Earnings Shocks, Expectations, and Spending*

Eungik Lee
New York University

Johan Sæverud
University of Copenhagen

November 1, 2023
(click here for the latest version)

Abstract

The impact of earnings shocks on workers' earnings expectations is a key determinant of subsequent consumption changes. Using a unique combination of an expectations panel survey and administrative data from Denmark, we identify earnings shocks, expectation changes, and consumption changes that are often challenging to observe. Simultaneously, prevailing assumptions in the income process literature, including perfect differentiation of permanent and transitory shocks, have limitations in explaining the empirical expectation changes upon earnings shocks. We introduce a new income process model in which workers possess partial information about the nature of earnings shocks. Our estimates show that workers distinguish only half of permanent and transitory shocks. We further investigate the implications of partial information on consumption changes both empirically and through the lens of a model. We find that workers' partial information about the earnings shocks is important to match the degree of consumption response upon earnings shocks. Moreover, we show that partial information better predicts the consumption insurance level observed in empirical literature than other conventional assumptions.

Keywords: earnings risk, income process, subjective expectations, consumption

JEL classification: D31, D84, E21, E24, J31

*Eungik Lee is indebted to his committee members Andrew Caplin, Guido Menzio, Christopher Flinn, Andrew Schotter, and Corina Boar for their continuous support. We are grateful to Soren Leth-Petersen, Victoria Gregory, Simon Gilchrist, Gian Luca Clementi, Diego Perez, Jarda Borovicka, Katka Borovickova, Jess Benhabib, Ricardo Lagos, Guillaume Frechette, David Cesarini, Syngjoo Choi, Elena Manresa, Juan M. Sanchez, Facundo Danza, Leon Guzman, seminar participants at NYU, Stockholm School of Economics, and Economics Graduate Student Conference for invaluable comments and discussion.

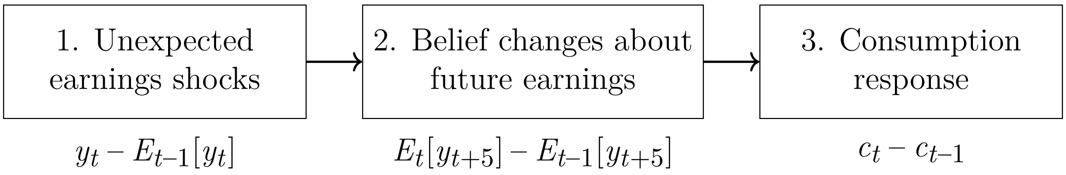
1 Introduction

How do workers adjust their consumption in response to unexpected earnings shocks? This is an important question for macro and labor economics, as it has implications for labor market policies and consumer behavior. Standard consumption-savings models describe consumption as a forward-looking decision highlighting the importance of expected earnings flows. This emphasizes changes in beliefs about future earnings flows upon earnings shocks as an important mechanism for understanding subsequent consumption changes. However, understanding how workers change their beliefs upon earnings shocks remains challenging because workers' beliefs about future earnings are usually not observed in the data.

The limited availability of data on workers' belief changes leads to two different strict assumptions in the income process literature (Lillard and Willis, 1978; MaCurdy, 1982). The first common assumption is that workers can perfectly differentiate permanent from transitory earnings shocks. Under this assumption, the belief changes are somewhat mechanical because workers can recognize the magnitude of permanent shocks and update their beliefs accordingly. However, as Blundell and Preston (1998) point out, workers' ability to differentiate between permanent and transitory earnings shocks has important implications for their consumption response to these shocks. To account for the possibility that workers may not be fully knowledgeable about shocks, the second opposite assumption is that workers do not have any information about the nature of earnings shocks, and they update their beliefs using a Kalman filtering approach (Guvenen, 2007; Guvenen and Smith, 2014). This assumption is similarly stringent, as it assumes that workers do not differentiate between permanent and transitory shocks at all.

In this paper, we use subjective earnings expectations to provide empirical evidence about belief changes. This approach allows us to directly measure workers' beliefs, avoiding the need for stringent assumptions and leading to a more realistic modeling of belief changes. Although the existing literature has confirmed the usefulness of subjective earnings expectations (Dominitz and Manski, 1997; Pistaferri, 2001; Manski, 2004; Kaufmann and Pistaferri, 2009; Stoltenberg and Uhlendorff, 2022; Koşar and Van der Klaauw, 2023; Almås et al., 2023), they are not typically available together with actual earnings realizations and consumption data, which likely limits their usefulness.[1] To overcome this concern, we utilize the Copenhagen Life Panel, a recently collected panel survey on subjective earnings expectations, matched with administrative data on realized earnings and

[1] Meghir and Pistaferri (2011) mention this data limitation and further discuss the possible issues with validations.



Note: This figure illustrates a key identification strategy for the paper. For notation, $E_{t-1}[y_t]$ stands for expected earnings in t evaluated in $t-1$. c_t is registry imputed consumption in t.

Figure 1: Identification strategy

imputed consumption.

We illustrate our key identification strategy using the expectations, realizations, and consumption in Figure 1. In our notation, y_t and c_t represent individuals' earnings and consumption in year t, respectively. First, we identify the unexpected earnings shocks as $y_t - E_{t-1}[y_t]$ using both earnings expectation in t evaluated one year before in $t-1$ ($E_{t-1}[y_t]$) from the survey and the corresponding earnings realization (y_t) from the registry. Unexpected earnings shocks refer to new information that workers did not anticipate in $t-1$. Second, we quantify the extent of belief changes by examining changes in long-term (five-year ahead) beliefs as $E_t[y_{t+5}] - E_{t-1}[y_{t+5}]$. Note that $E_{t-1}[y_{t+5}]$ is formed before they observe the unexpected earnings shocks, whereas $E_t[y_{t+5}]$ is established after they see the earnings shocks. Therefore, this belief change captures how much workers change their long-term expectations upon the earnings shock. Lastly, we investigate the effect of belief changes on subsequent consumption changes using imputed consumption (Browning and Leth-Petersen, 2003) in the registry as $c_t - c_{t-1}$.

We first show empirical evidence of how workers change their long-term beliefs in response to unexpected earnings shocks. We construct the "*belief change ratio*" which captures the degree to which workers internalize unexpected earnings shocks ($y_t - E_{t-1}[y_t]$) into changes in their $t+5$ expectations ($E_t[y_{t+5}] - E_{t-1}[y_{t+5}]$):

$$\text{Belief change ratio} = \frac{E_t[y_{t+5}] - E_{t-1}[y_{t+5}]}{y_t - E_{t-1}[y_t]}. \tag{1}$$

The belief change ratio reflects a proportion of how much workers internalize the earnings shocks into long-term ($t+5$) beliefs. If workers think that the shocks are mostly transitory, the measure will converge to 0. Otherwise, if workers believe that the shocks have permanent components, the measure will depart from 0. Therefore, it is an empirical measure capturing the degree to which workers perceive the earnings shock as permanent.

Using this new measure, we compare the two distinct assumptions in the income process literature, allowing us to generate testable predictions on the belief change ratio over the life cycle. Under the complete information assumption (Lillard and Willis, 1978), workers have full information about unexpected earnings shocks. Thus, they can perfectly distinguish between permanent and transitory shocks. With this assumption, the belief change ratio represents the proportion of realized permanent shocks to unexpected earnings shocks. If we assume constant shock parameters across the life cycle, as it is common in the literature, then the average belief change ratio is constant on average across the life cycle.

By contrast, the learning assumption (Guvenen, 2007) predicts a decreasing life cycle pattern of the belief change ratio. Guvenen (2007) models heterogeneity in the individual earnings growth and assumes workers have incomplete information about the individual-specific parameter and the persistence component in the income process. Under this assumption, workers do not distinguish at all between permanent and transitory earnings shocks and instead learn about their individual income process from their realized earnings using Kalman filtering. The individual earnings growth parameter is realized at the beginning of life and fixed over the life cycle. Therefore, individuals at different life stages will possess different amounts of information because older workers have a longer history of realized earnings, which are informative about their income process, especially their individual-specific parameters. This assumption implies the belief change ratio decreases over the life cycle because older workers have less uncertainty about their income process, which makes the new earnings shocks less informative.

We test these different predictions using the survey-measured belief change ratio. We construct an individual-level belief change ratio from the survey data and find a clear decreasing pattern of the average belief change ratio over the life cycle. This finding is consistent with the prediction of the learning assumption not only about the decreasing pattern but also about the levels across the life cycle.

Although the learning assumption closely fits the average belief change ratio over the life cycle in the survey, we show the belief change ratio in the survey is far more heterogeneous than the learning assumption predicts. The low heterogeneity from the learning assumption arises because workers do not have any visibility into the nature of earnings shocks. For a given worker's age, the learning assumption predicts the same belief change ratio level, which also implies age is the only predictor of heterogeneity in the belief change ratio. However, we find the standard deviation of the belief change ratio across individuals in the survey is more than six times higher than that predicted by the learning assumption. Overall, our empirical findings suggest that both the complete information and learning assumptions have limitations in explaining the belief change ratio in the survey. Specifically, the complete information assumption cannot explain the decreasing belief change ratio over the life cycle, and the learning assumption cannot capture the heterogeneity in the belief change ratio.

We propose a new partial information assumption, a hybrid of complete information and learning assumptions, to explain both the life cycle pattern and heterogeneity in the belief change ratio. The key idea is that workers know some parts of the nature of their earnings shocks. We decompose the variance of permanent shocks and transitory shocks into two different parts: known/unknown permanent shocks and known/unknown transitory shocks. In our model, workers can differentiate the known parts of these shocks, but they still solve the Kalman filtering problem for the unknown parts. To match the heterogeneity and life cycle patterns of the belief change ratio, our estimates show workers know 56% (41%) of permanent (transitory) shocks. We show that the partial information assumption is essential to match the size of empirical earnings shocks in the survey. The partial information assumption closely matches the mean squared earnings shocks in the data. In contrast, the complete information assumption underestimates the mean squared earnings shocks by 18% and the learning assumption overestimates them by 15%.

We further investigate the implications of partial information on consumption both empirically and through the lens of the model. We first empirically demonstrate that the belief change ratio is important in explaining the heterogeneous consumption response from earnings shocks. We estimate an average consumption elasticity to unexpected earnings shocks of 0.3. We then show this elasticity is larger for workers with a higher belief change ratio as a higher belief change ratio reflects the degree to which workers believe that unexpected earnings shocks are more permanent. For example, a one percentage point increase in the belief change ratio increases consumption elasticity by 0.17 percentage points. This empirical evidence underscores the importance of survey-measured belief changes in understanding the heterogeneity in consumption responses to unexpected earnings shocks.

We then show the empirical correlation between consumption elasticities across the belief change ratio is quantitatively consistent with the prediction of the partial information assumption. To show this, we simulate a standard life-cycle consumption saving model under three different assumptions: (1) complete information, (2) learning, and (3) partial information. We then show the partial information assumption closely matches empirical consumption elasticity across different levels of belief changes, whereas the other two assumptions do not match. This evidence demonstrates the usefulness of the partial information assumption in the consumption model.

Lastly, we investigate the implications of partial information for the degree of consumption smoothing (Blundell et al., 2008; Kaplan and Violante, 2010). Under the complete information assumption, consumption changes are predicted to be primarily driven by permanent shocks, whereas workers are predicted to largely smooth their consumption in response to transitory shocks. In contrast, under the learning assumption, workers are predicted to smooth their consumption to the same extent in response to both permanent and transitory shocks due to non-differentiability. The partial information assumption predicts workers' consumption smoothing is 17% higher for permanent shocks and 13% lower for transitory shocks compared to the complete information benchmark. The main reason for this is that workers do not know exactly the nature of the shocks. These new estimates are consistent with the empirical finding (for instance, Parker et al., 2013) that workers smooth their consumption less in response to transitory shocks. A better understanding of consumption smoothing has wide-ranging policy implications, including evaluating the aggregate impact of tax and labor market policy.

Related literature. First, this paper contributes to the growing literature using subjective expectations to empirically identify the information that workers possess. Since Dominitz and Manski (1997) and, more recently, Almås et al. (2023) highlighted the importance of survey expectations, there has been a surge of interest in using subjective expectations to identify model objects in the income process literature. For instance, Pistaferri (2001) and Kaufmann and Pistaferri (2009) both use the Italian Survey of Household Income and Wealth to identify the degree of transitory shocks. More recently, Stoltenberg and Uhlendorff (2022) use the same dataset and identify the amount of advanced information workers possibly have. Additionally, there is a strand of literature using the New York Fed's Survey of Consumer Expectations. Koşar and Van der Klaauw (2023) examine the heterogeneity in expected earnings growth and report the heterogeneity across the life stages and business cycle. Commault (2022) uses the same survey to identify the degree of the persistent component and demonstrate workers with higher persistent components are more responsive to transitory shocks. Lastly, Wang (2023) reports that perceived risk in the New York fed survey data is significantly lower than the calibrated risk. The Copenhagen Life Panel provides several advantages relative to the other datasets in the literature. It is matched with registry data; therefore, it allows access to earnings realization and imputed consumption on top of expectations. Using these expectations, realizations, and consumption, we first validate the collected expectations data.[2] We also make a direct link between expectations and consumption decisions which is a key focus of this paper.

Second, a strand of research models workers' incomplete information on the income process. The discussion on the possibility of workers having incomplete information about their income process was initiated by Pischke (1995). Pischke (1995) introduces a model in which workers are assumed to have incomplete information compared to econometricians regarding the macroeconomic states of the economy. Blundell and Preston (1998) and Cunha et al. (2005) emphasize the importance of the imperfect differentiation between permanent and transitory shocks in consumption response estimators. Guvenen (2007) is one of the first papers to combine modeling incomplete information with belief updating about individual income processes. A key innovation of this paper is the incorporation of individual-specific earnings growth rates, allowing workers to learn about the underlying individual-specific parameters gradually. Guvenen (2007) shows incorporating incomplete information is important in explaining cross-sectional consumption inequality over the life cycle. Guvenen and Smith (2014) estimate the degree of partial insurance and incomplete information within this framework. In a more recent study, Druedahl and Jørgensen (2020) also model incomplete information and learning within the income process, but without incorporating individual-specific earnings growth parameters as in Blundell et al. (2008). Their main contribution lies in modeling private signals that are informative about the persistence parameter. We contribute to this body of literature by empirically examining belief changes without making stringent assumptions about underlying learning structures. Subsequently, we propose a new partial information model to match these empirical survey beliefs. We then simulate consumption based on the partial information and further validate it by comparing it with the actual consumption.

Third, this paper adds to the existing literature on understanding consumption responses to earning shocks. Blundell et al. (2008) use imputed consumption in PSID data and empirically estimate the degree of consumption insurance to permanent and transitory shocks. There are attempts to generalize the structural model of Blundell et al. (2008). For instance, Commault (2021) allows the pass-through of transitory shocks and explains significant and positive consumption responses to transitory shocks. Kaplan and Violante (2010) simulate an incomplete market model and compare the degree of consumption smoothing with the empirical counterpart in Blundell et al. (2008). One of the goals of these strands of papers is to correctly estimate the consumption response upon the

[2] Caplin et al. (2023) report many validation exercises between the survey-measured earnings risk and registry counterpart. They also report validations for the job transition expectations.

permanent and transitory earnings shocks. It is not explicitly discussed how workers perceive earnings shocks in the literature, even though this is crucial for understanding subsequent consumption decisions. This paper fills this gap by providing empirical data on workers' perceptions of earnings shocks.

Lastly, a strand of papers shows the impact of earnings shocks on consumption responses using natural experiments.[3] Parker et al. (2013) and Misra and Surico (2014) show the positive marginal propensity to consume (MPC) in response to the government tax rebate. Fagereng et al. (2021) find a positive MPC from lottery shocks. Agarwal and Qian (2014) demonstrate that growth dividend in Singapore and confirm very high MPC. Lastly, Fuster et al. (2021) use the NY Fed Survey of Consumer Expectations to ask respondents about their consumption plans from \$500 paychecks. The literature broadly confirms that the MPC out of the transitory shocks is significantly greater than zero. Recently, there have been efforts to identify heterogeneity in the consumption response to earnings shocks. Lewis et al. (2019) find substantial heterogeneity in the marginal propensity to consume (MPC) in response to government stimulus during the 2008 financial crisis. Bunn et al. (2018) find a smaller MPC for positive shocks than for negative shocks. Crawley and Kuchler (2023) highlight the level of wealth as a source of heterogeneity in MPC. We contribute to the literature by demonstrating that the empirical belief changes can serve as an important explanatory factor for understanding heterogeneity in the consumption response to an earnings shock.

The paper is organized as follows. Section 2 introduces the survey and the registry instruments and validates the survey responses. Section 3 shows how workers change their beliefs upon earnings shocks. Section 4 describes the partial information model to explain the empirical pattern in the survey. Section 5 empirically describes the consumption response from earnings shocks and its interaction with the belief changes. Through the lens of the model, we also outline the incomplete market simulation across different belief change assumptions. Section 6 concludes.

