# Heterogenous Income Risk and Consumption Response

Kwanjai Yoo *†

July 7, 2024

Link to the latest version

**Abstract**

This paper examines heterogeneous income risk and its impact on consumption insurance. A novel framework is proposed wherein household-specific variances of both persistent and transitory income shocks are considered. Using the Panel Study of Income Dynamics (PSID), this study empirically investigates several key aspects of heterogeneous income risk. First, the distribution of household-fixed income volatility exhibits right-skewness with a fat tail, posing a challenge for conventional income dynamics literature. Second, a considerable portion of the heterogeneity in income risk remains unexplained by observable characteristics, highlighting latent factors. Third, households experiencing more volatile transitory risk tend to exhibit less consumption response. I estimate an income process under the assumption of reliable parametric income shocks to capture the income volatility distribution and consumption response to heterogeneous income risk. Quantitatively, I use a standard life-cycle incomplete market model to demonstrate that the model's predictions align consistently with empirical estimates.

**Keywords**: Income volatility, heteroscedasticity, consumption insurance, precautionary saving, life cycle

**JEL Classification Code**: C23, E21, J31

*Acknowledgements: I am immensely grateful to David Wiczer and Juan Carlos Conesa, for invaluable guidance and support. I am grateful to Eva Carceles-Poveda, Alexis Anagnostopoulos, Gabriel Mihalache, Steven Stern, and all participants at the macro workshop at Stony Brook University and at the presentation at Atlanta Fed. All errors are mine.

†Contact information: Department of Economics, Stony Brook University, USA. Email: kwanjai.yoo@stonybrook.edu

# 1 Introduction

Income risk is a key factor in household consumption and saving decisions and a fundamental element in understanding inequality in various dimensions, such as consumption, income, and wealth, and the impact of public policies like income tax and insurance. This paper studies the heteroskedasticity of income risk in which household-specific variance of income shock is imposed. I use the Panel Study of Income Dynamics (PSID) and find that household income volatility distribution is skewed with a fat right tail. In other words, most households experience minor fluctuations in income, whereas a minority experience significant variations over time. Neither the canonical income dynamics nor recent studies with non-Gaussian and non-linear features can capture the skewed volatility distribution.

I address two primary questions to fill the gap between the empirical evidence and previous studies. The first question is how much households have heterogeneous income risk in terms of fluctuation and, if so, whether the difference can capture the skewed distribution of income volatility. In addition, I assess how much the heterogeneity of income shock is explained by demographics or labor market outcomes, or remains latent. To answer the question, I propose heterogeneous income risk such that variances of both persistent and transitory shock is household-specific. It is supported by the findings that most heterogeneity of income volatility remains unexplained, so unobserved latent heterogeneity is captured by the shock component which is ex-ante heterogeneous in the second moment. I find only about 10 percent of income volatility heterogeneity is explained by observable characteristics.

The second one is how much consumption response changes with household-specific variance at aggregate and micro levels. With different levels of income volatility, households should self-insure differently. For example, imagine two household with different volatility of income. Precautionary savings motive make a household with more volatile income save more than the other with less volatile income against same level of shock because the former wants to have more buffer stock for the hard time that is more likely to come in the future. Therefore, the model implications for degree of consumption response change with or without heterogeneous income risk in micro levels. In aggregate perspective, heterogeneous income risk helps to predict how much aggregate consumption increases by the fiscal transfers more accurately.

This paper has three main contributions. First, I propose a joint estimation of income dynamics and consumption response with a new parametric assumption that helps to capture heterogeneous income risk. By assuming household-specific variance of income innovation distributed asymmetrically, both cross-sectional autocovariance of income and consumption change and the distribution of household income volatility are explained well. Technically, I use the simulated method of moments with a particle swarm global optimization algorithm that enables efficient solution finding.

The second contribution is to provide empirical evidence of consumption response heterogeneity arising from heteroskedasticity of income risk. Using the PSID, which covers more than 50 years, I jointly track the long history of each household's income fluctuation and consumption change. Furthermore, I exploit comprehensive information on demographics such as age, education, household composition, and labor market outcomes like firm size, unemployment, occupation, and industry. This enables me to answer how much the observables capture the heterogeneous income volatility.

Finally, I use a quantitative incomplete market life-cycle model and find consistent consumption response implications with the estimates. The quantitative analysis helps to conduct counterfactual experiments to explore potential outcomes of various government insurance policies such as unemployment insurance or progressive income tax.

The main results of the paper are as follows. One is heterogeneous income risk in the second-order dimension, where the household is different across variances in risk. Besides, the most heterogeneity comes from transitory risk. Second, a large proportion of heterogeneous income risk remains latent after examining the source of the heterogeneity. A few observable variables have a significant relationship with the heterogeneity. For instance, the old cohort has relatively small volatility, but the marginal difference is almost negligible. Households with self-employment experience have volatile income. Nevertheless, other variables, such as unemployment duration, occupation, industry, or education, do not significantly impact volatility.

Third, households with more volatile incomes present more extensive consumption response against transitory shock. Even after controlling liquid wealth, which usually plays the role of buffer stock, there is a difference in consumption response. It is consistent with the precautionary savings motive of risk-averse agents, which is amplified in this study by differentiating the second-order moment of income risk.

This paper is in line with two big strands of the literature. One is to study income dynamics of households or individuals. The literature has developed from canonical permanent and transitory decomposition (e.g., Geweke and Keane, 2000; MaCurdy, 1982) to novel developments with rich specification to match higher-order moments or time-varying features (Arellano et al., 2017; De Nardi et al., 2020; Guvenen et al., 2021; Karahan and Ozkan, 2013). All this literature, however, focuses on cross-sectional distributions in the data. I contribute to the literature by documenting the importance of household-specific fluctuation moments and suggesting a heterogeneous income risk idea to match the household-specific moment.

This paper is one of many to discuss the heteroscedasticity of income risk. Meghir and

Pistaferri, 2004 estimate the conditional variance of income shocks and find education—and time-specific heterogeneity and the remaining unobserved heterogeneity. Almuzara, 2020; Botosaru, 2023 provide sufficient conditions for nonparametric identification and estimation. This paper is similar to theirs regarding the household-specific variance of income risk but is dissimilar in its identification strategy and incorporation of consumption response.

The other branch is the literature studying consumption response against income shock. Blundell et al., 2008 (BPP, hereafter) estimates partial insurance of consumption for permanent and transitory income shock with semi-structural model. The semi-structural approach has become the primary method of jointly studying consumption and income dynamics. I follow this approach at a basic level with the extension of heteroskedastic income risk. Arellano et al., 2023 documents unobserved heterogeneity in consumption response. The heteroscedasticity of income risk enables me to account for the heterogeneity in unobserved consumption.

# 2 Empirical Facts

## 2.1 Data

The primary data I use is the PSID 1977-2019. The PSID started in 1968 as an annual survey, and since 1999, it has collected data biennially with more detailed variables of consumption expenditure and wealth. It is the most comprehensive longitudinal dataset in the U.S. that measures income, consumption, and wealth jointly. The primary sample is the PSID core sample without the Survey of Economic Opportunity (SEO) families who are low-income households. I select households with male heads aged between 25 and 60 because most cases with female heads are single, divorced, or widowed, and they are not the interest of this study.

Along with the sex of heads, family composition is limited to married couples with or without children. These criteria mostly overlap with the male head restriction. Also, households with no change or minor change in members other than head or wife in their family composition remain. Once a household has experienced a significant change in its family composition, such as divorce or the death of the head or wife, the household is treated as a new unit. It is necessary to rule out the potential that a large change in income and consumption caused by a change in family composition is misinterpreted as a risk because the change is the outcome of endogenous family composition choice. I follow Blundell et al., 2016 for the technical details to sort out those families.

I have imposed other criteria, such as heads' working hours and minimum earnings requirements. The minimum annual working hours are set to 520 hours per year, at least 10 hours per week every 52 weeks. For the earnings requirement, the minimum labor income of the head is calculated by half the federal minimum wage multiplied by 520, the minimum annual working hours as mentioned above. It takes an unemployed spell of male heads out so that sizable persistent income shock caused by an extensive margin is excluded. Also, these minimum earnings and working hours requirements help me overcome problems with zeroes when I take logarithms on labor income. Also, I exclude samples with more than 5,110 hours of work, which is 14 hours of working every day in a year. In addition to labor income restriction, I drop the observations with more than $20 million of net worth. The last two requirements drop outliers out of the sample.

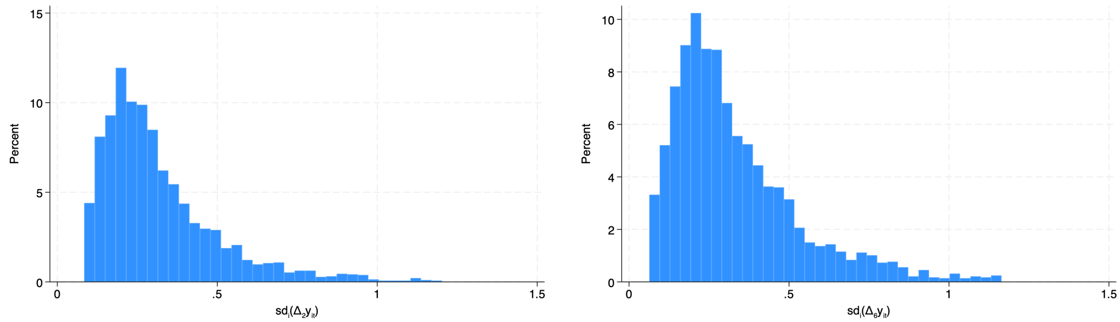
Lastly, I impose a duration requirement that households must satisfy all restrictions for more than ten interviews. Since the frequency of the dataset is mixed annually and biannually, the duration might vary across households but is guaranteed at least ten years. The duration requirement is necessary for this study since I focus on household-fixed income volatility, so a long enough income history is inevitable. The final sample contains 41,487 number of household-year observations with 2,922 unique households.

