THE UNIVERSITY OF CHICAGO

ESSAYS IN LABOR ECONOMICS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

KENNETH C. GRIFFIN DEPARTMENT OF ECONOMICS

BY
JACK DAVID ARMSTRONG LIGHT

CHICAGO, ILLINOIS
JUNE 2023

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I am extremely grateful to my advisors, Magne Mogstad, Stéphane Bonhomme, and Thibaut Lamadon. Chapter 1 is individual work. In addition to my advisors I thank Lucy Msall, Kevin Lee and seminar participants at the Econometrics and Labor working groups at the University of Chicago for detailed comments. Chapter 2 is joint work with Manuel Arellano, Richard Blundell and Stéphane Bonhomme.

ABSTRACT

This dissertation consists of two parts. The first part studies the effects of flexible work scheduling policies typically used by employers to efficiently manage their staffing operations in the context of variable customer demand. In particular, I study how such policies impact the quality of frontline work arrangements, the extent to which they are associated with higher employee turnover and quantify the value of such arrangements to employees. To this end, I construct and analyze a matched employer-employee dataset with precise information on employee scheduling arrangements from 10 million worked shifts. I motivate my empirical analyses with a model of workforce scheduling in which managers internalize the fact that employees jointly consider their work arrangements and wages when evaluating whether to remain with their current employer. I use my results to simulate the effects of shocks to customer demand on employee turnover.

In the second part, we study the link between household consumption decisions and earnings dynamics. We use the enhanced consumption data in the Panel Survey of Income Dynamics (PSID) from 2005-2017 to explore the transmission of income shocks to consumption. We build on the nonlinear quantile framework introduced in Arellano, Blundell and Bonhomme (2017). Our focus is on the estimation of consumption responses to persistent nonlinear income shocks in the presence of unobserved heterogeneity. To reliably estimate heterogeneous responses in our un-balanced panel, we develop Sequential Monte Carlo computational methods. We find substantial heterogeneity in consumption responses, and uncover latent types of households with different life-cycle consumption behavior. Ordering types according to their average log-consumption, we find that low-consumption types respond more strongly to income shocks at the beginning of the life cycle and when their assets are low, as standard life-cycle theory would predict. In contrast, high-consumption types respond less on average, and in a way that changes little with age or assets. We examine various mechanisms that might explain this heterogeneity.

CHAPTER 1

FRONTLINE WORK ARRANGEMENTS AND EMPLOYEE TURNOVER

1.1 Introduction

Current management practices for hourly employees are associated with schedules that are both highly variable and unpredictable ([2], [3], [4]). Surveys indicate that of the 73 million hourly workers in the United States (BLS, 2021) almost 40% receive less than a week's notice about when they are required to work ([4]) and as many as 80% of these workers have fluctuations in their weekly hours that exceed a full day of pay each month ([4]).

Such findings have made policymakers and researchers increasingly interested in understanding both the nature of employer scheduling practices as well as the broader impacts of such practices on employee and firm outcomes. In the United States so-called 'Fair Workweek' legislation has been introduced in 6 cities and one state since 2014, with the stated goal of establishing universal scheduling standards for hourly workers ([5]). Related initiatives have been introduced in the the EU through the Directive on Transparent and Predictable Working Conditions (2019) and in Australia though the Fair Work Act (2009). Similar interest is seen amongst firms seeking to reduce employee separations in a market environment with record high quit rates and rising wages.

This gives rise to several questions of both policy and practical importance. First, what are the precise mechanisms through which manager scheduling policies relate to the quality of worker schedules? Second, to what extent is the quality of schedules valued by workers? Third, does a failure to provide high quality schedules cause higher turnover within a firm? The goal of this paper is to answer these questions using a unique dataset obtained from a global provider of workforce management software with precise information from almost 10 million shifts worked by 100,000 workers in relevant frontline industries.

We start by developing a simple model of workforce scheduling at the weekly level. Importantly, we show how preferences depend on schedule quality and allow for the possibility that some variability may be desirable if chosen by workers. Our modeling incorporates several features which are common across modern scheduling applications. Workers have access to a shift-swapping technology which allows them to trade shifts with co-workers. This provides a mechanism through which variability in hours can be attributed to worker preference shocks. Managers have access to demand forecasting and scheduling algorithms which allow them to scale aggregate worker hours in response to changes in customer demand which follows an ARCH process. This provides a mechanism through which variability in worker hours can be attributed to the variance of weekly customer demand facing the firm.

