and labor income. We interpret a bigger percent change to the level of income resulting from a model-estimated shock to indicate higher income risk.

[33] We omit the estimates of the year factor loadings for brevity, but the full table of results in available in Appendix $A-9$.

[34] The qualitative results are similar when we choose a different household as our measure of comparison.

[35] The results in terms of the comparison between business and labor income are also qualitatively similar when we examine permanent or transitory shocks to income separately.

Table 7 shows the comparison of risk between business and labor income. The table includes the implied percent change to the level of business income and to the level of labor income resulting from a positive one standard deviation shock estimated from each one of the models (i)-(iv). For example, an estimated positive one standard deviation shock in the fixed effect, random walk, AR(1) model increases the level of business income of our benchmark average household by 179%, whereas it increases the level of labor income of that same household by 52%. The ratio of these percent changes is approximately 3.4, i.e. business income is about three and a half times riskier than labor income. The table shows that, depending on the model, this ratio ranges between 2.7 and 3.6, implying that business income is about 3 to 4 times riskier than labor income.

[Table 7 about here.]

6.3 Separating Risk from Heterogeneity

We use our results from the error-components models to decompose the variance of income into two components: heterogeneity across households and risk facing a particular household. Specifically, we use the theoretical covariance calculated as a function of the model parameters:

$$cov(\xi^i_{a,t}, \xi^i_{a+k,t+k}) = \lambda_t \cdot \lambda_{t+k} \cdot \left(\sigma^2_\alpha + (\sigma^2_r \cdot a)\right) + \rho^k var(z^i_t). \tag{11}$$

We then find the cross-sectional variance for year t by setting $k = 0$ and calculate a weighted average over a using the empirical distribution of primary filer age in year t. The cross-sectional variance is then bifurcated into heterogeneity, $\lambda^2_t \sigma^2_\alpha$ and risk, $\lambda^2_t \left(\sigma^2_r \cdot a\right) + \rho^k var(z^i_t)$.

Figure 9 shows the total cross sectional variance over the sample period and the corresponding fractions due to risk and heterogeneity as implied by the most flexible model specification, FE_RW_AR. On average, heterogeneity accounts for about 23% of the cross sectional variance in business income. This fraction has remained relatively stable over time. The remaining 77% of the cross-sectional variance is due to the residual risk faced by households. Contrast this with the variance of labor income in Figure 10. The variance in labor income is much lower and a larger share of it is due to heterogeneity across households rather than within household risk. In particular, heterogeneity accounts for 44% of the variance in labor income, though this share has fluctuated over the sample period, peaking in the late 1990s. That heterogeneity is important for explaining the dispersion of labor income is consistent with the work on Meghir and Pistaferri (2004) and Guvenen (2009) and others.

These results underscore the risk that business owners face. Business income is more dispersed than labor income and less of the dispersion is accounted for by heterogeneity across households. While heterogeneity may be something that can be accounted for when entering the business sector, unanticipated income shocks represent the bulk of the risk that business owners face. Thus, business owners face more income risk in the true sense.

[Figure 9 about here.]

[Figure 10 about here.]

As a robustness test, we calculate the fractions of the cross sectional variance due to risk and heterogeneity for models (ii) and (iii) (the remaining models) using fixed effects. Table 8 presents the results. Across model specifications, the fraction of the total variance that is due to heterogeneity is relatively constant. For business income, this fraction ranges from 16% to 31%. The same fraction falls between 44% and 47% for labor income.

[Table 8 about here.]

6.4 Decomposing Risk

We can further decompose risk into it's permanent and transitory components, as in DeBacker et al. (2013a) and Lochner and Shin (2014). Such a decomposition can tell us whether permanent or temporary shocks to income are driving the cross-sectional variance in income and which components are responsible for any trends. Knowing whether temporary or permanent shocks are relatively more important has direct implications for policy and the welfare analysis of the dispersion of income.

Figure 11 shows the decomposition of the cross-sectional variance of business income and Figure 12 shows the same for labor income. Transitory risk is that driven by the AR(1) process, while permanent risk is that driven by the random walk component to the income process. Total risk is the sum of the permanent and transitory components. Transitory risk accounts for about two-thirds of the total risk faced by business owners The shocks to the permanent component of income account for a similar share of the total variance as heterogeneity, about 25%. In contrast, the shocks to the permanent component of income account for a much smaller share of the total labor income variation. Heterogeneity represents the largest contributor to the cross-sectional variation in labor income, followed by shocks to the transitory component of labor income. As DeBacker et al. (2013a) find, we also provide evidence that the increase in the dispersion of labor income from the mid-1980's through present is due to more permanent factors. Both the permanent component of income and heterogeneity play increasing roles in describing the overall cross-sectional variance of labor income, while the contribution of shocks to the variance of the transitory component remain relatively constant.

[Figure 11 about here.]

[Figure 12 about here.]

7 The Business Ownership Puzzle

We have provided extensive evidence of the risk inherent in business ownership relative to labor income. Figure 2 also documented that the average labor income among employees is greater than the average business income among business owners. This leads one to question why some households pursue business ownership rather than employment. Similar facts and questions about the choice of business ownership are documented by Carrington et al. (1996), Hamilton (2000), and Kawaguchi (2002). Hamilton (2000) finds that the self-employed earn less in median than the paid-employed, have income concentrated in the upper and lower tails of the income distribution, and that all but the most successful would have would have earned more if they had switched into paid employment. This puzzle is related and similar to that documented by Moskowitz and Vissing-Jørgensen (2002), regarding the puzzling investment of equity into privately held businesses rather than a diversified portfolio. Moskowitz and Vissing-Jørgensen (2002) explain some of this puzzle by considering preferences for skewness, irrational expectations, and non-pecuniary benefits. Hamilton (2000) and Hurst and Pugsley (2011) suggest that non-pecuniary benefits to self-employment drive decisions into business ownership. de Meza and Southey (1996), and others, argue that the self-employed do not have accurate expectations about their earnings prospects, which leads them to choose self-employment. We don't deny that these may be important factors driving the decisions to take on business ownership, but given the income processes we estimate, we show that the business ownership puzzle may be at least partly explained by selection effects and the pecuniary benefits to business ownership. That is, while the simple descriptive statistics show the returns to business ownership to be less attractive than labor provision, this is not the appropriate comparison. A more meaningful comparison asks how a business owner would fare if she were instead were a labor provider. That is, to compare the returns of business ownership to the returns of labor provision *for business owners*. The challenge then, is creating an appropriate counterfactual. We do so by calibrating and simulating a life-cycle model that allows for endogenous labor/business ownership participation. We calibrate the model using our previous results from the error components models to construct a counterfactual of business owners participating in the labor income process and evaluate the contribution of selection to the business ownership puzzle.

7.1 The Model

The household's objective is to maximize expected, discounted, lifetime utility from consumption:

$$\max_{\{c_t\}_{t=0}^{T}} E_0 \sum_{t=0}^{T} \beta^t u(c_t) \tag{12}$$

We take β to be the household's (constant) discount factor and assume that period utility is given by a constant relative risk aversion function, $u(c) = \frac{c^{1-\gamma}}{1-\gamma}$. Expectations are taking over future income realizations, y_t^b for business income and y_t^l for labor income. The individual's period budget constraint is given by $c_t = a_t + y_t - \frac{a_{t+1}}{R_t}$, where c_t is consumption, a_t are asset holdings, and R_t is the gross real interest rate. Agents are subject to a constraint on borrowing, thus we restrict $a_t \geq 0 \; \forall t$. We further assume that $R_t < \frac{1}{\beta}$ for all t.

Earned income evolves according to the following process:

$$y_t^{\theta,i,j} = \bar{y}^j + \xi_t^{i,j}, \;\; j = b, l \tag{13}$$

As above, the term $y_t^{\theta,i,j}$ represents transformed income. We use the IHS transformation with $\theta = 0.7$ for business income and the log of labor income. The superscript j can take on the value b or l to represent business or labor income, respectively and $\bar{y}^j$ is the part of income that is common to all households in the model. For parsimony of the state space, we assume that the error term in the Equation 13 follows an AR(1) process. Specifically, $\xi_t^{i,j} = \alpha^{i,j} + z_t^{i,j}$, where z_t^i evolves according to $z_t^{i,j} = \rho z_{t-1}^{i,j} + \eta_t^{i,j}$, $\eta_t^{i,j} \sim N(0, \sigma_{j,z}^2)$. Both business income and labor income follow a process described by Equation 13, though the parameters of the processes differ.

For simplicity, households in the model choose to earn labor or business income, but not both. Our approach to modeling households in this regard is similar Cagetti and De Nardi (2006). When continuing in the business income or labor income sector, their income process is described as above. In the case of a transition between sectors, households draw $\xi_t^{i,j}$ from an unconditional distribution.

Given these assumptions, the household has five state variables at any point in time: business owner/employee status, assets, permanent labor income, permanent business income, and transitory income in the sector she is working in. Written recursively, the problem of a household who is currently realizing labor income and has knowledge of her permanent income in both sectors is:

$$V_t^l(a_t, \alpha^l, \alpha^b, z_t^l) = \max_{c_t} u(c_t) + \beta \max[E_{z_{t+1}^l | z_t^l} V_{t+1}^l(a_{t+1}, \alpha^l, \alpha^b, z_{t+1}^l), E_{z_{t+1}^b} V_{t+1}^b(a_{t+1}, \alpha^l, \alpha^b, z_{t+1}^b)]. \tag{14}$$

Similarly, the value of a household currently realizing business income is:

$$V_t^b(a_t, \alpha^l, \alpha^b, z_t^b) = \max_{c_t} u(c_t) + \beta \max[E_{z_{t+1}^b | z_t^b} V_{t+1}^b(a_{t+1}, \alpha^l, \alpha^b, z_{t+1}^b), E_{z_{t+1}^l} V_{t+1}^l(a_{t+1}, \alpha^l, \alpha^b, z_{t+1}^l)]. \tag{15}$$

The income processes in our model are calibrated using the estimation results from the FE_AR1 model described in Section 6. It is important to note that the income processes we calibrate the model to are the income processes *conditional on participation*. The parameters for these income process are likely to be different for the population as a whole than they are for those who chose to participate. Furthermore, there may be a non-zero covariance between these income processes. For example, households with a high business income fixed effect may be generally skilled and have a high labor income fixed effect as well. DeBacker, Panousi and Ramnath (2013b) estimate such joint income processes using error components models and control for selection effects. We abstract from the joint income process here to keep the model parsimonious. While we do not estimate a selection model to get population representative income processes, we do note that if there are selection effects at work in the data, they would make the model's counterfactuals stronger. For example, if it is the case that those with high proclivity for business ownership and low returns as labor providers tend to be business owners, then the labor income process for those who are business owners in the data would show lower returns to labor provision than the model calibration (which is based off the sample of those who receive labor income).[36]

The model is solved recursively over a finite state space, where we solve for the optimal consumption in the last period, then work backwards in time to solve for the decision rules (choice of employment/business ownership and consumption/savings) at each period given the household's state. For the discretized state space used in the numerical solution, the AR(1) process is approximated based on methods described in Rouwenhorst (1995). When transitioning between sectors, we assume that the household draws a $z_t^j \sim N(0, \sigma_{j,z}^2)$ for $j = l, b$. To set $\bar{y}^l$ and $\bar{y}^b$, we use the coefficients from the first stage regression shown in Equation 10. These parameters are set to the values for a married household two children. Consistent with Attanasio, Banks, Meghir and Weber (1999), we choose a value of 1.5 for the coefficient of relative risk aversion, γ. Following Carroll (2009), we set $\beta = 0.96$ and the risk free real interest rate to 4%. Each model period corresponds to one calendar year and we model households with primary filers aged 30 to 60. After solving our model, we simulate the decisions of 100,000 households for $T = 31$ periods each (corresponding to ages 30-60). From these simulations, we calculate the first and second moments from the endogenous income distributions.

A principle reason for separating risk and heterogeneity as we do is to separate variation in income that is household specific, and therefore may be known by the economic agents, and that which is within household variation. But heterogeneity may contribute to the risks faced by business owners to the extent that there is uncertainty over their time-invariant skills. Such uncertainty will also affect selection into the labor market or business ownership. As an example of individuals learning about the income process consider Guvenen (2007) who models individuals who update their priors regarding their own labor income process over time in a Bayesian framework. One could imagine similar learning about the income processes facing business owners. Thus, we also present results of relaxing the assumption that each household has a full understanding of their permanent income (i.e., heterogeneity) to observe how heterogeneity impacts the distribution of income under no information.

In our model, households' decisions to switch between sectors will be driven by their information about, and realization of, their fixed effect in each sector and the size and persistence of the transitory income shock. For example, a household with a high fixed effect for business income

[36]Preliminary results by DeBacker et al. (2013b) show that such a negative correlation between income across sectors is a feature of the data. These results are previewed in Appendix A-10.

may select into business ownership, but if this household receives a low income shock that has strong persistence, the household may transition from business ownership to employment. In this way the model corresponds to the transitions into and out of business ownership that we document in Section 5.2.

7.2 Baseline Model

After solving the model for the household's decision rules, we simulate distributions of labor and business income. The means and variances of the income for all agents and for those who are ever business owners are reported in the first two lines of Table 9. We next simulate the model without endogenous choice of labor supply or business ownership. The results of exogenous participation decisions are reported on line three. The results reported are for those who chose business ownership at some point during their life under endogenous participation decisions (that is, the sample of agents, is the same in lines two and three). Comparing lines two and three of Table 9 shows the counterfactual of those who opt into business income when the choice is endogenous instead working as employees for their working life. Line four relates the percent change in the mean and variance in average annual income between the lifetime income of business owners when they can choose business ownership and when they are predetermined to be employees (lines two and three). We see that, though business owners have a higher variance in annual income when they are able to choose business ownership, they also have a higher mean income. Thus, the business ownership puzzle can be at least partly explained by selection effects. Rational, forward-looking agents choose business ownership because it gives them higher average annual incomes; enough to compensate for the additional risk.[37] As in the data, business owners on average have less income than employees, but the counterfactual is even worse for these business owners as their earnings as employees would have been lower than if they are able to transition between business ownership and employment.

[Table 9 about here.]

7.3 No Information Model

Because of the importance of selection effects, it's important to consider changes in the information set of agents. For example, what happens to the decisions of agents to enter self-employment if they have more uncertainty about their skill as a business owner? To consider the effects of knowledge of one's own fixed effect on career decisions and the subsequent impact on the business ownership puzzle, we modify the baseline households' problems above to consider changes in information regarding their permanent incomes. In particular, we simulate the model under an information treatment, which we call the "no information case", where agents do not know their fixed effect

[37]Note that when the model is simulated under higher levels of risk aversion, the quantitative results differ somewhat, but are qualitatively similar at least under reasonable parameterizations.

as either employees or business owners. Our baseline case from above is the "full information case".[38] In the full information case, agents know their permanent income as a laborer and as a business owner (i.e., they know both α_i^l and α_i^b). In the no information case, households do not know their permanent income in either sector (α_i^l or α_i^b). These changes are introduced through the expectations operator that affects the continuation values. In the case of no information, the Bellman equation for a labor provider becomes:

$$V_t^l(a_t, \alpha^l, \alpha^b, z_t^l) = \max_{c_t} u(c_t) + \beta \max[E_{\alpha^l, \alpha^b, z_{t+1}^l | z_t^l} V_{t+1}^l(a_{t+1}, \alpha^l, \alpha^b, z_{t+1}^l), E_{\alpha^l, \alpha^b, z_{t+1}^b} V_{t+1}^b(a_{t+1}, \alpha^l, \alpha^b, z_{t+1}^b)] \tag{16}$$

The equations for a business owner are written similarly. Households consume out of today's known income realization, but are uncertain of their future income due to uncertainty over (potentially) both the permanent and transitory components of income.