I use net disposable household income as a benchmark for income since that is the primary source of consumption choice. Since PSID does not provide disposable income or the detailed tax amount to be paid during the periods, I use gross household income and subtract total tax expenditures computed from TAXSIM. The version of TAXSIM I use is TAXSIM35. Total taxes include property tax provided directly from PSID, federal and state income taxes, and taxes under the Federal Insurance Contributions Act (FICA) computed from TAXSIM. Transfers include unemployment insurance, disability insurance, workers' compensation, among others. In the appendix, I will present the results using different income measurements, such as gross household income, male earnings, and total earnings of the head and wife.

The consumption variable I use is the sum of nondurable consumption and service expenditures. Nondurable consumption includes expenditures on food and gasoline. Expenditures on services include healthcare-related spending, utilities, spending on vehicles, public transportation, and education, such as tuition and rent. Homeowners' equivalent rent is 6% of home value following Blundell et al., 2016. All variables are deflated by the Consumer Price Index (CPI) of all urban consumers in the U.S. from the Bureau of Labor Statistics.

Following the standard approach to studying income and consumption dynamics, I use the residual log variable after ruling out observable components. I regress logarithm of income and consumption on a polynomial of age, year dummies, and a set of demographics dummy variables such as year of birth, education, family size, region, race, income recipients outside

Figure 1: Distribution of household-specific income volatility




(a) Transitory volatility (b) Persistent volatility

*Note:* Household-specific income volatility is measured by standard deviation of income growth over time by each household.

the household, and an indicator of the wife's employment. The polynomial of age captures common life-cycle components of income and consumption. Year dummies are to remove aggregate time effects such as cyclical components. A set of demographic dummy variables is included to extract heterogeneity among households and make the residuals comparable between different households.

## 2.2 Empirical Findings

**Fact 1: Household income volatility is distributed with fat tail to the right.** Denote $y_{it}$ as a residual log income of household $i$ at age $t$ and $\Delta_2 y_{it}$ as 2-year log income growth. I use 2-year growth as a benchmark due to the biennial frequency of the data. I measure household income volatility as $s.d._i(\Delta y_{it})$, the standard deviation of residual income growth at $t$ for every household $i$. It measure how much the income growth fluctuates over life-cycle by each household.

The histogram of household income volatility is presented in Figure 1. The distribution of volatility is skewed to the right. This means that the majority of households face little fluctuation in their income, but for some households, their income growth changes over time a lot. I take various income process models to examine whether the skewed distribution of household income volatility is captured or not. Specifically, I use two alternative income dynamics as true data-generating processes and simulate the income sequence of the same number of individuals with the data to calculate the same measure I use for income volatility.

One is the most canonical linear model of permanent and transitory shocks decomposition of income.[1]

$$
\begin{aligned}
y_{it} &= p_{it} + \varepsilon_{it} \\
p_{it} &= \rho p_{it-1} + \eta_{it} \\
\eta_{it} &\overset{iid}{\sim} N(0, \sigma_\eta^2), \;\; \varepsilon_{it} \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)
\end{aligned} \tag{1}
$$

The residual log-earnings of household $i$ at age $t$, $y_{it}$, is the sum of persistent component $p_{it}$ and transitory shock $\varepsilon_{it}$. The persistent component follows AR(1) process with the persistent shock $\eta_{it}$ and the transitory component is white noise. Note that both shocks $\eta$ and $\varepsilon$ are normally distributed and every household faces identical distribution that is independent over any dimension. It is effective to describe the exogenous income in quantitative model with simple Markov approximation to incorporate. However, it is too restrictive to capture higher-order empirical moments of earnings growth distribution that are documented lately.

The other is the Gaussian mixture model proposed by Guvenen et al., 2021 that targets higher-order cross-sectional moments of earnings growth and unemployment duration. The specification is

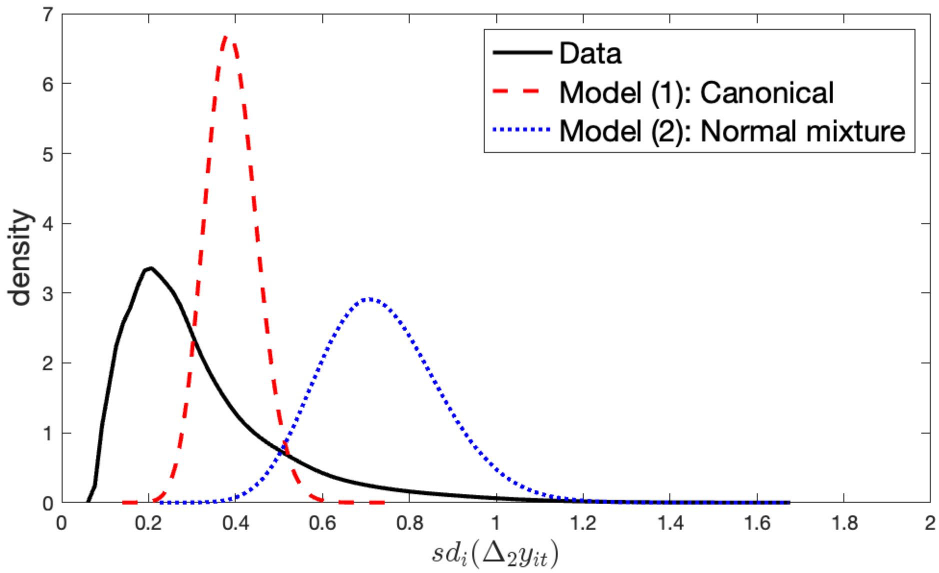
$$
\begin{aligned}
Y_{it} &= (1-\nu_{it}) \exp\left(\alpha_i + \beta_i t + p_{it} + \varepsilon_{it}\right) \\
p_{it} &= \rho p_{it-1} + \eta_{it} \\
\eta_{it} &\sim \begin{cases} N(\mu_{\eta,1}, \sigma_{\eta,1}^2) & \text{with prob. } \; p_\eta \\ N(\mu_{\eta,2}, \sigma_{\eta,2}^2) & \text{with prob. } \; 1-p_\eta \end{cases} \\
\varepsilon_{it} &\sim \begin{cases} N(\mu_{\varepsilon,1}, \sigma_{\varepsilon,1}^2) & \text{with prob. } \; p_\varepsilon \\ N(\mu_{\varepsilon,2}, \sigma_{\varepsilon,2}^2) & \text{with prob. } \; 1-p_\varepsilon \end{cases} \\
\nu_{it} &\sim \begin{cases} 0 \quad \text{with prob. } \; p_\nu(t, p_{it}) \\ \min\{1, \exp(\lambda)\} \quad \text{with prob. } \; 1-p_\nu(t, p_{it}). \end{cases}
\end{aligned} \tag{2}
$$

Let $Y_{it}$ denote residual gross earnings without taking logarithm. It is affected by $\nu$, an unemployment duration with probability $p_\nu(t, p_{it})$ which depends on age $t$ and persistent component $z_{it}$. Both persistent and transitory shocks are drawn from normal mixture distribution. The normal mixture shocks provide nonnormality of the earnings growth distribution

The comparison with various income dynamics and data aims to examine whether or not the model captures the skewed distribution of household-specific income volatility. In Figure 2, I plot the kernel density function of household income volatility. The solid black plot is the same distribution as the histogram in Figure 1. Note that both the canonical linear Gaussian

[1] See Meghir and Pistaferri, 2011 for a survey review.

Figure 2: Comparison of household-specific income volatility distribution




*Note:* Each plot shows the kernel density of the household-specific income volatility from different specifications. I simulate income sequences over the life cycle for the same number of households with the data and calculate the standard deviation of income growth by household. The solid line is same as the histogram in Figure 1.

model and recently developed models have difficulty capturing the skewness of the volatility distribution. The right tail of the data is larger than that of all the other models. Table 1 shows the descriptive statistics of Figure 2. Both models (1) and (2) somewhat capture the mean and median but have a difficulty to reproduce the skewness, as seen in Figure 2.

Table 1: Descriptive statistics of income volatility distribution

| Statistics | Data | Model (1) | Model (2) |
|---|---|---|---|
| Mean | 0.316 | 0.391 | 0.740 |
| Median | 0.269 | 0.389 | 0.735 |
| Min | 0.084 | 0.216 | 0.255 |
| Max | 1.202 | 0.611 | 1.401 |
| Skewness | 1.609 | 0.209 | 0.379 |

**Fact 2: The skewed income volatility is driven by latent heterogeneity.** Next, I look for answers to whether this asymmetric household-specific income volatility is caused by heterogeneous income risk and how much heteroskedasticity is explained by observable characteristics. To do so, I implement group fixed-effect multinomial logit to determine which observable covariates significantly predict the group assignment and how much overall as-

signment is explained by the observables.[2] The method is proposed by Bonhomme et al., 2022 and useful to find fixed effects by group. Due to the asymmetrically distributed volatility, there is likely a non-linear relationship between observable characteristics and income volatility.