2 Survey and matched administrative data

In this section, we present the survey data from the Copenhagen Life Panel and its registry data counterpart. We also outline the key identification strategies to be employed in the main analysis and we conduct a baseline validity check of survey responses.

[3] Havranek and Sokolova (2020) conduct a meta-analysis of previous papers on the consumption response to earnings change.

2.1 Copenhagen Life Panel

This subsection outlines the survey questionnaires and sample characteristics. We use the Copenhagen Life Panel (CLP), an annual panel survey that focuses on earnings expectations. Our sampling strategy involves selecting 100,000 individuals between the ages of 20 and 70 from the population registry and inviting them to participate in the online survey. The first round of the CLP (CLP1) was administered in January 2020 and the second round (CLP2) took place in January 2021. It is worth noting that the CLP is an ongoing panel, with four waves conducted as of January 2023.

The surveys were administered online, yielding response rates of approximately 15%. We have 6,621 repeaters who completed both CLP1 and CLP2. Furthermore, we narrow down our sample to individuals aged between 20 and 60 who reported being currently employed in CLP1 resulting in a final sample size of approximately 5,847 individuals.

Our questionnaire in CLP1 is as follows. In Jan-2020, after we ask about their 2019 (last year) earnings, we first ask about their one-year ahead (2020) and then subsequent five-year ahead (2025) earnings expectations. The exact questionnaire is as follows.

- We are now going to ask you about your beliefs about your future earned income before tax. Assume no inflation such that a dollar in the future is worth the same as a dollar today in terms of how much you can buy for it.

1. *Think about your potential earned income during 2020. How much do you expect?*

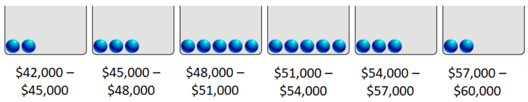
2. *Think about your potential earned income during 2025. How much do you expect?*

One year later in Jan-2021, we ask again for their upcoming year (2021) and the subsequent five years (2025), using an identical introductory preamble as before. The key design of CLP involves matching the 2025 expectations between CLP1 and CLP2 to capture empirical belief changes for the same long-term year, 2025.

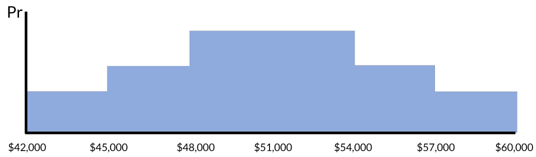
3. *Think about your potential earned income during 2021. How much do you expect?*

4. *Think about your potential earned income during 2025. How much do you expect?*

We elicit the whole distribution of their beliefs about future earned income using "balls in bins" elicitation device that is visually oriented (Delavande and Rohwedder, 2008). In our specific implementation, respondents are first asked to state the minimum and maximum values for possible future earnings, as in the pioneering work of Dominitz and Manski (1997). Then the range between the stated minimum and maximum was divided into six



(a) Elicitation



(b) Distributions

Note: Panel (a) shows the sample screen for the elicitation and Panel (b) shows how we interpret the distribution of the answer in Panel (a) as a mixture of uniform distributions.

Figure 2: Balls in Bins

equally sized bins. Respondents were then instructed to move 20 balls into the six bins to reflect how likely earnings are to fall in each of the ranges. Figure 2(a) shows an example of a response in the form of what respondents see while filling out the survey. Delavande and Rohwedder (2008) show that this method elicits usable and reasonable answers.[4] We interpret the distribution as a combination of uniform distributions where each ball represents 5% probability and the probability density is equally distributed within the bin. Fig 2(b) shows an interpretation of Fig 2(a). The labels of each bin are transformed into a logarithmic scale for analysis.

2.2 Danish administrative data

One of the distinctive features of CLP is its matched administrative data, which provides a means of validating the survey responses by giving access to the realized earnings and imputed consumption, which enriches our analysis.

We describe how we construct the registry earnings and imputed consumption. We have access to monthly earnings data from all sources reported for tax purposes in the monthly income-tax registry. To construct annual earnings from the registry, we sum up the monthly earnings from all employers within a year.[5]

For consumption, we follow the imputing method of Browning and Leth-Petersen (2003) and Fagereng and Halvorsen (2017). We use the information on non-housing and non-pension wealth (NW), pension contributions (P), and (after tax and transfer) disposable earnings (Y^{disp}). Then we can impute the consumption for individual i and year t as a

[4] Goldstein and Rothschild (2014) show that bins and balls elicitation increases the accuracy of reported distribution compared to other non-graphical elicitation methods.

[5] We compare this total earnings sum with earnings from the main employer and find that 93% of respondents have one main employer.

gap between disposable earnings and changes in net wealth and pension contribution.

$$C_t^i = Y_t^{\text{disp},i} - (\Delta NW_t^i + P_t^i) \tag{2}$$

The idea behind this imputation method is to approximate workers' savings by analyzing changes in their net wealth and pension contributions over the course of the year. In Denmark, financial institutions are mandated to provide reports on the value of their client's accounts as of December 31st each year. Given that the tax reporting year spans from January 1st to December 31st, the data on income and wealth reported in the tax returns align with the information required to utilize this approach to impute consumption. A lot of validation efforts have been done to assess the reliability and accuracy of the imputation method. For instance, Abildgren et al. (2018) demonstrate a strong correspondence between imputed consumption and survey reported consumption for the period spanning 2002 to 2015.[6] Baker et al. (2018) also demonstrate that the measurement gap of the imputation method compared to self-reported consumption is relatively small in the German dataset. In the Online Appendix A.3, we provide a more detailed discussion of potential issues related to the imputation methods used in this study.

Table 1 presents the sample characteristics of our study compared to the population in the registry. As previously mentioned, our survey includes employed individuals aged between 20 and 60. To match this criterion in the registry, we restrict the samples to individuals with earnings greater than 24,000 DKK in 2019 and age between 20 and 60. It is worth noting that the distribution of females in the survey closely aligns with the population, with both groups comprising approximately 50%. Regarding age, we observe that, on average, survey participants are 4.1 years older than the population. However, it is important to highlight that the survey sample still exhibits a wide range of age groups, capturing the diverse life cycle stages.

Survey participants exhibit higher levels of education compared to the general population, with 48% of the survey sample having education beyond the college level, compared to 35% in the population. Additionally, survey respondents have higher earnings, as evidenced by their average annual earnings before tax being approximately 13% higher, average disposable earnings being 11.2% higher, and average imputed consumption being 7% higher compared to the population.

[6] For the survey comparison, Abildgren et al. (2018) use Danish household budget surveys. The definition of consumption in the survey follows the standard national account.

Table 1: Demographics

	Survey Sample	Registry
N	5,867	2,877,935
Female	0.49	0.5
Age		
20-29	0.19	0.25
30-39	0.21	0.25
40-49	0.25	0.24
50-60	0.35	0.26
Education		
Above college	0.48	0.35
Annual earnings (DKK)		
Mean	418,745	367,721
S.D.	272,111	309,593
Disposable earnings (DKK)		
Mean	293,531	262,346
S.D.	228,312	249,473
Imputed consumption (DKK)		
Mean	264,866	231,091
S.D.	178,346	196,322
Liquid wealth (DKK)		
Mean	341,594	214,663
S.D.	674,954	2,734,836
Liquid constrained	0.32	0.40

Note: The table presents a comparison of the average demographic characteristics of the survey sample and the Danish population as observed in the administrative data in 2019. The column "Survey Sample" comprises repeaters and employed respondents from both Copenhagen Life Panel 1 and 2. The "Registry" column comprises all individuals aged 20-60 within the Danish population who earned a minimum of 24,000 DKK in 2019. We also dropped workers who have a self-employed income of more than 24,000 DKK in 2019. The liquid wealth is the value at Dec-31st of 2018. In Jan-2020, the exchange rate for 1 US Dollar was approximately 7 Danish Krone (DKK).

1. Expectations	2. Realizations	3. Consumption
• $E_{19}[y_{20}]$ (CLP1) • $E_{19}[y_{25}]$ (CLP1) • $E_{20}[y_{25}]$ (CLP2)	• y_{20} (Registry)	• c_{19} (Registry) • c_{20} (Registry)

Figure 3: Data summary

Furthermore, the average liquid wealth (excluding housing) of survey participants is 16% higher. When considering liquid constraints, defined as the ratio of liquid wealth to annual earnings being less than 0.2, approximately 32% of survey participants are liquid constrained. This percentage is comparable to the registry data, where 40% are identified as liquid constrained. To account for these differences, we apply population weights to the results. Further details on the construction of these weights are in Online Appendix A.1. We also provide a comparison of sample characteristics for non-repeaters in the Online Appendix A.2 where we confirm that, in most characteristics, it is well aligned with the repeaters' demographics.

2.3 Data summary and identification

We now provide a summary of the empirical data collected and an outline for identifying key variables. We have individual-level expectations, realization, and consumption decisions summarized by Fig 3. For the expectations, in CLP1, we collected individual-level data on 2020 and 2025 earnings expectations, which were evaluated in early January 2020. To avoid confusion in notation, we denote these variables as $E_{19}[y_{20}]$ and $E_{19}[y_{25}]$ to indicate that they were evaluated in 2019 instead of January 2020. Likewise, in CLP2 collected in January 2021, we measured $E_{20}[y_{25}]$. For the realizations, we have y_{20} from the registry value. Lastly, for consumption decisions, we have imputed consumptions from the registry c_{19} and c_{20}.

We introduce how we use these components to identify three key objects: 1) unexpected earnings shocks, 2) belief changes, and 3) consumption changes. First, we define the unexpected earnings shock for individual i in 2020 as the difference between the realized log earnings from the registry data and the expected log earnings: $y_{20}^i - E_{19}[y_{20}^i]$. The key tension in defining shocks in the literature revolves around the advanced information that workers may possess, which is not directly observable to the researcher. By using subjective expectations, we can identify shocks more effectively because we directly measure the unobserved expectations that workers hold.

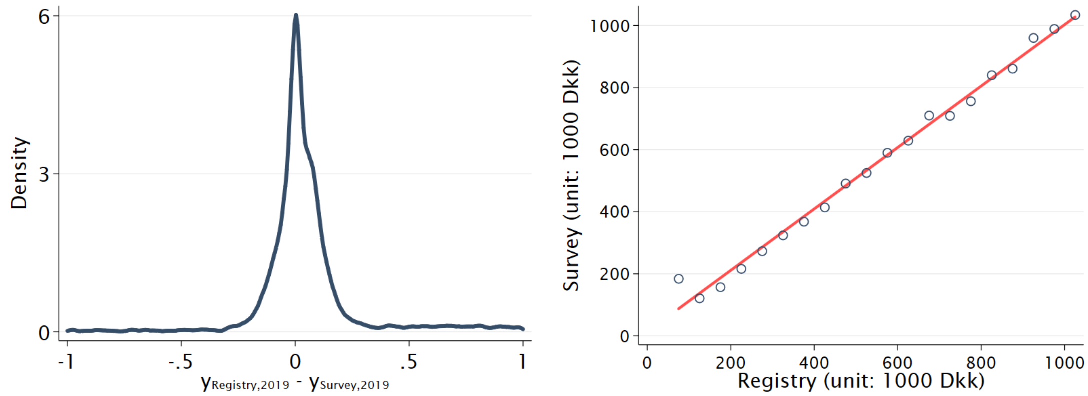
The second object of interest is workers' belief changes, where we focus on how workers adjust their long-term earnings expectations. To capture this adjustment, we align the timing between CLP1 and CLP2, specifically in the year 2025. This allows us to measure the extent to which workers modify their earnings expectations based on the earnings shocks they experienced $(y^i_{20} - E_{19}[y^i_{20}])$. We utilize $E_{20}[y^i_{25}] - E_{19}[y^i_{25}]$ as the degree to which workers incorporate permanent components within the earnings shocks. We expect as workers incorporate a greater proportion of permanent components, their adjustments to the 2025 expectations become more pronounced.

The third object is consumption change $c^i_{20} - c^i_{19}$. We are interested in how workers' consumption changes upon the earnings shocks. Moreover, we aim to establish the correlation between the degree of consumption response and belief change. Our empirical data is valuable in directly examining the relationship between belief and subsequent consumption changes, as many consumption-saving models emphasize the significance of beliefs in driving such changes.

2.4 Comparison between survey and registry

As Meghir and Pistaferri (2011) point out, the validation is crucial to ensure the reliability of the survey expectations data. More recently, Almås et al. (2023) also highlight the importance of effort in validating new measurements to ensure that they capture the phenomena they are meant to. In this subsection, we conduct a first-order validation of survey responses using the unique matched survey and registry structure.

To begin, we check whether workers have accurate information about their past earnings. This serves as a fundamental first-order check, which would typically be challenging to ascertain without the use of the matched registry and survey structure. Prior to asking the earnings expectations question, we ask about their earnings in the last (2019) year in CLP1. We compare the self-reported earnings in 2019 from the survey with the corresponding earnings value recorded in the registry. Fig 4 shows the distribution of the log-scaled individual-level difference between the registry $(y^i_{r,19})$ and survey self-reported values $(y^i_{s,19})$ of earned income for 2019. Fig 4(a) shows the distribution of the gap between $y^i_{r,19}$ and $y^i_{s,19}$. Approximately 75% of respondents fall within the ± 0.1 boundary, indicating a high level of correspondence and confirming that workers understand the meaning of earned income and follow their earnings in the last year. In Fig 4(b), we plot the mean of survey reported earnings $y^i_{s,19}$ against 20 different levels of registry earnings $y^i_{r,19}$. The plot shows a strong linear relationship along the 45-degree line, indicating a high level of correspondence across the different levels of earnings. This



(a) Distributions

(b) Levels

Note: Panel (a) illustrates the distributions of the gap between self-reported 2019 earnings and their registry counterparts. Panel (b) shows binned-scatter plots depicting the mean of self-reported survey earnings across different levels of registry earnings.

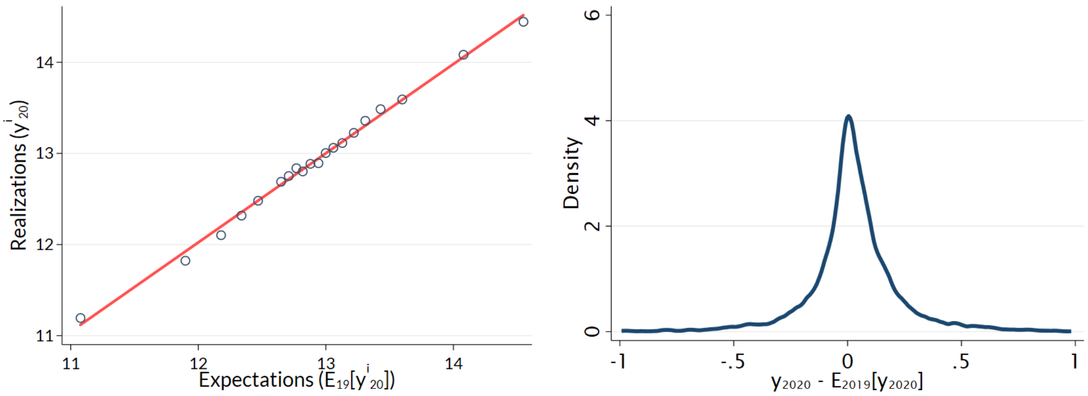
Figure 4: Last year's earnings comparison

high correspondence shows the workers follow how much they earned last year which is important to give baseline credibility to their expectation response.

As a second validation, we compare expectations and realizations cross-sectionally. If the expectations data is correctly measured, we expect it to match the actual realization at the cross-sectional level. Fig 5(a) first compares the average level of expectations (X-axis) and realizations (Y-axis). We divide the sample into 20 equally sized groups based on their expected levels and then display the average expected and observed levels within each group. The scatter plots closely follow the 45-degree line, indicating a strong alignment between expectations and realizations across various levels. This validates the survey expectations, showing that it is hard to find systematic optimism or pessimism in workers' expectations.

In Fig 5(b), we show the distribution of unexpected earnings shocks ($y^i_{20} - E_{19}[y^i_{20}]$). Even though the average level of expectations corresponds to the average realizations, there are substantial variations in the distribution. We find that 44% respondents are in the ± 0.1 boundary and that the distribution of earnings shocks is centered around zero. In Online Appendix A.4, we show that the distribution of earnings shocks is stable around 0 across the different levels of current earnings (y^i_{19}), life cycle, education, gender, and job transitions which confirms the earnings shock is not biased conditional on various demographics and labor market experience.

Finally, we show the first-order evidence that the cross-sectional variations in earnings expectations are correlated with the levels of consumption. The precautionary saving mo-



(a) Levels (b) Distribution of earnings shocks

Note: Panel (a) shows the expected earnings ($E_{19}[y_{20}^{i}]$) on the X-axis and the realizations in the registry data y_{20}^{i} on the Y-axis. We construct the 20 equal-sized groups based on the X-axis and plot the empirical mean of each group. Panel (b) displays the distribution of earning shocks in 2020 ($y_{20}^{i} - E_{19}[y_{20}^{i}]$).