Table 10 shows the results for the no information case. Note that business owners do better when they have the choice of business ownership as compared to being employees. That is, even with no information about the fixed effect by sector, agents are able to select into the higher income role as they are hit with persistent shocks.[39] Furthermore, one can compare across Tables 9 and 10 and see that as information on the agents' fixed effects increases, selection effects are stronger and thus incomes are higher in the endogenous participation case (summarized by the percentage increase in income in line four, column one). In short, heterogeneity in permanent income across sectors can help explain the business ownership puzzle. As information on these fixed effects increases, business ownership becomes less puzzling as business ownership is driven by relative differences in individuals income across sectors.

[Table 10 about here.]

8 Conclusion

Our paper is the first to document and quantify the properties of income risk from privately held businesses in the US. Using a panel of individual income tax returns from 1987-2009, we find that business income is generally riskier than labor income. This increased risk is attributed not only to higher probability of household exit from business endeavors, but also to higher income fluctuations, conditional on no exit. We show that business income is less persistent, and is characterized by

[38] Note that a model where agent's learn their income process such as that in Guvenen (2007) becomes extremely computationally burdensome when agents have sector specific income processes and have the choice to switch sectors. Thus, we simply our analysis and only consider the three cases outlined, where agents either have knowledge of their income processes or not. These results still give useful comparative statics for the effects of information on the distribution of income, which is what we are interested in.

[39] In the no information case, and with perfectly transitory shocks, agents would be no better off with endogenous participation - they would always choose employment.

higher probabilities of extreme upward and downward movements, compared to labor income. The distribution of business income risks is more dispersed, with less mass in the middle and more mass in the tails, compared to labor income.

Results from estimating error components models of income dynamics also indicate that business income risk is substantially higher than labor income risk. We find that a positive one standard deviation model-estimated shock leads to a percentage increase in the level of business income that is about 3-4 times larger than the corresponding increase for labor income. Furthermore whereas heterogeneity accounts for about 45% of the cross-sectional variance in labor income, we find that about 25% of the cross-sectional variance in business income is due to individual heterogeneity and that this is relatively stable over our sample period. This suggests that a large portion of the variance found in business income is attributed to risk rather than differences in individual-specific ability.

Given the substantial risk involved in earning business income relative to labor income, and lower mean income, we address the puzzle of why individuals choose to pursue business endeavors. Using a calibrated life-cycle model that includes endogenous business participation, we show that selection into business ownership occurs by those most likely to realize a better return from business ownership. A life-cycle model using the income processes we uncover from administrative data, is therefore able to account for the apparent business ownership puzzle, even with rational, forward-looking agents.

References

Albanesi, Stefania, "Selection and Incentives: Optimal Taxation with Occupational Choice and Private Information," Technical Report 2012.

Altonji, Joseph G and Lewis M Segal, "Small-Sample Bias in GMM Estimation of Covariance Structures," *Journal of Business & Economic Statistics*, July 1996, *14* (3), 353–66.

Andreoni, James, Brian Erard, and Jonathan Feinstein, "Tax Compliance," *Journal of Economic Literature*, June 1998, *36* (2), 818–860.

Angeletos, George-Marios, "Uninsured Idiosyncratic Investment Risk and Aggregate Saving," *Review of Economic Dynamics*, January 2007, *10* (1), 1–30.

___ **and Laurent-Emmanuel Calvet**, "Idiosyncratic production risk, growth and the business cycle," *Journal of Monetary Economics*, September 2006, *53* (6), 1095–1115.

___ **and Vasia Panousi**, "Revisiting the supply side effects of government spending," *Journal of Monetary Economics*, March 2009, *56* (2), 137–153.

___ **and** ___ , "Financial integration, entrepreneurial risk and global dynamics," *Journal of Economic Theory*, 2011, *146* (3), 863 – 896. Incompleteness and Uncertainty in Economics.

Attanasio, Orazio, James Banks, Costas Meghir, and Guglielmo Weber, "Humps and Bumps in Lifetime Consumption," *Journal of Business & Economic Statistics*, January 1999, *17* (1), 22–35.

Benhabib, Jess and Shenghao Zhu, "Age, Luck, and Inheritance," NBER Working Papers 14128, National Bureau of Economic Research, Inc June 2008.

Bitler, Marianne P., Tobias J. Moskowitz, and Annette Vissing-Jørgensen, "Testing Agency Theory with Entrepreneur Effort and Wealth," *Journal of Finance*, 04 2005, *60* (2), 539–576.

Blanchflower, David G and Andrew J Oswald, "What Makes an Entrepreneur?," *Journal of Labor Economics*, January 1998, *16* (1), 26–60.

Cagetti, Marco and Mariacristina De Nardi, "Entrepreneurship, Frictions, and Wealth," *Journal of Political Economy*, October 2006, *114* (5), 835–870.

Carrington, William J, Kristin McCue, and Brooks Pierce, "The Role of Employer-Employee Interactions in Labor Market Cycles: Evidence from the Self-Employed," *Journal of Labor Economics*, October 1996, *14* (4), 571–602.

Carroll, Christopher D., "The Buffer-Stock Theory of Saving: Some Macroeconomic Evidence," *Brookings Papers on Economic Activity*, 1992, *23* (2), 61–156.

____ , "Portfolios of the Rich," NBER Working Papers 7826, National Bureau of Economic Research, Inc August 2000.

____ , "Precautionary saving and the marginal propensity to consume out of permanent income," *Journal of Monetary Economics*, September 2009, *56* (6), 780–790.

Cooper, Michael and Matthew Knittel, "The Implications of Tax Asymmetry for U.S. Corporations," *National Tax Journal*, March 2010, *63* (1), 33–61.

Davis, Steven J., John Haltiwanger, Ron Jarmin, and Javier Miranda, "Volatility and Dispersion in Business Growth Rates: Publicly Traded versus Privately Held Firms," in "NBER Macroeconomics Annual 2006, Volume 21" NBER Chapters, National Bureau of Economic Research, Inc, December 2007, pp. 107–180.

____ , ____ , Ronald S. Jarmin, C.J. Krizan, Javier Miranda, Alfred Nucci, and Kristin Sandusky, "Measuring the Dynamics of Young and Small Businesses: Integrating the Employer and Nonemployer Universes," in "Producer Dynamics: New Evidence from Micro Data" NBER Chapters, National Bureau of Economic Research, Inc, December 2009, pp. 329–366.

de Meza, David and Clive Southey, "The Borrower's Curse: Optimism, Finance and Entrepreneurship," *Economic Journal*, March 1996, *106* (435), 375–86.

De Nardi, Mariacristina, Phil Doctor, and Spencer D. Krane, "Evidence on entrepreneurs in the United States: data from the 1989–2004 survey of consumer finances," *Economic Perspectives*, 2007, (Q IV), 18–36.

DeBacker, Jason, Bradley Heim, Vasia Panousi, Shanthi Ramnath, and Ivan Vidangos, "Rising Inequality: Transitory or Persistent? New Evidence from a Panel of U.S. Tax Returns," *Brookings Papers on Economic Activity*, 2013, *46* (1 (Spring), 67–142.

____ , Vasia Panousi, and Shanthi Ramnath, "The joint income processes of business owners," Working paper 2013.

Evans, David S and Boyan Jovanovic, "An Estimated Model of Entrepreneurial Choice under Liquidity Constraints," *Journal of Political Economy*, August 1989, *97* (4), 808–27.

____ and Linda S Leighton, "Some Empirical Aspects of Entrepreneurship," *American Economic Review*, June 1989, *79* (3), 519–35.

Feldman, Naomi E. and Joel Slemrod, "Estimating tax noncompliance with evidence from unaudited tax returns," *Economic Journal*, 03 2007, *117* (518), 327–352.

Gentry, William M. and R. Glenn Hubbard, "Entrepreneurship and Household Saving," NBER Working Papers 7894, National Bureau of Economic Research, Inc September 2000.

Gourinchas, Pierre-Olivier and Jonathan A. Parker, "Consumption Over the Life Cycle," *Econometrica*, January 2002, *70* (1), 47–89.

Guvenen, Fatih, "Learning Your Earning: Are Labor Income Shocks Really Very Persistent?," *American Economic Review*, June 2007, *97* (3), 687–712.

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

Hamilton, Barton H., "Does Entrepreneurship Pay? An Empirical Analysis of the Returns to Self-Employment," *Journal of Political Economy*, June 2000, *108* (3), 604–631.

Heaton, John and Deborah Lucas, "Portfolio Choice and Asset Prices: The Importance of Entrepreneurial Risk," *Journal of Finance*, 06 2000, *55* (3), 1163–1198.

Holmes, Thomas J and Jr Schmitz James A, "A Theory of Entrepreneurship and Its Application to the Study of Business Transfers," *Journal of Political Economy*, April 1990, *98* (2), 265–94.

____ **and** ____, "On the Turnover of Business Firms and Business Managers," *Journal of Political Economy*, October 1995, *103* (5), 1005–38.

____ **and** ____, "Managerial Tenure, Business Age, and Small Business Turnover," *Journal of Labor Economics*, January 1996, *14* (1), 79–99.

Holtz-Eakin, Douglas, David Joulfaian, and Harvey S Rosen, "Sticking It Out: Entrepreneurial Survival and Liquidity Constraints," *Journal of Political Economy*, February 1994, *102* (1), 53–75.

Hurst, Erik and Benjamin Wild Pugsley, "What do Small Businesses Do?," *Brookings Papers on Economic Activity*, 2011, *43* (2 (Fall)), 73–142.

____**, Geng Li, and Benjamin Pugsley**, "Are Household Surveys Like Tax Forms: Evidence from Income Underreporting of the Self Employed," NBER Working Papers 16527, National Bureau of Economic Research, Inc November 2010.

Hyslop, Dean R., "Rising U.S. Earnings Inequality and Family Labor Supply: The Covariance Structure of Intrafamily Earnings," *American Economic Review*, September 2001, *91* (4), 755–777.

Kawaguchi, Daiji, "Compensating Wage Differentials among Self-Employed Workers:Evidence from Job Satisfaction Scores," ISER Discussion Paper, Institute of Social and Economic Research, Osaka University 2002.

Knittel, Matthew, Susan Nelson, Jason DeBacker, John Kitchen, James Pearce, and Richard Prisinzano, "Methodology to Identify Small Businesses and Their Owners," OTA Technical Paper 4, U.S. Department of the Treasury 2011.

Lochner, Lance and Youngki Shin, "Understanding Earnings Dynamics: Identifying and Estimating the Changing Roles of Unobserved Ability, Permanent and Transitory Shocks," NBER Working Papers 20068, National Bureau of Economic Research, Inc April 2014.

Lucas, Robert E., "On the Size Distribution of Business Firms," *Bell Journal of Economics*, 1978, *9* (2), 508–523.

Meghir, Costas and Luigi Pistaferri, "Income Variance Dynamics and Heterogeneity," *Econometrica*, 01 2004, *72* (1), 1–32.

Moore, Kevin, "Comparing the earnings of employees and the self-employed," Technical Report 2004.

Moskowitz, Tobias J. and Annette Vissing-Jørgensen, "The Returns to Entrepreneurial Investment: A Private Equity Premium Puzzle?," *American Economic Review*, September 2002, *92* (4), 745–778.

Panousi, Vasia, "Capital Taxation with Entrepreneurial Risk," Technical Report 2012.

Parker, Jonathan A. and Annette Vissing-Jørgensen, "Who Bears Aggregate Fluctuations and How?," *American Economic Review*, May 2009, *99* (2), 399–405.

Pence, Karen M., "The Role of Wealth Transformations: An Application to Estimating the Effect of Tax Incentives on Saving," *The B.E. Journal of Economic Analysis & Policy*, July 2006, *5* (1), 1–26.

Piketty, Thomas and Emmanuel Saez, "Income Inequality in the United States, 1913-1998," *Quarterly Journal of Economics*, February 2003, *118* (1), 1–39.

Plumley, Alan H., "The Determinants of Individual Income Tax Compliance: Estimating The Impacts of Tax Policy, Enforcement, and IRS Responsiveness," Publication 1916 (Rev. 11-96), Internal Revenue Service 1996.

Polkovnichenko, Valery, "Human Capital and the Private Equity Premium," *Review of Economic Dynamics*, October 2003, *6* (4), 831–845.

Quadrini, Vincenzo, "Entrepreneurship, Saving and Social Mobility," *Review of Economic Dynamics*, January 2000, *3* (1), 1–40.

Roussanov, Nikolai, "Diversification and Its Discontents: Idiosyncratic and Entrepreneurial Risk in the Quest for Social Status," *Journal of Finance*, October 2010, *65* (5), 1755–1788.

Rouwenhorst, K G, "Asset Pricing Implications of Equilibrium Business Cycle Models," in Thomas F. Cooley, ed., *Frontiers of Business Cycle Research*, Vol. 3, Princeton University Press, 1995, chapter 41, pp. 294–330.

Schulhofer-Wohl, Sam and Andriy Norets, "Heterogeneity in income processes," Technical Report 2009.

Slemrod, Joel, "Did the Tax Reform Act of 1986 Simplify Tax Matters?," *Journal of Economic Perspectives*, Winter 1992, *6* (1), 45–57.

——, "Cheating Ourselves: The Economics of Tax Evasion," *Journal of Economic Perspectives*, Winter 2007, *21* (1), 25–48.

Wang, Chong, Neng Wang, and Jinqiang Yang, "Dynamics of Entrepreneurship under Incomplete Markets," NBER Working Papers 16843, National Bureau of Economic Research, Inc March 2011.

Zhu, Shenghao, "Wealth distribution under idiosyncratic investment risk," Working paper 2010.

Appendix A

A-1 Panel summary statistics

[Table 11 about here.]

[Table 12 about here.]

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

[Table 13 about here.]

A-3 Probit model estimates, alternative sample selection

[Table 14 about here.]

A-4 Mobility robustness tests

In order to examine the robustness of the income mobility results, we perform the following tests.

First, to remove income variation that is due to observables and therefore cannot be termed risk, we conduct the analysis using residuals of income from a the regression described by Equation 2. In particular, it has been shown that predictable variation in income over the lifecycle matters a great deal for labor income. We then take the resulting income residuals and split them into deciles, separately for each year, so as to again construct one year transition matrices. The resulting matrices present the same qualitative picture for the comparison of business and labor income risk as do Tables 1 and 2.

Second, in order to address the issue of potential changes in household composition, and though this issue was also partially addressed by the first robustness test above, we restrict our samples to continuously married households and repeat the transition matrix analysis. The results about the qualitative comparison of business and labor income riskiness are again unchanged.

Third, we examine whether the higher riskiness of business income is preserved over longer horizons. To that end, we also construct three year business and labor income transition matrices, following a similar procedure to that outlined in 5.1. Here, both business and labor income exhibit higher mobility than in their corresponding one year matrices, but the comparison between business and labor income is again qualitatively preserved, in that business income has a lower immobility ratio (lower persistence), higher probabilities of transitioning to deciles far apart, and higher probabilities of transitioning from the bottom to the top of the distribution.