The $k$-means clustering is an unsupervised learning algorithm and a data-driven rule of splitting observations into discrete numbers of groups. I discretize $N$ number of households into $J$ number of groups over $K$ household-specific *time-invariant* moments ($K = 2$: standard deviations of 2-year log-income growth $s.d_i(\Delta_2 y_{it})$ and 6-year log-income growth $s.d._i(\Delta_6 y_{it})$). Each household $i$ is assigned to group $j$ which represents time-invariant volatility type. The assignment is to minimize the Euclidean distance from the group center for all observations. The optimal assignment $j^*(i)$ for all household $i$ is

$$\begin{aligned} j^*(i) &= \arg\min_{j(i)} \sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{k=1}^{K} \mathbf{1}[j = j(i)](m_{k,i} - m^*_{k,j})^2, \\ s.t. \quad m^*_{k,j} &= \frac{\sum_{i=1}^{N}\sum_{j=1}^{J} \mathbf{1}[j = j(i)] m_{k,i}}{\sum_{i=1}^{N}\sum_{j=1}^{J} \mathbf{1}[j = j(i)]}, \end{aligned} \tag{3}$$

where $m_{k,i}$ is the moment $k$ of the household $i$ and $m^*_{k,j}$ is the centroid moment within the group $j$.[3]

Clustering methods require the choice of the number of $J$ and the optimal number of groups. I combine two approaches to decide the number of groups: the elbow method and the Caliński-Harabasz pseudo-$F$ index.[4] The elbow method is a visualization method based on the with-in cluster sum of squares (WCSS), which is the objective function of $k$-means clustering in equation 3. WCSS is increasing in the number of clusters by construction. Using the plot of WCSS or logarithm of WCSS, the elbow method finds a kink or "elbow" where the marginal decrease of WCSS from adding one more group is not as great as the compared group. In addition, the coefficient of determination or the proportional reduction of error (PRE) is used together with WCSS. The coefficient of determination is similar to the $R^2$ in that it calculates the ratio of explained variation by the cluster average to the total variation. The PRE illustrates marginal decrease of WCSS by adding one more number of group.

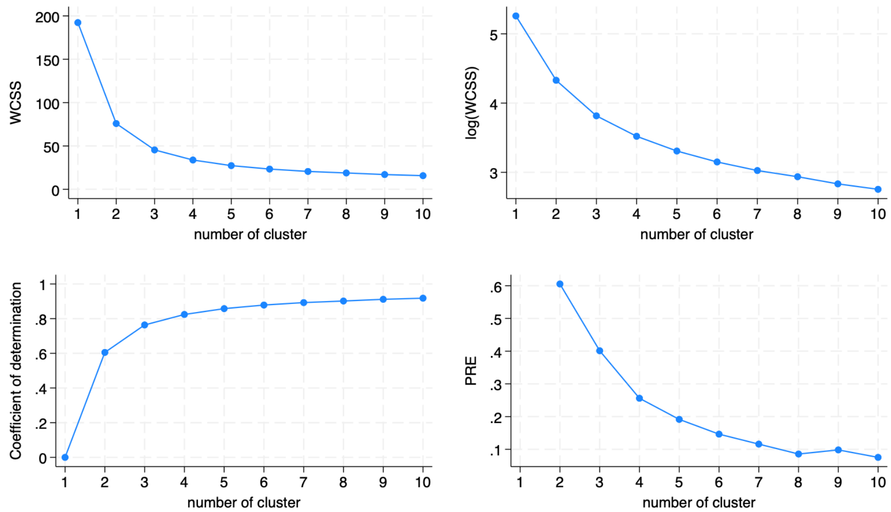
Note that the objective of the elbow method is to find $J$, the optimum number of groups, by finding the cutoff point on the graphs. On the top-left panel in Figure 3, for instance, WCSS decreases concavely, and the marginal decrease with more groups gets smaller with

[2] In Appendix

[3] I use Stata's iterative algorithm to find group indicators $j(i), \forall i = 1, \cdots, N$

[4] The details of the measures used for the elbow method and the Caliński-Harabasz pseudo-$F$ index are in Appendix A.

Figure 3: Number of groups using the elbow method




*Note:* In order to find the optimal number of clusters in $k$-means clustering, the elbow method provides visualization of the with-in cluster sum of squares (WCSS) and additional metrics. The goal is to find a kink point on the graphs where the marginal contribution of adding one more cluster is insignificant.

3 or 4 number of groups. The elbow method can easily observe the slope change in the measures from the graphs and predict the optimum $J$. Nevertheless, the exact kinked point does not exist due to the monotonically decreasing feature of WCSS by construction, and the "elbow" is ambiguous in determining only based on the plots.

The Caliński-Harabasz pseudo-$F$ index (CHI) complements to the elbow method in providing the optimum $J^*$ with the highest value. The CHI measures the ratio of the between-cluster sum of squares (BCSS) to the WCSS. The BCSS measures how far each cluster is separated away from each other. Therefore, the BCSS penalizes an increase in the number of groups, and it helps to find the optimal number of groups balanced between the compactness of each cluster and the distinctiveness between clusters. In Table 2, the CHI is the highest when the population is divided into three subgroups.

With the $J = 3$, I use multinomial logit regression to find how the observable demographics and labor market outcome contribute to the group assignment. I call each cluster a low, medium, or high volatility group for convenience. Table 3 shows the descriptive statistics of the volatility of income by each group. The low volatility group is the largest, with about 58 percent of households in the data. The medium and high groups comprise 32 and 10 percent of the population. Note that groups with higher volatility have more considerable variation within the group. This is because the volatility distribution has a long right tail, as seen in

Table 2: Optimal number of groups

| Number of groups $J$ | Caliński-Harabasz psuedo-$F$ index (CHI) |
|---|---|
| 2 | 4331.72 |
| 3 | 4561.68 |
| 4 | 4409.88 |
| 5 | 4256.51 |
| 6 | 4084.11 |

*Note:* A higher value of CHI indicates a better clustering because it means that the data points are more spread out between clusters than they are within clusters. In this case, the optimal number of group $J^*$ is 3 with the highest CHI.

Figure 1.

Table 3: Descriptive statistics of income volatility by group

| | Low | Medium | High |
|---|---|---|---|
| transitory volatility (2-year growth) | | | |
| Mean | 0.21 | 0.40 | 0.69 |
| Variation | 0.0637 | 0.0984 | 0.1708 |
| persistent volatility (6-year growth) | | | |
| Mean | 0.21 | 0.42 | 0.76 |
| Variation | 0.0684 | 0.0945 | 0.1513 |
| Number of households | 1,653 | 897 | 276 |
| Share of households | 0.58 | 0.32 | 0.10 |

Next, I run group fixed effects multinomial logit regression of group assignments on various observables. Following Gregory et al., 2021, I examine how much heterogeneity in income volatility is explained by looking at the marginal likelihood contribution of covariates on each group assignment. The regressors contain demographics, occupations, industries, and labor market outcomes. Demographics include birth year, education (less than high school, high school, college and higher), and race (white, black, others). Occupations are categorized into six groups following Autor and Dorn, 2013 using Census classification. For industries, I use 12 aggregate groups following the 1970 NAICS code. Labor market outcome variables are the number of weeks unemployed and out-of-labor force, job tenure, and whether an individual is enrolled in a union. If both head and wife's information is available, I use all of them.

Table 4 presents the results of the multinomial logit regression. To maintain readability, only variables with significant coefficients in at least one group assignment are included. [5]

[5]For the full results, see Appendix.

The baseline outcome is the low volatility group, so the estimates are interpreted as how more or less likely households belong to the middle or high volatility groups.

Table 4 reveals several crucial findings. First and foremost, only about 13 percent of group assignments can be explained by the observables, as indicated by the the pseudo-$R^2$. Despite the inclusion of a wide variety of covariates that could potentially be correlated with income volatility, the majority of heterogeneity in income volatility remains unexplained. This finding suggests the possibility of heterogeneity as a form of unexplained components, such as the heterogeneous income risk proposed in this paper.

Second, most demographic variables have significant predictions and are qualitatively reasonable for the medium volatility group assignment. For instance, households with younger heads are likely to be low volatility groups. Higher-educated individuals are less likely to have medium volatility; in other words, they tend to have less volatile income than less-educated people. There is a racial difference between black and white households in that black households are likely to have more volatile income. Households with self-employment experience, whether they fail or not, are more likely to be the medium volatility group. It has the most considerable likelihood contribution among other demographics.

In contrast, only a few demographics and industry variables are significant for the high volatility group assignment prediction. In other words, the extremely high volatile income is rarely explained by the observable characteristics. Still, self-employment experience and the mining industry present a significantly positive increase in the likelihood of being in the high group.

Third, occupations are not a significant factor in capturing the volatility of income. Note that there is no single occupation variable in Table 4 even though they are included in the regressors. When I run the same regression without industry variables, occupations significantly predict the volatility group. For instance, service occupations are likely to have the highest income volatility, whereas managerial and professional, production, or technical, sales and administrative support occupations are the opposites. However, the occupational effects turn out to be industry effects once industries are included in the regression. Out of 12 categories, four industries, agriculture, mining, finance, and entertainment, report a positive effect of being the medium volatility group, and only the mining industry is the significant factor of the high volatility group.