Figure 5: Comparing expectations and realizations

tive suggests that workers with higher mean in expected earnings tend to consume more, while those with higher levels of uncertainty in future earnings tend to spend less due to self-insurance motives. If expectations data is valid, we expect the expectations data to be correlated with their variations in consumption decisions. To show the validations we run the following regression:

$$c_{19}^{i} = \beta_1 y_{19}^{i} + \beta_2 E_{19}[y_{20}^{i}] + \beta_3 SD_{19}[y_{20}^{i}] + \beta_0 + X_i'\gamma + \epsilon_i \tag{3}$$

The dependent variable, c_{19}^{i}, represents the log-scaled consumption of individual i during 2019. The explanatory variables include the individual's current level of log earnings (y_{19}^{i}), mean expected earnings ($E_{19}[y_{20}^{i}]$), and standard deviation in expected earnings ($SD_{19}[y_{20}^{i}]$). Control variables (X_i) include age, age-squared, gender, and education.

Table 2 presents the regression results. In column (1) where no controls are added, we observe a positive and statistically significant coefficient at the 1% level for $E_{19}[y_{20}^{i}]$, indicating that workers with higher expected earnings tend to have higher levels of consumption. Furthermore, the coefficient of $SD_{19}[y_{20}^{i}]$ is negative and statistically significant at the 5% level suggesting that workers facing greater uncertainty about their future earnings tend to spend less, consistent with the precautionary saving motive. Column (2) shows the results are robust with the various control variables. This result again confirms the validity of the data, as it demonstrates that the cross-sectional variations in expectations data are correlated with variations in consumption decisions, consistent with the precautionary saving motive.

Table 2: Cross-sectional consumption and expectations

	(1)	(2)
	Dependent variable: c^i_{19}	
y^i_{19}	0.385***	0.273***
	(0.036)	(0.033)
$E_{19}[y^i_{20}]$	0.453***	0.391***
	(0.064)	(0.059)
$SD_{19}[y^i_{t+1}]$	-0.488**	-0.479**
	(0.203)	(0.210)
Controls	N	Y
R-sq	0.271	0.314
N	5,867	5,867

Note: Standard errors are clustered at the individual level. The dependent variable is log scale consumption. Control variables include age, age-squared, gender, and education.

In this section, we introduce and validate our survey by comparing survey responses with matched administrative data. It is worth highlighting the significance of having access to matched survey and registry data for conducting important validations. Such data integration is crucial for performing even the most fundamental validations, ensuring the robustness of measured expectations data.

3 Belief changes upon earnings shock

Building on the validity in the previous section, in this section, we show how workers change their beliefs upon unexpected earnings shocks. We first make a testable prediction based on two important assumptions (complete information and learning) about belief changes in the heterogeneous income process framework. The heterogeneous income process is a class of income process models that specifically model individual heterogeneity in earnings growth (Lillard and Willis, 1978; Hause, 1980; Guvenen, 2009). This income process is especially relevant for the learning assumption because the learning assumption describes workers learning about their individual income processes.[7] Therefore, we first outline how a heterogeneous income process models earnings realizations. Subsequently, we introduce how the complete information and learning assumptions model belief changes differently upon the same earnings realization process.

[7] We note that the prediction on the complete information holds in a more general structure including the restricted income process. We have a discussion in Online Appendix A.7.

3.1 Earnings realization process

The heterogeneous income process assumes that the log earnings (y_t^i) dynamics for individual i at age t can be decomposed into three distinct components: the observable component, the individual component, and the stochastic component.[8]

$$y_t^i = \underbrace{g(X_t^i)}_{\text{observable}} + \underbrace{\beta_i t}_{\text{individual}} + \underbrace{(z_t^i + \epsilon_t^i)}_{\text{stochastic}} \tag{4}$$

In equation (4), first the observable component, X_t^i represents the individual's observable characteristics such as age, gender, and education. The function $g(X_t^i)$ captures the average predicted dynamics of earnings based on X_t^i.

The second individual component captures heterogeneity in earnings growth rates. The parameter β_i is an individual-specific parameter determined at the beginning of labor market participation from the normal distribution $\mathcal{N}(0, \sigma_\beta^2)$. The distribution of β captures the heterogeneity in individual earnings growth rate. For instance, the variation in β can arise from differences in human capital and heterogeneity in skill acquisition which makes differences in earnings growth over the life cycle.

Lastly, the stochastic components include the standard AR(1) permanent and transitory shocks components, where $z_t^i = \rho z_{t-1}^i + \eta_t^i$ and η_t^i and ϵ_t^i are from $\mathcal{N}(0, \sigma_\eta^2)$ and $\mathcal{N}(0, \sigma_\epsilon^2)$ respectively that are independent and identically distributed (i.i.d.) over time and across individuals. In summary, there are 4 different parameters that describe the earnings dynamics: σ_β^2, σ_η^2, σ_ϵ^2, and ρ.

We estimate the parameters for earnings realizations in a standard minimum distance method (Guvenen, 2009) using the registry. The description of the steps is as follows. First, to identify observable parts, we run a regression with log earnings as the dependent variable and age, age-squared, gender, and education as independent variables. Second, we calculate earnings residuals for each worker from 2015 to 2020. Using these values, we construct the empirical variance and covariance matrix of earnings residual across the life cycle. Third, we write down the model-predicted variance and covariance matrix as a function of the four parameters. For instance, the variance of residualized earnings at age t can be represented as follows: $\text{var}(y_t^i) = \sigma_\beta^2 t^2 + \sigma_\eta^2 \sum_{j=0}^{t-1} \rho^{2j} + \sigma_\epsilon^2$. After we write down this variance and covariance matrix, we use the minimum distance method with the identity weighting matrix and find four parameters minimizing the distance from

[8] In the survey, there are no time variations across survey respondents because we use the repeaters in CLP1 and 2. To avoid the use of notation, we interchange age and time.

Table 3: Estimated parameters

	Sample	ρ	σ^2_β	σ^2_η	σ^2_ϵ
(1)	All	0.921	0.00084	0.028	0.026
		(0.004)	(0.00002)	(0.003)	(0.001)
(2)	Male, University or above	0.910	0.00113	0.024	0.036
		(0.019)	(0.00010)	(0.005)	(0.002)
(3)	Male, Below University	0.933	0.00054	0.029	0.026
		(0.008)	(0.00006)	(0.009)	(0.005)
(4)	Female, University or above	0.841	0.00095	0.032	0.019
		(0.023)	(0.00014)	(0.007)	(0.004)
(5)	Female, Below University	0.885	0.00063	0.040	0.030
		(0.018)	(0.00008)	(0.009)	(0.011)

Note: This table shows estimated parameters across each sample criterion. Standard errors are in parentheses. "University or above" refers to workers possessing education beyond the university level, while "below university" indicates workers with education levels lower than a university degree.

the empirical counterpart. In Online Appendix B, we have a more detailed estimation description.

Table 3 presents the estimated parameters for the income process. The first specification shows the estimated parameters for all samples. Notably, the persistence parameter (ρ) is estimated to be less than 1, indicating that the degree of permanent shock is relatively mild. The parameter estimates for the variance of permanent and transitory shocks are around 0.028 and 0.026, which is lower than the estimates for the U.S. (Guvenen, 2009) which possibly comes from a relatively stable labor market and unemployment insurance system in Denmark. Lastly, we note that the variance of β_i is significantly greater than 0. Considering that it affects earnings with a multiplicative value of t, the magnitude is not small and bigger than the estimates in Guvenen (2007).

The second-to-last rows in Table 3 present the subgroup analysis across gender and education. We find that the persistence parameter and the variance of β_i are lower for females compared to males. However, the variance for the permanent shock is higher for females than for males. Additionally, among both males and females, the group with a university education or above exhibits a higher level of variance in β_i, while the variance for the permanent shock is lower in this group.[9]

[9] In the Online Appendix B, we provide an alternative estimation method that allows for variations in age following the approach of Karahan and Ozkan (2013). We will later use these specifications will be used for robustness checks of the results.

3.2 Assumptions about belief changes

In this subsection, we compare the assumptions of complete information and learning in the existing literature, focusing on how each assumption describes earnings shocks and the following belief changes.

First, the complete information assumes that worker i at any age t has complete information on the realized value of individual-specific parameter β_i, the current level persistence component z_t^i, the amount of realized permanent and transitory shock η_t^i, and ϵ_t^i. The complete information assumption describes earnings shocks $(y_t^i - E_{t-1}[y_t^i])$ as the sum of the realized permanent shock (η_t^i) and transitory shock (ϵ_t^i). Workers perfectly see every component inside of the earnings shocks and they only use the permanent shock in belief changes.

Second, the learning assumption (Guvenen, 2007) models workers enter the labor market without knowing the realized value of β_i and have prior beliefs about their β_i as $\mathcal{N}(0, \sigma_\beta^2)$ and z_1^i as $\mathcal{N}(0, \sigma_\eta^2)$. Unlike the complete information, the workers only observe the realized earnings (y_t^i) as an informative signal about the income process but do not see the components in it. Therefore workers do not see the nature of earnings shocks at all. Instead, using the parameters of the income process, workers change their beliefs about β_i and z_t^i as a Kalman filtering problem. The key feature of this filtering problem is workers are learning a certain proportion of earnings shocks.

$$E_t[y_{t+5}^i] - E_{t-1}[y_{t+5}^i] = G(t) \times (y_t^i - E_{t-1}[y_t^i]) \tag{5}$$

Equation (5) describes how workers update their beleifs about $t+5$ earnings. On the left-hand side of the equation, we see how much worker i changes her mean beliefs about earnings y_{t+5}^i as she observes one more realization y_t^i. The right-hand side of the equation shows that workers learn in *proportion* to earnings shocks. In the learning assumption, the proportion $G(t)$ is determined by their age t. The $G(t)$ is a function of Kalman gain and the level of G changes across age t because of different amounts of information across life stages as workers are getting older, they have more earnings realizations which is more informative about their individual earnings process. We put the exact state representation equations with examples in the Online Appendix C.

The key difference between learning and complete information comes from that workers are learning about the individual-specific parameter (β_i) over the life cycle, even though it is already realized and remains the same within the individual. In the learning assumption, workers accumulate information about their β_i across the life cycle because they have more earnings realizations that make different amounts of information different across the life stage. Consequently, on average, as the worker's age increases, the mean beliefs about β_i are closer to the true realized value. On the contrary, the complete information assumes workers already know their realized β_i from the beginning and workers have the same amount of information about the underlying process across life stages.

Based on this key difference, we construct a measure called the *"belief change ratio"* that captures the extent to which workers incorporate their earnings shocks into long-term belief changes. As a measure of the extent to which workers internalize into long-term beliefs, we use the difference in expectations about earnings in $t + 5$, represented as $E_t[y^i_{t+5}] - E_{t-1}[y^i_{t+5}]$. We normalize this by dividing it by the size of the earnings shock at time t, represented as $y^i_t - E_{t-1}[y^i_t]$. Then the belief change ratio G defined below represents the proportion of how much workers internalize the earnings shock into the 5 years out expectations.

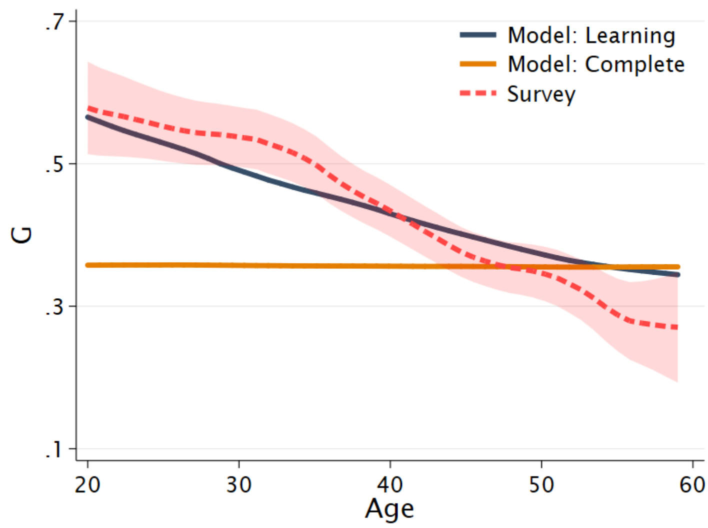
$$G = \frac{E_t[y^i_{t+5}] - E_{t-1}[y^i_{t+5}]}{y^i_t - E_{t-1}[y^i_t]} = \frac{\text{Expectations revision t+5}}{\text{Earnings shock t}} \tag{6}$$

For instance, if workers interpret the earnings shocks as solely originating from transitory factors, then G will be close to 0. Conversely, if workers perceive the shocks to have permanent components, G will be a positive value.

The complete information assumption interprets G as the proportion of permanent shocks among the earnings shocks $\frac{\eta^i_t}{\eta^i_t + \epsilon^i_t} \times \rho^5$. The income process parameters are assumed to be constant across the life cycle, therefore, on average G should be relatively flat across the life cycle because the shocks come from the same set of parameters. We also note that other income process such as the restricted income process also predicts the flat G because the definition of the measure is identical to the above.[10]

However, the learning assumption leads to a different life cycle pattern of G. The Kalman gain determines the extent to which workers learn from earnings shocks. As workers accumulate more earnings realizations, the variance in beliefs about β_i steadily decreases over the life cycle. As workers are more certain about the income process, the amount of the Kalman gain from earnings shock decreases. Therefore G which is a direct mapping from the Kalman gain also decreases over the life cycle. The magnitude of the decrease in G over the life cycle depends on the interaction between the parameters σ^2_β, σ^2_η, and σ^2_ϵ. As σ^2_β is larger, the decreasing pattern in G over the life cycle becomes more

[10] The detailed discussion about the difference between heterogeneous and restricted income processes is in Guvenen (2009).



Note: The blue and orange lines show the local linear predictions of the simulated belief change ratio across the life cycle. The red dotted line is the survey belief change ratio with 95% confidence intervals. The lines are a local regression fitted line.

Figure 6: Average belief changes over the life cycle

salient because the relative degree of learning about the individual profile is getting more salient. In the Online Appendix C, we provide simulation results across different sets of parameters.

3.3 Average belief change ratio over the life cycle

To compare the predictions of the two different assumptions with the survey-measured belief changes, we conducted two separate simulations of 100,000 workers' beliefs based on the parameters in the first row of Table 3. We construct the average level of G across the life cycle in each assumption. On the survey side, we measure $E_{20}[y^i_{25}] - E_{19}[y^i_{25}]$ and $y^i_{20} - E_{19}[y^i_{20}]$. Based on these data, we construct G for individual i as G_i as follows.

$$G_i = \frac{E_{20}[y^i_{25}] - E_{19}[y^i_{25}]}{y^i_{20} - E_{19}[y^i_{20}]} = \frac{\text{Expectations revision 25}}{\text{Earnings shock 20}} \tag{7}$$

Fig 6 presents a comparison between the average survey G_i and simulated G in each assumption across the life cycle. We use local regression smoothed lines plotted for survey G_i, learning G, and complete info G. There is a clear pattern that the average of survey G_i decreases over the life cycle. Notably, the average of learning G closely matches this decreasing life cycle pattern, both in terms of shape and level across the life cycle which confirms the validity of the learning assumption. In contrast, the average pattern of complete info G is relatively flat across the life cycle and the level is lower than the survey G_i. This finding highlights the importance of modeling incomplete information and validating the Kalman updating in the learning assumption.

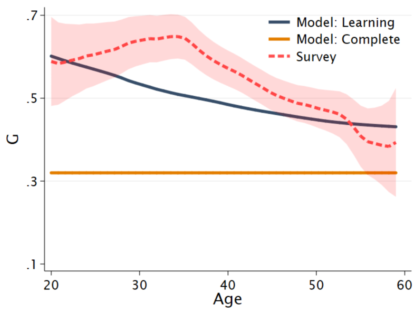
It is important to note that both the underlying income process and assumptions about belief changes affect the simulated G. To ensure the robustness of our results to the underlying specification of the earnings realization process, we try different subgroup analyses allowing different subgroup parameters in Table 3. Fig 7(a) and 7(b) show subgroup analysis for males and across different education levels. In both Fig 7(a) and Fig 7(b), there are clear patterns that the average of survey G_i is decreasing, while the average level of complete information G still predicts flat levels across the life cycle. This again confirms the previous result that the average learning G is closer to the average value of survey G_i. On the other hand, the complete information G deviates from the average survey G_i. Similarly, Fig 7(c) and 7(d) depict the results for females with different levels of education. The patterns observed are once again very similar, with survey G_i decreasing over the life cycle, and the learning assumption effectively capturing the downward slope patterns in both groups. This finding highlights the robustness of our results, indicating that the underlying earnings realization process does not drive the findings but rather reflects the process of belief updating. In Online Appendix A.5, we first allow age-dependent parameters following the method of Karahan and Ozkan (2013) and then simulate again learning and complete information G. We find that the complete information G assumption predicts a U-shaped pattern and fails to incorporate the decreasing pattern.[11]

Result 1. *The average belief changes over the life cycle observed in the survey data align with the learning assumption, while the complete information assumption fails to capture the pattern.*

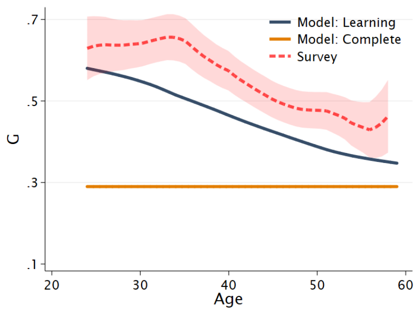
3.4 Heterogeneity in belief change ratio

While the learning assumption is useful in explaining the average pattern of survey G_i, its underlying assumption that all workers are unaware of the nature of earnings shocks may be overly restrictive and questionable. We further investigate the heterogeneity of G_i by comparing the predictions of the complete information G and learning assumptions G. The complete information assumes that workers can observe the nature of earnings shocks perfectly. As we mentioned, the complete information explains G as $\eta_t^i/(\eta_t^i + \epsilon_t^i) \times \rho^5$.