Fourth, we also construct transition matrices using fixed income bins, so as to fix the starting points across distributions. In particular, we first de-trend each income measure, and we then run the residuals through fixed income bins. For example, the first bin includes (residuals of) income more negative than -$25,000, the second bin goes from -$25,000 to zero, and so on until the last bin, which includes income residuals higher than $1 million. The qualitative comparison between business and labor income is again similar to that in Tables 1 and 2.

Fifth, we construct transition matrices for our "drop 0's" business income panel. That is, we drop the years in which those who ever have business income above our threshold realize zero business income (i.e., the have exited the business sector). The results are almost identical to those in Table 1, with the exception of the zero state.

Sixth, to examine whether our results are different when we focus on households who have more business experience, as captured by more non-zero business income years, we construct transition matrices for business income in a panel that drops households who have had non-zero business income for less than 8 years. The results are overall similar to those in Table 1, indicating that business income is riskier than labor income, even among more "experienced" business households.

Finally, we construct income residuals from a modified version of Equation 2. In particular, for each type of income, we run a pooled regression with time and household fixed effects, plus our other controls. We then run those residuals through our transition matrices and proceed as before. The results about the qualitative comparison between business and labor income continue to hold.

A-5 Percent changes in income robustness tests

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

First, we conduct the analysis using levels of income, as opposed to residuals from a first stage regression. As we note in Section 5.3, the results are qualitatively very similar to those using residuals from a first-stage regression of income on demographics.

Second, we use alternative definitions of percent changes. In particular, we use either $\frac{abs(x_t+x_{t-1})}{2}$ or $\frac{abs(x_t)+abs(x_{t-1}))}{2}$ in the denominator of Equation 3. We use the average in order to smooth out some of the most extreme fluctuations in business income. We use the absolute value of the sum in order to capture the idea that a business making +\$15,000 in one year and -\$5,000 in the next year is a business of size +\$10,000. We use the sum of the absolute values to address the problem that the previous definition would encounter when the business makes +\$5,000 in one year and -\$5,000 in the next year. When using the sum of the absolute values, the idea is that a business making +\$5,000 in one year and -\$5,000 in the next year is a business of absolute size \$5,000. Hence, size absorbs part of the risk in this definition. Overall, regardless of the definition used, the results about the qualitative comparison between business and labor income are similar to those in Section 5.3.

Third, we address the concern that the larger percentage changes in business income are due to business income changes around zero. In other words, when business income changes from "small" positive to "small" negative values (or vice versa), this will register as a very large percent change, and might unduly influence the analysis. In addition, the concern might be that small, in absolute value, business income observations may indicate a household that is not truly involved in business endeavors. To that end, we truncate income based on some (arbitrary) thresholds. For example, we recreated the results using thresholds $\pm\$1,000$, $\pm\$2,500$, $\pm\$10,000$,and $\pm\$20,000$.[40] In addition, we have experimented with dropping entirely the observations in between the positive and the symmetric negative threshold. Regardless of treatment, the results remain qualitatively similar to those in Section 5.3 . Hence, the larger risk associated with business income, vis-á-vis labor income, is not due to the influence of income observations around zero.

Fourth, we examine whether the dispersion of the distribution of percent changes is higher for business income when the horizon is longer. To that end, we construct three and five year percent

[40] For a justification of these thresholds, see Knittel et al. (2011). Additionally, note that, in real terms, one-fourth of a full-year, full-time minimum wage in 2005 was about \$2,575. Though our thresholds will necessarily be somewhat arbitrary, they nonetheless make the point that our results are not due to business income changes around zero.

changes, using the corresponding adaptation of Equation 3. We find that the distribution of percent changes for business income is more dispersed than that for labor income, even when using more "long run" measures of income.

Fifth, calculate percentage changes in residuals where we include zero observations. That is, we form a business income panel composed of all filers who ever receive business income over the sample. We include these individuals even in year when they receive no business income. We create a labor income panel similarly. Using these alternative samples, we calculate the distribution of percentage changes and the percentile ratios. Both are very similar to our baseline sample, although the mass in the tails of the business income distribution increases.

A-6 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 5-8 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 $t = 1987, ..., 2009$. The empirical variance and autocovariances are computed from our panel of tax returns for $a = 1, ..., 31$, $t = 1987, ..., 2009$, and $k = 0, ..., 22$. There are 6,532 of these variances and autocovariances using in the estimation. We weight each moments using the inverse of the number of observations used to compute each empirical statistical moment.[41]

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

A-7 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 5-8 are as follows:

$$cov(\xi^i_{a,t}, \xi^i_{a+k,t+k}) = \lambda_t \cdot \lambda_{t+k} \cdot \left(\sigma^2_\alpha + (\sigma^2_r \cdot a)\right) + \rho^k var(z^i_t) \tag{A.7.1}$$

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

$$var(z^i_{a,1987}) = \sigma^2_z \frac{1-\rho^{2a}}{1-\rho^2} \tag{A.7.2}$$

For $1987 \leq t \leq 2009$, $a = 1$:

$$var(z^i_{1,t}) = \pi^2_t \sigma^2_z \tag{A.7.3}$$

For $1988 \leq t \leq 2009$, $2 \leq a \leq 36$:

$$var(z^i_{a,t}) = \rho^2 var(z^i_{a-1,t-1}) + \pi^2_t \sigma^2_z \tag{A.7.4}$$

To obtain identification, we impose the normalization $\lambda_t = \pi_t = 1$ for all calendar years $t \leq 1987$, where 1987 is the first year in the sample. We also impose $\pi_{2009} = \pi_{2008}$, since 2009 is the last year in the sample.

A-8 Inverse Hyperbolic Sine (IHS) estimation

In order to estimate the location parameter, θ, for the Inverse Hyperbolic Sine (IHS), we assume that, for some unknown θ, the transformed variables y^θ satisfy the full set of normal theory assumptions:

$$y^\theta = X\beta + \epsilon, \quad \epsilon \sim iid\, N(0, \sigma^2)\ . \tag{A.8.5}$$

Then, the likelihood function with respect to the original observations, y, is:

$$L(\theta, \beta, \sigma^2; X, y) = \prod_{i=1}^{N} f(\epsilon_i) J(\theta; y_i)\ , \tag{A.8.6}$$

where

$$f(\epsilon_i) = (2\pi\sigma^2)^{-1/2} exp\{-\frac{\epsilon_i^2}{2\sigma^2}\} \ , \tag{A.8.7}$$

$$J(\theta; y_i) = |\frac{dy_i^\theta}{dy_i}| = (1 + \theta^2 y_i^2)^{-1/2} \ . \tag{A.8.8}$$

Inspection of $(A.8.6)$ indicates that, but for a constant (or, equivalently, for each given θ) L is the likelihood for a standard least squares problem. To formalize this idea, note that one of the parameters we wish to estimate, namely σ^2, has a best estimate that is a function of the other two parameters, θ and β. In other words, the best estimate for the variance of the errors from the first stage regression in $A.8.5$ is:

$$\hat{\sigma}^2 = \frac{1}{N}\epsilon'\epsilon = \frac{1}{N}(y^\theta - X\beta)'(y^\theta - X\beta) = \frac{1}{N}\sum_{i=1}^{N}\epsilon_i^2 \ . \tag{A.8.9}$$

Using this, we can re-write the likelihood in $(A.8.6)$ as a function of θ and β only, called the concentrated log-likelihood:

$$l(\theta, \beta; X, y) = -\frac{N}{2} log(\sum_{i=1}^{N}\epsilon_i^2) - \frac{1}{2}\sum_{i=1}^{N} log(1 + \theta^2 y_i^2) \ . \tag{A.8.10}$$

In the concentrated log likelihood, the first term comes from the normal distribution and is the log sum of squared residuals from the linear model in $(A.8.5)$. The second term comes from the Jacobian of the transformed observations in $(A.8.8)$. We therefore find the maximum likelihood estimates as follows. First, for a given θ, the maximum likelihood estimates of β are the least squares estimates for the dependent variable y^θ and the estimate of σ^2, denoted, for a fixed θ, by $\hat{\sigma}^2(\theta)$. We denote the maximized concentrated log likelihood as:

$$l_{max}(\theta) = -\frac{N}{2} log(\sum_{i=1}^{N}\hat{\epsilon}_i^2) - \frac{1}{2}\sum_{i=1}^{N} log(1 + \theta^2 y_i^2) \ . \tag{A.8.11}$$

Second, we search over all possible values of θ to find the one that maximizes the concentrated log likelihood, $l_{max}(\theta)$. We call that value of θ "optimal" and we note that it is sample specific.

A-9 Error component model parameter estimates, full set of parameter estimates

[Table 15 about here.]

[Table 16 about here.]

A-10 Joint income processes and income risk across sources

To see the extent of the correlation between risk and heterogeneity across income sources, we not consider a model of joint income processes. As a first pass, we calculate the household-level correlations between business and labor income, which is -0.20. That is, households have negative co-movement between business and labor income. This may be because households substitute between income sources with their limited labor supply or it may be due to negatively correlated shocks (risk).

Next, we run a regression as in Equation 10, on our IHS-transformed income measures. We add to this regression equation household fixed effects (while dropping covariates that are time-invariant) and run separate regressions for business and labor income. We then calculate the correlation between the household fixed effects across sectors. The correlation is -0.53. The strong negative correlation suggests that selection into income sources may be important. The strong negative correlation between these fixed effects suggests that household most often specialist in one sector. Those with high permanent income as a business owner, for example, may be more likely to sort into business ownership. Of course, this could also mean that people who are largely business owners over the sample (and thus who have high average business income) put little effort towards labor provision and thus have low labor income).

To more precisely decompose the negative household-level correlations between business and labor income, we now turn to our baseline error components model (FE_RW_AR1), but allow for the components to be correlated across income sources. The methodology is inspired by Hyslop (2001), who estimates error components models for the joint income processes of spouses.

A-10.1 Error components model

To estimate the joint processes for residual business and labor income, we first transform income levels for both income sources using the IHS transformation with parameter $\theta = 1$, as described in Equation 9. Second, we run the first stage regression as in Equation 10, separately for each income sources. Finally, we estimate the following error components model, which allows for a household fixed effect, α_i, and a random walk and AR(1) component. The model similar to the FE_RW_AR1 model that best fit the data in the above error component models and allowed for the richest parameterization of the data. However, in this case, we allow the income processes to be correlated. Specifically, we define the residual of transformed income as:

$$\xi^{j,i}_{a,t} = \lambda^j_t \cdot (\alpha^{j,i} + r^{j,i}_{a,t}) + z^{j,i}_{a,t} \tag{A.10.12}$$

where

$$r^{j,i}_{a,t} = r^{j,i}_{a-1,t-1} + \epsilon^{j,i}_{a,t} \tag{A.10.13}$$

$$z^{j,i}_{a,t} = \rho_j z^{j,i}_{a-1,t-1} + \pi^j_t \eta^{j,i}_{a,t} \tag{A.10.14}$$

$$\alpha^{j,i} \sim \ iid\ (0,\ \sigma^2_{j,\alpha}),\ \epsilon^{j,i}_{a,t} \sim \ iid\ (0,\ \sigma^2_{j,r}),\ \eta^{j,i}_{a,t} \sim \ iid\ (0,\ \sigma^2_{j,z}), \tag{A.10.15}$$

where $j = b, l$ for business and labor income, respectively. We parameterize the correlations between the shocks by $\rho_\alpha = corr(\alpha_{b,i}, \alpha_{l,i})$, $\rho_r = corr(\epsilon_{b,i,t}, \epsilon_{l,i,t})$, and $\rho_z = corr(\eta_{b,i,t}, \eta_{l,i,t})$. To aid in identification, we assume that the shocks to the permanent and transitory components across income types are not correlated over time. That is, $cov(\epsilon^{b,i}_{a,t}, \epsilon^{l,i}_{a+k,t+k}) = 0$ and $cov(\eta^{b,i}_{a,t}, \eta^{l,i}_{a+k,t+k}) = 0\ \forall k \neq 0$. The model is then estimated using a Generalized Method of Moments (GMM) methodology that minimizes the distance between the models theoretical moments and their empirical counterparts.[42]

[42]For more details on the estimation methodology and the theoretical moments, see Appendix Sections $A-11$ and $A-12$.

A-10.2 Results

Tables 17 displays the results.[43] This sample differs a from that used in Section 6, in particular, it includes the "zero" observations for those who do not participate in business ownership or labor supply. The parameter estimates may change from these differences in the sample selection criteria and the transformation of income. However, note that the qualitative importance of the permanent and transitory factors are largely the same. Interestingly, we find both the heterogeneity components (the individual fixed effect) and the risk components to be negatively correlated across income sources.

[Table 17 about here.]

A-11 Estimation of error-components model for joint income processes

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, and income source (business or labor), $j$the error-components models described by equations $A.10.12$-$A.10.15$ imply the parametric form for each autocovariance of residual, transformed income such as $cov(\xi^j_{a,t}, \xi^j_{a+k,t+k})$. As an example, if $a = 5$, $t = 2006$, $k = 0$, $j = b$ for business income, the counterpart to the theoretical moment is the variance (since $k = 0$) of business income for 35 year olds (since normalized age is 5) in year 2006. The theoretical variances and autocovariances ($cov(j, a, t, k)$) are functions of the model parameters: $\sigma^2_{j,\alpha}, \sigma^2_{j,r}, \sigma^2_{j,z}, \rho_j, \rho_\alpha, \rho_r, \rho_z, \lambda^j_t, \pi^j_t$, for $t = 1987, ..., 2009$ and $j = b, l$. The empirical variance and autocovariances are computed from our panel of tax returns for $a = 1, ..., 31$, $t = 1987, ..., 2009$, $k = 0, ..., 22$, and $j = b, l$. There are 26,128 of these variances and autocovariances using in the estimation. We weight each moments using the inverse of the number of observations used to compute each empirical statistical moment.[44]

[43] The full results, including the calendar year factor loadings, are available in Appendix Table 18.

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

A-12 Moment conditions for joint income processes

Let a be "normalized age", defined as a = age-30+1, or years since age 30. Then, the theoretical moments implied by our joint income error-components model in equations $A.10.12$-$A.10.15$ are as follows. For the autocovariances within an income category ($j = b, l$):

$$cov(\xi_{a,t}^{j,i}, \xi_{a+k,t+k}^{j,i}) = \lambda_t^j \lambda_{t+k}^j [\sigma_{j,\alpha}^2 + (a \times \sigma_{j,r}^2)] + \rho_j^k var(z_{a,t}^{j,i}) \tag{A.12.16}$$

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

$$var(z_{a,1987}^{i,j}) = \sigma_{j,z}^2 \frac{1 - \rho_j^{2a}}{1 - \rho_j^2} \tag{A.12.17}$$

For $1987 \leq t \leq 2009$, $a = 1$:

$$var(z_{1,t}^{j,i}) = \pi_t^j)^2 \sigma_{j,z}^2 \tag{A.12.18}$$

For $1988 \leq t \leq 2009$, $2 \leq a \leq 36$:

$$var(z_{a,t}^{j,i}) = \rho_j^2 var(z_{a-1,t-1}^{j,i}) + (\pi_t^j)^2 \sigma_{j,z}^2 \tag{A.12.19}$$

Across income categories, have:

$$cov(\xi_{a,t}^{b,i}, \xi_{a+k,t+k}^{l,i}) = \lambda_t^b \lambda_{t+k}^l [\rho_\alpha \sigma_{b,\alpha} \sigma_{l,\alpha} + (a \times \rho_r \sigma_{b,r} \sigma_{l,r})] + \rho_z \sigma_{b,z} \sigma_{l,z} \sum_{n=0}^{a} \rho_l^n \rho_b^n \pi_{t-n}^b \pi_{t-n}^l \tag{A.12.20}$$

and:

$$cov(\xi_{a,t}^{l,i}, \xi_{a+k,t+k}^{b,i}) = \lambda_t^l \lambda_{t+k}^b [\rho_\alpha \sigma_{l,\alpha} \sigma_{b,\alpha} + (a \times \rho_r \sigma_{l,r} \sigma_{b,r})] + \rho_z \sigma_{l,z} \sigma_{b,z} \sum_{n=0}^{a} \rho_b^n \rho_l^n \pi_{t-n}^l \pi_{t-n}^b \tag{A.12.21}$$

To obtain identification, we impose the normalization $\lambda_t^j = \pi_t^j = 1$ for all calendar years $t \leq 1987$, where 1987 is the first year in the sample. We also impose $\pi_{2009}^j = \pi_{2008}^j$, since 2009 is the last year in the sample.