Fourth, labor market outcomes predict in a qualitatively expected way. The longer the heads are unemployed, the more likely they are to have higher income volatility, but not extremely. A wife's duration out of the labor force significantly increases the chance of having higher volatility. Job tenure matters because the longer period with one employer gives less volatile income. The longer job tenure means more stability in the job with less

Table 4: Volatility group prediction by observables

| | Middle | High |
|---|---|---|
| **Demographics** | | |
| birth year | -0.017** | -0.042** |
| | (-3.02) | (-4.91) |
| High school graduate | -0.337* | 0.011 |
| | (-2.45) | (0.05) |
| College graduate | -0.400*** | 0.167 |
| | (-2.75) | (0.71) |
| Black | 0.356*** | 0.359 |
| | (2.97) | (1.71) |
| Self-employment experience | 1.131*** | 1.568*** |
| | (9.88) | (8.93) |
| **Industry** | | |
| Agriculture | 1.340* | 1.545 |
| | (2.24) | (1.54) |
| Mining | 1.108** | 2.301*** |
| | (2.74) | (3.92) |
| Finance, Insurance, and Real Estate | 0.563* | 0.022 |
| | (2.15) | (0.04) |
| Entertainment | 0.986*** | 0.743 |
| | (2.76) | (1.19) |
| **Labor market outcomes** | | |
| # of weeks of head's unemployment | 0.0455* | 0.058 |
| | (2.07) | (1.81) |
| # of weeks of wife's unemployment | 0.028 | 0.031 |
| | (1.40) | (0.99) |
| # of weeks of head's out-of-labor force | 0.063 | 0.101* |
| | (1.88) | (2.08) |
| # of weeks of wife's out-of-labor force | 0.020*** | 0.039*** |
| | (3.57) | (5.03) |
| # of weeks of head's job tenure | -0.003*** | -0.009*** |
| | (-3.65) | (-5.38) |
| # of weeks of wife's job tenure | -0.006*** | -0.013*** |
| | (-5.12) | (-5.66) |
| Pseudo $R^2$ | 0.13 | |
| $N$ | 897 | 276 |

$t$ statistics in parentheses

*Note:* This table shows the results of group fixed effect multinomial logit regression of group assignment on observables. Only variables with significant coefficients at least one group assignment are presented here.

unemployment or negative risk as well as less upward mobility, which potentially induces an earnings jump.

# 3 Estimation

I introduce a new income process capturing the heterogeneous household income volatility documented in the previous section. The key assumption is a household-specific distribution of the income shock, where every household has different variances in both persistent and transitory shocks. The log of real income and consumption of household $i$ at age $t$ is denoted by $y_{it}$ and $c_{it}$, respectively. Both variables are residualized with the same controls to extract residuals for both income and consumption, as I described in the previous section.

**Heterogeneous risk income process** The log income $y_{it}$ is decomposed into a persistent component and a transitory shock,

$$y_{it} = p_{it} + \varepsilon_{it}, \;\; \varepsilon_{it} \sim N(0, \sigma^2_{\varepsilon,i}) \tag{4}$$

$$p_{it} = \rho p_{it-1} + \eta_{it}, \;\; \eta_{it} \sim N(0, \sigma^2_{\eta,i}). \tag{5}$$

I denote the persistent component by $p_{it}$, which follows the AR(1) process with the persistence $\rho$, the persistent shock by $\eta_{it}$, and the transitory shock by $\varepsilon_{it}$. Note that both $\eta_{it}$ and $\varepsilon_{it}$ are drawn from household $i$-specific distribution with the variance $\sigma^2_{\eta,i}$ and $\sigma^2_{\varepsilon,i}$, respectively. Therefore, the income shock is not identical across households. I assume a lognormal distribution for the variance of shocks:

$$\sigma^2_{\eta,i} \sim Lognormal(\tilde{\mu}_\eta, \tilde{\sigma}^2_\eta), \quad \sigma^2_{\varepsilon,i} \sim Lognormal(\tilde{\mu}_\varepsilon, \tilde{\sigma}^2_\varepsilon). \tag{6}$$

Note that both $\tilde{\mu}$ and $\tilde{\sigma}$ govern the dispersion of the variance of the shock. If $\tilde{\sigma}$ goes to zero or $\tilde{\mu}$ is small, all shocks are homogeneous across households and, if $\tilde{\sigma}$ diverges to infinity or $\tilde{\mu}$ is big, every single household has significantly heterogeneous in terms of shock variance.

The heterogeneous variance of the shock is the main contribution of the paper and plays a role of capturing unobserved heterogeneity in income volatility across households. A choice of lognormal distribution is motivated by right-skewness of income volatility distribution. The statistical features I look for are strictly positive support and right-skewness. It is parsimonious but also effective to target the empirical findings. I assume the skewed distribution of income volatility comes solely from heterogeneity in variance of income shocks. The rationale is backed up by the fact that most income volatility remains latent after examining potential observed sources in the previous section. There are other parametric assumption on variance of shocks such as gamma or Pareto distribution, which have similar statistical characteristics with lognormal distribution. Lognormal distribution turns out the best fitting assumption on the estimation target. I will provide estimation results from the different choices of parametric assumption in appendix.

There are recent papers documenting heterogeneous income risk nonparametrically with similar features. Botosaru, 2023 assumes time-varying and individual-specific unobserved heterogeneity in earnings shocks and provides nonparametric identification results of the density of the variances. Also, she find that despite time-varying density assumption, the estimates of the density shows relatively constant over time for most percentiles. Almuzara, 2020 proposes heterogeneous transitory risk and allows for potential correlation between permanent component and transitory shock. My results show that persistent shock is almost identical, consistent with Almuzara, 2020, though I allow for heterogeneity in persistent shocks, too.

**Consumption response equation** Another goal of the study is to investigate consumption response with heterogeneous income risk. To do so, I follow the partial insurance specification of BPP,

$$\Delta c_{it} = \phi_{it}\eta_{it} + \psi_{it}\varepsilon_{it} + \zeta_{it}. \tag{7}$$

where the income shocks $\eta_{it}$ and $\varepsilon_{it}$ change consumption with the factors $\phi_{it}$ and $\psi_{it}$, pass-through coefficients, and $\zeta_{it}$ captures any change in consumption independent of income. In the equation (7), the pass-through coefficients are characterized by the following formula:

$$\phi_i = \frac{Cov(\Delta c_{it}, \eta_{it})}{\sigma^2_{\eta,i}}, \quad \psi_i = \frac{Cov(\Delta c_{it}, \varepsilon_{it})}{\sigma^2_{\varepsilon,i}}. \tag{8}$$

The interpretation of the coefficients is the share of the variance of the shock translated into consumption change (Kaplan and Violante, 2010). With household-specific variances, the denominators in equation 8 are different by every unit. If all households change consumption equally against same level of shock regardless of the variance, then those with more volatile income necessarily have low consumption response. However, it might not be the case. Intuitively, households with volatile income are likely to have higher precautionary savings motive since they need larger buffer stock to self-insure than others with less volatile income. Then, their covariance between consumption change and income shock, the numerator, is larger than the other group.

Heterogeneous income risk can be a potential reason of unexplained heterogeneity in consumption response. For instance, Arellano et al., 2023 find that consumption responses vary substantially with unobserved types. They examine potential sources of the heterogeneity such as preference, latent pattern of asset accumulation, and inter-generational links. Lewis et al., 2019 also document latent heterogeneiety in the marginal propensity to consume (MPC). They explore various observables such as homeownership, mortgage remainder, salary, and demographics for potential drivers of the MPC variation and find only about 6% of variation is captured. The heterogeneous income risk I propose in this paper can be a significant cause of the heterogeneity in consumption.

## 3.1 Identification of parameters

The parameters to identify are $\rho, \tilde{\mu}_\eta, \tilde{\sigma}_\eta, \tilde{\mu}_\varepsilon, \tilde{\sigma}_\varepsilon, \phi, \psi$ and $\sigma_\zeta$. To retrieve the parameters of lognormal distributions of the shocks $\tilde{\mu}_\eta, \tilde{\mu}_\varepsilon, \tilde{\sigma}_\eta$ and $\tilde{\sigma}_\varepsilon$, the second and fourth central moments of $\sigma_{\eta,i}$ and $\sigma_{\varepsilon,i}$ have to be available from the data. Formally, from the lognormal assumption 6,

$$\tilde{\mu}_j = \ln\left(\frac{\left(\mathbb{E}\left[\sigma_{j,i}^2\right]\right)^2}{\mathbb{E}\left[\sigma_{j,i}^4\right]}\right), \quad \tilde{\sigma}_j = \ln\left(\frac{\mathbb{E}\left[\sigma_{j,i}^4\right]}{\left(\mathbb{E}\left[\sigma_{j,i}^2\right]\right)^2}\right) \tag{9}$$

for $j = \eta, \varepsilon$. I exploit the second and fourth central moments of income growth to find the corresponding moments of the standard deviations.

Here I temporarily assume a random walk persistent component for simplicity.[6] The income growth is characterized by

$$\Delta y_{it} = \eta_{it} + \Delta\varepsilon_{it} \tag{10}$$

and due to the household-specific distribution assumption, the law of iterated expectation is necessary to find the moment conditions. First, household-$i$ specific variances of shocks are identified as

$$\sigma_{\eta,i}^2 = \mathbb{E}\left[\Delta y_{it} \times \left(\Delta y_{it-1} + \Delta y_{it} + \Delta y_{it+1}\right)|i\right] \tag{11}$$

$$\sigma_{\varepsilon,i}^2 = -\mathbb{E}\left[\Delta y_{it} \times \Delta y_{it+1}|i\right]. \tag{12}$$

Taking unconditional expectations on the both sides yields the mean of $\sigma_{\eta,i}^2$ and $\sigma_{\varepsilon,i}^2$ equals to the unconditional expectation of the income growths on the right hand side.