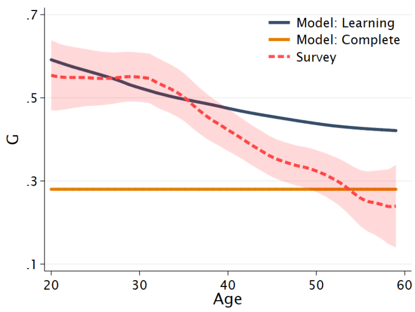
[11] In Online Appendix A.6, we also compare the median, instead of the mean, and find a very consistent pattern with the results in Fig 6.



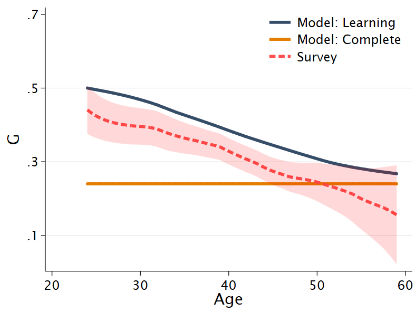
(a) Male, Below university



(b) Male, University or above



(c) Female, Below university



(d) Female, University or above

Note: Panel (a) and (b) show the results for males across each education group and panel (c) and (d) show the results for females across each education group. The line is a local regression fitted line with 95% confidence intervals.

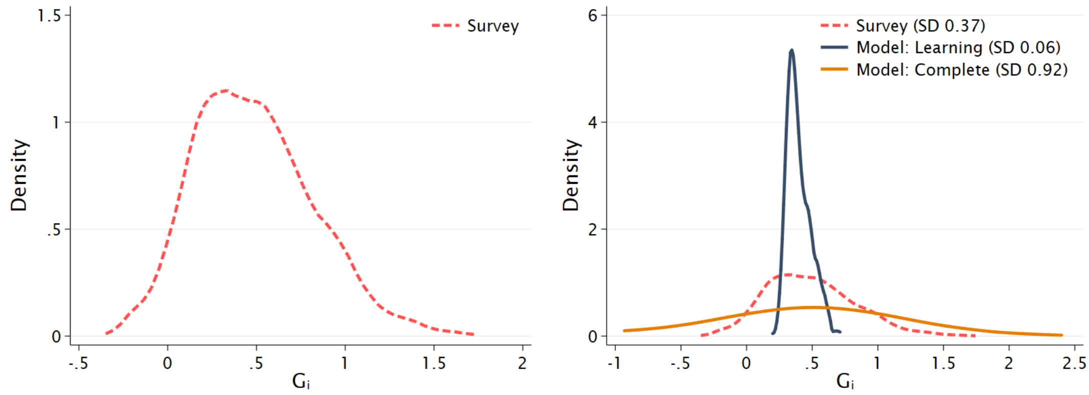
Figure 7: Robustness check: average belief change ratio over the life cycle

Heterogeneity in G among workers can arise due to the different realizations of η_t^i and ϵ_t^i even when drawn from the same parameter of the distribution. However, the learning assumption describes workers have no information about the earnings shocks which implies that workers of the same age have the same level of Kalman gain. Consequently, the only factor that can differentiate the level of G is age, and there is no heterogeneity of G among workers of the same age.

Fig 8(a) displays the cross-sectional distribution of survey G_i and Fig 8(b) overlays the same distribution of survey G_i with the simulated distribution of G from two different assumptions. In Fig 8(a), we observe that the standard deviation of G_i is approximately 0.37. About 88% of the samples exhibit positive values of G_i, indicating that the majority of workers adjust their 2025 earnings expectations in the same direction as the earnings shocks in 2020. In Fig 8(b), we plot the distribution of survey G_i with the simulated G. The most notable pattern is that the heterogeneity of learning G is much lower and more concentrated compared to the heterogeneity in the survey G_i. The standard deviation of learning G is 0.06, whereas the standard deviation of G_i in the survey data is 0.37, which is approximately six times higher in the data. By construction, learning G should be positive because workers adjust their 5-year ahead expectations following the direction of earnings shocks because the components of earnings shocks are not verifiable. This observation suggests that workers have heterogeneous interpretations of the earnings shock, and assuming a constant level of G based on age alone may be insufficient to explain this empirical heterogeneity.

On the other hand, the complete information assumption predicts a much larger degree of heterogeneity in G than the survey G_i. This comes from the fact that the realizations of η_t^i and ϵ_t^i are different in the simulated data set. Especially, the standard deviation of complete information G is 0.92 which is more than twice heterogenous than the survey G_i. The proportion of the negative G is around 32% a lot higher than the counterpart in the survey (12%). This result suggests that the degree of heterogeneity in survey G_i is relatively moderate than the two assumptions predict.

To examine the robustness of our findings, we conduct additional analyses to explore heterogeneity across different education and gender groups. We estimate separate parameters for each group and simulate the heterogeneity of the learning G based on these parameters. We then compute the 25th, median, and 75th percentiles of G as measures of its distribution within each group. We also compute the corresponding moments in the survey G_i and construct the same moments. We put the estimated parameters in Table 3. Fig 9 shows the moments of G across gender and education groups. The scatter



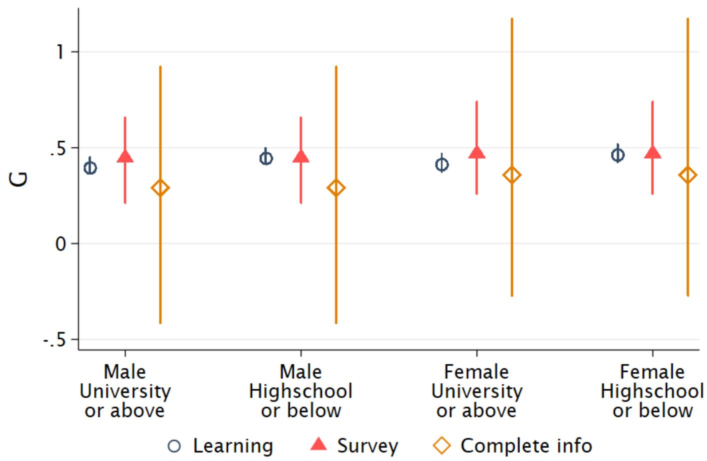
(a) Distribution G_i (b) Comparison with assumptions

Note: Panel (a) shows the distribution of G_i in the survey. Panel (b) shows the comparison of the distribution plot between the survey G_i and simulated G in learning and complete information assumption respectively. We excluded the top and bottom 2.5% of outliers from the analysis in each dataset.

Figure 8: Heterogeneity in belief change ratio

plot shows the median value of G respectively and the bar shows the 25-75 percentile of G. For the male and university or above group, which constitutes 21% of our survey respondents, the 25th to 75th percentile moments of G_i range from 0.64 to 0.25. In contrast, the corresponding moments predicted by the learning assumption range from 0.46 to 0.34. The complete information assumption, on the other hand, predicts a much wider range of 0.92 to -0.41, which deviates significantly from the survey predictions. We find a very consistent pattern in all demographic groups that the learning assumption predicts much less heterogeneity, while complete information overshoots the degree of heterogeneity.

We additionally demonstrate that the belief changes ratio exhibit correlations with demographic characteristics and job transition experience. There are distinct life cycle patterns of heterogeneity in G_i. Fig 10(a) shows the distribution of G_i across three age groups: 20-34, 35-49, and 50-60. Notably, younger workers' learning is much more heterogenous in the younger age group (standard deviation: 0.54), while the learning of the older age group is much less heterogenous (standard deviation: 0.30). We also show that the level of G_i is correlated with the job transition which shows that workers have a higher level of belief change ratio after the job change. Using monthly employer-employee registry data, we identify workers who were separated from their employers during 2020 (transition group) and those who remained with the same employer throughout the year (stay group). We found that 79% of the workers belong to the stay group, while 21% were classified as the transition group. Fig 10(b) shows that workers who experienced job transitions have a higher level of G_i which means they change the beliefs more from the



Note: This graph shows the distribution of the belief change ratio across different subgroups. The scatter plot shows the median value and the bar represents 25-75 percentile moments. For the survey, we use the mean of $\pm 1\%$ around the moments.

Figure 9: Robustness check: Heterogeneity in belief changes

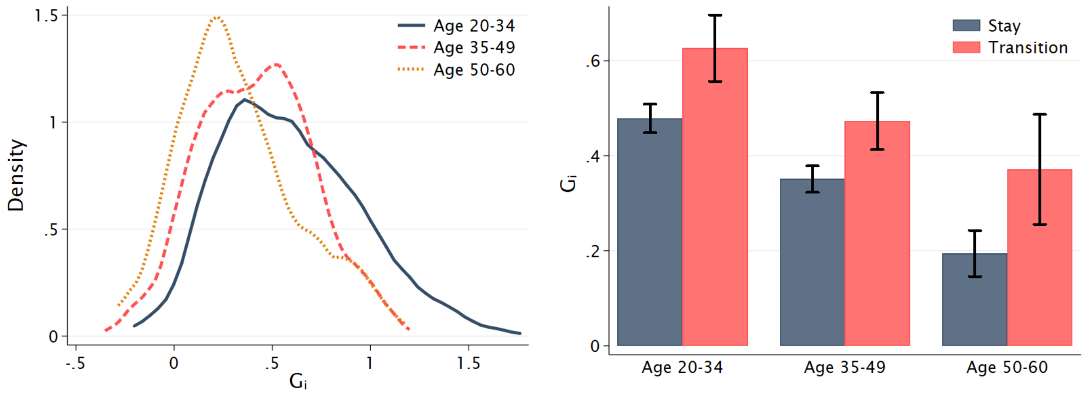
same amount of earnings shock in all age groups. This finding suggests in the case of job transitions workers possibly know more about the nature of earnings shock which is not predicted from the learning assumption.

Result 2. *The learning assumption underestimates the degree of heterogeneity in the belief change ratio, while the complete information assumption overestimates it.*

This section shows that workers' average belief change closely approximates the learning assumption, but there is substantial heterogeneity in empirical belief changes that is difficult to explain within the learning assumption.

4 Modelling partial information

The previous empirical findings show that neither the complete information assumption nor the learning assumption can adequately explain the empirical belief change ratio. To match the empirical belief change ratio, we introduce a new (hybrid of the two assumptions) income process model to explain the empirical findings. We model partial information that accounts for workers' partial knowledge regarding the nature of their earnings shocks. Later in section 5, we connect this income process model to the incomplete market consumption simulations.



(a) Life cycle (b) Job transitions

Note: Panel (a) presents the distribution of survey G_i across three age groups: 20-34, 35-49, and 50-60. Panel (b) displays the average level of survey G_i for two groups: stay and transition. The stay group consists of workers who remained with the same employer throughout the year, while the transition group includes workers who experienced at least one job separation.

Figure 10: Heterogeneity in belief change ratio

4.1 Partial information income process

Each agent works i with age t earns stochastic earnings y_t^i. The earnings realization process follows the heterogeneous income process we introduced in Section 3. The key difference from the previous model is that we decompose the permanent and transitory shocks into two distinct components: known (K) and unknown (Uk) shocks. The known shock represents the part workers can distinguish, while the unknown shock represents the part workers cannot distinguish which is represented below.

$$
\begin{aligned}
y_t^i &= \beta_i t + z_t^i + (\epsilon_{t,\mathrm{K}}^i + \epsilon_{t,\mathrm{Uk}}^i) \\
z_t^i &= \rho z_{t-1}^i + (\eta_{t,\mathrm{K}}^i + \eta_{t,\mathrm{Uk}}^i) \\
\beta_i &\sim \mathcal{N}(0, \sigma_\beta^2) \\
\epsilon_{t,\mathrm{K}}^i &\sim \mathcal{N}(0, \sigma_{\epsilon,\mathrm{K}}^2) \quad \text{and} \quad \epsilon_{t,\mathrm{Uk}}^i \sim \mathcal{N}(0, \sigma_{\epsilon,\mathrm{Uk}}^2) \\
\eta_{t,\mathrm{K}}^i &\sim \mathcal{N}(0, \sigma_{\eta,\mathrm{K}}^2) \quad \text{and} \quad \eta_{t,\mathrm{Uk}}^i \sim \mathcal{N}(0, \sigma_{\eta,\mathrm{Uk}}^2)
\end{aligned} \tag{8}
$$

For instance, instead of ϵ_t^i in the previous equation, we add a layer about differentiability about the nature of the shock, $\epsilon_t^i = \epsilon_{t,\mathrm{K}}^i + \epsilon_{t,\mathrm{Uk}}^i$. This implies that workers possess partial information on the nature of shocks. For the unknown part of the shocks, the problem goes back to the filtering problem where workers update their beliefs based on the remaining unknown parts of the shocks. To isolate the impact of known and unknown shocks only on belief dimensions, we partition the aggregated shock parameter into two distinct components: one representing the known shocks and the other representing the unknown shocks.

$$\sigma^2_\eta = \sigma^2_{\eta,\mathrm{K}} + \sigma^2_{\eta,\mathrm{Uk}} \quad \text{and} \quad \sigma^2_\epsilon = \sigma^2_{\epsilon,\mathrm{K}} + \sigma^2_{\epsilon,\mathrm{Uk}} \tag{9}$$

The above equation preserves the aggregate level of permanent shocks and transitory shocks affecting earnings realizations. The workers then use residual earnings $\tilde{y}^i_t = y^i_t - \epsilon^i_{t,\mathrm{K}} - \eta^i_{t,\mathrm{K}}$ to filtering their beleifs about β_i and z^i_t. We assume that all possible shocks are i.i.d. each other. If we allow the correlation between known and unknown shocks then workers may have more information than the i.i.d. benchmark because they can infer about the size of unknown shocks from the realized known shocks. This framework nests the complete information and learning assumption. For instance, if $\sigma^2_\eta = \sigma^2_{\eta,\mathrm{K}}$ and $\sigma^2_\epsilon = \sigma^2_{\epsilon,\mathrm{K}}$, then this model converges to the complete information assumption because workers will know the realized value of β_i right away in $t = 1$ and no further learning about individual profile. On the other hand, if $\sigma^2_\eta = \sigma^2_{\eta,\mathrm{Uk}}$ and $\sigma^2_\epsilon = \sigma^2_{\epsilon,\mathrm{Uk}}$, this model converges to the learning assumption.

We now describe the state representation of this partial information income process. The underlying state equation is as follows.

$$\underbrace{\begin{bmatrix} \beta_i \\ z^i_{t+1} \end{bmatrix}}_{S^i_{t+1}} = \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & \rho \end{bmatrix}}_{F} \underbrace{\begin{bmatrix} \beta_i \\ z^i_t \end{bmatrix}}_{S^i_t} + \begin{bmatrix} 0 \\ \eta^i_{t+1,\mathrm{Uk}} + \eta^i_{t+1,\mathrm{K}} \end{bmatrix} \tag{10}$$

The log earnings (y^i_t) signal equation is expressed as a linear function of the underlying hidden state and the sum of both known and unknown transitory shocks.

$$y^i_t = \begin{bmatrix} t & 1 \end{bmatrix} \begin{bmatrix} \beta_i \\ z^i_t \end{bmatrix} + (\epsilon^i_{t,\mathrm{K}} + \epsilon^i_{t,\mathrm{Uk}}) = H'_t S^i_t + (\epsilon^i_{t,\mathrm{K}} + \epsilon^i_{t,\mathrm{Uk}}) \tag{11}$$

Unlike the learning assumption, workers also observe the known permanent and transitory shocks. The net log earnings $\tilde{y}^i_t$ are defined by removing the known components from the log earnings y^i_t. In summary, each worker i at age t observes $(\tilde{y}^i_t, \eta^i_{t,\mathrm{K}}, \epsilon^i_{t,\mathrm{K}})$.