When manager imposed variability is undesirable to workers we expect higher turnover rates to impose additional costs which should be internalized when making scheduling decisions. We microfound these turnover costs in the spirit of on-the-job search ([6]). Within each period employed workers receive outside utility offers drawn from some exogenous distribution and accept offers yielding higher expected utility than at their current employer, where utility is given as a linear combination of wages, amenities and realized schedule quality. Fluctuations in customer demand imply that schedule quality will vary over time and so optimal worker behavior will depend on how expectations are formed. If workers have adaptive or rational expectations about the future path of realized schedule quality then current realized values can have a direct impact on turnover. Solving for the manager's optimal policy taking this behavior as given reveals that the realized quality of schedules in a team will depend on the conditional variance of team-level demand and the extent to which workers get disutility from poor quality schedules. The model unambiguously predicts that schedule quality will be decreasing in the variance of customer demand but that, ceteribus paribus, managers will compensate workers for lower schedule quality in the form of wages.

To investigate the model empirically we leverage our unique dataset to map variables of interest into observable counterparts. First, we combine timestamps from when employees were notified about work with detailed time-tracking records from actual worked shifts to construct measures of schedule quality which precisely capture both variability in an employee's weekly hours as well as the predictability of their schedule. Our data show that from week-to-week a worker employed in the average team receives 5.96 days advance notice about upcoming work and has hours that fluctuate by approximately 25%, with meaningful heterogeneity observed across both measures. Second, we reconstruct the demand forecasts which were rendered to managers when making their scheduling decisions and can be used to infer shocks to demand at the team level. We document meaningful heterogeneity in the nature of demand across teams and show that, consistent with our model predictions, schedule quality is decreasing in the variance of team-level demand forecasts.

Our main empirical analyses attempt to identify the causal effect of schedule quality on employee turnover and to quantify worker WTP for improved schedule quality. Turnover is of particular interest for industries in our current setting where annual turnover rates typically exceed 100%. Similar causal parameters are studied by [7] in a Home Care setting and by [8] in the retail and food service sectors. In addition to direct turnover costs, which are surveyed in [9], a large literature documents the negative economic effects of turnover on the long-run performance of firms ([10], [11], [12], [13]).

Although a simple theory of compensating differentials implies that there must exist positive wage differentials if unpredictable and variable schedules are perceived negatively by employees ([14]) it is well known that hedonic prices need not coincide with worker preferences if there exist market frictions when searching for jobs ([15], [16]). Accordingly, our identification strategy leverages information from employee separations in the spirit of [17]. Intuitively, relative worker valuations for schedule quality and wages are revealed through the probability of separating from their current employer.

We apply this strategy using the variation in schedule quality, wages and separation rates observed within a team over time. Our model highlights two reasons why within-team variation is preferable to cross-sectional comparisons. Firstly, the stochastic process for customer demand may be directly correlated with team amenities. This induces a correlation between schedule quality and amenities in the cross-section because it is the realized values of customer demand which determine observed schedule quality. Recent empirical work documents positive correlations between the productivity distribution of firms and the level of amenities that they offer ([18], [19]). Secondly, the total compensation chosen by managers depends directly on the level of amenities in the team. As a result cross-sectional comparisons may underestimate the causal effects of schedule quality and wages on account of high (low) amenity teams who, ceteribus paribus, have lower (higher) levels of separations and for whom it is optimal to offer lower (higher) compensation.

To further account for the fact that some of the observed within-team variation in schedule quality will reflect employee preference shocks we leverage instruments constructed from the demand forecasts which were rendered to managers. For estimation we treat team-level amenities as additional parameters to be estimated jointly under an asymptotic where both the number of time periods and the number of teams grows. We adjust our reported coefficients to correct for the asymptotic bias which arises with incidental parameters in these settings ([20], [21], [22], [23]). In addition to our own estimates we also present results using specifications based on identification strategies that have been proposed in the existing literature. We use our model to provide an economic and econometric interpretation of these existing strategies and make comparisons to our baseline estimates.