The fourth central moments $\mathbb{E}\left[\sigma_{\eta,i}^4\right]$ and $\mathbb{E}\left[\sigma_{\varepsilon,i}^4\right]$ are identified by taking unconditional

[6] The non-unit persistence complicates algebra without adding meaningful intuition at this point. The full identification is in Appendix.

expectations on the following equations

$$3\sigma_{\eta,i}^4 + 8\sigma_{\eta,i}^2\sigma_{\epsilon,i}^2 + 4\sigma_{\epsilon,i}^4 = \mathbb{E}\left[(\Delta y_{i,t})^2 \times (\Delta y_{i,t-1} + \Delta y_{i,t} + \Delta y_{i,t+1})^2 | i\right] \tag{13}$$

$$\sigma_{\eta,i}^4 + 4\sigma_{\eta,i}^2\sigma_{\epsilon,i}^2 + 4\sigma_{\epsilon,i}^4 = \mathbb{E}\left[(\Delta y_{i,t})^2 \times (\Delta y_{i,t+1})^2 | i\right] \tag{14}$$

with $\mathbb{E}\left[\sigma_{\eta,i}^2\right]$ and $\mathbb{E}\left[\sigma_{\varepsilon,i}^2\right]$ identified which are provided above.

The pass-through coefficients of income shocks are identified as

$$\phi_i = \frac{\mathbb{E}\left[\Delta c_{it} \times (\Delta y_{it-1} + \Delta y_{it} + \Delta y_{it+1})\, | i\right]}{\mathbb{E}\left[\Delta y_{it} \times (\Delta y_{it-1} + \Delta y_{it} + \Delta y_{it+1})\, | i\right]} \tag{15}$$

$$\psi_i = \frac{\mathbb{E}\left[\Delta c_{it} \times \Delta y_{it} | i\right]}{\mathbb{E}\left[\Delta y_{it} \times \Delta y_{it} | i\right]}. \tag{16}$$

Note that the coefficients are household-$i$ specific.

## 3.2 Estimation details

I estimate the income process parameters as well as consumption pass-through coefficients jointly using the method of simulated moments (MSM). Denote $\Theta = [\rho, \tilde{\mu}_\eta, \tilde{\sigma}_\eta^2, \tilde{\mu}_\varepsilon, \tilde{\sigma}_\varepsilon^2, \psi, \phi, \sigma_\zeta^2]$ as a vector of parameters to estimate.

There are two groups of moments to target. One is cross-sectional auto-covariances of income and consumption growth over the life cycle, and the other is the cumulative distribution of household-specific income volatility. The first group focuses on how income and consumption change inequality evolves over age. In specific,

$$E[\Delta y_{it}\Delta y_{it+k}|t], E[\Delta c_{it}\Delta c_{it+k}|t], E[\Delta c_{it}\Delta y_{it+k}|t] \text{ for } k = 0, 1, \cdots, T-t \text{ at every age } t = 25 \sim 60$$

For the second group, I target the first to the fourth central moments as well as the 10th, 25th, 75th, and 90th percentiles of the distribution. Denote $F(\cdot)$ as the cumulative distribution, then the moments to target are from

$$F(x) \text{ where } x \text{ includes } Var(\Delta_2 y_{it}|i), Var(\Delta_6 y_{it}|i).$$

Note that each $x$ represents the volatility of transitory changes and persistent changes, respectively.

I minimize the weighted sum of squared deviations between the data and simulated moments. In detail, I draw the same number of model agents with the data with initial draw of parameter vector $\Theta$. Then, I simulate the sequences of income and consumption growth $\{\Delta y_{it}\}_{t=25}^{t=60}, \{\Delta c_{it}\}_{t=25}^{t=60}$ for each agent $i$ based on the data generating process equation (4) to

(7). I construct the vector $d_j$ for every $j$-th moment such that

$$d_j = \frac{1}{R}\sum_{r=1}^{R} \hat{m}_j^r(\hat{\Theta}) - m_j, \forall j$$

where $\hat{m}_j^r(\hat{\Theta})$ is a model generated $j$-th moment at $r$-th simulation and $m_j$ is the corresponding moment from the data.

When it comes to a weighting matrix, I assign equal weight between two sets of moments. Within the first cross-sectional moments set, each moment gains different weight by the number of observation from the data that is used to calculate the specific moment. An equal weight is assigned to every moment within the second set of moments because each moment represents the cumulative distribution equally. Denote $\mathbf{W}$ a block diagonal matrix such that

$$\mathbf{W} = \begin{bmatrix} \mathbf{W_1} & 0 \\ 0 & \mathbf{W_2} \end{bmatrix}$$

where $\mathbf{W_1}$ and $\mathbf{W_2}$ are the weighting matrices for each group of moments. $\mathbf{W_1}$ is a triangular matrix and $\mathbf{W_2}$ is a diagonal matrix. Finally, the estimates are defined by

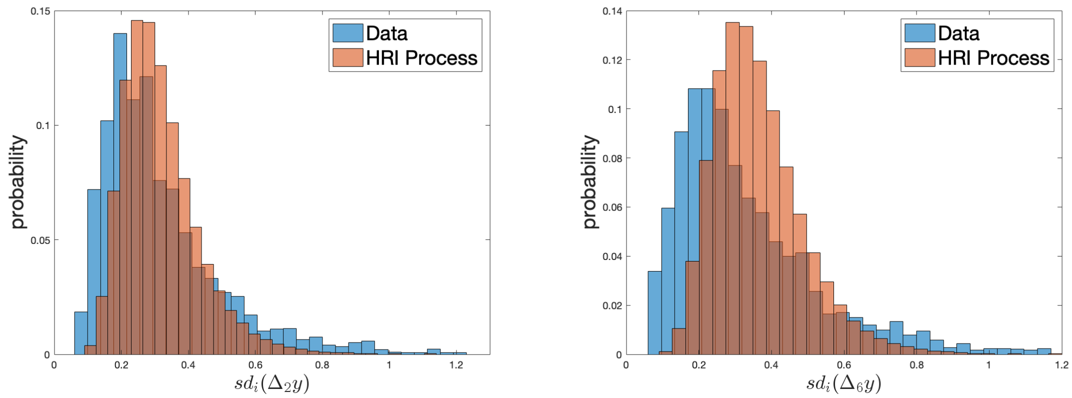
$$\Theta^* = \arg\min_{\Theta} \mathbf{d}' W \mathbf{d}.$$

where $\mathbf{d}$ is a vector containing all $d_j$'s.

There are a total 682 number of moment conditions consisting of the quadratic objective function. The objective is high-dimensional with potentially numerous local minima. I use the particle swarm optimization (PSO) algorithm to find the global minimum. The PSO is a part of randomized search methods in stochastic optimization. It is similar to a multi-start algorithm in that both algorithms begin with many random (or quasi-random) points, called particles in the PSO. The difference is that the PSO is derivative-free, so once the objective is evaluated at the initial particles, the next iteration is updated not by gradient but by velocity. Velocity is determined by the weighted sum of the previous velocity, the difference between the current location and the best location out of the previous location, and the difference between the current location and the best location in the current neighborhood. So, each swarm's updates are guided not only by its own information but also by the information gathered from the entire swarm.

The biggest advantage of the PSO is that it prevents the solution from getting stuck in the local optimum and helps achieve the global optimum. This is enabled by using the entire swarm's information. Moreover, the PSO is computationally more efficient than the

Figure 4: Estimation Fit




(a) Transitory volatility

(b) Persistent volatility

*Note:* The orange histogram on the both panels are generated from Monte Carlo simulation of the estimates of heterogeneous risk income process. The blue histogram is the empirical measures of household specific income volatility which is same as Figure 1.

gradient-based method since it does not require differentiation during the iteration. The potential issue of the PSO or stochastic optimization is that the global optimum is not necessarily guaranteed. To overcome this, I use a traditional gradient-based optimization as a hybrid function once the PSO algorithm completes the solution search.

## 3.3 Estimation results

Figure 4 presents the estimation fit of the distribution of income volatility, which is the second set of moments to target. The estimation match well the right-skewness of the household-specific income volatility.

There are several findings to point out from the estimation result in Table 5. First, persistent shock is almost homogeneous whereas transitory shock is heterogeneous across households. Note that the dispersion of $\sigma_\eta^2$ is almost zero but that of $\sigma_\varepsilon^2$ is 0.0017. It means that the transitory shock accounts for most heterscedasticity. The heterogeneous transitory shock is consistent with the recent findings in income dynamics literature. Almuzara, 2020 estimate the density of household-specific variance of transitory shock and document latent heterogeneity. Arellano et al., 2017 find that the persistent component is relatively Gaussian distribution, but the transitory shock looks far from normal distribution and presents high kurtosis and fat tails. Household-specific variance can explain the high kurtosis by large population with small variance of transitory shock. The fat tails are the extremely opposite group who has very volatile transitory income.

Table 5: Estimates of the heterogeneous income risk process and consumption pass-through

| Parameters | Remarks | Results |
|---|---|---|
| $\rho$ | persistence | 0.992 |
| | | (0.0065) |
| $\tilde{\sigma}_\eta$ | $Var(\sigma_\eta^2) = 0.0001$ | 0.706 |
| | | (0.0220) |
| $\tilde{\sigma}_\varepsilon$ | $Var(\sigma_\varepsilon^2) = 0.0017$ | 0.772 |
| | | (0.0203) |
| $\tilde{\mu}_\eta$ | median $\sigma_\eta^2 = 0.0085$ | -4.766 |
| | | (0.0744) |
| $\tilde{\mu}_\varepsilon$ | median $\sigma_\varepsilon^2 = 0.0341$ | -3.379 |
| | | (0.0724) |
| $\phi$ | persistent shock pass-through | 0.917 |
| | | (0.0066) |
| $\psi$ | transitory shock pass-through | 0.692 |
| | | (0.0111) |

*Note:* The numbers in parentheses are standard errors calculated from bootstrap sampling.