Tables

Table 1: One-year Transition Matrix Across Deciles of Business Income Distribution

From/To	0	1	2	3	4	5	6	7	8	9	10
0	0.83	0.02	0.03	0.03	0.03	0.02	0.02	0.01	0.01	0.01	0.00
1	0.11	0.47	0.15	0.06	0.04	0.03	0.02	0.02	0.03	0.03	0.03
2	0.15	0.14	0.33	0.16	0.07	0.05	0.04	0.02	0.02	0.02	0.01
3	0.21	0.06	0.14	0.30	0.12	0.07	0.04	0.03	0.02	0.02	0.01
4	0.18	0.04	0.06	0.13	0.25	0.15	0.08	0.05	0.03	0.02	0.01
5	0.14	0.03	0.05	0.07	0.16	0.25	0.15	0.08	0.05	0.02	0.01
6	0.11	0.02	0.03	0.04	0.07	0.15	0.27	0.17	0.08	0.03	0.01
7	0.07	0.03	0.02	0.03	0.05	0.08	0.15	0.31	0.18	0.06	0.02
8	0.05	0.03	0.02	0.02	0.03	0.04	0.07	0.16	0.36	0.19	0.04
9	0.04	0.03	0.01	0.01	0.02	0.02	0.03	0.06	0.16	0.47	0.15
10	0.02	0.04	0.01	0.01	0.01	0.01	0.01	0.02	0.03	0.13	0.74

Table 1 shows the transition one-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 2: One-year Transition Matrix Across Deciles of Labor Income Distribution

From/To	0	1	2	3	4	5	6	7	8	9	10
0	0.75	0.13	0.04	0.02	0.01	0.01	0.01	0.00	0.01	0.00	0.00
1	0.07	0.57	0.22	0.07	0.03	0.02	0.01	0.00	0.00	0.00	0.00
2	0.02	0.18	0.49	0.20	0.06	0.03	0.01	0.01	0.00	0.00	0.00
3	0.01	0.07	0.14	0.50	0.19	0.05	0.02	0.01	0.00	0.00	0.00
4	0.01	0.04	0.05	0.13	0.51	0.19	0.05	0.02	0.01	0.00	0.00
5	0.01	0.02	0.03	0.04	0.12	0.51	0.20	0.05	0.02	0.01	0.00
6	0.00	0.01	0.01	0.02	0.04	0.13	0.51	0.21	0.04	0.02	0.00
7	0.00	0.01	0.01	0.01	0.02	0.04	0.14	0.53	0.20	0.03	0.01
8	0.00	0.01	0.00	0.01	0.01	0.02	0.04	0.14	0.58	0.18	0.02
9	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.03	0.13	0.66	0.14
10	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.02	0.11	0.84

Table 2 shows the transition 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 3: Exit from business: Probit results

	(1)	(2)
log(positive business income$_{t-1}$)	-0.044***	-0.040***
	(0.000)	(0.000)
log(negative business income$_{t-1}$)	0.042***	0.040***
	(0.000)	(0.000)
log(labor income$_{t-1}$)		0.012***
		(0.001)
Number of consecutive years with biz income	-0.016***	-0.016***
	(0.001)	(0.001)
Number of years w/ biz income	0.001	0.001
	(0.000)	(0.001)
Age 35-39		0.006*
		(0.004)
Age 40-44		0.007*
		(0.004)
Age 45-49		0.008**
		(0.004)
Age 50-54		0.010**
		(0.004)
Age 55-60		0.010**
		(0.005)
Has children		0.010***
		(0.003)
Single male		0.060***
		(0.004)
Single female		0.048***
		(0.004)
Married female		-0.022**
		(0.011)
Observations	88,990	88,990

[a] $*p < 0.05$, $**p < 0.01$, $***p < 0.001$
[b] Standard errors in parentheses below parameter estimates.

Table 3 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 to construct percent changes in income for each household is our benchmark 1987-2009 panel, which drop filers who never report income outside the $\pm\$5,000$ threshold at any point in the sample. We restrict ages to 30-60, and exclude farmers.

Table 4: Ratio of Percentiles of the Distribution of One Year Percent Changes

$p5$	$p10$	$p20$	$p25$	$p30$	$p40$	$p60$	$p70$	$p75$	$p80$	$p90$	$p95$	$p99$
1.71	1.77	2.00	2.12	2.16	2.22	1.95	1.83	1.78	1.68	1.68	1.82	1.78

Table 4 presents the ratio of the percentiles of the distribution of one year percent changes for business income, divided by the corresponding percentiles of the distribution of one year percent changes for labor income. The panels used to construct percent changes in income for each household are our benchmark 1987-2009 panels, which drop zero income observations, restrict ages to 30-60, and exclude farmers. The one year percent changes are calculated from income residuals from the first stage regression in (2), see text for details. The calculations of percent changes use the formula (3). The distributions of percent changes are pooled across all sample years.

Table 5: Error Component Model Parameter Estimates, Business Income

	FE_AR1_RW	FE_AR1	FE_RW_WN	RW_WN
ρ	0.741	0.932	0.000	0.000
	(0.013)	(0.003)	-	-
σ_α^2	0.989	1.795	1.301	0.000
	(0.116)	(0.212)	(0.093)	-
σ_r^2	0.075	0.000	0.092	0.158
	(0.009)	-	(0.006)	(0.012)
σ_z^2	0.865	0.348	1.018	1.366
	(0.122)	(0.030)	(0.231)	(0.251)
Root MSE	1.140	1.151	1.190	1.294

Table 5 presents the results of model estimation using our benchmark 1987-2009 business income panels where we drop observations inside 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 6: Error Component Model Parameter Estimates, Labor Income

	FE_AR1_RW	FE_AR1	FE_RW_WN	RW_WN
ρ	0.645	0.846	0.000	0.000
	(0.006)	(0.003)	-	-
σ_α^2	0.191	0.218	0.204	0.000
	(0.004)	(0.007)	(0.005)	-
σ_r^2	0.004	0.000	0.005	0.019
	(0.000)	-	(0.000)	(0.001)
σ_z^2	0.174	0.082	0.264	0.280
	(0.007)	(0.004)	(0.012)	(0.028)
Root MSE	0.051	0.059	0.067	0.163

Table 6 presents the results of model estimation uses our benchmark 1987-2009 labor income panels 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.

Table 7: Percent Changes to Levels of Income

Model	Business Income	Labor Income
FE_RW_AR1	179%	52%
FE_AR1	89%	33%
FE_RW_WN	204%	68%
RW_WN	265%	73%

Table 7 presents the percent changes to the levels of business and of labor income resulting from a positive one standard deviation model-estimated shock to each type of income. Model estimation uses our benchmark 1987-2009 panels where we drop business income observations in the $\pm\$5,000$ range and labor income 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. Our benchmark average household has a primary filer who is male, married, aged 35, and with two kids. For that household, the percent change in the level of income with and without the model-estimated shock is calculated using (3).

Table 8: Robustness: Risk vs. Heterogeneity by Model

Model	Business Income		Labor Income	
	Heterogeneity	Risk	Heterogeneity	Risk
FE_RW_AR1	23%	77%	44%	56%
FE_AR1	16%	84%	45%	55%
FE_RW_WN	31%	69%	47%	53%

Table 8 presents the average percentage of the cross-sectional variance that is due to household fixed effects (heterogeneity) or risk (over calendar years of the sample). Model estimation uses our benchmark 1987-2009 panels where we drop business income observations in the $\pm\$5,000$ range and labor income 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.

Table 9: Life-cycle Model Results, Counterfactual Income Distributions Under Full Information

	Mean Income	Variance in Income
Endogenous		
All	66,933.21	29,206.17
Business Owners	64,613.73	29,019.60
Exogenous		
As Employees	58,675.53	22,702.78
% Difference for Business Owners	10.12%	27.82%

Table 9 presents mean and variance of annual income (in \$1,000s) by business ownership status and by the endogeneity of employment decisions. Lines one and two show the distribution of income for all model agents and those that are business owners at some point during their lifetime. Lines three and four show the income distribution for those who select into business ownership when the decision is endogenous, but consider the case where these agents are exogenously determined to be business owners (line three) or employees (line four) for their entire lifetime. Line five then gives the percentage changes between lines two and four, which is the counterfactual we are interested in. Income processes in the life-cycle model are calibrated using the estimates from Section 6.

Table 10: Life-cylce Model Results, Counterfactual Income Distributions Under No Information

	Mean Income	Variance in Income
Endogenous		
All	63,719.03	27,283.83
Business Owners	70,554.29	31,359.35
Exogenous		
As Employees	66,689.34	26,792.48
% Difference for Business Owners	5.80%	17.05%

Table 10 presents mean and variance of annual income (in $1,000s) by business ownership status and by the endogeneity of employment decisions. Lines one and two show the distribution of income for all model agents and those that are business owners at some point during their lifetime. Lines three and four show the income distribution for those who select into business ownership when the decision is endogenous, but consider the case where these agents are exogenously determined to be business owners (line three) or employees (line four) for their entire lifetime. Line five then gives the percentage changes between lines two and four, which is the counterfactual we are interested in. Income processes in the life-cycle model are calibrated using the estimates from Section 6.

Table 11: Descriptive Statistics for Benchmark Labor Income Panel

Year	Mean	Median	Std. Dev.	Observations
1987	37,451	27,151	108,762	97,035
1988	37,862	26,743	93,298	83,197
1989	37,185	26,118	67,210	80,545
1990	37,322	26,152	78,982	80,445
1991	36,854	25,721	62,570	88,594
1992	37,913	26,236	102,531	73,129
1993	37,549	26,022	83,287	73,807
1994	37,873	26,332	71,233	77,201
1995	38,622	26,254	79,481	83,273
1996	39,283	26,514	93,449	91,161
1997	40,635	27,128	113,496	90,752
1998	42,419	27,862	136,490	122,424
1999	43,776	28,414	163,787	132,280
2000	45,262	28,857	228,864	146,356
2001	44,990	29,334	152,565	142,398
2002	44,481	29,398	117,726	127,392
2003	44,580	29,497	136,528	131,634
2004	45,266	29,498	156,074	144,870
2005	45,362	29,154	168,327	217,938
2006	45,961	29,195	168,898	240,291
2007	45,896	28,788	183,620	241,505
2008	45,743	28,726	164,078	236,104
2009	44,989	28,446	133,012	209,675

Table 11 presents descriptive statistics for our benchmark labor income panel where we drop observations where income is less than $2,575. The sample selection keeps households with primary filer aged 30-60 and drops households where the primary filer is a farmer (filing Schedule F). The mean, median and standard deviation are in real 2005 dollars.

Table 12: Descriptive Statistics for Benchmark Business Income Panel

Year	Mean	Median	Std. Dev.	Observations
1987	15,391	52,957	6,758	1,766
1988	18,976	66,232	7,238	1,894
1989	17,727	62,428	6,388	2,046
1990	19,199	65,144	5,984	2,081
1991	18,375	61,241	5,402	2,086
1992	20,349	67,552	5,870	2,161
1993	20,386	60,069	6,038	2,288
1994	21,440	61,228	6,672	2,351
1995	22,478	68,839	6,382	2,381
1996	24,722	73,915	6,309	2,340
1997	23,756	74,442	6,681	2,472
1998	23,565	80,134	6,660	2,522
1999	27,629	87,784	7,010	2,517
2000	26,051	80,946	6,970	2,522
2001	26,984	86,169	7,679	2,623
2002	27,479	103,530	6,638	2,678
2003	24,769	102,544	6,337	2,784
2004	27,935	123,757	6,829	2,878
2005	30,616	126,873	7,616	2,911
2006	30,041	137,554	7,951	2,948
2007	27,976	119,115	7,578	2,982
2008	25,165	102,480	7,352	2,938
2009	23,678	100,485	7,608	2,823

Table 12 presents descriptive statistics for our benchmark business income panel where we drop filers who never report income outside the $\pm\$5,000$ range over the entire panel. The statistic are calculated after dropping zero-income observations from this sample. The sample selection keeps households with primary filer aged 30-60 and drops households where the primary filer is a farmer (filing Schedule F). The mean, median and standard deviation are in real 2005 dollars.

Table 13: Exit from business: Probit results

	(1)	(2)
log(positive business income$_{t-1}$)	-0.044***	-0.040***
	(0.000)	(0.000)
log(negative business income$_{t-1}$)	0.042***	0.040***
	(0.000)	(0.000)
log(labor income$_{t-1}$)		0.012***
		(0.001)
Number of consecutive years with biz income	-0.016***	-0.016***
	(0.001)	(0.001)
Number of years w/ biz income	0.001	0.001
	(0.000)	(0.001)
1988	0.000	0.000
	(0.000)	(0.000)
1989	-0.002	-0.002
	(0.007)	(0.007)
1990	0.016**	0.014**
	(0.007)	(0.007)
1991	0.030***	0.028***
	(0.007)	(0.007)
1992	0.023***	0.021***
	(0.007)	(0.007)
1993	0.012*	0.010
	(0.007)	(0.007)
1994	0.021***	0.020***
	(0.007)	(0.007)
1995	0.026***	0.022***
	(0.007)	(0.007)
1996	0.026***	0.022***
	(0.007)	(0.007)
1997	0.026***	0.022***
	(0.007)	(0.007)
1998	0.027***	0.022***
	(0.007)	(0.007)
1999	0.038***	0.033***
	(0.007)	(0.007)
2000	0.040***	0.034***
	(0.007)	(0.007)
2001	0.039***	0.033***
	(0.007)	(0.007)
2002	0.024***	0.018**
	(0.007)	(0.007)
2003	0.008	0.002
	(0.007)	(0.007)
2004	0.013*	0.007
	(0.008)	(0.008)
2005	0.013*	0.007
	(0.008)	(0.008)
2006	0.013*	0.006
	(0.008)	(0.008)
2007	0.019**	0.012
	(0.008)	(0.008)
2008	0.017**	0.010
	(0.008)	(0.008)
2009	0.030***	0.023***
	(0.008)	(0.008)

	(1)	(2)
Age 35-39		0.006*
		(0.004)
Age 40-44		0.007*
		(0.004)
Age 45-49		0.008**
		(0.004)
Age 50-54		0.010**
		(0.004)
Age 55-60		0.010**
		(0.005)
Has children		0.010***
		(0.003)
Single male		0.060***
		(0.004)
Single female		0.048***
		(0.004)
Married female		-0.022**
		(0.011)
Observations	88,990	88,990

[a] $*p < 0.05$, $**p < 0.01$, $***p < 0.001$
[b] Standard errors in parentheses below parameter estimates.