Across all of our specifications we document that improved schedule quality has a significant and negative effect on the probability of worker separations. Our preferred estimates indicate that a 1% increase in schedule quality as measured by a reduction in the variability of hours results in a decrease in the probability that a worker separates the following week of 0.01 percentage points. When comparing these effects to those which correspond to an increase in wages our estimates suggest that workers would be willing to accept a reduction in wages of 1.2% for a 1% increase in schedule quality.

The insights from our paper relate first and foremost to a recent and growing literature which studies the scheduling arrangements of hourly workers and to which we make three contributions. First, we provide a simple model that explains the mechanisms through which schedule quality is determined as a function of shocks to both team-level demand and worker preferences. The model highlights the challenges that arise when observable quality measures that can be constructed from data are used as proxies for the schedule quality perceived by workers.

Second, we provide empirical estimates of both worker valuations for improved quality as well as the causal effect of schedule quality on separations. Our estimates are based on detailed personnel data that can precisely construct the observable measures of schedule quality which are typically referenced in the existing literature. In contrast to our work, many existing studies have typically relied on qualitative interviews and surveys ([24], [25], [26], [27], [28]) and so much research focuses on the potential mechanisms through which schedule quality is valued, focusing on the potentially negative effects of variability and unpredictability on work-life conflict ([29]), parent-child interactions ([24]) and resultant earnings volatility ([30]).

Third, we use our model to provide an economic and econometric interpretation of identification strategies using two-stage least squares or fixed effects which have been proposed in the existing literature ([8], [7]). Whilst two-stage least squares may approximate the causal effect of schedule quality on separations under an assumption that the proposed instrument is orthogonal to employee preference shocks we show that fixed effects strategies will generally not yield consistent estimates and provide expressions that approximate the resultant bias.

Our paper also contributes to a large literature in economics studying how employees may trade-off higher wages in exchange for better job characteristics ([31], [32], [33], [34]). Most recently a range of studies have focused on the importance of flexibility and control over one's working hours ([35], [36], [37] [38]). We complement this work by documenting related empirical findings across a large number of teams in relevant frontline industries with job characteristics that are particularly salient for hourly workers.

Finally, we also contribute to an active operations research literature studying firm scheduling decisions. Within this literature researchers are interested in understanding both how scheduling decisions may be optimized in order to improve overall firm performance ([39], [40]) as well as understanding the various mechanisms through which scheduling decisions may be linked to various performance drivers ([41], [42], [43]). Relative to existing work our paper highlights how managers may internalize the costs of downstream effects when making their scheduling decisions. Of particular interest to this current paper is work examining the interaction of manager performance with the use of algorithmic tools ([44]).

The outline of the paper is follows: Section 1.2 outlines our data and presents some descriptive statistics which motivate our subsequent modelling choices. Section 1.3 then presents our model of workforce scheduling. Section 1.4 presents our empirical strategy and identification arguments. Section 1.5 uses the model to provide an economic and econometric interpretation of identification strategies that have been proposed in the existing literature. Section 1.8 presents our main results which are used to simulate impulse responses in Section 1.9. Section 1.10 concludes.

1.2 Data

This section describes our dataset, sample restrictions and presents summary statistics describing the sample. We then explain how we use our data to construct the main variables of interest used in our analysis. We present descriptive evidence of heterogeneity in both the average quality of schedules and the characteristics of demand forecasts across teams and report team-level correlations documenting a negative relationship between schedule quality and the variance of demand forecasts. These descriptive findings motivate our subsequent modelling choices.

1.2.1 Overview

To perform our analyses we obtained data from a global provider of workforce management software serving businesses in frontline service industries such as retail, QSR, restaurants, hotels, staffing and healthcare. By the cutoff date used for data collection the provider was operating across a total of 21 countries including Australia, the United Kingdom and the United States. Businesses use workforce management tools as a means of forecasting customer demand which can then used as an input to scheduling algorithms and templates offered by the provider to generate optimized employee schedules. The software also offers additional features to facilitate the tracking of employee time and attendance, co-ordinating payroll and communicating with employees. As such their data offers a unique opportunity to study questions related to frontline work arrangements.

Our dataset is constructed in the form of a worker-week panel by combining data from four separate sources which we detail below: planned schedules; time-stamped activity logs; forecasting integrations; and HR records. The HR records provide only a limited set of demographic information about age and employment history and so in order to perform additional analyses studying heterogeneous effects we performed an imputation exercise to predict an employee's gender based on their reported first and last name.