Second, the dispersion of persistent shock is very low on average. The median $\sigma_\eta^2$ equals 0.0085 and considering the persistent shock variance is homogeneous, most households have moderate persistent shock. Compared to the value typically used in the literature, it is relatively small. For instance, Krueger et al., 2016 find the variance of persistent shock equals to 0.0384.

Third, the transitory pass-through coefficient is almost 70%, which is very high. The dense population with small variance of transitory shock is the cause to boost up the aggregate consumption response.

# 4 Implications for Consumption and Savings Model

Given that exogenous income is an essential ingredient in macroeconomic models, it is important to see whether the quantitative model with the heterogeneous risk income process has consistent implications for consumption response to income shocks with the empirical estimates from the previous section. In this section, I use a standard life-cycle incomplete market model, a workhorse model to study consumption and savings, and find the consumption pass-through.

## 4.1 Model

Households start working at age 0, retire at $t_R$, and live until $T$ with unconditional survival probability $s_t$ at every age $t$. Households have constant relative risk aversion (CRRA) utility from consumption with a discount factor $\beta$. There is a risk-free asset $A_{it}$ only with a time-invariant interest rate $r$ and households can borrow $\bar{A}$ at most. I assume every household is born with zero asset. I impose the wedge between net borrowing and savings. The wedge makes a large number of households hold zero wealth and have a high marginal propensity to consume (MPC). It has been well documented that many households have MPC of almost 1 and most are restrained by liquidity constraints (C. Carroll et al., 2017; Kaplan et al., 2018; Koşar et al., 2023). The interest wedge helps the model to replicate the empirical distribution of MPC and wealth correctly, in particular the share of households holding net negative wealth.

It is a partial equilibrium model, so the interest rate is taken directly from the data and is not determined by the asset market clearing. It abstracts the general equilibrium effect of savings. However, the true essence of the quantitative experiment lies in its ability to mirror the aggregate consumption pass-through of households, shedding light on how households with varying income volatility respond to the same level of shock. The partial equilibrium approach is helpful to achieve the goal.

Households labor income $Y_{it}$ corresponds households' disposable income net of taxes and transfers for a consistent measure with the empirical counterpart. $Y_{it}$ is decomposed into a deterministic age-profile of income common across households, $\kappa_t$, and a stochastic component, $y_{it}$. The stochastic component follows the same specification and estimates as the estimated income process in the previous section. After retirement, households receive social security income $\tau Y_{it_R}$ until death. In sum, the households maximization problem is

$$\max_{\{C_{it}\}_{t=0}^{T}} E_{i,0} \sum_{t=0}^{T} \beta^t s_t \frac{C_{it}^{1-\gamma} - 1}{1-\gamma}$$

$$\textit{subject to}$$

$$C_{it} + A_{it+1} = (1+r)A_{it} + Y_{it}$$

$$A_{it} \geq -\bar{A}, \;\; A_{i0} = 0$$

$$Y_{it} = \begin{cases} exp(\kappa_t + y_{it}), & \textit{for } t \leq t_R \\ \tau Y_{it_R}, & \textit{for } t > t_R \end{cases}$$

$$y_{it} = p_{it} + \varepsilon_{it}, \;\; \varepsilon_{it} \sim N(0, \sigma^2_{\varepsilon,i}), \;\; \sigma^2_{\varepsilon,i} \sim Lognormal(\tilde{\mu}_{\varepsilon}, \tilde{\sigma_{\varepsilon}}^2)$$

$$p_{it} = \rho p_{it-1} + \eta_{it}, \;\; \eta_{it} \sim N(0, \sigma^2_{\eta,i}), \;\; \sigma^2_{\eta,i} \sim Lognormal(\tilde{\mu}_{\eta}, \tilde{\sigma}^2_{\eta})$$

Though the model is the standard version of an incomplete market, it is effective to examine the implication of heterogeneous income risk on consumption. The main channel is to amplify the role of the precautionary saving motive. Risk-averse households want to save enough buffer stock to self-insure against future income shock. Households facing a large variance of income shock need greater buffer stock savings than those with small future income fluctuations. In other words, the precautionary saving motives of those with volatile income are more potent than those of households with non-volatile income.

## 4.2 Calibration

I assume households start their life at age 25 and retire at 60 for sure. There is no early retirement decision. It is a consistent working individual assumption as the estimation part. Households can live at most 90 years old. So, $t_R = 35$ and the terminal period is $T = 65$. Conditional survival rate $s_t$ is calculated by interpolating male's probabilities of death from Table 8, calendar year 2018 in Bell and Miller, 2005.

The common life-cycle income profile $\kappa_t$ is estimated by regressing the net disposable household income on the fourth-order polynomial of age. The variables and samples are from the same dataset used for the estimation in the previous sections. For the stochastic component of income $y_{it}$, I use the same income process estimates as in Table 5.

The interest rate for net savers is 1.7% and it is average after-tax rate of return for net worth in 2001 Survey of Consumer Finances (SCF). The borrowing limit is 74% of the annual income. Both the interest rate and the borrowing limit are taken from Kaplan and Violante, 2014. For the pension replacement ratio, I take 70% of last period's income from Hryshko and Manovskii, 2022.

I calibrate the remaining parameters $\gamma$, a risk aversion coefficient of households, $\beta$, a discount factor, and $r_{borrowing}$, the borrowing rate for net borrowers, to match net liquid wealth distribution and share of negative net liquid wealth households in 2019 SCF. The reason why I use net liquid wealth, not total net worth, is my focus is the quantitative implication of consumption response and the major wealth source to consume out is liquid wealth mostly. Net liquid wealth is defined by liquid asset net of unsecured debt and credit card balances. Liquid asset includes money market accounts (MMA), checking, savings, and call accounts, prepaid cards, directly held pooled investment funds (NMMF), stocks, bonds. Table 6 presents the summary of parameters and the upper section of Table 7 shows how the model target well the data counterparts. The net liquid wealth distribution is targetted well in general, except the top 10%. It is well known that the incomplete market model is difficult to reproduce the fat right tail of the wealth distribution (e.g., Aiyagari, 1994; Huggett, 1993). It is because the main mechanism of building wealth in the Bewley-Aiyagari-Huggett economy is precautionary savings motive to self-insure against idiosyncratic shock, but for the top wealthy, this motive is weak.[7] Nevertheless, the top wealthy is out of scope of this paper and their contribution to aggregate consumption response is trivial so I abstract from it.

Table 6: Calibrated Parameters

| Parameters | Description | Value | Source |
|---|---|---|---|
| calibrated outside the model | | | |
| $\rho$ | persistence | 0.992 | |
| $\tilde{\sigma}_\eta$ | $Var(\sigma^2_\eta) = 0.0001$ | 0.706 | |
| $\tilde{\sigma}_\varepsilon$ | $Var(\sigma^2_\varepsilon) = 0.0017$ | 0.772 | Table 5 |
| $\tilde{\mu}_\eta$ | median $\sigma^2_\eta = 0.0085$ | -4.766 | |
| $\tilde{\mu}_\varepsilon$ | median $\sigma^2_\varepsilon = 0.0341$ | -3.379 | |
| $\kappa_t$ | life-cycle income profile | in text | author's calculation |
| $s_t$ | conditional survival rate | in text | Bell and Miller, 2005 |
| $r_{saving}$ | saving interest rate | 0.017 | Kaplan and Violante, 2014 |
| $\bar{A}$ | borrowing limit | $0.74 \cdot Y_{it}$ | Kaplan and Violante, 2014 |
| $\tau$ | pension replacement ratio | 0.7 | Hryshko and Manovskii, 2022 |
| calibrated in the model | | | Target (in 2019 SCF) |
| $\gamma$ | risk aversion coefficient | 2.0 | |
| $\beta$ | discount factor | 0.9 | |
| $r_{borrowing}$ | borrowing interest rate | 0.047 | |

*Note:* $\kappa$ is estimatied by the author using the same dataset for the estimation. The interest wedge of borrowing is 3% annually.

To solve the model, first I draw 100 types of households with different $(\sigma_{\eta,i}, \sigma_{\varepsilon,i})$. Both $\sigma_{\eta,i}$ and $\sigma_{\varepsilon,i}$ are drawn from the estimated lognormal distribution of household-specific variance. Then, I discretize persistent and transitory income shocks into 11-state and 17-state Markov chains by Rouwenhorst method. The model is solved by the endogenous grid method proposed by C. D. Carroll, 2006. I simulate 10,000 households for each type, total 1 million number of households.

## 4.3 Results

Remind that the objective of the quantitative analysis is whether the model can generate the consistent consumption response with empirical estimates. So I conduct the same estimation

[7] Various sources have been studied to capture the top wealth distribution. For instance, entrepreneurship, intergenerational links and bequest motive, persistent heterogeneous rate of return on capital, or skewed earnings are potential factors (e.g., Benhabib et al., 2019; Cagetti and De Nardi, 2009; Fagereng et al., 2020; Nardi, 2004). For the further discussion, refer to Benhabib and Bisin, 2018.

with simulated data of income and consumption. To make comparable estimation, I take 43% of total consumption in the model as nondurable consumption which is the measure I use in empirical estimation. Then, I regress simulated log income and log consumption on polynomial of age to extract residual. The advantage of simulated data is availability of directly observed income shock $\eta$ and $\varepsilon$. Instead of joint estimation of income and consumption with the data, I can use the observed shock to predict the $\phi$ and $\psi$.

At the bottom two rows in Table 7, the right column shows the estimates with the simulated income shocks, and the left column is taken from Table 5, the estimates with the data. The model reproduces the consumption pass-through of persistent shock well.