The belief updating equation is as follows. The prior belief of each individual about (β^i, z^i_1) is represented by a multivariate normal distribution with mean $(\hat{\beta}_{1|0}, \hat{z}_{1|0})$ and the covariance matrix $P_{1|0} = \begin{bmatrix} \sigma^2_\beta & 0 \\ 0 & \sigma^2_{z,0} \end{bmatrix}$. After observing $(\tilde{y}^i_1, ..., \tilde{y}^i_t)$, $(\eta^i_{1,\mathrm{K}}, ..., \eta^i_{t,\mathrm{K}})$, and $(\epsilon^i_{1,\mathrm{K}}, ..., \epsilon^i_{t,\mathrm{K}})$, the posterior beleif about S^i_t is normally distriubted with mean parameter vector $\hat{S}^i_{t|t-1}$ and the covariance matrix $P_{t|t-1}$. The recursive Kalman updating formulas are given below.

$$\hat{S}^i_{t|t} = \hat{S}^i_{t|t-1} + K_t(\tilde{y}^i_t - H'_t\hat{S}^i_{t|t-1}) + \begin{bmatrix} 0 \\ \rho\eta^i_{t,\mathrm{K}} \end{bmatrix}$$
$$\hat{S}^i_{t+1|t} = F\hat{S}^i_{t|t} \tag{12}$$

where $K_t = P_{t|t-1}H_t[H'_tP_{t|t-1}H_t + \sigma^2_{\epsilon,\mathrm{Uk}}]^{-1}$ is the Kalman gain. We note that $H'_t\hat{S}^i_{t|t-1}$ represents $E_{t-1}[\tilde{y}_t]$ and workers learn in a proportion from their forecast error. The variance-covariance matrix $P_{t|t}$ has the following recursive formula.

$$P_{t|t} = P_{t|t-1} + K_tH'_tP_{t|t-1}$$
$$P_{t+1|t} = FP_{t|t}F' + \begin{bmatrix} 0 & 0 \\ 0 & \sigma^2_{\eta,\mathrm{Uk}} \end{bmatrix} \tag{13}$$

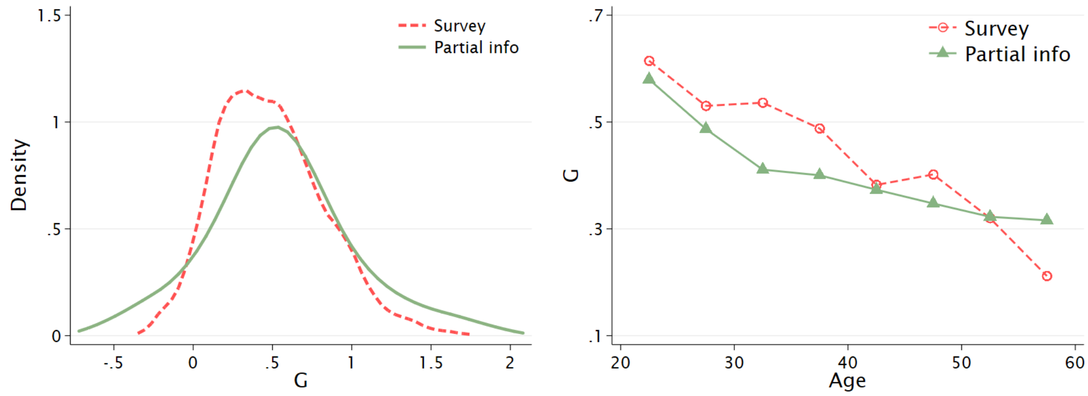
Note that the covariance matrix evolves independently of the signal realizations and remains deterministic in this environment, given that H_t is deterministic. In the end, the one period ahead of earnings expectation conditional on the $\hat{S}_{t|t}$ is distributed as $y^i_{t+1}|\hat{S}^i_{t|t} \sim \mathcal{N}(H'_{t+1}\hat{S}^i_{t+1|t}, H'_{t+1}P_{t+1|t}H_{t+1} + \sigma^2_{\epsilon,\mathrm{Uk}})$.

4.2 Measuring the size of known and unknown shocks

In this subsection, based on the income process parameters specified in Table 3, we decompose the proportion of known shocks using the survey moments while ensuring that the sum of the known and unknown shock parameters remains constant as below.

$$\sigma^2_\eta(= 0.028) = \sigma^2_{\eta,\mathrm{K}} + \sigma^2_{\eta,\mathrm{Uk}}$$
$$\sigma^2_\epsilon(= 0.025) = \sigma^2_{\epsilon,\mathrm{K}} + \sigma^2_{\epsilon,\mathrm{Uk}} \tag{14}$$

We aim to match two specific survey moments in our analysis: the heterogeneity in survey belief changes ratio (G_i) and the life cycle mean of belief changes ratio (G_i). For the life cycle mean, we construct an empirical mean of G_i across 5 years age bin and targeted those as moments. The calibrated parameters for our model are $\sigma^2_{\eta,\mathrm{K}} = 0.016$ and $\sigma^2_{\eta,\mathrm{Uk}} = 0.012$. This indicates that workers are able to differentiate 52% of the permanent shocks, as $\sigma^2_{\eta,\mathrm{K}}/\sigma^2_\eta = 0.52$. The calibrated values for $\sigma^2_{\epsilon,\mathrm{K}}$ and $\sigma^2_{\epsilon,\mathrm{Uk}}$ are 0.011 and 0.014, respectively. This indicates that workers differentiate approximately 43% of the transitory shocks, as $\sigma^2_{\epsilon,\mathrm{K}}/\sigma^2_\epsilon = 0.43$. Fig 11(a) shows the simulated distribution of G alongside the distribution of survey-measured G_i, revealing a strong alignment between



(a) Heterogeneity (b) Life cycle mean of G

Note: Panel (a) shows the empirical distribution of survey G and the distribution of G simulated under partial information. Panel (b) show the comparison of the life cycle pattern of survey G and simulated G. We use 5 years age bins and construct the mean of corresponding G in each age bin.

Figure 11: Targeted moments

the simulated and empirical distributions. Fig 11(b) shows the life cycle mean of G_i and the simulated G also well matches the empirical moments. In the Online Appendix D, we have a detailed discussion about the identification of the targeted moments.

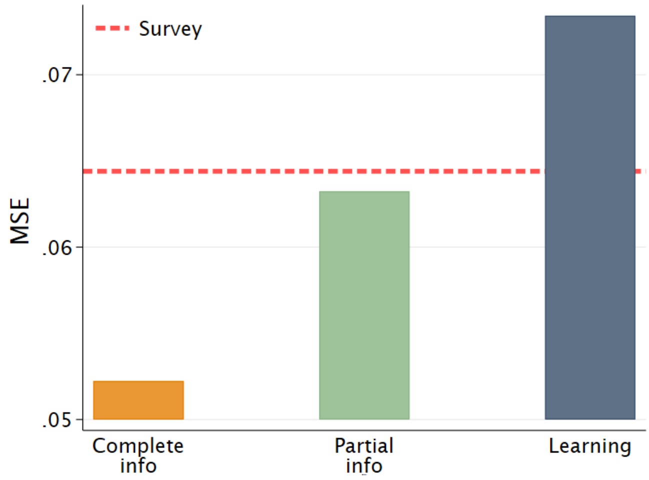
This calibration result confirms that assuming partial information in the belief changes can effectively explain the heterogeneity in the belief change ratio and its life cycle patterns.

4.3 Matching the size of earnings shocks

It is very important to correctly understand the possible size of earnings shocks that workers could face. At the same, the three assumptions have different implications for the size of earnings shocks because the expectation is based on the amount of information they have. On the survey side, we have empirical earnings shocks, so we compare which of the three assumptions most closely matches the empirical size of earnings shocks. To quantify the level of earnings shocks, we introduce the measure mean squared earnings shocks defined below.

$$\text{MSE} = \frac{1}{N}\sum_i (y_t^i - E_{t-1}[y_t^i])^2 \tag{15}$$

Fig 12 shows the empirical MSE from the survey with the counterparts in each assumption. The horizontal bar represents the empirical mean of MSE in the survey. The value of the empirical mean squared earnings shock is 0.064. The bars represent the predic-



Note: This figure shows the level of mean squared earnings shocks in each assumption. The horizontal line is the empirical mean squared earnings shocks and the bars show the prediction of MSE across assumptions.

Figure 12: Mean squared earnings shocks

tion of MSE across assumptions. First, the complete information predicts 0.053 which is 17% lower than the survey MSE. Second, the learning assumption predicts MSE as 0.073 which is 13% higher than the survey MSE. Lastly, the partial information assumption very closely matches the level of MSE. The value of MSE in partial information is 0.062. This result shows that assuming partial information is very important to correctly understand the size of earnings shocks that workers possibly face. Moreover, it is also important to note that assuming complete information could possibly underestimate the level of earnings shocks by 17% which is important to estimate further consumption changes.

In this section, we introduce the partial information partial income process. Our estimates show that workers know around half of the nature of earnings shocks. We further show that partial information closely matches the size of earnings shocks that workers face.

5 Consumption responses and partial information

Different assumptions hold different subsequent implications for consumption. We now investigate the implications of partial information on subsequent consumption responses, both empirically and through the lens of the consumption-saving model. We first empirically estimate the consumption elasticity and the interaction between the belief change ratio. This empirical finding holds particular significance due to the limited evidence connecting survey-measured beliefs and consumption behavior. Based on this empirical finding, we systematically compare the three different belief change assumptions (complete information, learning, and partial information) using an incomplete market model.

5.1 Consumption elasticity and belief change ratio

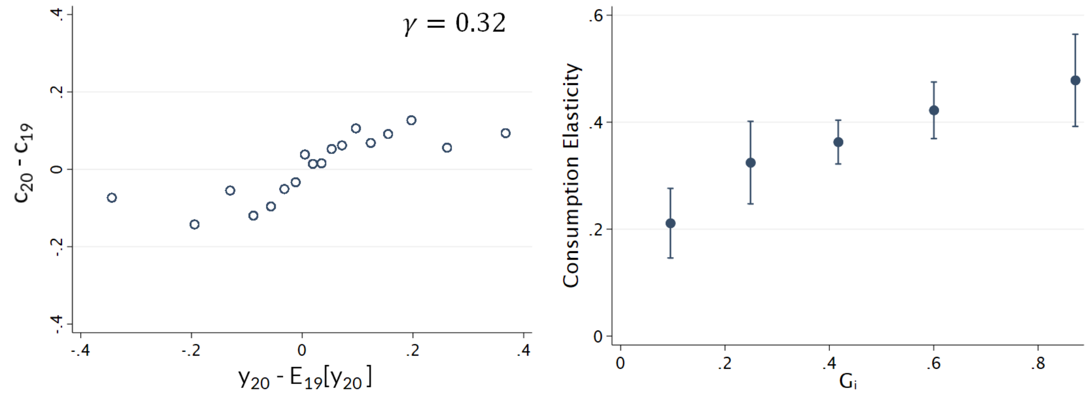
In this subsection, we empirically estimate the elasticity between consumption changes and earnings shocks. We further examine how this relationship varies across different levels of survey belief changes ratio (G_i). We expect if workers perceive earnings shocks as primarily transitory (indicated by G_i being close to 0), the consumption elasticity from earnings shocks would be lower. Conversely, if workers think that the earnings shocks are mainly driven by the permanent shocks (G_i increases), we would expect a higher consumption elasticity.

To estimate the consumption elasticity from earnings shocks (γ), we set a dependent variable as consumption growth in 2020 and unexpected earnings shock as an explanatory variable. For individual characteristics, we again control age, age-squared, gender, and education.

$$\Delta c_{20}^{i} = \gamma(y_{20}^{i} - E_{19}[y_{20}^{i}]) + X_i'\beta + \epsilon \tag{16}$$

Fig 13(a) presents the correlation between consumption growth and earnings shock. We present a binned scatter plot, where the X-axis displays the unexpected earnings shock on a log scale, divided into 20 equally-sized bins. The Y-axis shows the mean of consumption growth, and each dot represents the empirical mean of the respective bin. The regression coefficient for γ is 0.315, which is statistically significant at the 1% level. We put the regression table in Online Appendix A.8. This finding indicates that a 10% positive earnings shock leads to approximately 3.2% increase in consumption.

We now show that the belief change ratio in the survey (G_i) is correlated with the elasticity of consumption (γ). Fig 13(b) shows the estimated elasticity (dot) and the corresponding confidence intervals (bar) of consumption after dividing the sample into 5 equal-size groups depending on their levels of G_i. There is a clear pattern that as G_i increases, γ increases. Specifically, as the belief change ratio increases by 1%, the elasticity of consumption increases by 0.28%. This highlights the significance of the belief change ratio as a crucial mechanism in explaining the heterogeneity in consumption elasticity. It also underscores the importance of accurately modeling the belief change ratio to comprehend the heterogeneous consumption response. We note that the learning assumption, which predicts a limited range of the belief change ratio (G), fails to fully



(a) Pooled (b) Heterogeneity across G_i

Note: Panel (a) shows the pooled sample consumption elasticity from the earnings shocks. Panel (b) shows the separately estimated consumption elasticity across different levels of G_i.

Figure 13: Consumption elasticity and belief change ratio

explain the pattern observed in Fig 13(b). Consequently, in the next section, we develop a model that captures this moderate degree of the belief change ratio and its impact on consumption elasticity.

Result 4. *Empirically, as the belief change ratio increases, the consumption elasticity from unexpected earnings shocks increases.*

5.2 Consumption simulation across belief changes ratio

In this subsection, we set up an incomplete market/life-cycle/partial equilibrium model without aggregate risks to simulate consumption. The model structure is very standard but the key exercise is we simulate the consumption with three different belief change assumptions (complete information, learning, and partial information). We control the sequence of realized earnings but only allow beliefs to be different across three different sets of simulations.

5.2.1 Consumer's problem

We first describe the consumption-savings problem. An individual works for the first T years of her life, and lives up until L ($> T$) years. Preference over consumption follows the CRRA utility function with the parameter ϕ. The constant interest rate of the asset is r and δ denotes a time discount factor. The cash-on-hand (asset and earnings) is represented as x_t^i, and the vector of mean beliefs is as $\hat{S}_t^i = (\hat{\beta}_t^i, \hat{z}_t^i)$. Therefore, the value function is as follows.

$$
\begin{aligned}
V_t^i(x_t^i, \hat{S}_t^i) = \max_{c_t^i, a_{t+1}^i} & \left\{ \frac{c_t^{i\,1-\phi}}{1-\phi} + \delta \mathbb{E}_t [V_{t+1}^i(x^i, \hat{S}_{t+1|t}^i)] \right\} \\
\text{s.t.} \quad & c_t^i + a_{t+1}^i = x_t^i \\
& x_t^i = (1+r)a_t^i + exp(y_t^i) \\
& a_{t+1}^i \geq \underline{a}_t, \text{ and } \hat{S}_t^i \text{ follows belief changes assumption}
\end{aligned} \tag{17}
$$

for $t = 1, ..., T-1$, where V_t^i is the value function of individual with age t. For the income process, For borrowing constraints ($\underline{a}_t$), we allow age-dependent natural borrowing limit a_t which will be explained later.

We use the same stochastic earnings realization and follow the heterogeneous income process we introduced in Section 4. We separately simulate the workers' belief changes in three different ways. It's important to emphasize that different simulations affect solely the belief dimension and not the actual earnings realization. Here we lay the three different assumptions using the known and unknown shocks parameters.

- Complete information assumes $\sigma_\eta^2 = \sigma_{\eta,K}^2$ and $\sigma_\epsilon^2 = \sigma_{\epsilon,K}^2$. This assumption leads to workers having a belief in the true underlying state with no uncertainty.
- Learning assumes $\sigma_\eta^2 = \sigma_{\eta,Uk}^2$ and $\sigma_\epsilon^2 = \sigma_{\epsilon,Uk}^2$. This assumption makes the simulation correspond to Guvenen (2007).
- Partial information assumes both known and unknown terms are positive. This means $\sigma_\eta^2 = \sigma_{\eta,K}^2 + \sigma_{\eta,Uk}^2$ and $\sigma_\epsilon^2 = \sigma_{\epsilon,K}^2 + \sigma_{\epsilon,Uk}^2$. We use the calibrated values in Section 4.2.

During retirement, workers receive a fixed amount of annual social security ss^i which is a function of their retirement income y_T^i with no uncertainty. $g(\cdot)$ is a mapping from y_T^i to ss^i.

$$
\begin{aligned}
V_t^i(x_t^i, ss^i) = \max_{c_t^i, a_{t+1}^i} & \left\{ \frac{c_t^{i\,1-\phi}}{1-\phi} + \delta \mathbb{E} [V_t^i(x^i, ss^i)] \right\} \\
\text{s.t.} \quad & c_t^i + a_{t+1}^i = x_t^i \\
& x_t^i = (1+r)a_t^i + exp(ss^i) \\
& ss^i = g(y_T^i)
\end{aligned} \tag{18}
$$

for $t = T, .., L$ with $V_{L+1} = 0$. We note that we use the exact same structure for complete information and learning assumptions except for the part with the belief changes.

5.2.2 Baseline Parameterization

This section presents the collection of model parameters used in our simulation. An overview of these parameters is in Table 4.

Income process. We use the same set of parameters in the first row of Table 3. These parameters are estimated from the registry. We use the values we calibrated in the survey moments in Section 4 for the partial information.