The planned schedules provide information on the businesses, teams and locations in which individual employees are assigned as well as the weekly schedules set by their managers. These records include details as to when individual employees were supposed to start and finish work each day as well as information on any scheduled breaks or days off. These logs also provide us with records as to when employees were last notified about any updates to their schedule. Whilst researchers have long been interested in studying the effects of having short notice of schedule changes and last-minute shift cancellations they have typically had to rely on survey-based datasets to infer the prevalence of such practices ([8]).

The time-stamped activity logs are obtained from geo-fenced time clocks which are installed on employee mobile devices and are used by employers to track the hours worked by their employees. They provide information as to when each individual actually punched in and out for a shift, any breaks taken and any unscheduled absences. The logs also provide information on the total wage costs incurred by the firm for each shift which we use to determine employees' hourly earnings.

The forecasting integrations allow us to observe the time series data which businesses use to inform their day-to-day staffing operations. For example, the data for a restaurant or hospitality business might provide details on the number of sales and transactions through each of the cash registers in the organization; a retail business may use total footfall; and a hotel might use total room bookings. In addition to the realized values for these time series we were also provided with details related to the forecasting algorithms used by the software provider to display forecasts to managers when making scheduling decisions. Importantly, prior to 2021 these algorithms comprised of a set of deterministic heuristics which allow demand forecasts to be reconstructed exactly.

1.2.2 Sample

In order to focus our analyses on teams using the software as part of their day-to-day staffing operations we restricted our sample to teams that we observe for at least 3 months. This restriction also ensures that our data excludes organizations that may be in the process of migrating across from other software vendors and whose data may not yet offer an accurate insight into their current operations. We also excluded firms who the software provider indicated were using the software as part of an on-going pilot or sales process. At the employee level we restrict ourselves to a sample of workers with non-missing data who have been continuously working in a team for at least four weeks. Such a restriction allows us to focus on a subset of workers most likely to be employed by the firm and for whom our variables of interest are well defined. For example, it is common for businesses to use a small number of trial shifts as part of their recruitment process which we wish to exclude from our analyses.

Our final sample contains detailed records for the period 2012-2021 and covers a total of 2,123,228 worker weeks for 83,095 employees working across 5,274 teams within 790 separate organisations. Sampled organisations are typically medium-sized enterprises with an average of 70 employees. We present a set of summary statistics in Table 1.1 and highlight some important features. Most workers are young and would not typically be considered full-time. The average employee is 21.08 ($\sigma = 6.87$) and typically works 15.07 hours per week ($\sigma = 9.10$) across 3.13 separate shifts ($\sigma = 1.31$). We see no meaningful disparities in the gender composition of our sample which is 57% male ($\sigma = 0.49$).

1.2.3 Variables of Interest

We describe below how our main variables of interest are calculated using our data. In constructing these measures we join a recent body of research attempting to measure the quality of shift-based work using personnel data collected as part of the management of a firm's day-to-day workforce operations ([45], [44], [42], [7]). This work adds to earlier research making use of employee interviews and surveys ([24], [25], [26], [27] [28]).

Separations — Our primary outcome of interest is the separation of employee i from team j in week t, denoted as S_{ijt}. A separation is defined as the first week in which an employee with a continuous employment history is no longer observed within a firm. This definition automatically excludes temporary separations related to vacations and absences as well as separations which occur as a result of internal mobility between teams within the same firm. To avoid falsely including separations which arise when a firm closes or changes software providers we exclude separations that occur within 4 weeks of a firm exiting our sample.

Table 1.1: Descriptive statistics

	Mean	Standard Deviation
Workers		
Proportion male	0.57	0.49
Age at entry	21.08	6.87
Shifts per week	3.13	1.31
Hours per shift	4.70	1.67
Hours per week	15.07	9.10
Log wage	3.02	0.37
Average tenure at separation (weeks)	26.81	31.78
Firms		
Teams per firm	6.67	19.93
Employees per team	10.10	7.37
Separation rate (weekly)	0.02	0.06
Log output per worker	8.81	1.34
Sample Size		
Firms		790
Teams		5,274
Employees		83,095
Worker weeks		2,123,228
Shifts		6,645,704

Although our definition doesn't enable us to separately distinguish between voluntary and involuntary separations it is well documented that voluntary separations typically account for the vast majority of total separations. For example, the JOLTS report 12/31/2022 indicates that 70% of all separations are voluntary (US [46]). We expect this number to be higher for the industries covered in our sample.