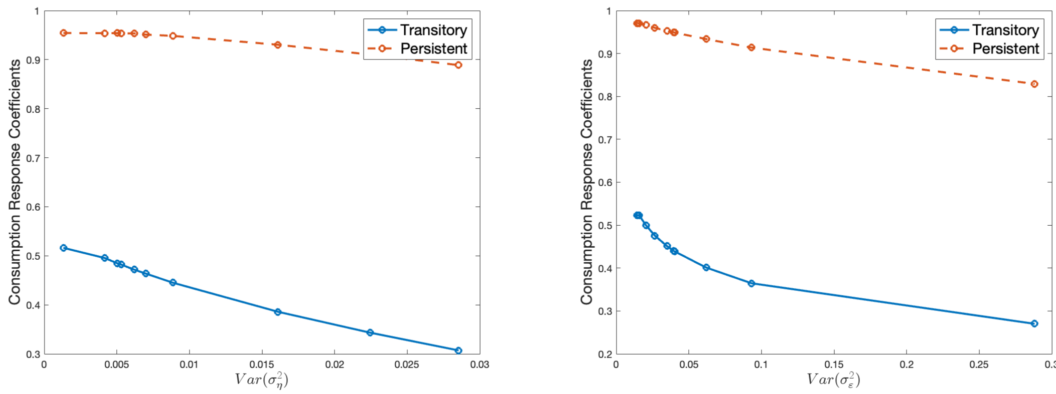
Table 7: Model results

| Target moments | Data | Model |
|---|---|---|
| Net liquid wealth distribution ($1000 2019 USD) | | |
| P10 | -4.3 | -1.4 |
| P25 | 0.03 | 0 |
| P50 | 3.1 | 3.4 |
| P75 | 28.5 | 23.4 |
| P90 | 117 | 71.0 |
| Share of neg. liq. wealth (%) | 21.52 | 13.32 |
| Consumption pass-through | | |
| $\phi$ (persistent shock) | 0.917 | 0.923 |
| $\psi$ (transitory shock) | 0.692 | 0.356 |

*Note:* The consumption pass-through coefficients in the first column are the estimates in Table 5.

Another advantage of the quantitative model is using discretized variance of shock for each type of household from simulation to observe the heterogeneity in consumption response across different variance of income shocks. Figure 5 shows that households with more volatile shock have less response of consumption. Remind that I randomly draw 10 number of variances for each persistent and transitory shocks. Panel (a) and (b) show the average consumption response coefficients of every variance of persistent and transitory shock, respectively. The consumption responses look remarkable heterogeneous across variances of the shocks for both persistent and transitory component, but there is more significant decline in transitory pass-through. The elasticity of pass-through is about 40% greater for transitory shock than for persistent shock.

Figure 5: Consumption response heterogeneity




(a) Persistent shock

(b) Transitory shock

*Note:* Each panel presents the heterogeneity of consumption response over the variance of persistent and transitory shock, respectively. The coefficients are calculated by simulated shocks in the model. Every dot on the graph represents consumption responses of simulated households with the specific variance of income shock. Note that ranges of x-axis are different across two panels. It is because the average and dispersion of persistent shock variances are smaller than those of transitory shock variances.

# 5 Conclusion

I propose heterogeneous income risk based on empirical evidence and develop an income process with a household-specific variance of shocks. With the PSID, I exploit a long history of households to find the distribution of the income volatility by each household, which is skewed with a fat right tail. It implies the existence of heteroskedastic income risk across households. To find the source of the heterogeneity, I tested various observables, including demographics, occupation, industry, and labor market outcomes. However, only about 13% of the volatility measure is explained by the observables. This implies that heterogeneity in income volatility remains unexplained largely, and latent heterogeneity shows the potential for heterogeneous income risk.

I estimate a new income process with randomly distributed household-specific variance drawn from a parametric assumption and target cross-sectional moments of income and consumption growth as well as the distribution of income volatility by households. Most heterogeneity in income risk comes from transitory risks, whereas the persistent shock is relatively homogeneous across households. Furthermore, the aggregate response of consumption on transitory income change is greater than the estimates in other literature without heteroskedastic income risk. Most households with a low variance of innovation boost consumption since their stable income over their lifetime makes them consume more instead of self-insuring for precautionary savings. In contrast, a few households rarely adjust consumption because they have a strong self-insurance motive against their lumpy income. However, despite their extremely small consumption response, they consist of a small population. As a result, the aggregate consumption response becomes greater. I evaluate the empirical results with a standard quantitative life cycle model and prove the consumption response heterogeneity caused by different variances of income shocks across households.

There is more work to be done. First, the source of heterogeneity in income risk still needs to be examined further. One thing that might be related to the heterogeneous fluctuation of workers' income but is not studied in this paper is the firm size due to the lack of data variable. Employer-employee matched data is preferable to examine the link between the firm and household income. Second, it would be interesting to investigate the relationship between the higher-order moments in earnings change and heteroskedatic income risk. It is well documented that the cross-sectional earnings growth distribution is away from Gaussian and is negatively skewed and leptokurtic.

Nevertheless, it has not yet been answered whether the non-Gaussian features are attributed to extreme shock that might occur to everyone with low probability or fixed type, which is close to income risk heterogeneity in this paper. I provide the potential implications of heterogeneous income risk in determining the cross-sectional non-Gaussian features of income growth from the identification. The importance of the higher-order moments is indirectly embedded in identifying the lognormal distribution of the heterogeneous variances.

Lastly, the outcome of public policies could change with heterogeneous income risk. For instance, public insurance such as progressive income tax, unemployment insurance, or stimulus checks is less effective for those with high volatile income since they are likely to be self-insured well enough, so there might not be room for public policies to improve their welfare. The other groups with a moderate income can benefit more from social insurance than the other way around. With or without the second-order dimension of household income risk, the welfare improvements of social insurance can vary.

# References


Aiyagari, S. R. (1994). Uninsured idiosyncratic risk and aggregate saving. *The Quarterly Journal of Economics*, *109*(3), 659–684. https://doi.org/10.2307/2118417

Almuzara, M. (2020). *Heterogeneity in transitory income risk*.

Arellano, M., Blundell, R., & Bonhomme, S. (2017). Earnings and consumption dynamics: A nonlinear panel data framework. *Econometrica*, *85*(3), 693–734. https://doi.org/10.3982/ECTA13795

Arellano, M., Blundell, R., Bonhomme, S., & Light, J. (2023). Heterogeneity of consumption responses to income shocks in the presence of nonlinear persistence. *Journal of Econometrics*.

Autor, D. H., & Dorn, D. (2013). The growth of low-skill service jobs and the polarization of the US labor market. *American Economic Review*, *103*(5), 1553–1597. https://doi.org/10.1257/aer.103.5.1553

Bell, F. C., & Miller, M. L. (2005). Life tables for the united states social security area 1900-2100. *Social Security Administration, Office of the Chief Actuary*.

Benhabib, J., & Bisin, A. (2018). Skewed wealth distributions: Theory and empirics. *Journal of Economic Literature*, *56*(4), 1261–1291. https://doi.org/10.1257/jel.20161390

Benhabib, J., Bisin, A., & Luo, M. (2019). Wealth distribution and social mobility in the US: A quantitative approach. *American Economic Review*, *109*(5), 1623–1647. https://doi.org/10.1257/aer.20151684

Blundell, R., Pistaferri, L., & Preston, I. (2008). Consumption inequality and partial insurance. *American Economic Review*, *98*(5), 1887–1921. https://doi.org/10.1257/aer.98.5.1887

Blundell, R., Pistaferri, L., & Saporta-Eksten, I. (2016). Consumption inequality and family labor supply. *American Economic Review*, *106*(2), 387–435. https://doi.org/10.1257/aer.20121549

Bonhomme, S., Lamadon, T., & Manresa, E. (2022). Discretizing unobserved heterogeneity. *Econometrica*, *90*(2), 625–643. https://doi.org/10.3982/ECTA15238

Botosaru, I. (2023). Time-varying unobserved heterogeneity in earnings shocks. *Journal of Econometrics*, *235*(2), 1378–1393. https://doi.org/10.1016/j.jeconom.2022.08.012

Cagetti, M., & De Nardi, M. (2009). Estate taxation, entrepreneurship, and wealth. *American Economic Review*, *99*(1), 85–111. https://doi.org/10.1257/aer.99.1.85

Carroll, C., Slacalek, J., Tokuoka, K., & White, M. N. (2017). The distribution of wealth and the marginal propensity to consume [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/QE69 *Quantitative Economics*, *8*(3), 977–1020. https://doi.org/10.3982/QE694

Carroll, C. D. (2006). The method of endogenous gridpoints for solving dynamic stochastic optimization problems. *Economics Letters*, *91*(3), 312–320. https://doi.org/10.1016/j.econlet.2005.09.013

De Nardi, M., Fella, G., & Paz-Pardo, G. (2020). Nonlinear household earnings dynamics, self-insurance, and welfare. *Journal of the European Economic Association*, *18*(2), 890–926. https://doi.org/10.1093/jeea/jvz010

Fagereng, A., Guiso, L., Malacrino, D., & Pistaferri, L. (2020). Heterogeneity and persistence in returns to wealth. *Econometrica*, *88*(1), 115–170. https://doi.org/10.3982/ECTA14835

Geweke, J., & Keane, M. (2000). An empirical analysis of earnings dynamics among men in the PSID: 1968–1989. *Journal of Econometrics*, *96*(2), 293–356. https://doi.org/10.1016/S0304-4076(99)00063-9

Gregory, V., Menzio, G., & Wiczer, D. G. (2021, April). The alpha beta gamma of the labor market. https://doi.org/10.3386/w28663

Guvenen, F., Karahan, F., Ozkan, S., & Song, J. (2021). What do data on millions of u.s. workers reveal about lifecycle earnings dynamics? *Econometrica*, *89*(5), 2303–2339. https://doi.org/10.3982/ECTA14603