Life cycle. The model is set at an annual frequency. The working age spans 20 to 65 ($T = 45$) and the agent dies with certainty at the age of 80 ($L = 60$).

Retirement income. For retirement income, we assume that the amount of Social Security ss^i is the following function of y^i_T and the cross-sectional mean of $\bar{y}_T$. We adopt a formulation for the pension system inspired by the salient features of the Social Security system, as described in Guvenen (2007). Let $\tilde{y}^i_T = y^i_T/\bar{y}_T$ as relative earned income in the last working period, where $\bar{y}_T$ represents the cross-sectional mean of earnings at age T. π is the scaling parameter. The formula for the pension system is as follows:

$$ss^i = \pi \begin{cases} 0.9\tilde{y}_T & \text{if} \quad \tilde{y}^i_T < 0.3 \\ 0.27 + 0.32(\tilde{y}_T - 0.3) & \text{if} \quad \tilde{y}^i_T \in [0.3, 2] \\ 0.81 + 0.15(\tilde{y}_T - 2) & \text{if} \quad \tilde{y}^i_T \in [2, 4.1] \\ 1.1 & \text{if} \quad \tilde{y}^i_T > 4.1 \end{cases}$$

Borrowing constraint. For the borrowing constraints, we adopt the natural borrowing limit. The natural borrowing limit ensures full repayment in the final periods, even if the household experiences the lowest possible income realizations for the remaining years. Therefore, our borrowing constraint is defined as follows, where $\min(y_\tau)$ represents the minimum value of cross-sectional realized earnings at age τ.

$$\underline{a}_t = \sum_{\tau=1}^{T-t} \delta^\tau \min(y_\tau)$$

We also have a robustness check for zero-lower bound and it gives almost similar main results in Online Appendix E.

Preference. The coefficient of the relative risk aversion parameter, ϕ, is set to the

Table 4: Baseline parameterization

Block	Parameter	Values	Source
income process	ρ	0.921	Registry estimated
income process	σ^2_η	0.028	Registry estimated
income process	σ^2_ϵ	0.026	Registry estimated
income process	σ^2_β	0.00084	Registry estimated
income process	$\sigma^2_{z_0}$	0.210	Registry estimated
partial info	$\sigma^2_{\eta,K}$	0.016	Matching Survey G_i
partial info	$\sigma^2_{\eta,Uk}$	0.012	Matching Survey G_i
partial info	$\sigma^2_{\epsilon,K}$	0.011	Matching Survey G_i
partial info	$\sigma^2_{\epsilon,Uk}$	0.015	Matching Survey G_i
life cycle	T	45	standard assumption
life cycle	L	60	standard assumption
preference	ϕ	2	standard assumption from literature
preference	δ	0.97	standard assumption from literature
asset return	r	0.03	standard assumption from literature
retirement income	π	0.79	Registry estimated
tax-rate slope	γ_1	0.69	Registry estimated
tax-rate intercept	γ_2	0.043	Registry estimated

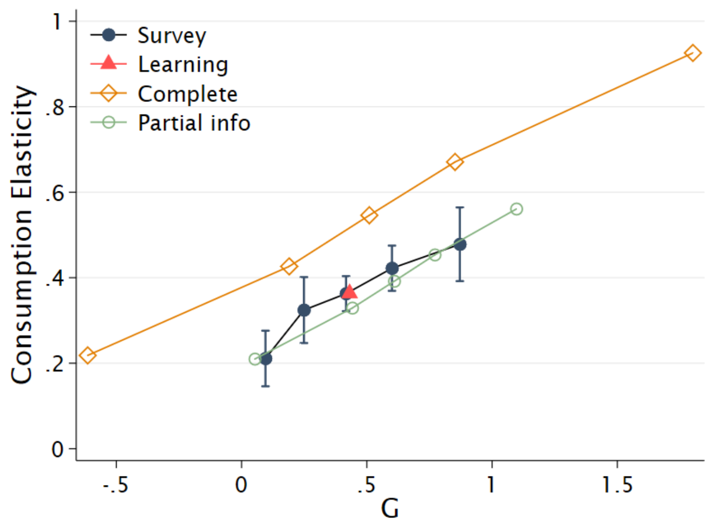
commonly used value of 2.0 in the literature. The discount factor, δ, is set to 0.97, which is also a standard value in the literature. Table 4 summarizes the parameters used in the calibration of the baseline model.

Tax. We incorporate the Danish tax system using the registry. We regress the simple linear tax system by regressing after-tax log earnings on the before-tax log earnings at the population level in 2020.

We simulate 100,000 workers and construct three different simulated annual panel datasets based on each assumption. To investigate the difference only in the belief dimension, we control the earnings realizations as the same across belief changes assumptions. After we simulate the permanent and transitory shocks using σ^2_η and σ^2_ϵ and for the partial information, we simulate the known shocks and then impute the unknown shocks as a remainder.

5.3 Result: Matching empirical consumption elasticity

We empirically showed that the survey G_i is important to explain the consumption elasticity in Section 5.1. In this section, we present the predictions of simulated consumption elasticity across G in each assumption. The blue line represents the degree of consumption elasticity across G_i in the survey, which corresponds to the values shown in Fig 13(b).



Note: On the X-axis, we divide the sample depending on the level of G and plot the estimated consumption elasticity from unexpected earnings shocks on the Y-axis.

Figure 14: Non-targeted moments: consumption elasticity

For the simulated data, we use workers' ages between 20 and 60, consistent with the age range in the survey. In each assumption, we construct G and then we divide the sample into five groups based on different levels of G then we estimate consumption elasticity within each quintile of G.

Fig 14 shows the result of consumption elasticity across three different assumptions. The complete information assumption predicts a much wider range of G compared to the other two assumptions, as we have previously observed. It also overshoots the degree of consumption elasticity because the complete information assumption does not account for uncertainty about β, leading to a lower level of uncertainty compared to the other two assumptions and predicts higher consumption elasticity on average. The learning assumption predicts a very narrow range of G and does not accurately capture the variations in consumption elasticity across G. In contrast, the partial information model closely aligns with the average level of consumption elasticity observed across G_i, highlighting the usefulness of this model in explaining consumption behavior. This result confirms that the partial information we estimated using survey data is valid to explain the actual empirical consumption pattern.

5.4 Policy relevance: Consumption insurance

In this section, we investigate how the partial information model predicts different levels of self-insurance compared to the complete information and learning assumptions. To do

Table 5: Consumption insurance coefficient

Assumptions	(1) Complete information	(2) Learning	(3) Partial information
Permanent shock (η_t^i)	0.23	0.65	0.41
Known ($\eta_{t,\mathrm{K}}^i$)			0.27
Unknown ($\eta_{t,\mathrm{Uk}}^i$)			0.62
Transitory shock (ϵ_t^i)	0.92	0.65	0.79
Known ($\epsilon_{t,\mathrm{K}}^i$)			0.90
Unknown ($\epsilon_{t,\mathrm{Uk}}^i$)			0.62

Note: This table presents the consumption insurance coefficients calculated from the simulated data under three assumptions. The formula for calculating the consumption insurance coefficient is provided in Equation (19).

this, we compute the consumption insurance coefficient following Kaplan and Violante (2010).

$$1 - \frac{cov(\Delta c_t^i, x_t^i)}{var(x_t^i)}\ , \text{ where } x \in \{\epsilon, \eta\} \tag{19}$$

In the above equation, Δc_t^i is the consumption growth of agent i and time t. x_t^i is the simulated value of shocks and x could be permanent shock (η) or transitory shock (ϵ). The variance and covariance are taken cross-sectionally over the entire population in the simulated dataset. The insurance coefficient has an intuitive interpretation; it is the share of the variance of the x shock that does not translate into consumption growth. If the insurance coefficient is close to 1, it indicates that consumption growth is perfectly smoothed from the permanent (or transitory) shocks because the covariance is 0. On the other hand, if the insurance coefficient is close to 0, it suggests that the consumption varies with the shock which means the degree of consumption insurance is lower.

Table 5 shows the insurance coefficients, with each column representing different assumptions. In column (1), which presents the results for complete information, the insurance coefficient for permanent shocks is approximately 0.23, while for transitory shocks it is 0.92. This indicates that workers insure their consumption against transitory shocks to a much greater extent, compared to permanent shocks. These coefficients are consistent with Kaplan and Violante (2010) which also use the complete information assumption. Column (2) shows the coefficients for the learning assumption which is in line with Guvenen (2007). According to this assumption, workers don't distinguish between permanent and transitory shocks, yielding an identical consumption insurance coefficient of 0.65.

This implies that workers self-insure uniformly regardless of shock type, due to their lack of visibility into the shocks' nature. This result shows very different implications for consumption insurance compared to that of complete information. Lastly, in column (3), we show the result for the partial information assumption. For comparison with other benchmarks, we estimate the coefficient after adding up the known and unknown shocks (for instance, $\eta_t^i = \eta_{t,\mathrm{K}}^i + \eta_{t,\mathrm{Uk}}^i$). We then compute again the coefficient for the permanent shock as 0.41 and the transitory shock as 0.79, which indicates a moderate difference between the permanent and transitory shocks. If we decompose these shocks into known and unknown shocks, our estimate for known permanent shock is 0.27 and known transitory shock is 0.90. These are very consistent with the estimates in the complete information assumption in column (1). For both unknown shocks, the coefficient is 0.62 which is in line with column (2).

In summary, the complete information assumption predicts the largest difference in consumption insurance between permanent and transitory shocks, while the learning assumption predicts no difference. The partial information assumption predicts a moderate degree of difference in the degree of consumption insurance. Many empirical studies have highlighted that the level of consumption insurance against transitory shocks is often lower (estimates around 0.6-0.8) than what the complete information assumption predicts (for instance, Parker et al., 2013 for tax rebates and Fagereng et al., 2021 for lottery winnings). Considering these results, the consumption insurance parameters in the partial information are more aligned with empirical findings, offering a possible explanation.

6 Conclusion

We investigate the empirical relationship between unexpected earnings shocks, belief changes, and consumption decisions using a uniquely collected survey and matched registry data in Denmark. A key feature of our survey is that we measure the long-term expectations in the panel so we can track the changes in long-term expectations workers have. We introduce the "belief change ratio" to quantify the degree to which workers internalize earnings shocks into their long-term expectations. We further build a testable hypothesis around the belief change ratio considering two assumptions: complete information and the learning assumption. We show that the learning assumption closely fits the average belief change ratio over the life cycle. This underscores the significance of incorporating incomplete information into the literature on income processes. We further show that the learning assumption is somewhat limited in explaining the heterogeneity in the belief change ratio. This suggests a need for a hybrid partial information assumption where workers have partial information about the nature of the earnings shocks and estimate they only know half of the nature of earnings shocks.

We further connect this belief change ratio to explain the heterogeneous degree of consumption responses from the earnings shocks. Empirically, we show evidence that the belief change ratio is correlated with consumption elasticity from the earnings shocks. Through the lens of the model, we show simulation evidence that the partial information closely matches the actual consumption elasticity in empirical data.

More broadly, our findings highlight the value of using survey-based measures of earnings expectations to understand heterogeneity in important decision making including consumption and saving behavior. We richly use the earnings expectations with the administrative data jointly to directly identify the model objects. An interesting avenue for future research agenda is about understanding the labor market search expectations (Caplin et al., 2023) and connecting it to the precautionary savings and labor market search behavior.

References

Abildgren, K., A. Kuchler, A. S. L. Rasmussen, and H. S. Sørensen (2018). Consistency between household-level consumption data from registers and surveys. Technical report, Danmarks Nationalbank Working Papers.

Agarwal, S. and W. Qian (2014). Consumption and debt response to unanticipated income shocks: Evidence from a natural experiment in singapore. *American Economic Review 104*(12), 4205–4230.

Almås, I., O. Attanasio, and P. Jervis (2023). Economics and measurement: New measures to model decision making. Technical report, National Bureau of Economic Research.

Baker, S. R., L. Kueng, S. Meyer, and M. Pagel (2018). Measurement error in imputed consumption. Technical report, National Bureau of Economic Research.

Blundell, R., L. Pistaferri, and I. Preston (2008). Consumption inequality and partial insurance. *American Economic Review 98*(5), 1887–1921.

Blundell, R. and I. Preston (1998). Consumption inequality and income uncertainty. *The Quarterly Journal of Economics 113*(2), 603–640.

Browning, M. and S. Leth-Petersen (2003). Imputing consumption from income and wealth information. *The Economic Journal 113*(488), F282–F301.

Bunn, P., J. Le Roux, K. Reinold, and P. Surico (2018). The consumption response to positive and negative income shocks. *Journal of Monetary Economics 96*, 1–15.

Caplin, A., V. Gregory, E. Lee, S. Leth-Petersen, and J. Sæverud (2023). Subjective earnings risk. Technical report, National Bureau of Economic Research.

Commault, J. (2021). Does consumption respond to transitory shocks? *American Economic Journal: Macroeconomics, forthcoming.*

Commault, J. (2022). How do persistent earnings affect the response of consumption to transitory shocks?

Crawley, E. and A. Kuchler (2023). Consumption heterogeneity: Micro drivers and macro implications. *American Economic Journal: Macroeconomics 15*(1), 314–41.

Cunha, F., J. Heckman, and S. Navarro (2005). Separating uncertainty from heterogeneity in life cycle earnings. *oxford Economic papers 57*(2), 191–261.

Delavande, A. and S. Rohwedder (2008). Eliciting subjective probabilities in internet surveys. *Public Opinion Quarterly 72*(5), 866–891.

Dominitz, J. and C. F. Manski (1997). Using expectations data to study subjective income expectations. *Journal of the American statistical Association 92*(439), 855–867.

Druedahl, J. and T. H. Jørgensen (2020). Can consumers distinguish persistent from transitory income shocks? *The Economic Journal 130*(632), 2410–2437.

Fagereng, A. and E. Halvorsen (2017). Imputing consumption from norwegian income and wealth registry data. *Journal of Economic and Social Measurement 42*(1), 67–100.

Fagereng, A., M. B. Holm, and G. J. Natvik (2021). Mpc heterogeneity and household balance sheets. *American Economic Journal: Macroeconomics 13*(4), 1–54.

Fuster, A., G. Kaplan, and B. Zafar (2021). What would you do with 500?*spendingresponsestogains*, *losses*, *news*, *andloans*. *TheReviewofEconomicStudies* 88(4), 1760– –1795.

Goldstein, D. G. and D. Rothschild (2014). Lay understanding of probability distributions. *Judgment & Decision Making 9*(1).

Guvenen, F. (2007). Learning your earning: Are labor income shocks really very persistent? *American Economic Review 97*(3), 687–712.

Guvenen, F. (2009). An empirical investigation of labor income processes. *Review of Economic dynamics 12*(1), 58–79.

Guvenen, F. and A. A. Smith (2014). Inferring labor income risk and partial insurance from economic choices. *Econometrica 82*(6), 2085–2129.

Hause, J. C. (1980). The fine structure of earnings and the on-the-job training hypothesis. *Econometrica: Journal of the Econometric Society*, 1013–1029.

Havranek, T. and A. Sokolova (2020). Do consumers really follow a rule of thumb? three thousand estimates from 144 studies say "probably not". *Review of Economic Dynamics 35*, 97–122.

Kaplan, G. and G. L. Violante (2010). How much consumption insurance beyond self-insurance? *American Economic Journal: Macroeconomics 2*(4), 53–87.

Karahan, F. and S. Ozkan (2013). On the persistence of income shocks over the life cycle: Evidence, theory, and implications. *Review of Economic Dynamics 16*(3), 452–476.

Kaufmann, K. and L. Pistaferri (2009). Disentangling insurance and information in intertemporal consumption choices. *American Economic Review 99*(2), 387–392.

Koşar, G. and W. Van der Klaauw (2023). Workers' perceptions of earnings growth and employment risk. *FRB of New York Staff Report* (1056).

Lewis, D. J., D. Melcangi, and L. Pilossoph (2019). Latent heterogeneity in the marginal propensity to consume. *FRB of New York Staff Report* (902).

Lillard, L. A. and R. J. Willis (1978). Dynamic aspects of earning mobility. *Econometrica 46*(5), 985.

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.

Manski, C. F. (2004). Measuring expectations. *Econometrica 72*(5), 1329–1376.

Meghir, C. and L. Pistaferri (2011). Earnings, consumption and life cycle choices. In *Handbook of labor economics*, Volume 4, pp. 773–854. Elsevier.

Misra, K. and P. Surico (2014). Consumption, income changes, and heterogeneity: Evidence from two fiscal stimulus programs. *American Economic Journal: Macroeconomics 6*(4), 84–106.

Parker, J. A., N. S. Souleles, D. S. Johnson, and R. McClelland (2013). Consumer spending and the economic stimulus payments of 2008. *American Economic Review 103*(6), 2530–2553.

Pischke, J.-S. (1995). Individual income, incomplete information, and aggregate consumption. *Econometrica: Journal of the Econometric Society*, 805–840.