Our data indicate that the average employment spell of an employee within any given team is short. The average tenure at separation is 26.81 weeks ($\sigma = 31.78$) which at the team level manifests itself in a weekly separation rate of close to 2% ($\sigma = 0.06$). This separation rate is consistent with numbers typically reported for businesses in the retail, leisure and hospitality sectors. These numbers suggest that the costs of turnover will be of first-order importance to the firm even though weekly separation rates may be small on a per-worker basis.

Schedule Quality — We use our data to construct a baseline measure of schedule quality V_{ijt} as well as two additional measures V_{ijt}^{Daily} and $V_{ijt}^{Predictability}$ which we use to highlight the robustness of our results. These measures are commonly used in the existing literature and are motivated by existing research from the University of Chicago's Employment Instability, Family Well-being, and Social Policy Network ([25]) and the Harvard Kennedy School's SHIFT project ([26]). This work highlights the central roles played by variability and unpredictability in determining the quality of a worker's schedule. In Section 1.3 we describe the relationship between our observable quality measures and the primitives of worker preferences.

For worker i working in team j in week t our primary measure of schedule quality V_{ijt} is constructed as follows

$$V_{ijt} = -\Big(H_{ijt} - \bar{H}_{ij}\Big)^2 \tag{1}$$

where H_{ijt} denotes i's total hours worked in week t and $\bar{H}_{ij}$ denotes worker i's average weekly hours. Our V_{ijt}^{Daily} measure of schedule quality captures variability in daily hours and is constructed as

$$V_{ijt}^{Daily} = -\Big(S_{ijt} - \bar{S}_{ij}\Big)^2$$

where S_{ijt} denotes worker i's average shift length in t and $\bar{S}_{ij}$ denotes worker i's average shift length. Our $V_{ijt}^{Predictability}$ measure captures the predictability of a schedule and is calculated as

$$V_{ijt}^{Predictability} = \frac{1}{K_{ijt}} \sum_{s=1}^{K_{ijt}} M_{ijts}$$

where K_{ijt} denotes the number of shifts worked by worker i in week t and M_{ijts} denotes the amount of notice that was provided to worker i in advance of each of those shifts.

To help motivate the policy relevance of our subsequent findings it is useful to explain how our observable measures of schedule quality may be linked to common features of 'Fair Workweek' legislation which has been proposed in multiple US jurisdictions. For example, our V and V^{Daily} measures of schedule quality are often formalized by a requirement that, prior to or upon employment, employers provide written estimates of the usual number of days and hours each employee can expect to work. Our $V^{Predictability}$ measure of schedule quality is often formalized by a requirement that employers pay an hourly wage premium after making changes to employees' shifts at short notice. In many cases proposed legislation will link predictability and variability by making the aforementioned wage premia dependent on whether or not shift changes resulted in a loss or increase of hours.

Customer Demand — We measure customer demand A_{jt} at the team level by reconstructing the demand forecasts which were rendered to managers when making their scheduling decisions. We divide by the team size to obtain per-worker measurements. Whilst in practice we cannot rule out the possibility that managers have access to additional information when making scheduling decisions we believe these forecasts to be a reasonable proxy.

1.2.4 Descriptive Statistics

We document below three descriptive features of our data which motivate our subsequent modelling decisions.

Heterogeneity in Quality Between Teams — Previous work suggests there may be meaningful heterogeneity in the provision of schedule quality across teams and firms. For example, several studies highlight that low predictability and high variability are typically more prevalent for workers in certain industries, such as retail and service sectors ([47], [26], [25]). To investigate this idea Figure 1.1 provides a first illustration of the extent to which our measures of schedule quality vary across teams. The left panel reports team-level medians of our V measure and documents noticeable heterogeneity. The average value of these team-specific measures is given by -55.65 which corresponds to a deviation from week-to-week of 7.42 hours or approximately 25% of total weekly hours. The standard deviation across teams is 39.67 and the distribution displays negative skew suggesting the presence of certain teams with significantly greater variability in hours.