Table 13 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 to construct percent changes in income for each household is our benchmark 1987-2009 panel, which drop filers who never report income outside the $\pm\$5,000$ threshold at any point in the sample. We restrict ages to 30-60, and exclude farmers.

Table 14: Exit from business: Probit results, ever have non-zero business income sample

	(1)	(2)
log(positive business income$_{t-1}$)	-0.044***	-0.041***
	(0.000)	(0.000)
log(negative business income$_{t-1}$)	0.042***	0.042***
	(0.000)	(0.000)
log(labor income$_{t-1}$)		0.013***
		(0.001)
Number of consecutive years with biz income	-0.016***	-0.020***
	(0.001)	(0.001)
Number of years w/ biz income	0.001	-0.001
	(0.000)	(0.000)
1988	0.000	0.000
	(0.000)	(0.000)
1989	-0.002	0.001
	(0.007)	(0.006)
1990	0.016**	0.009
	(0.007)	(0.006)
1991	0.030***	0.025***
	(0.007)	(0.006)
1992	0.023***	0.021***
	(0.007)	(0.006)
1993	0.012*	0.018***
	(0.007)	(0.006)
1994	0.021***	0.028***
	(0.007)	(0.006)
1995	0.026***	0.028***
	(0.007)	(0.006)
1996	0.026***	0.028***
	(0.007)	(0.006)
1997	0.026***	0.030***
	(0.007)	(0.006)
1998	0.027***	0.032***
	(0.007)	(0.006)
1999	0.038***	0.036***
	(0.007)	(0.006)
2000	0.040***	0.044***
	(0.007)	(0.006)
2001	0.039***	0.040***
	(0.007)	(0.006)
2002	0.024***	0.028***
	(0.007)	(0.006)
2003	0.008	0.011*
	(0.007)	(0.006)
2004	0.013*	0.013**
	(0.008)	(0.006)
2005	0.013*	0.022***
	(0.008)	(0.006)
2006	0.013*	0.018***
	(0.008)	(0.006)
2007	0.019**	0.017***
	(0.008)	(0.006)
2008	0.017**	0.016**
	(0.008)	(0.007)
2009	0.030***	0.025***
	(0.008)	(0.007)

	(1)	(2)
Age 35-39		0.013***
		(0.003)
Age 40-44		0.015***
		(0.003)
Age 45-49		0.019***
		(0.003)
Age 50-54		0.021***
		(0.004)
Age 55-60		0.018***
		(0.004)
Has children		0.013***
		(0.002)
Single male		0.061***
		(0.003)
Single female		0.046***
		(0.003)
Married female		-0.020***
		(0.008)
Observations	131,495	131,495

[a] $*p < 0.05$, $**p < 0.01$, $***p < 0.001$
[b] Standard errors in parentheses below parameter estimates.

Table 14 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 to construct percent changes in income for each household is a 1987-2009 panel, where we drop filers who do not report non-zero income at any point in the sample. We restrict ages to 30-60, and exclude farmers.

Table 15: Error Component Model Parameter Estimates, Business Income

	FE_AR1_RW	FE_AR1	FE_RW_WN	RW_WN
ρ	0.741	0.932	0.000	0.000
	(0.013)	(0.003)	-	-
σ^2_α	0.989	1.795	1.301	0.000
	(0.116)	(0.212)	(0.093)	-
σ^2_r	0.075	0.000	0.092	0.158
	(0.009)	-	(0.006)	(0.012)
σ^2_z	0.865	0.348	1.018	1.366
	(0.122)	(0.030)	(0.231)	(0.251)
π_{1987}	1.000	1.000	1.000	1.000
	-	-	-	-
π_{1988}	1.073	1.307	0.976	0.983
	(0.161)	(0.182)	(0.166)	(0.131)
π_{1989}	0.972	0.844	1.078	1.033
	(0.142)	(0.241)	(0.157)	(0.121)
π_{1990}	1.051	1.594	1.099	1.037
	(0.131)	(0.125)	(0.153)	(0.122)
π_{1991}	0.821	0.725	1.083	1.004
	(0.168)	(0.273)	(0.161)	(0.130)
π_{1992}	0.928	1.227	1.003	0.951
	(0.143)	(0.155)	(0.156)	(0.124)
π_{1993}	0.864	1.001	1.041	0.966
	(0.137)	(0.172)	(0.148)	(0.116)
π_{1994}	0.615	1.090	0.948	0.890
	(0.154)	(0.149)	(0.136)	(0.107)
π_{1995}	0.723	0.854	0.999	0.902
	(0.130)	(0.187)	(0.136)	(0.104)
π_{1996}	1.009	1.284	1.138	0.988
	(0.113)	(0.129)	(0.149)	(0.109)
π_{1997}	1.063	1.154	1.179	1.051
	(0.117)	(0.158)	(0.155)	(0.116)
π_{1998}	0.968	1.097	1.121	1.017
	(0.123)	(0.171)	(0.152)	(0.117)
π_{1999}	1.183	1.171	1.299	1.154
	(0.113)	(0.152)	(0.161)	(0.119)
π_{2000}	0.930	1.234	1.233	1.093
	(0.122)	(0.149)	(0.155)	(0.115)
π_{2001}	1.045	1.370	1.113	1.011
	(0.121)	(0.148)	(0.151)	(0.118)
π_{2002}	1.180	1.467	1.226	1.113
	(0.123)	(0.141)	(0.162)	(0.127)
π_{2003}	0.735	0.000	1.046	0.984
	(0.175)	(3628.805)	(0.159)	(0.121)
π_{2004}	1.134	1.487	1.037	0.977
	(0.126)	(0.171)	(0.158)	(0.126)
π_{2005}	0.660	0.648	0.981	0.922
	(0.178)	(0.400)	(0.152)	(0.124)
π_{2006}	0.617	0.953	0.914	0.870
	(0.174)	(0.260)	(0.146)	(0.118)
π_{2007}	0.844	1.277	1.002	0.913
	(0.129)	(0.192)	(0.145)	(0.113)
π_{2008}	1.054	1.564	1.156	1.017
	(0.097)	(0.116)	(0.146)	(0.106)
π_{2009}	1.054	1.564	1.156	1.017
	-	-	-	-

	FE_AR1_RW	FE_AR1	FE_RW_WN	RW_WN
λ_{1987}	1.000	1.000	1.000	1.000
	-	-	-	-
λ_{1988}	0.947	0.868	1.001	1.001
	(0.068)	(0.084)	(0.047)	(0.057)
λ_{1989}	0.890	0.834	0.941	0.943
	(0.066)	(0.072)	(0.044)	(0.052)
λ_{1990}	0.821	0.507	0.899	0.903
	(0.063)	(0.076)	(0.040)	(0.048)
λ_{1991}	0.927	0.660	0.944	0.968
	(0.070)	(0.075)	(0.041)	(0.050)
λ_{1992}	0.887	0.530	0.918	0.929
	(0.067)	(0.070)	(0.040)	(0.048)
λ_{1993}	0.864	0.473	0.879	0.893
	(0.065)	(0.070)	(0.038)	(0.044)
λ_{1994}	0.918	0.447	0.885	0.904
	(0.064)	(0.065)	(0.036)	(0.043)
λ_{1995}	0.930	0.471	0.869	0.895
	(0.064)	(0.063)	(0.036)	(0.042)
λ_{1996}	0.824	0.197	0.804	0.841
	(0.057)	(0.056)	(0.033)	(0.040)
λ_{1997}	0.820	0.305	0.828	0.845
	(0.058)	(0.066)	(0.035)	(0.042)
λ_{1998}	0.860	0.436	0.870	0.878
	(0.060)	(0.073)	(0.037)	(0.043)
λ_{1999}	0.730	0.204	0.795	0.804
	(0.054)	(0.071)	(0.034)	(0.040)
λ_{2000}	0.790	0.213	0.833	0.851
	(0.056)	(0.064)	(0.034)	(0.041)
λ_{2001}	0.803	0.196	0.871	0.880
	(0.058)	(0.070)	(0.036)	(0.043)
λ_{2002}	0.786	0.125	0.884	0.885
	(0.056)	(0.068)	(0.036)	(0.042)
λ_{2003}	0.961	0.659	0.987	0.988
	(0.068)	(0.079)	(0.041)	(0.048)
λ_{2004}	0.895	0.505	0.976	0.972
	(0.063)	(0.094)	(0.040)	(0.048)
λ_{2005}	0.962	0.606	0.980	0.980
	(0.065)	(0.090)	(0.039)	(0.046)
λ_{2006}	1.025	0.665	0.993	0.994
	(0.066)	(0.087)	(0.039)	(0.046)
λ_{2007}	1.011	0.610	0.967	0.979
	(0.065)	(0.084)	(0.039)	(0.045)
λ_{2008}	0.953	0.482	0.924	0.944
	(0.061)	(0.080)	(0.038)	(0.045)
λ_{2009}	0.888	0.303	0.871	0.894
	(0.061)	(0.088)	(0.038)	(0.044)
N	6532	6532	6532	6532
Root MSE	1.140	1.151	1.190	1.294

Table 15 presents the results of model estimation using our benchmark 1987-2009 business income panels where we drop observations inside 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 16: Error Component Model Parameter Estimates, Labor Income

	FE_AR1_RW	FE_AR1	FE_RW_WN	RW_WN
ρ	0.645	0.846	0.000	0.000
	(0.006)	(0.003)	-	-
σ^2_α	0.191	0.218	0.204	0.000
	(0.004)	(0.007)	(0.005)	-
σ^2_r	0.004	0.000	0.005	0.019
	(0.000)	-	(0.000)	(0.001)
σ^2_z	0.174	0.082	0.264	0.280
	(0.007)	(0.004)	(0.012)	(0.028)
π_{1987}	1.000	1.000	1.000	1.000
	-	-	-	-
π_{1988}	1.008	1.107	0.968	1.000
	(0.037)	(0.076)	(0.032)	(0.076)
π_{1989}	1.007	1.160	0.950	0.994
	(0.035)	(0.067)	(0.034)	(0.079)
π_{1990}	0.993	1.213	0.930	0.981
	(0.034)	(0.064)	(0.030)	(0.077)
π_{1991}	0.936	1.089	0.899	0.969
	(0.034)	(0.068)	(0.033)	(0.075)
π_{1992}	0.953	1.072	0.927	0.999
	(0.039)	(0.072)	(0.034)	(0.076)
π_{1993}	0.890	1.072	0.897	0.972
	(0.038)	(0.070)	(0.032)	(0.075)
π_{1994}	0.879	1.071	0.887	0.968
	(0.041)	(0.073)	(0.033)	(0.075)
π_{1995}	0.880	1.126	0.865	0.953
	(0.039)	(0.069)	(0.032)	(0.074)
π_{1996}	0.847	1.056	0.865	0.955
	(0.039)	(0.073)	(0.031)	(0.075)
π_{1997}	0.769	0.855	0.829	0.918
	(0.043)	(0.091)	(0.031)	(0.074)
π_{1998}	0.836	1.098	0.838	0.935
	(0.041)	(0.071)	(0.031)	(0.076)
π_{1999}	0.813	0.904	0.849	0.940
	(0.037)	(0.083)	(0.027)	(0.071)
π_{2000}	0.884	1.108	0.871	0.960
	(0.039)	(0.068)	(0.031)	(0.073)
π_{2001}	0.909	1.084	0.882	0.965
	(0.035)	(0.066)	(0.030)	(0.074)
π_{2002}	0.929	1.055	0.882	0.961
	(0.031)	(0.066)	(0.028)	(0.071)
π_{2003}	1.034	1.214	0.930	1.003
	(0.032)	(0.061)	(0.030)	(0.074)
π_{2004}	0.980	1.024	0.915	0.982
	(0.038)	(0.075)	(0.034)	(0.076)
π_{2005}	0.974	1.039	0.908	0.968
	(0.040)	(0.084)	(0.035)	(0.077)
π_{2006}	1.031	1.162	0.934	0.987
	(0.040)	(0.075)	(0.035)	(0.080)
π_{2007}	0.968	1.056	0.932	0.976
	(0.037)	(0.075)	(0.035)	(0.081)
π_{2008}	1.015	1.280	0.977	0.984
	(0.026)	(0.043)	(0.028)	(0.066)
π_{2009}	1.015	1.280	0.977	0.984
	-	-	-	-

	FE_AR1_RW	FE_AR1	FE_RW_WN	RW_WN
λ_{1987}	1.000	1.000	1.000	1.000
	-	-	-	-
λ_{1988}	1.035	1.020	1.067	1.041
	(0.015)	(0.023)	(0.017)	(0.035)
λ_{1989}	1.058	1.005	1.105	1.073
	(0.015)	(0.023)	(0.018)	(0.034)
λ_{1990}	1.071	0.965	1.124	1.089
	(0.015)	(0.022)	(0.018)	(0.033)
λ_{1991}	1.110	0.988	1.151	1.104
	(0.016)	(0.024)	(0.018)	(0.033)
λ_{1992}	1.120	0.998	1.154	1.097
	(0.016)	(0.024)	(0.018)	(0.032)
λ_{1993}	1.129	0.987	1.153	1.096
	(0.016)	(0.024)	(0.017)	(0.032)
λ_{1994}	1.153	1.002	1.166	1.103
	(0.016)	(0.026)	(0.017)	(0.032)
λ_{1995}	1.171	1.003	1.180	1.109
	(0.016)	(0.025)	(0.017)	(0.031)
λ_{1996}	1.199	1.030	1.196	1.120
	(0.017)	(0.026)	(0.017)	(0.031)
λ_{1997}	1.239	1.100	1.218	1.144
	(0.017)	(0.028)	(0.017)	(0.032)
λ_{1998}	1.242	1.093	1.220	1.136
	(0.017)	(0.028)	(0.017)	(0.032)
λ_{1999}	1.247	1.125	1.219	1.135
	(0.017)	(0.029)	(0.017)	(0.031)
λ_{2000}	1.223	1.094	1.206	1.118
	(0.016)	(0.027)	(0.017)	(0.031)
λ_{2001}	1.215	1.091	1.209	1.119
	(0.016)	(0.027)	(0.016)	(0.030)
λ_{2002}	1.217	1.108	1.221	1.132
	(0.016)	(0.027)	(0.017)	(0.031)
λ_{2003}	1.172	1.061	1.203	1.107
	(0.016)	(0.026)	(0.017)	(0.031)
λ_{2004}	1.159	1.067	1.199	1.106
	(0.016)	(0.026)	(0.018)	(0.031)
λ_{2005}	1.164	1.088	1.206	1.119
	(0.017)	(0.028)	(0.018)	(0.032)
λ_{2006}	1.156	1.089	1.206	1.120
	(0.016)	(0.025)	(0.018)	(0.033)
λ_{2007}	1.162	1.105	1.200	1.117
	(0.016)	(0.025)	(0.019)	(0.033)
λ_{2008}	1.160	1.086	1.182	1.120
	(0.016)	(0.023)	(0.018)	(0.033)
λ_{2009}	1.180	1.100	1.174	1.121
	(0.016)	(0.024)	(0.018)	(0.034)
N	6532	6532	6532	6532
Root MSE	0.051	0.059	0.067	0.163

Table 16 presents the results of model estimation using our benchmark 1987-2009 labor income panels 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.