Hryshko, D., & Manovskii, I. (2022). How much consumption insurance in the u.s.? *Journal of Monetary Economics*, *130*, 17–33. https://doi.org/10.1016/j.jmoneco.2022.05.003

Huggett, M. (1993). The risk-free rate in heterogeneous-agent incomplete-insurance economies. *Journal of Economic Dynamics and Control*, *17*(5), 953–969. https://doi.org/10.1016/0165-1889(93)90024-M

Kaplan, G., Moll, B., & Violante, G. L. (2018). Monetary policy according to HANK. *American Economic Review*, *108*(3), 697–743. https://doi.org/10.1257/aer.20160042

Kaplan, G., & Violante, G. L. (2010). How much consumption insurance beyond self-insurance? *American Economic Journal: Macroeconomics*, *2*(4), 53–87. https://doi.org/10.1257/mac.2.4.53

Kaplan, G., & Violante, G. L. (2014). A model of the consumption response to fiscal stimulus payments. *Econometrica*, *82*(4), 1199–1239. https://doi.org/10.3982/ECTA10528

Karahan, F., & Ozkan, S. (2013). On the persistence of income shocks over the life cycle: Evidence, theory, and implications. *Review of Economic Dynamics*, *16*(3), 452–476. https://doi.org/10.1016/j.red.2012.08.001

Koşar, G., Melcangi, D., Pilossoph, L., & Wiczer, D. (2023). Stimulus through insurance: The marginal propensity to repay debt.

Krueger, D., Mitman, K., & Perri, F. (2016). Macroeconomics and household heterogeneity. In *Handbook of macroeconomics* (pp. 843–921, Vol. 2). Elsevier. https://doi.org/10.1016/bs.hesmac.2016.04.003

Lewis, D. J., Melcangi, D., & Pilossoph, L. (2019). Latent heterogeneity in the marginal propensity to consume. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3489434

MaCurdy, T. E. (1982). The use of time series processes to model the error structure of earnings in a longitudinal data analysis. *Journal of Econometrics*, *18*(1), 83–114. https://doi.org/10.1016/0304-4076(82)90096-3

Meghir, C., & Pistaferri, L. (2004). Income variance dynamics and heterogeneity. *Econometrica*, *72*(1), 1–32. https://doi.org/10.1111/j.1468-0262.2004.00476.x

Meghir, C., & Pistaferri, L. (2011). Earnings, consumption and life cycle choices. In *Handbook of labor economics* (pp. 773–854, Vol. 4). Elsevier. https://doi.org/10.1016/S0169-7218(11)02407-5

Nardi, M. D. (2004). Wealth inequality and intergenerational links. *REVIEW OF ECONOMIC STUDIES*.

# Appendices

## A Determination of Optimal Number of Clusters

For the group fixed effect multinomial logit regression, the optimal number of clusters have to be specified. Two methods are combined to determine the optimal number of clusters: the elbow method and the Caliński-Harabasz pseudo-$F$ index. The elbow method uses visualization of three measures, with-in cluster sum of squares (WCSS), the proportional reduction of error (PRE), and the coefficient of determination. The coefficienet of determination is

$$1 - \frac{WCSS(J)}{TSS} = 1 - \frac{\sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{k=1}^{K}\mathbf{1}[j = j(i)](m_{k,i} - m^*_{k,j})^2}{\sum_{i=1}^{N}\sum_{k=1}^{K}(m_{k,i} - \bar{m}_k)^2}, \quad \forall J \in \{1, 2, 3, \cdots\}. \tag{17}$$

The PRE illustrates marginal decrease of WCSS by adding one more number of group such that

$$PRE = \frac{|WCSS(J) - WCSS(J-1)|}{WCSS(J-1)}, \quad \forall J \in \{2, 3, 4, \cdots\}. \tag{18}$$

The Caliński-Harabasz pseudo-$F$ index (CHI) complements the elbow method in providing the optimum $J^*$ with the highest value. The metric is

$$CHI = \frac{BCSS/(J-1)}{WCSS/(N-J)}, \text{ where } BCSS = \sum_{k=1}^{K}\sum_{j=1}^{J} N_j(m^*_{j,k} - \bar{m}_k)^2.$$

where the BCSS, between-cluster sum of squares, is the weighted sum of Euclidean distances between the centroid of group $j$, $m^*_{j,k}$ and the unconditional centroid $\bar{m}_k$ with the weight of the size of the group $N_j$.

## B Linear Regression of Income Volatility on Observables

As a supplementary practice of finding observable sources of large income volatility, I take the volatility as a continuous measure instead of discrete group. First, I use principal component analysis (PCA) to combine persistent and transitory volatility and to predict a unique measure of income volatility. PCA is an orthogonal linear transformation used to reduce the dimension of data. It generates a new variable by predicting a linear combination of the raw variables that contains most of the variance. In this case, I use PCA to predict a measure combining persistent and transitory volatilities. Then, I run a linear regression of the predicted volatility on the same observables used in the group fixed effect multinomial logit regression. The results are in Table 8 below.

Table 8: Linear regression of income volatiltiy on observables

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Birth year | 0.00562* | 0.00457* | 0.00446* | 0.00421 | -0.0164*** |
| | (2.50) | (1.98) | (1.99) | (1.82) | (-5.70) |
| High school graduate | -0.154* | -0.108 | -0.117 | -0.0947 | -0.0418 |
| | (-2.11) | (-1.50) | (-1.60) | (-1.30) | (-0.59) |
| College graduate | -0.257*** | -0.153* | -0.0936 | -0.0797 | -0.0405 |
| | (-3.77) | (-2.19) | (-1.22) | (-1.04) | (-0.55) |
| Black | 0.149* | 0.148* | 0.0647 | 0.0898 | 0.174** |
| | (2.43) | (2.44) | (1.03) | (1.43) | (2.85) |
| Self-employment experience | 1.110*** | 0.973*** | 1.084*** | 0.975*** | 0.819*** |
| | (20.73) | (17.15) | (20.06) | (17.16) | (13.92) |
| Industry | | | | | |
| Agriculture | | 1.072*** | | 0.975*** | 0.537 |
| | | (3.74) | | (3.37) | (1.76) |
| Mining | | 1.450*** | | 1.338*** | 1.213*** |
| | | (8.11) | | (7.02) | (6.58) |
| Manufacturing | | 0.460** | | 0.435** | 0.305* |
| | | (3.25) | | (2.89) | (2.08) |
| Business services | | 0.511** | | 0.505** | 0.334* |
| | | (3.11) | | (3.05) | (2.08) |
| Entertainment | | 0.524** | | 0.413* | 0.320 |
| | | (2.85) | | (2.19) | (1.76) |
| Occupation | | | | | |
| Managerial and professionals | | | -0.365*** | -0.139 | -0.123 |
| | | | (-4.58) | (-1.54) | (-1.39) |
| Technical, sales and admin supports | | | -0.251* | -0.0505 | -0.00958 |
| | | | (-2.09) | (-0.39) | (-0.08) |
| Service | | | 0.349* | 0.414** | 0.341* |
| | | | (2.35) | (2.60) | (2.21) |

| | | | | | |
|---|---|---|---|---|---|
| Production, craft, and repair | | | -0.529*** | -0.313* | -0.225 |
| | | | (-3.72) | (-2.13) | (-1.58) |
| Labor market outcomes | | | | | |
| # of weeks of head's unemployment | | | | | 0.0156 |
| | | | | | (1.40) |
| # of weeks of wife's unemployment | | | | | 0.0119 |
| | | | | | (1.13) |
| # of weeks of head's out-of-labor force | | | | | 0.0465** |
| | | | | | (2.75) |
| # of weeks of wife's out-of-labor force | | | | | 0.0189*** |
| | | | | | (6.74) |
| # of weeks of head's job tenure | | | | | -0.00260*** |
| | | | | | (-6.70) |
| # of weeks of wife's job tenure | | | | | -0.00374*** |
| | | | | | (-7.37) |
| Observations | 2826 | 2826 | 2821 | 2821 | 2821 |
| Adjusted $R^2$ | 0.139 | 0.167 | 0.150 | 0.170 | 0.227 |

$t$ statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

# C Estimation with Various Income Variable

| Parameters | Results | | | |
|---|---|---|---|---|
| | net income | gross income | male earnings | total earnings |
| $\rho$ | 0.992 | 0.856 | 0.851 | 0.864 |
| | (0.0065) | (0.0067) | (0.0066) | (0.0065) |
| $\tilde{\sigma}_\eta$ | 0.706 | 0.344 | 0.444 | 0.468 |
| | (0.0220) | (0.0207) | (0.0215) | (0.0208) |
| $\tilde{\sigma}_\varepsilon$ | 0.772 | 0.474 | 0.480 | 0.479 |
| | (0.0203) | (0.0197) | (0.0210) | (0.0217) |
| $\tilde{\mu}_\eta$ | -4.766 | -2.754 | -2.951 | -2.882 |
| | (0.0744) | (0.0710) | (0.0756) | (0.0749) |
| $\tilde{\mu}_\varepsilon$ | -3.379 | -1.264 | -1.237 | -1.397 |
| | (0.0724) | (0.0701) | (0.0719) | (0.0706) |
| $\phi$ | 0.917 | 0.936 | 0.959 | 1.000 |
| | (0.0066) | (0.0064) | (0.0070) | (0.0068) |
| $\psi$ | 0.692 | 0.162 | 0.157 | 0.143 |
| | (0.0111) | (0.0091) | (0.0109) | (0.0106) |

Table 9: Estimation Result