The right panel reports team-level medians of our $V^{Predictability}$ measure and indicates that there is also a large amount of variability in the amount of notice received by employees about their upcoming shifts. The average value of these team-specific measures of $V^{Predictability}$ is given by 5.96 days. The standard deviation of this measure across teams is 2.77 and the distribution displays positive skew. We document a sizeable number of teams requiring their employees to work at very short notice. For example, the bottom 10% of teams provide less than 3.56 days notice ahead of employee shifts.

Heterogeneity in Demand Between Teams — To investigate whether similar heterogeneity exists in the variability of customer demand we estimated the following ARCH specification separately for each team:

$$\begin{aligned} A_t &= \mu + \epsilon_t, \quad \mathbb{E}[\epsilon_t] = 0 \\ \epsilon_t &= \sigma_t e_t \quad e_t \sim \mathcal{N}(0,1) \\ \sigma_t^2 &= \rho_0 + \rho_1 \epsilon_{t-1}^2 \end{aligned} \tag{2}$$

where μ is a constant mean, ϵ_t is a time-varying shock, e_t is a standardized residual and σ_t^2 is the conditional variance of demand. The normality assumption is not essential but will allow us to generate closed form solutions for optimal manager behavior when incorporated into the model in Section 1.3.

Our results are illustrated in Figure 1.2. The left panel shows a histogram of the unconditional variance and shows substantial heterogeneity across teams. The average variance is given by 0.17 and the standard deviation across teams is given by 0.81. The estimated variances also document positive skew with certain demand streams exhibiting particularly high variance. The right panel shows a histogram for the absolute value of the estimated team-specific ARCH coefficients and also shows substantial heterogeneity. The average coefficient has an absolute value of 0.31 which is consistent with the presence of meaningful autoregressive conditional heteroskedasticity. The standard deviation across teams is 0.279.

Cross-Sectional Correlations — Table 1.2 reports estimates from team-level OLS regressions of average schedule quality on the variance of customer demand within a team. Estimates for each of our quality measures V, V^{Daily} and $V^{Predictability}$ are shown in Columns (1), (2) and (3) respectively. The estimates indicate that, on average, teams in which the variance of customer demand is higher are associated with lower quality schedules. Coefficients are statistically significant at conventional levels for both our V and V^{Daily} measures.

We note that these findings at the team level are consistent with our subsequent analysis using worker panel data.