Table 17: Error Component Model Parameter Estimates, Joint Income Processes

	Business Income	Labor Income
ρ_j	0.731	0.782
	(0.005)	(0.004)
$\sigma^2_{j,\alpha}$	0.305	0.399
	(0.013)	(0.011)
$\sigma^2_{j,r}$	0.031	0.018
	(0.001)	(0.001)
$\sigma^2_{j,z}$	0.584	0.336
	(0.027)	(0.014)
ρ_α	-0.716	-
	(0.010)	-
ρ_r	-0.257	-
	(0.008)	-
ρ_z	-0.353	-
	(0.005)	-
N	26,128	
Root MSE	0.203	

Table 12 presents descriptive statistics for our benchmark business income panel where we drop filers who never report income outside the $\pm\$5,000$ range over the entire panel. The statistic are calculated after dropping zero-income observations from this sample. The sample selection keeps households with primary filer aged 30-60 and drops households where the primary filer is a farmer (filing Schedule F). The mean, median and standard deviation are in real 2005 dollars.

Table 18: Error Component Model Parameter Estimates, Joint Income Processes

	Business Income	Labor Income
ρ_j	0.731	0.782
	(0.005)	(0.004)
$\sigma^2_{j,\alpha}$	0.305	0.399
	(0.013)	(0.011)
$\sigma^2_{j,r}$	0.031	0.018
	(0.001)	(0.001)
$\sigma^2_{j,z}$	0.584	0.336
	(0.027)	(0.014)
ρ_α	-0.716	-
	(0.010)	-
ρ_r	-0.257	-
	(0.008)	-
ρ_z	-0.353	-
	(0.005)	-
π^j_{1987}	1.000	1.000
	-	-
π^j_{1988}	1.093	1.123
	(0.052)	(0.053)
π^j_{1989}	0.982	1.049
	(0.049)	(0.055)
π^j_{1990}	1.043	1.105
	(0.043)	(0.051)
π^j_{1991}	0.975	0.984
	(0.045)	(0.046)
π^j_{1992}	0.960	0.961
	(0.046)	(0.047)
π^j_{1993}	0.883	0.917
	(0.047)	(0.050)
π^j_{1994}	0.895	0.881
	(0.047)	(0.051)
π^j_{1995}	0.866	0.936
	(0.047)	(0.047)
π^j_{1996}	0.832	0.763
	(0.051)	(0.055)
π^j_{1997}	0.898	0.855
	(0.048)	(0.052)
π^j_{1998}	0.846	0.810
	(0.057)	(0.053)
π^j_{1999}	0.893	0.867
	(0.057)	(0.053)
π^j_{2000}	0.898	0.977
	(0.055)	(0.049)
π^j_{2001}	0.995	1.014
	(0.048)	(0.048)
π^j_{2002}	0.988	1.048
	(0.048)	(0.045)
π^j_{2003}	0.921	1.069
	(0.051)	(0.045)
π^j_{2004}	1.006	1.068
	(0.047)	(0.046)
π^j_{2005}	0.906	1.063
	(0.049)	(0.049)
π^j_{2006}	0.951	1.077
	(0.046)	(0.049)
π^j_{2007}	0.928	1.042
	(0.047)	(0.051)
π^j_{2008}	1.088	1.126
	(0.034)	(0.033)
π^j_{2009}	1.088	1.126
	-	-

	Business Income	Labor Income
λ^j_{1987}	1.000	1.000
	-	-
λ^j_{1988}	0.993	1.001
	(0.027)	(0.019)
λ^j_{1989}	1.010	1.036
	(0.028)	(0.020)
λ^j_{1990}	0.972	1.018
	(0.027)	(0.019)
λ^j_{1991}	0.964	1.017
	(0.026)	(0.019)
λ^j_{1992}	0.985	1.055
	(0.027)	(0.019)
λ^j_{1993}	1.040	1.073
	(0.027)	(0.019)
λ^j_{1994}	1.072	1.092
	(0.027)	(0.020)
λ^j_{1995}	1.103	1.108
	(0.028)	(0.020)
λ^j_{1996}	1.088	1.132
	(0.027)	(0.020)
λ^j_{1997}	1.122	1.154
	(0.028)	(0.021)
λ^j_{1998}	1.140	1.169
	(0.029)	(0.021)
λ^j_{1999}	1.129	1.157
	(0.029)	(0.021)
λ^j_{2000}	1.115	1.130
	(0.028)	(0.020)
λ^j_{2001}	1.089	1.122
	(0.028)	(0.021)
λ^j_{2002}	1.076	1.091
	(0.027)	(0.020)
λ^j_{2003}	1.118	1.109
	(0.028)	(0.020)
λ^j_{2004}	1.146	1.101
	(0.028)	(0.019)
λ^j_{2005}	1.197	1.118
	(0.029)	(0.020)
λ^j_{2006}	1.180	1.106
	(0.028)	(0.020)
λ^j_{2007}	1.176	1.104
	(0.028)	(0.020)
λ^j_{2008}	1.084	1.084
	(0.025)	(0.019)
λ^j_{2009}	1.034	1.104
	(0.025)	(0.019)
N	26,128	
Root MSE	0.203	

Table 18 presents the results of model estimation using our the full1987-2009 panel where we restrict ages to 30-60 and exclude farmers. The error components model allows for individual fixed effects, a random walk, and an AR(1) process. Furthermore, the model allows for non-zero covariances in these income components across income sources. See text for details.

Figures

Figure 1: Aggregate Net Business Income by Business Entity Type

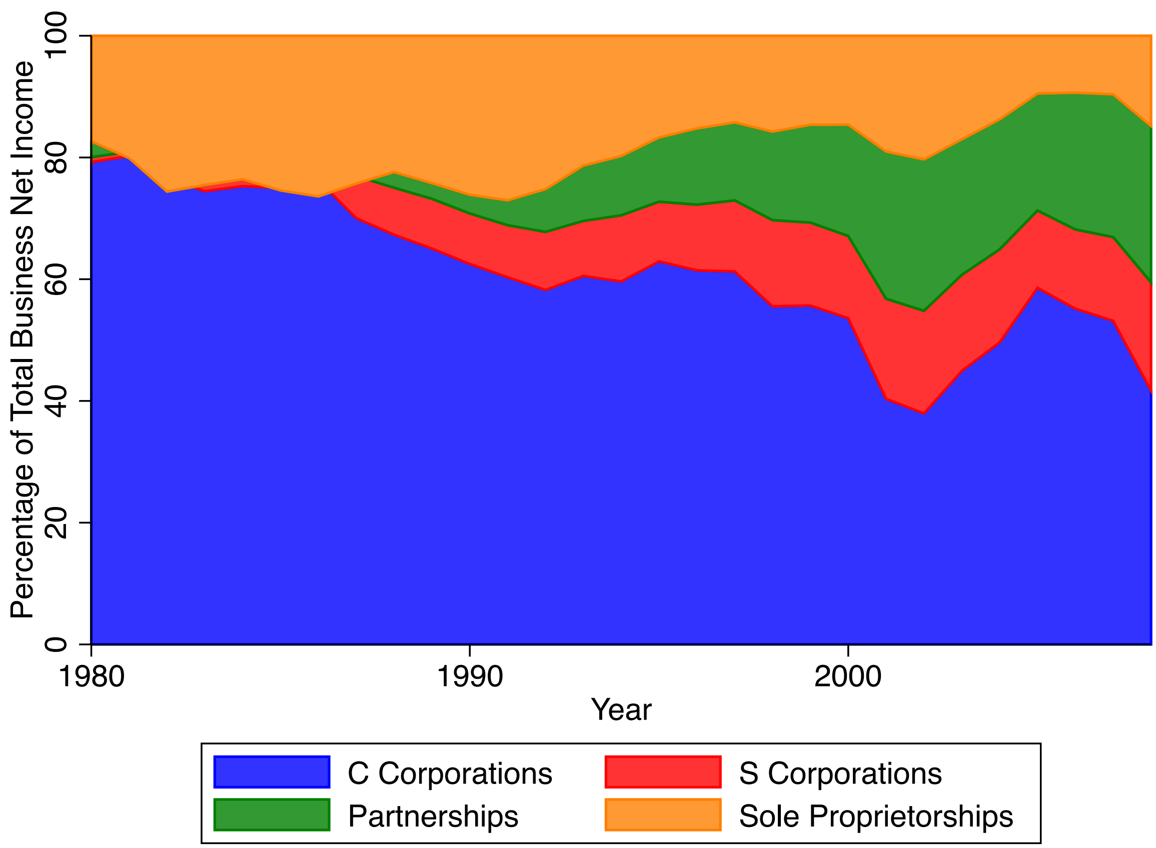

Figure 1 shows aggregate net business income by business entity type from 1980-2008. Data come from the IRS SOI Tax Stats Integrated Business Data.

Figure 2: Mean Business and Labor Income

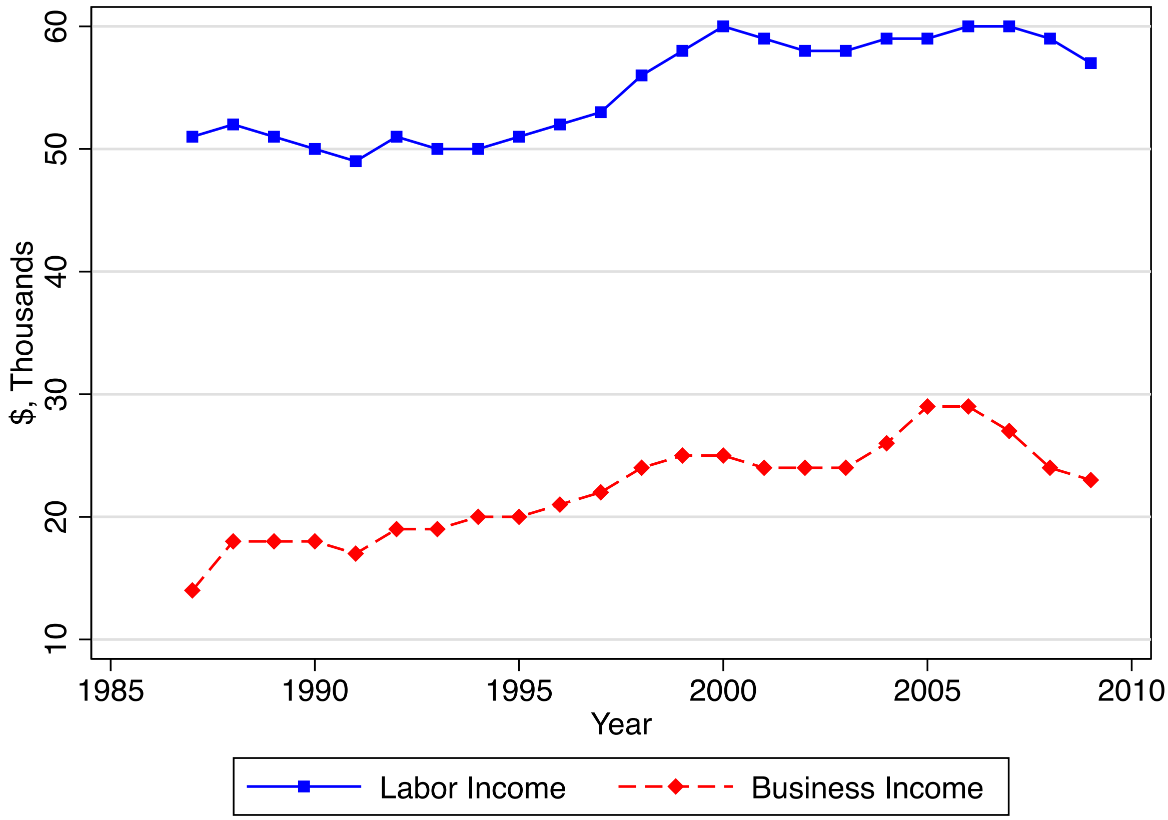

Figure 2 shows mean business income and labor income over time using the SOI cross-sections from 1987-2009. Mean income are calculated using only those with non-zero income. We restrict ages to 30-60, and exclude farmers. Sampling weights are used to create population means.

Figure 3: Business Income at the Top

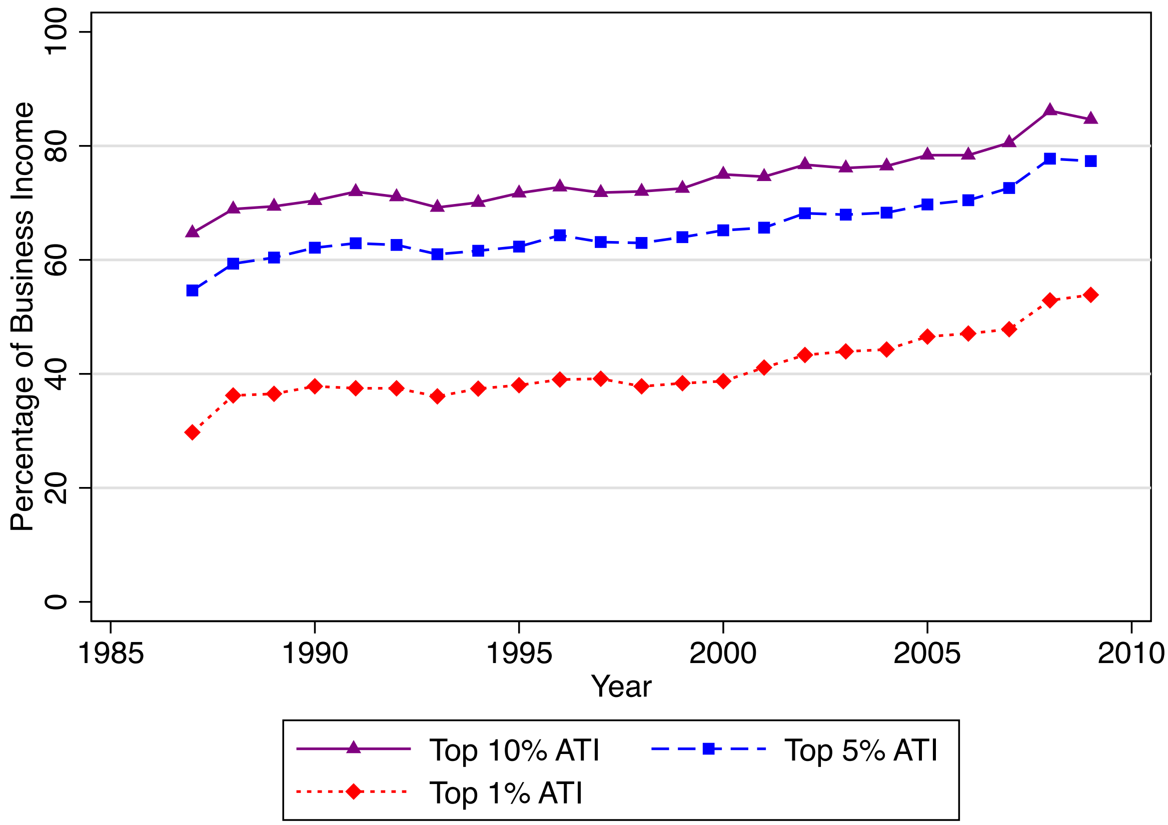

Figure 3 shows the concentration of business income at the top of the adjusted total income (ATI) distribution. The data are from our annual cross sections, 1987-2009. The red line indicates households at the top 1% of the ATI distribution. The blue line indicates households at the top 5% of the ATI distribution. The purple line indicates households at the top 10% of the ATI distribution. Time in calendar years is on the horizontal axis. Business income held at the top, as a fraction of aggregate business income, is on the vertical axis.