Pistaferri, L. (2001). Superior information, income shocks, and the permanent income hypothesis. *Review of Economics and Statistics 83*(3), 465–476.

Stoltenberg, C. A. and A. Uhlendorff (2022). Consumption choices and earnings expectations: empirical evidence and structural estimation.

Wang, T. (2023). Perceived versus calibrated income risks in heterogeneous-agent consumption models. Technical report, Working paper.

Online Appendix for "Earnings Shocks, Expectations, and Spending"

Eungik Lee
New York University

Johan Sæverud
University of Copenhagen

October 30, 2023

A Additional results

A.1 Survey weights construction

In the analysis, we scale all statistics by the relative population weights. To construct population weights, we the Danish population observed in 2019 in the administrative data. We estimate a probit regression with a survey participation dummy as the dependent variable and age, log earnings, education, and gender as explanatory variables. All these characteristics are available in the administrative data. Table A.1 shows the marginal effect on survey participation using probit regression. Our survey respondents are around 0.37% of the Danish population. We find that the selection of the survey is related to various demographics. For instance, as age increases by one unit, the probability of participation increases by 0.012%.

Table A.1: Marginal effect on participation

	Mean of Pr(participation):0.36%		
	$\frac{dy}{dx} \times 100$	z-statistics	p-value
age	0.013	41.22	<0.001
female	-0.029	-6.45	<0.001
log earnings	0.011	19.38	<0.001
above university	0.225	24.9	<0.001
	N: 2,634,812		
	Log-likelihood: -89,435		

Note: The table presents marginal effects from probit regressions where the dependent variable is a dummy variable for survey participation.

To obtain population weights for the survey, we use the inverse of the predicted probability of participating in the survey. Then we apply these population weights to the analysis and figures in the main text.

A.2 Repeaters vs. Non-repeaters

In this section, we compare the repeaters and non-repeaters samples. There are 7,734 respondents who only participated in CLP 1. In Table A.2, we show the demographics for the non-repeaters sample. It is well balanced across gender, lifecycle, earnings, and consumption level.

Table A.2: Demographics for non-repeaters

	Repaters	Non-repeaters
N	5,867	7,734
Female	0.49	0.5
Age		
20-29	0.19	0.22
30-39	0.21	0.24
40-49	0.25	0.23
50-60	0.35	0.31
Education		
Above college	0.48	0.50
Annual earnings (USD)		
Mean	63,678	65,271
S.D.	38,873	33,227
Disposable earnings (USD)		
Mean	41,933	43,878
S.D.	32,616	35,639
Consumption (USD)		
Mean	37,838	38,013
S.D.	25,478	29,046
Liquid constrained	0.32	0.31

Note: The table presents a comparison of the average demographic characteristics of the survey sample and the Danish population as observed in the administrative data in 2019. The column "Survey Sample" comprises repeaters and employed respondents from both Copenhagen Life Panel 1 and 2. The “Registry” column comprises all individuals aged 20-60 within the Danish population who earned a minimum of 24,000 DKK in 2019. We also dropped workers who have a self-employed income of more than 24,000 DKK in 2019. The liquid wealth is the value at Dec-31st of 2018. In Jan-2020, the exchange rate for 1 US Dollar was approximately 7 Danish Krone (DKK).

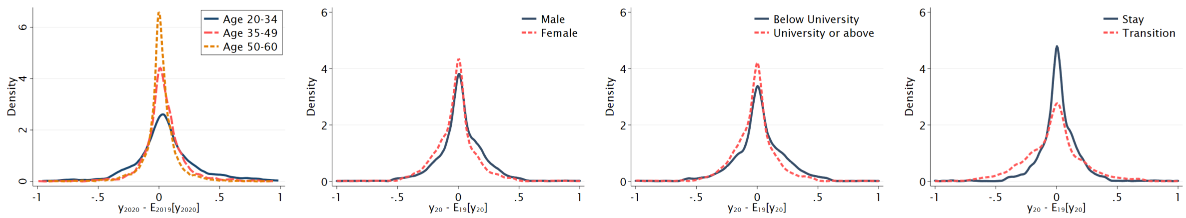
A.3 Discussion of consumption imputation

We discuss the possible limitations of this imputation method and how we address the issues. While the disposable income used for expenditure imputation includes all labor income and capital income; it does exclude capital gains. Changes in capital gains and losses, such as fluctuations in housing and stock prices, could affect consumption. However, these factors are not considered within the imputation method.

To address the issue of capital gains, we address it through the following empirical strategy. First, we exclude housing wealth and treat housing as an off-balance sheet asset, considering its significance as a major asset held by households in Denmark. Second, we exclude self-employed workers from our survey sample, which accounts for approximately 4% of the total sample. Self-employed workers often face challenges in distinguishing between consumption and their business investment, which could introduce large measurement errors (Crawley and Kuchler, 2023). Lastly, we note that the ownership of stocks among Danish households is relatively low, with only around 10% of households holding any stocks in 2020. Furthermore, for many of these households, stocks represent only a small proportion of their overall wealth.

A.4 Distributions of earnings shocks across demographics and labor market experience

In this subsection, we plot the earnings shocks ($y_t^i - E_{t-1}[y_t^i]$) across different demographics. In Fig A.1, we show the distribution of earnings shocks after we divide the sample into four different demographic variables. Fig A.1a shows the distribution of earnings shocks across three different age groups. There is a clear life cycle pattern that older workers experience lower levels of earnings shocks compared to the younger workers which possibly comes from the different labor market dynamics. Across gender and education, it is hard to find the difference between groups. In Fig A.1d, we divide the sample into two groups: stay group includes workers who stayed with the same employer for the entire year 2020 (81% of the sample), and another group that experienced at least one job transition (19% of the sample) from their employer during the year. We find that the earnings shocksis more heterogenous for the job transition group than the stay group. Our findings suggest that labor market transitions are the primary source of forecast errors for workers, as we observe a correlation between forecast errors and these transitions.



(a) Life cycle (b) Gender (c) Education (d) Job transitions

Note: Panel (a) shows the distribution of earnings shocks across three different age groups. Panel (b) and Panel (c) show the distribution of earnings shocks across gender and education respectively. Lastly, Panel (d) shows the distribution of earnings shocks between stay and transition groups. The stay group represents workers who don't experience a job separation during the year. The transition group is a group where workers experience a job transition.

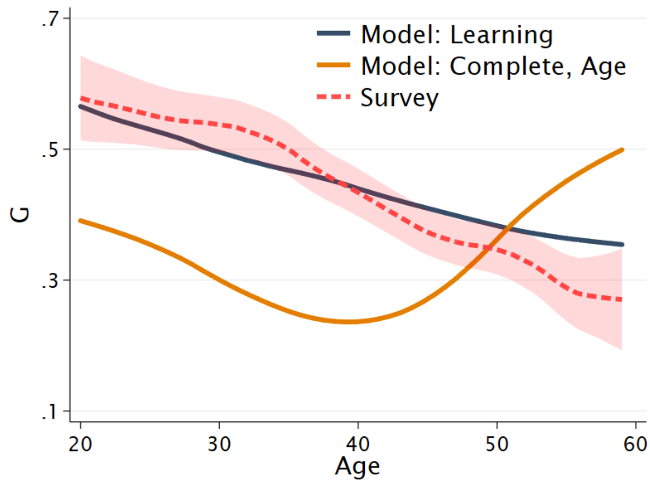
Figure A.1: Earnings shocks across demographics

A.5 Age-varying parameters and average belief change ratio

In this section, we conduct a robustness check by permitting income process parameters that vary with age. Initially, we estimate the income process using the approach by Karahan and Ozkan (2013). The idea is to impose a polynomial structure to the parameters and estimate the coefficients. The details of the estimation process are in Section B.1 and the summary of estimated parameters is as follows.

$$\begin{aligned}\rho =&0.813 + 0.009\text{age} - 0.00015\text{age}^2 \\ \sigma^2_\eta =&0.044 - 0.003\text{age} + 0.00009\text{age}^2 \\ \sigma^2_\epsilon =&0.039 + 0.004\text{age} - 0.00010\text{age}^2 \\ \sigma^2_\beta =&0.0051\end{aligned} \tag{1}$$

We simulate 100,000 workers' earnings realizations and beliefs based on the above parameters. Fig A.3 shows the results. It is very clear that the complete information predicts a U-shaped G mainly because of U-shaped permanent shocks and inverse U-shaped transitory shocks as age increases.



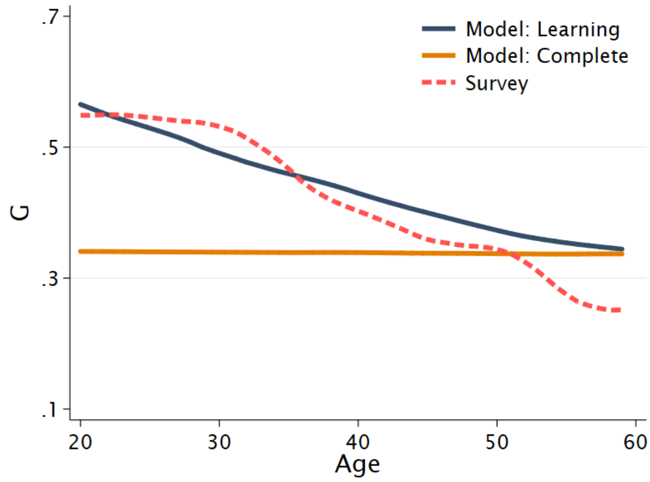
Note: The blue and orange lines show the local linear predictions of the simulated belief change ratio across the life cycle. The red dotted line is the survey belief change ratio with 95% confidence intervals. The lines are a local regression fitted line. We excluded the top and bottom 2.5% of outliers from the analysis in each dataset.

Figure A.2: Robustness check: Age-varying average belief change ratio

This result confirms that even after allowing different parameters across ages, it is hard to predict the monotonic decreasing patterns across ages.

A.6 Median belief change ratio

In this section, we conduct a robustness check of Fig 6 in the main text using median instead of average. For the survey G, we compute the median for each age group by averaging the values within a range of $\pm 2\%$ around the median. We find a very consistent pattern that the median is decreasing across ages in the survey G which is consistent with the learning assumption, while complete information predicts the flat line. We note that the learning assumption doesn't have a meaningful median because G is constant given age level.



Note: This graph shows the local linear smoothed line of the belief change ratio. We first construct the median (the mean of $\pm 2\%$ around the median in the survey) and plot the local regressions.

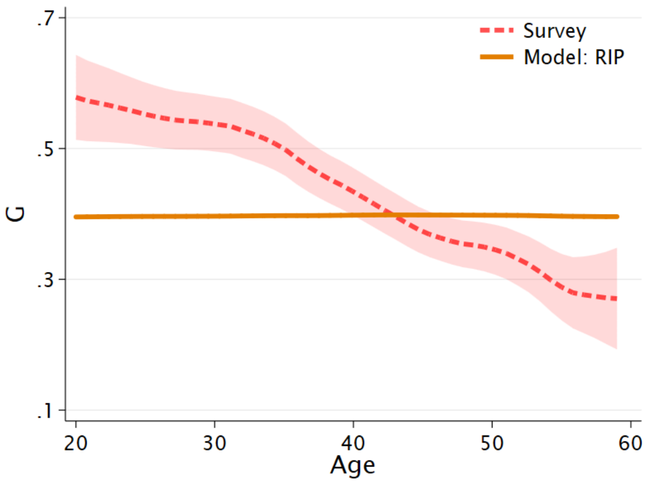
Figure A.3: Robustness check: Median belief change ratio

A.7 Discussion on restricted income process

In this subsection, we show that our main result is robust to the restricted income process, a widely used income process specification without individual heterogeneity.

$$\begin{aligned} y_t^i &= \underbrace{g(X_t^i)}_{\text{observable}} + \underbrace{(z_t^i + \epsilon_t^i)}_{\text{stochastic}} \\ z_t^i &= \rho z_t^i + \eta_t^i \end{aligned} \tag{2}$$

We estimated the restricted income process with the same minimum distance process. The estimated ρ, σ_η^2, and σ_ϵ^2 are 0.962, 0.037, and 0.041. We simulate the belief change ratio based on these estimated values. Fig A.4 shows the result. We confirm that the restricted income process still predicts the flat level of G far from the survey predictions.



Note: This graph shows the local linear smoothed line of the belief change ratio. The orange line is the prediction of the restricted income process and the red line is the prediction of the survey.

Figure A.4: Robustness check: Restricted income process

A.8 Regression for consumption elasticity

In this section, we show the regression results for the consumption elasticity. The regression specification is as follows. For the dependent variable we use consumption growth. For the control variable, we use age, age-sqaured, gender, and education.

$$\Delta c_{20}^{i} = \gamma(y_{20}^{i} - E_{19}[y_{20}^{i}]) + X_i'\beta + \epsilon \tag{3}$$

Table A.3: Consumption elasiticity

	(1)	(2)
Dependent variable: Δc_{20}^{i}		
$y_{20}^{i} - E_{19}[y_{20}^{i}]$	0.284^{***}	0.307^{***}
	(0.036)	(0.033)
Age		-0.002^{*}
		(0.001)
Age sqaured		0.00004
		(0.00423)
Female		-0.031^{**}
		(0.018)
University or above		-0.008^{***}
		(0.002)
Controls	N	Y
R-sq	0.271	0.314
N	5,867	5,867

Note: Standard errors are clustered at the individual level. The dependent variable is log scaled consumption growth. Control variables include age, age-squared, gender, and education.

B Estimation details

In this section, we describe a detailed process of income process estimations. Guvenen (2009) describe the earnings process as follows.

$$y_t^i = \underbrace{g(X_t^i)}_{\text{observable}} + \underbrace{\beta_i t}_{\text{individual}} + \underbrace{(z_t^i + \epsilon_t^i)}_{\text{stochastic}} \tag{4}$$

Step 1 residual earnings: The left-hand side of the equation represents worker i's log-transformed earnings at age t. On the right-hand side, the first function g captures the common variation in earnings across all individuals, especially from observable demographics. We run the following regression to capture the aggregated patterns of earnings over the life cycle:

$$y_t^i = \beta_0 + \beta_1 t_i + \beta_2 t_i^2 + \epsilon_t^i \tag{5}$$

The estimated value of $(\beta_0, \beta_1, \beta_2)$ is $(9.9438, 0.0537, -0.00087)$ and all the coefficients are significant at 1% level. We then use these estimates to calculate the residual earnings $\hat{y}_i^t$.

Step 2 covariance matrix: To estimate the four parameters ρ (persistence), σ_β^2 (variance for β_i), σ_η^2 (variance for permanent shock), and σ_ϵ^2 (variance for transitory shock), we write down the variance and We use the minimum distance method proposed by Guvenen (2009) to estimate the parameters. First, we write the model variance and covariance matrix of the income residual $\hat{y}_{i,t}$ as a function of the four parameters, across age.

$$\begin{aligned}
\mathrm{var}(\hat{y}_{i,t}) &= \sigma_\beta^2 t^2 + \mathrm{var}(z_{i,t}) + \sigma_\epsilon^2 \\
\mathrm{cov}(\hat{y}_{i,t}, \hat{y}_{i,t+n}) &= \sigma_\beta^2 t(t+n) + \rho^n \mathrm{var}(z_{i,t}) \\
\mathrm{var}(z_{i,t}) &= \rho^2 \mathrm{var}(z_{i,t-1}) + \sigma_\eta^2 \\
\mathrm{var}(z_{i,1}) &= \sigma_\eta^2
\end{aligned} \tag{6}$$

Based on the above representation, we can write down a 40 by 40 variance-covariance matrix from age 20-60 as a function of four parameters. We also obtain income residuals' empirical variance and covariance matrix across age. We use the earnings from 2015-2020 in the registry.

Step 3 minimum distance estimator: Our estimation strategy is based on minimizing the distance between the elements of these 40 by 40 empirical covariance matrix and the counterpart implied by the model. Let c_n be an element of the empirical covariance matrix of income residual where $n = 1, ..., N(= T(T+1)/2)$. Let d_n represent the corresponding model covariance using four parameters b. Define $F_n(b, \gamma_{in}) = [F_1(b, \gamma_{i,1}), ..., F_N(b, \gamma_{i,N})]$, where γ_i the vector of indicator functions where individual i belongs to the certain moments.

The minimum distance estimator b is the solution to the below equations. For the weighting matrix, we use the identity matrix.

$$\min_b [I^{-1} \sum_{i=1}^{I} F_n(b, \gamma_i)]' A_N [I^{-1} \sum_{i=1}^{I} F_n(b, \gamma_i)]$$

For the asymptotic covariance matrix, $\Sigma = (D'D)^{-1} D' \Omega D (D'D)^{-1}$, where D is Jacobian moments , $E[\partial F(b, \gamma_{in})/\partial b']$ and Ω is covariance matrix $E[F(b, \gamma_i) F(b, \gamma_i)']$.