Figure 4: Labor Income at the Top

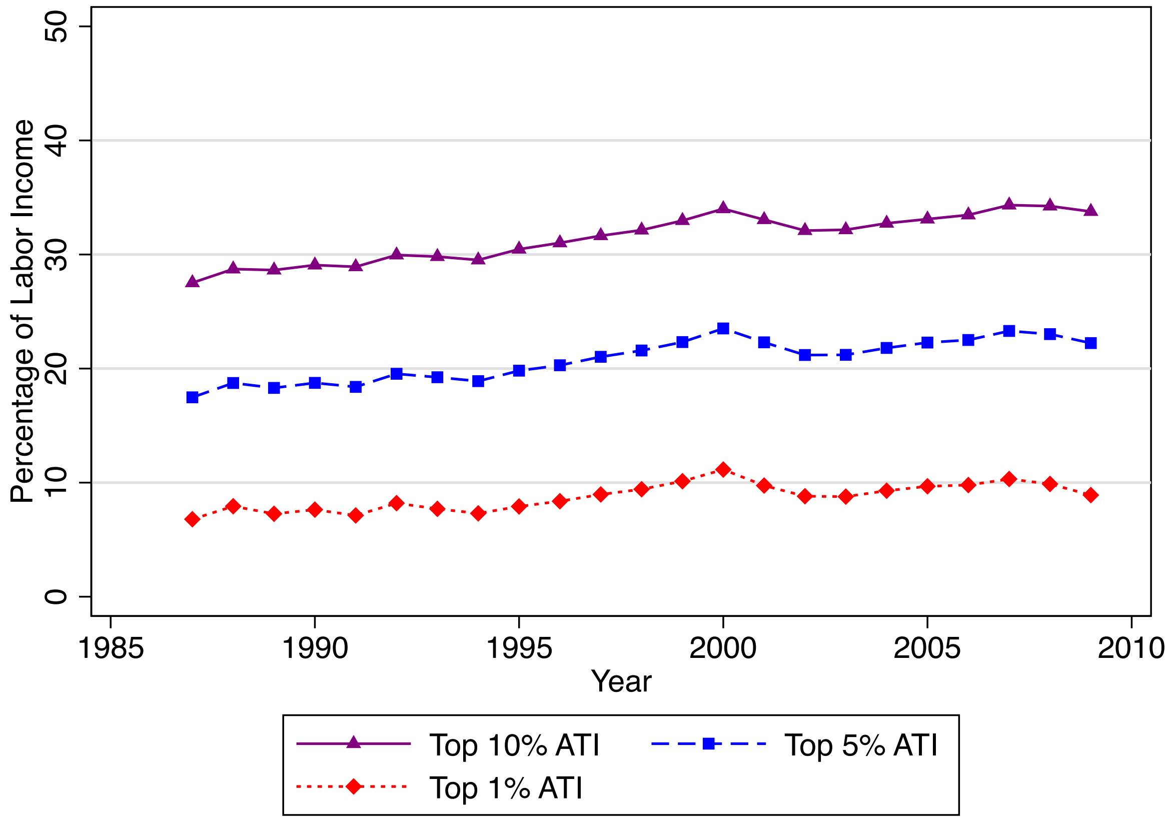

Figure 4 shows the concentration of labor income at the top of the adjusted total income (ATI) distribution. The data are from our annual cross sections, 1987-2009. The red line indicates households at the top 1% of the ATI distribution. The blue line indicates households at the top 5% of the ATI distribution. The purple line indicates households at the top 10% of the ATI distribution. Time in calendar years is on the horizontal axis. Labor income held at the top, as a fraction of aggregate labor income, is on the vertical axis.

Figure 5: Business Income Distribution, 2009

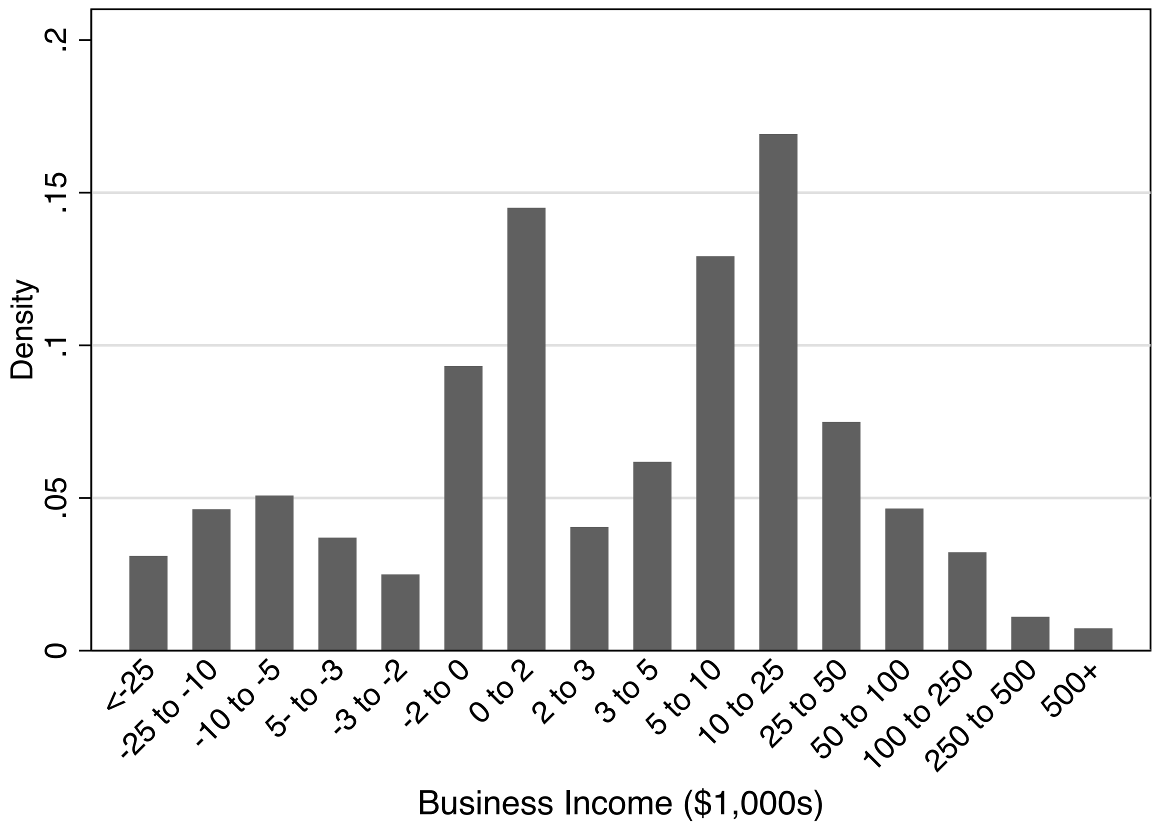

Figure 5 shows the distribution of business income from our 2009 cross sectional data file. Values are in real 2005 dollars. We restrict ages to 30-60, and exclude farmers. Sampling weights are used to represent the population.

Figure 6: Labor Income Distribution, 2009

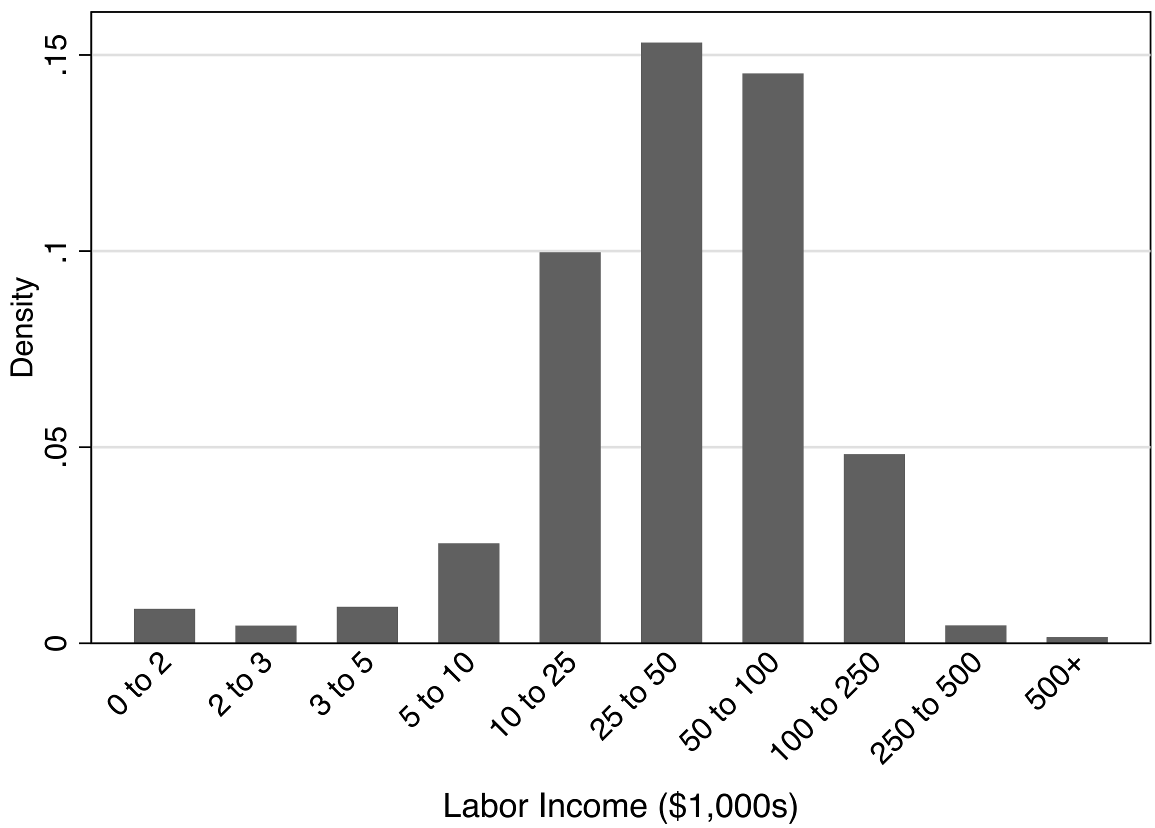

Figure 6 shows the distribution of labor income from our 2009 cross section data file. Values are in real 2005 dollars. We restrict ages to 30-60, and exclude farmers. Sampling weights are used to represent the population.

Figure 7: One Year Distribution of Percent Changes

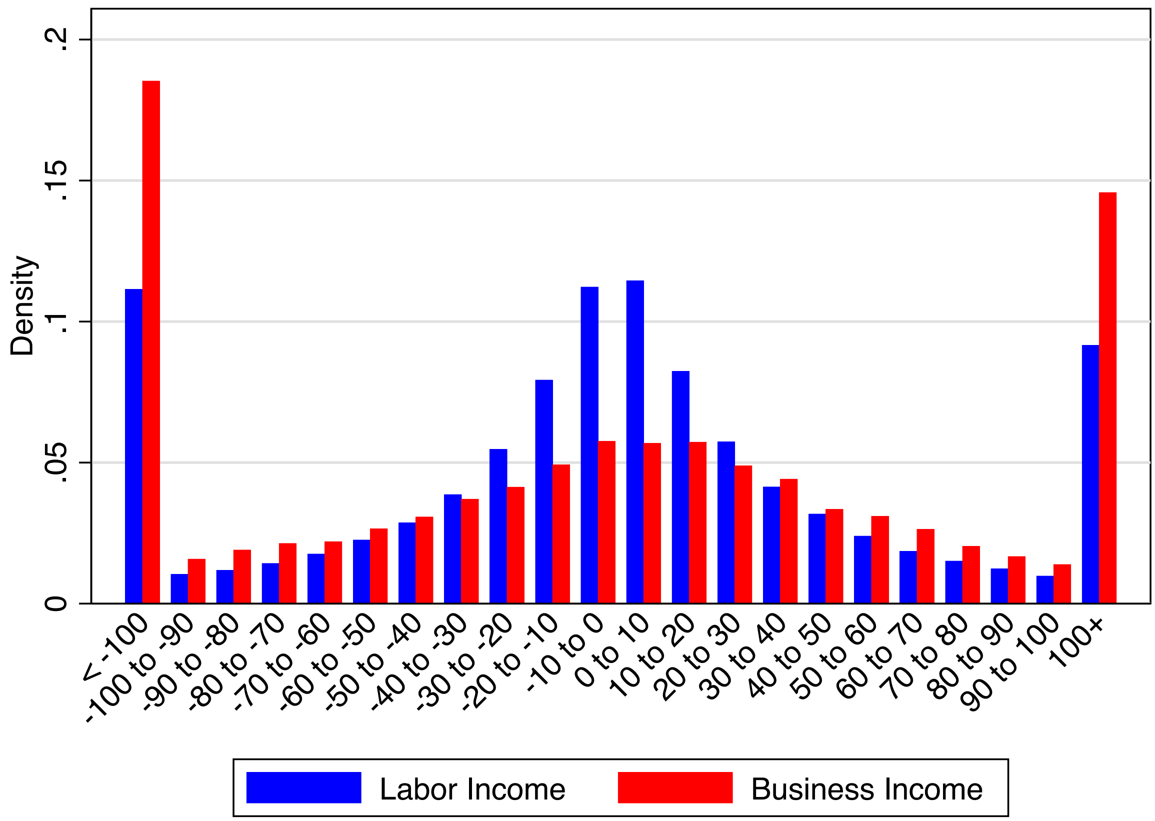

Figure 7 presents the (pooled) distribution of the size of the one year percent changes in income residuals, for business and for labor income. The panels used to construct percent changes in income for each household are our benchmark 1987-2009 panels, which drop zero income observations, restrict ages to 30-60, and exclude farmers. The one year percent changes are calculated from income residuals from the first stage regression in (2), see text for details. The calculations of percent changes use the formula (3). The blue bars indicate business income and the red bars indicate labor income. 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 residual income increased by more than 100%. The last bin on the left groups together all observations for which residual 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.

Figure 8: Inverse Hyperbolic Sine (IHS) vs. Log

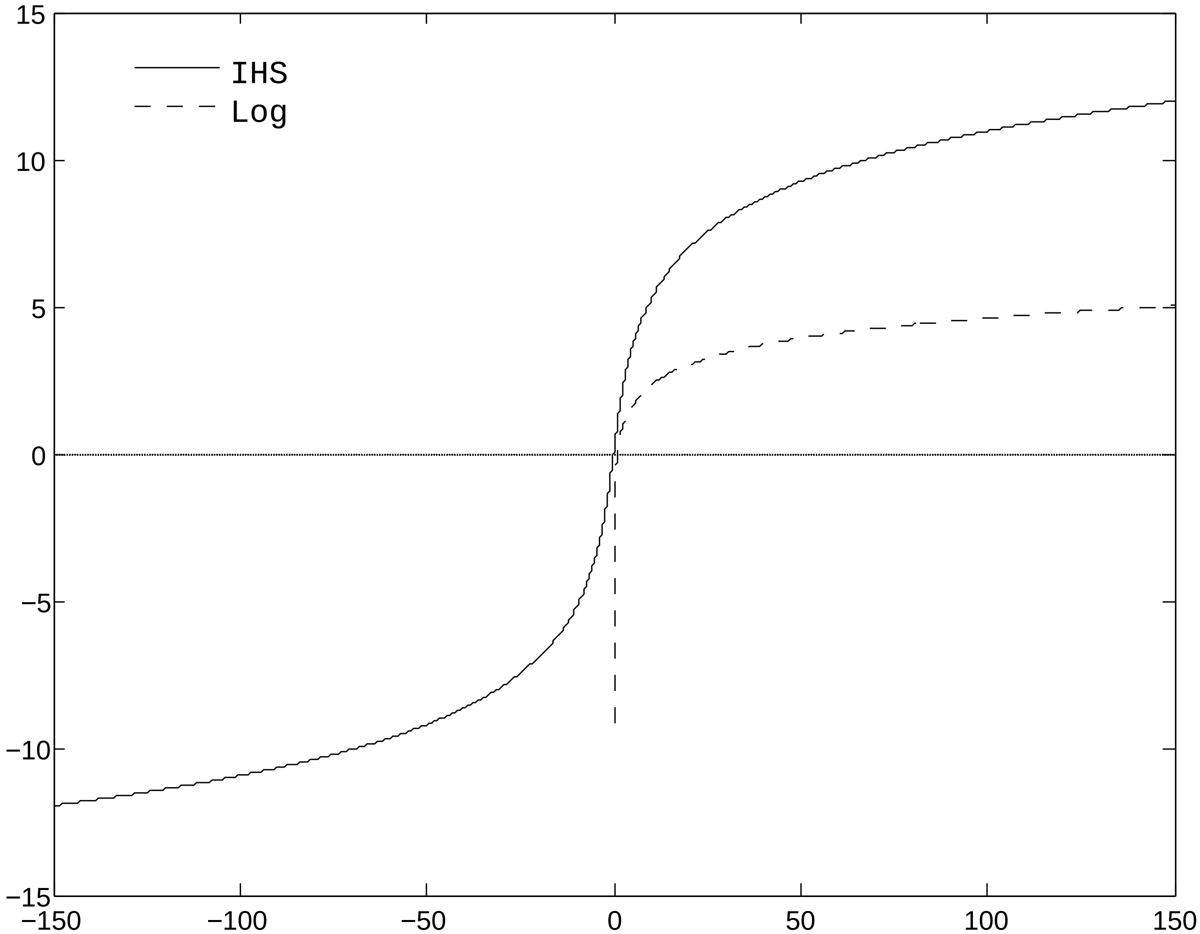

Figure 8 plots the Inverse Hyperbolic Sine (IHS), for a given location parameter, θ, versus the log function. Income, y, is on the horizontal axis. The log of income (dashed line) or the IHS of income (solid line) are on the vertical axis. The IHS of y is given in (9).

Figure 9: Risk vs. Heterogeneity, Business Income

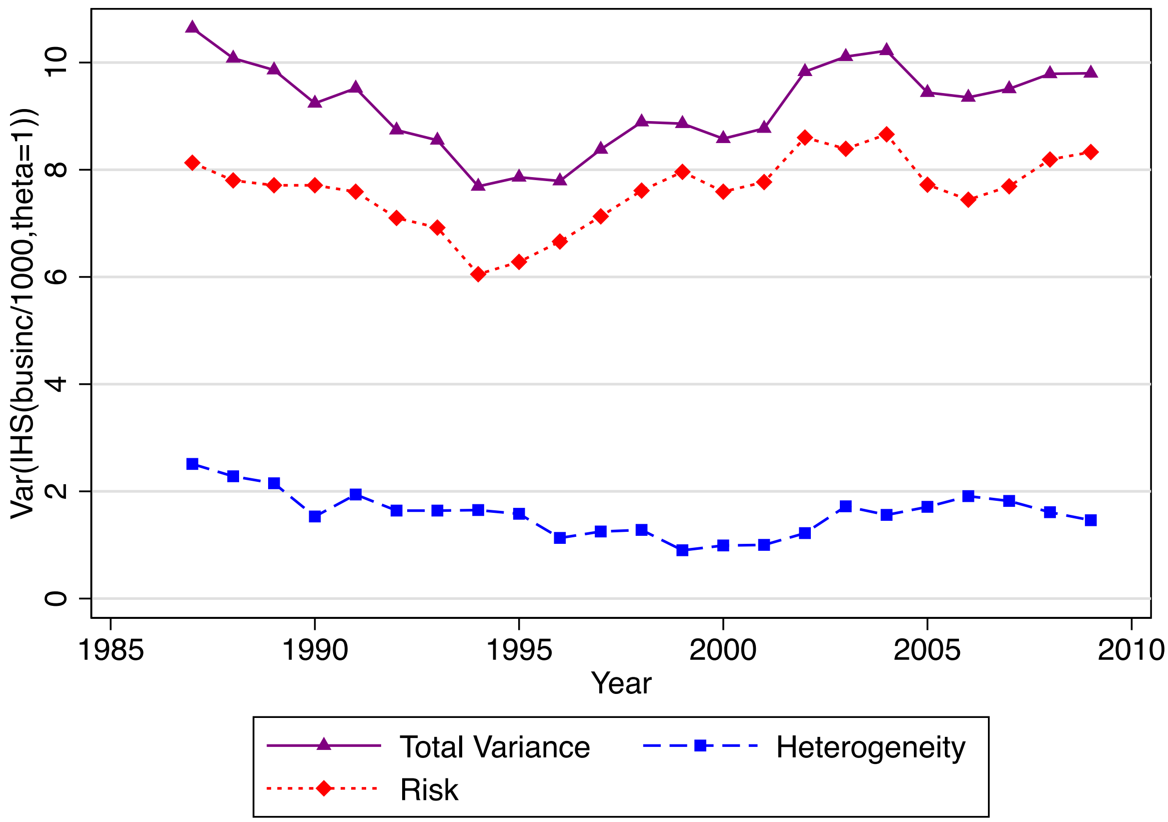

Figure 9 decomposes the total cross sectional variance into that due to the filer fixed effects (heterogeneity) that that due to residual risk by calendar year. The model used is the FE_RW_AR model, using parameter estimates from Table 5. See text for more details.

Figure 10: Risk vs. Heterogeneity, Labor Income

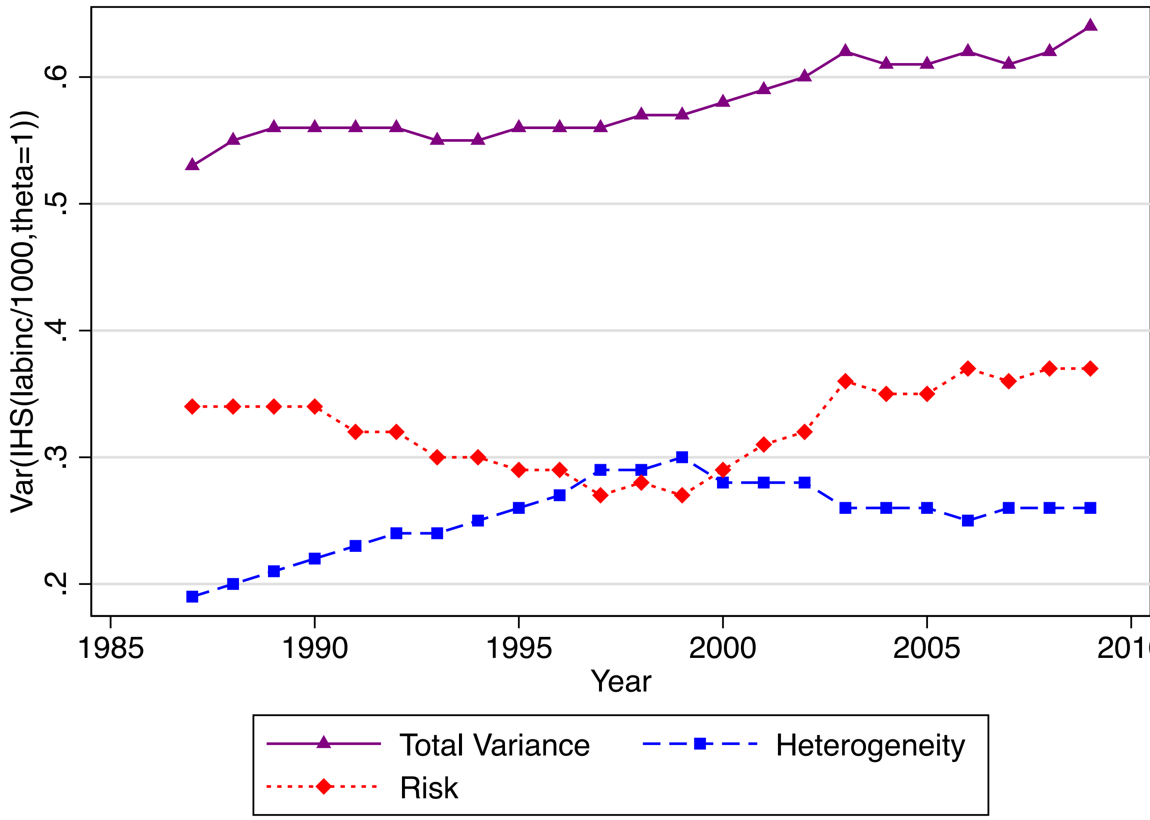

Figure 10 decomposes the total cross sectional variance into that due to the filer fixed effects (heterogeneity) that that due to residual risk by calendar year. The model used is the FE_RW_AR model, using parameter estimates from Table 6. See text for more details.

Figure 11: Decomposition of Variance, Business Income

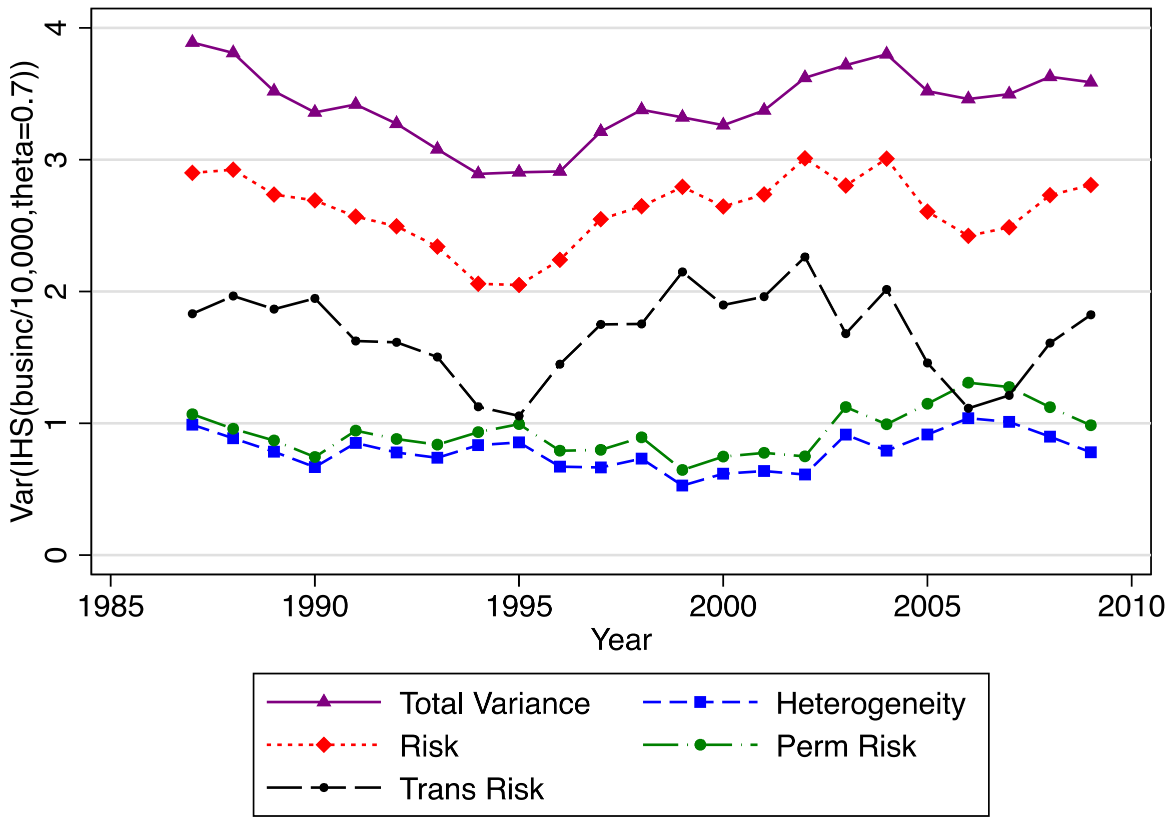

Figure 11 decomposes the total cross sectional variance into that due to the filer fixed effects (heterogeneity) and that due to permanent (random walk) and transitory risk (AR(1)) by calendar year. The model used is the FE_RW_AR model, using parameter estimates from Table 5. See text for more details.

Figure 12: Decomposition of Variance, Labor Income

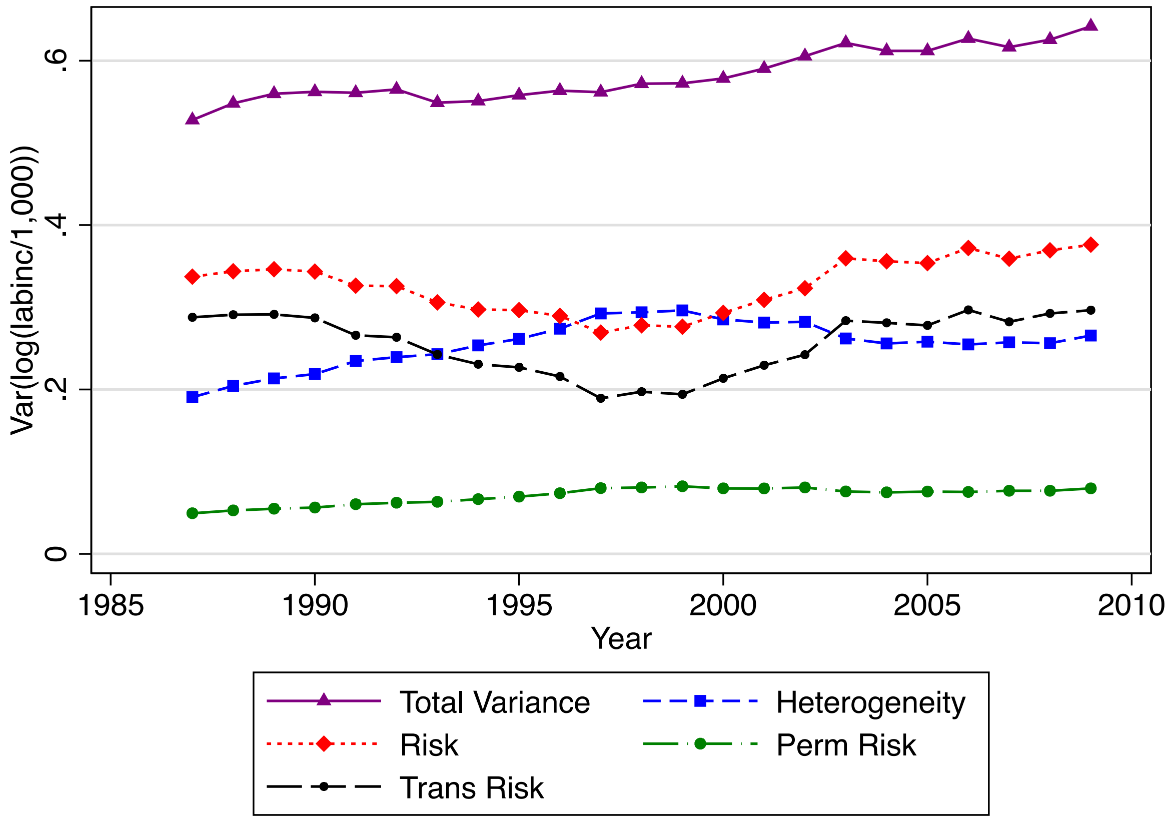

Figure 11 decomposes the total cross sectional variance into that due to the filer fixed effects (heterogeneity) and that due to permanent (random walk) and transitory risk (AR(1)) by calendar year. The model used is the FE_RW_AR model, using parameter estimates from Table 6. See text for more details.