B.1 Age-varying parameters

We further estimate following the method of Karahan and Ozkan (2013). We allow the three parameters $(\rho, \sigma^2_\eta, \sigma^2_\epsilon)$ are age-dependent $(\rho_t, \sigma^2_{\eta,t}, \sigma^2_{\epsilon,t})$.

$$\begin{aligned} \mathrm{var}(\hat{y}_{i,t}) &= \sigma^2_\beta t^2 + \mathrm{var}(z_{i,t}) + \sigma^2_{\epsilon,t} \\ \mathrm{cov}(\hat{y}_{i,t}, \hat{y}_{i,t+n}) &= \sigma^2_\beta t(t+n) + \rho_t \ldots \rho_{t+n} \mathrm{var}(z_{i,t}) \\ \mathrm{var}(z_{i,t}) &= \rho^2_t \mathrm{var}(z_{i,t-1}) + \sigma^2_{\eta,t} \\ \mathrm{var}(z_{i,1}) &= \sigma^2_{\eta,1} \end{aligned} \tag{7}$$

We further impose a polynomial as follows.

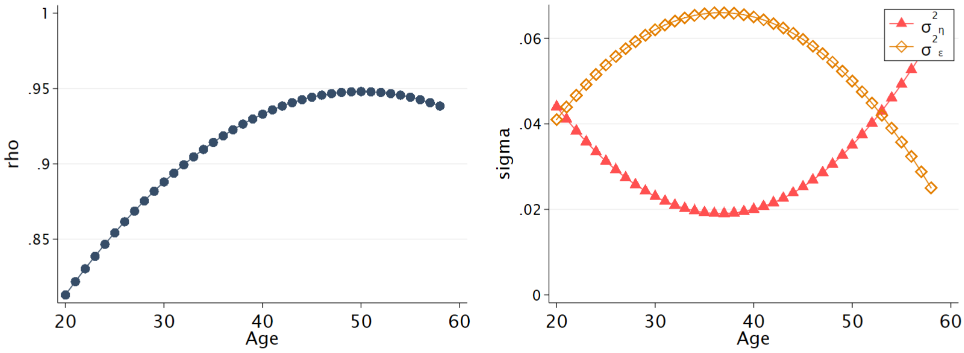
$$\gamma_t = \gamma_0 + \gamma_1 t + \gamma_2 t^2 \quad \text{where } \gamma \in \{\rho, \sigma^2_\epsilon, \sigma^2_\eta\} \tag{8}$$

The details of the estimation are the same as above.

Table B.1: Age-varying: estimated parameters

	γ_0	γ_1	γ_2
ρ	0.813	0.009	-0.00015
	(0.0004)	(0.0002)	(0.000082)
σ^2_ϵ	0.044	-0.003	0.00009
	(0.0001)	(0.0001)	(0.0000032)
σ^2_η	0.039	0.004	-0.00010
	(0.0008)	(0.00004)	(0.000027)

We further illustrate the parameters graphically. Figure B.1a shows the changes in the ρ parameter over the life cycle. It is inverse U-shaped and has a maximum value around age 45. Fig B.1b shows the pattern of σ_ϵ^2 and σ_η^2 across age levels. It is clear that σ_ϵ^2 is inverse U-shaped and σ_η^2 is U-shaped.



(a) ρ (b) σ_ϵ^2 and σ_η^2

Note: Panels (a) and (b) show the estimated parameters for the income process after imposing a polynomial structure. We show the estimated ρ across the life cycle in Panel (a). Panel (b) shows the estimated σ_η^2 and σ_ϵ^2 across the ages.

Figure B.1: Age-varying parameters

C Details of Guvenen (2007)

In this section, we describe the details of Guvenen (2007). It is essentially identical to the partial information with the unknown shocks part only. The general process of log earnings, y_t^i, of individual i who is t years old is given as follows.

$$\begin{aligned} y_t^i &= g(X_t^i, \theta^0) + f(X_t^i, \theta^i) + z_t^i + \epsilon_t^i \\ z_t^i &= \rho z_{t-1}^i + \eta_t^i, \qquad z_0^i = 0 \end{aligned} \tag{9}$$

where the function g and f denote two separate life-cycle components of earnings. The first function g captures the part common to all individuals and is imposed as polynomial as age (t) and age quadratic (t^2) part. The estimated parameters represent θ^0 which are common to all individuals. The second function f is the individual-specific component of life cycle earnings. We note that Guvenen (2007) first propose $f(X_t^i, \theta^i) = \alpha_i + \beta_i t$. α_i represents individual intercept parameter and β_i represents individual earnings growth (slope) paramter. The heterogeneous income process literature later evolves to focus on the earnings growth parameter β_i (for instance, Guvenen and Smith, 2014). We drop α^i from the description but the main result is robust regardless of learning about α_i. The individual specific parameter β_i is realized at the beginning of the labor market from the normal distribution $N(0, \sigma_\beta^2)$. Lastly, $z_t^i = \rho z_{t-1}^i + \eta_t^i$ and $\eta_t^i \sim N(0, \sigma_\eta^2)$.

The state equation representation is as follows.

$$\underbrace{\begin{bmatrix} \beta_i \\ z_{t+1}^i \end{bmatrix}}_{S_{t+1}^i} = \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & \rho \end{bmatrix}}_{F} \underbrace{\begin{bmatrix} \beta_i \\ z_t^i \end{bmatrix}}_{S_t^i} + \underbrace{\begin{bmatrix} 0 \\ \eta_{t+1}^i \end{bmatrix}}_{v_{t+1}^i} \tag{10}$$

Although the parameters of the income profile are not dynamic, including them in the state vector yields recursive updating formulas for beliefs using the Kalman filter. A second (observation) equation expresses the observable variables in the model—in this case, log income—as a linear function of the underlying hidden state and a transitory shock.

$$y_t^i = \begin{bmatrix} t & 1 \end{bmatrix} \begin{bmatrix} \beta_i \\ z_t^i \end{bmatrix} + \epsilon_t^i = H_t' S_t^i + \epsilon_t^i \tag{11}$$

The shocks are i.i.d. Normal distributions. To capture an individual's initial prior about (β_i, z_t^i), they assume normal prior with mean $\hat{S}_{1|0} = (0, 0)$ and variance-covariance matrix

$$\mathbf{P}_{1|0} = \begin{bmatrix} \sigma_\beta^2 & 0 \\ 0 & \sigma_\eta^2 \end{bmatrix} \tag{12}$$

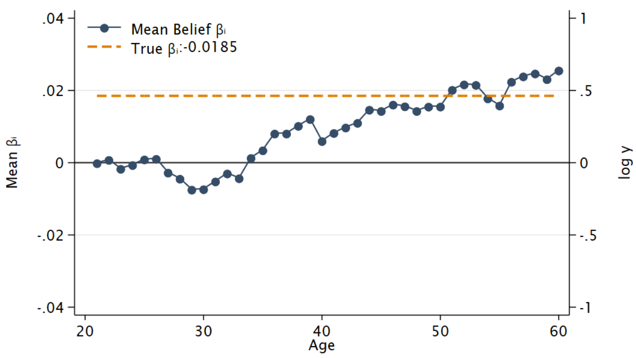
The evolution of the mean after observing signal y_t^i is as follows.

$$
\begin{aligned}
\hat{S}^i_{t|t} =& \hat{S}^i_{t|t-1} + K_t(y_t^i - H_t'\hat{S}^i_{t|t-1}) \\
\hat{S}^i_{t+1|t} =& F\hat{S}^i_{t|t}
\end{aligned} \tag{13}
$$

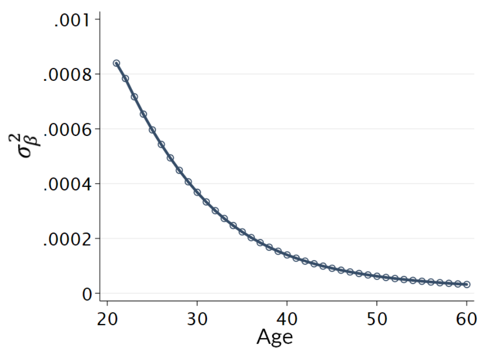
where $K_t = P_{t|t-1}H_t[H_t'P_{t|t-1}H_t + R]^{-1}$. The evolution of the variance and covariance matrix is as follows.

$$
\begin{aligned}
\mathbf{P}_{t|t} &= \mathbf{P}_{t|t-1} + K_t H_t' \mathbf{P}_{t|t-1} \\
\mathbf{P}_{t+1|t} &= F\mathbf{P}_{t|t}F' + \begin{bmatrix} 0 & 0 \\ 0 & \sigma_\eta^2 \end{bmatrix}
\end{aligned} \tag{14}
$$

Fig C.1a is an example of belief updating about individual-specific parameter β_i. The realized value for β_i is 0.0185, but the mean beliefs about β_i is 0. It is notable that the gap between the mean beliefs and the realized β_i is far when the worker is young. However, as age increases, the gap between the mean beliefs and actual realization decreases. The key mechanism behind this is as workers observe the realizations $(y_1^i, ..., y_t^i)$ their information about the individual parameter is getting more accurate. Likewise Fig C.1b shows the example of a decrease in variance about individual parameters. Again as age increases, it is very clear that the uncertainty about individual specific parameters decreases which means that workers are more accurate about their income process.



(a) Mean



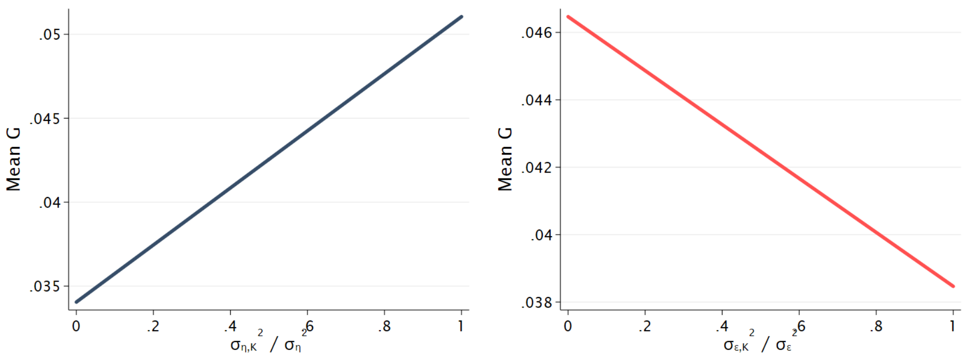
(b) Variance

Note: Panels (a) and (b) show the example of learning in both moments. We show the learning in mean in Panel (a). Panel (b) shows the variance of β_i.

Figure C.1: Example in learning

D Identification of partial information

In this section, we describe how the known and unknown shocks, can match the distribution and life cycle pattern of G. We show how the relative size of known permanent and transitory shocks $\sigma^2_{\eta,K}$ and $\sigma^2_{\epsilon,K}$ affect both the mean and standard deviation of G. The following figure shows the level of simulated mean of G. Figure D.1 shows the results. In Figure D.1a, on the X-axis, we plot the proportion of known shocks ($\frac{\sigma^2_{\eta,K}}{\sigma^2_{\eta}}$) and on the Y-axis, we plot the simulated mean of G fixing the known proportion of transitory shock as 0.5. As the known proportion of permanent shocks increases, the mean increases. In contrast, Fig D.1b shows the results for the proportion of known transitory shocks ($\frac{\sigma^2_{\epsilon,K}}{\sigma^2_{\epsilon}}$). We fixed the proportion of known permanent shock as 0.5. As the known proportion of transitory shocks increases, the mean decreases. Workers increase their G as they have more information about permanent shocks, while G decreases as they have more information about transitory shocks inside of the earnings shocks.



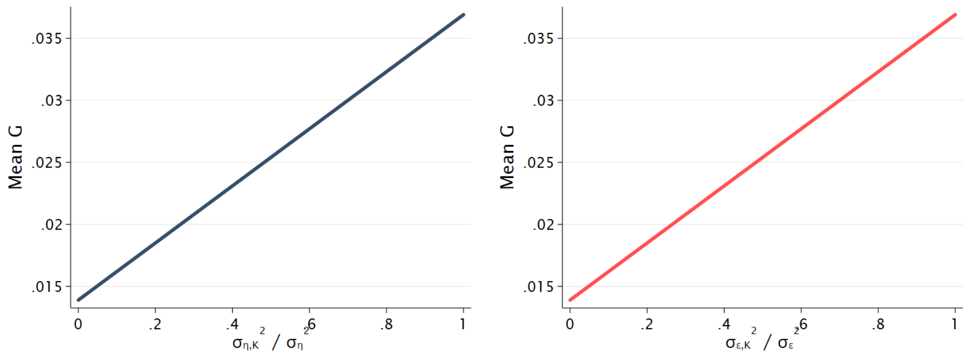
(a) Known permanent (b) Known transitory

Note: Panels (a) shows the proportion of known permanent shocks on the mean G and Panel (b) shows the proportion of known transitory shocks on the mean G.

Figure D.1: Effect on mean G

We also show the results for the standard deviation. For the standard deviation, those two shocks have a similar effect. As the proportion of known shocks increases, the standard deviation of G increases. Therefore, the different effect on the mean is the key to identifying the levels while controlling for the standard deviation of G.

For the life cycle patterns, we plot two different levels of $\sigma^2_{\eta,K}$ given the size of $\sigma^2_{\epsilon,K}$ in Fig c. As the known permanent shocks increase the life cycle pattern increases. This is because now the earnings shocks is more informative about the β. On the other hand, Fig d shows the two different levels of $\sigma^2_{\epsilon,K}$ given $\sigma^2_{\eta,K}$. In this case, the life cycle pattern decreases as the $\sigma^2_{\eta,K}$ increases. This is because the earnings shocks are less informative about the individual profile.



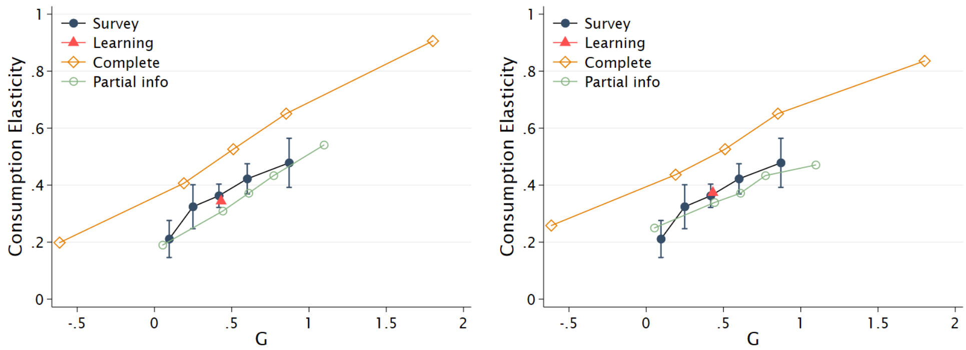
(a) Known permanent (b) Known transitory

Note: Panels (a) shows the proportion of known permanent shocks on the standard deviation G and Panel (b) shows the proportion of known transitory shocks on the standard deviation G.

Figure D.2: Effect on standard deviation G

E Discussion on consumption simulation

This section discusses the robustness of our consumption simulation results. We perform two robustness checks. First, we show that our result is robust to the no-borrowing constraint. We use the natural borrowing limit in the main text. We resimulate consumption with the no-borrowing constraint and confirm similar results. Second, we perform a robustness check using a more realistic Danish tax system. Kreiner et al. (2016) describes the Danish tax system, which has two tiers: above median income and below median income. To define the median income, we simulate the population and find the median. Then we apply the tax rate of 0.40 for the below-median group and 0.49 for the above-median group, as reported tax rate in Kreiner et al. (2016). Fig E.1 shows the results. Fig E.1a is the result for no borrowing constraint. Again the partial information well fits the empirical data pattern as in the main results. On the other hand, the complete information and learning assumption again fails to fit the empirical data. Fig E.1b shows the robustness result for the tax system, while the consumption elasticity across G is lower than the main result but we find very consistent results with the main text.



(a) No borrowing constraint (b) Tax-system

Note: Panels (a) shows the robustness result for the no borrowing constraint instead of natural borrowing limit. Panel (b) shows the result for tax system as in Kreiner et al. (2016).

Figure E.1: Robustness check on consumption elasticity across G

References

Crawley, E. and A. Kuchler (2023). Consumption heterogeneity: Micro drivers and macro implications. *American Economic Journal: Macroeconomics 15*(1), 314–41.

Guvenen, F. (2007). Learning your earning: Are labor income shocks really very persistent? *American Economic Review 97*(3), 687–712.

Guvenen, F. (2009). An empirical investigation of labor income processes. *Review of Economic dynamics 12*(1), 58–79.

Guvenen, F. and A. A. Smith (2014). Inferring labor income risk and partial insurance from economic choices. *Econometrica 82*(6), 2085–2129.

Karahan, F. and S. Ozkan (2013). On the persistence of income shocks over the life cycle: Evidence, theory, and implications. *Review of Economic Dynamics 16*(3), 452–476.

Kreiner, C. T., S. Leth-Petersen, and P. E. Skov (2016). Tax reforms and intertemporal shifting of wage income: Evidence from danish monthly payroll records. *American Economic Journal: Economic Policy 8*(3), 233–257.