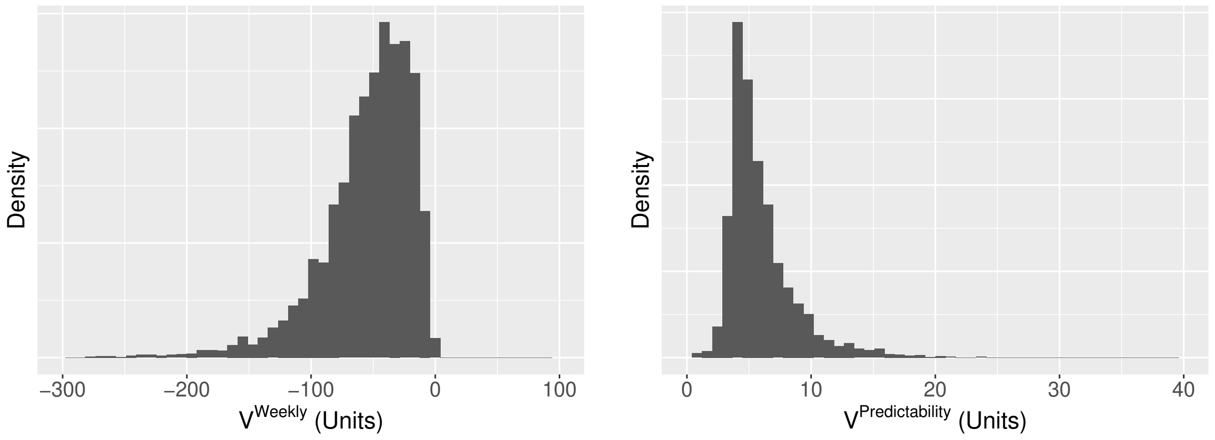


Figure 1.1: Distribution of average schedule quality across teams

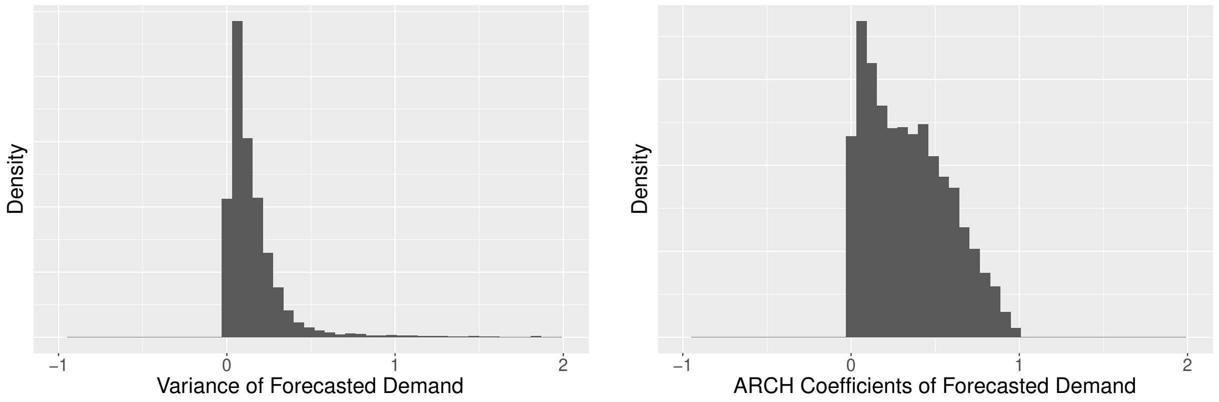


Figure 1.2: Distribution of forecast demand across teams

	(1)	(2)	(3)
(Intercept)	-2.3195^{***}	2.0420^{***}	4.7369^{***}
	(0.0173)	(0.0227)	(0.0083)
Var(A_{jt})	-0.9603^{***}	-1.6397^{***}	-0.0354
	(0.1218)	(0.1599)	(0.0583)
Quality Measure	V	V^{Daily}	$V^{Predictability}$
Num. obs.	5274	5274	5274

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Team level regressions of average schedule quality on the variance of demand forecasts. Standard errors in parentheses.

Table 1.2: Relationship between variance of customer demand and schedule quality

1.3 Model

Time is discrete and runs from $t = 1, ..., T$. We consider a representative manager who is responsible for a team comprised of a large set of N workers indexed by $i = 1, ..., N$. We begin by formalizing how worker preferences depend on schedule quality and show how variable scheduling may lead to higher turnover rates. Next, we present a simple model of team aggregation in which shift swaps between workers can be used to derive a mapping between worker level utilization rates and the aggregate variables that are chosen by the manager. Lastly we study the optimal behavior of the manager who takes worker choices as given and internalizes the effect of variable scheduling on turnover costs when choosing aggregate variables.

1.3.1 Workers

We assume that in period t an employed worker i receives the following flow utility from wages W, a set of time-invariant team amenities ξ and hours H:

$$u_{it}(W, H, \xi) = \alpha_w W - \alpha_v \Big(H - H^*_{it}\Big)^2 + \xi + \nu_{Eit},$$

where H^*_{it} represents worker i's ideal total hours in period t and ν_{Eit} represents a shock to the utility that worker i receives from their continued employment in the team. We assume that H^*_{it} and ν_{Eit} are identically and independently distributed across both individuals and time and introduce the following distributional assumptions

$$H^*_{it} \sim \mathcal{N}(1, \sigma^2_H) \qquad \nu_{E_{it}} \sim T1EV$$

Schedule quality is defined at the worker level using the term

$$V_{it}^* \equiv -\left(H_{it} - H_{it}^*\right)^2 \tag{3}$$

so that the quality of a worker's schedule depends on the extent to which their hours deviate relative to their desired level. Our definition is motivated by a large sociology literature suggesting that poor quality schedules make it difficult for employees to effectively plan activities and meet responsibilities outside work ([47], [48]). Such findings will be consistent with the model only if $\alpha_v > 0$ so that utility is decreasing in the squared difference between actual and ideal hours.[1]

Exposing the worker to variability in hours will be useful for the manager because it will allow them to scale hours worked in response to fluctuations in customer demand. When $\alpha_v > 0$ we also expect there to be additional costs associated with variable scheduling that arise on account of higher employee turnover. To microfound these turnover costs Appendix 1.A.1 presents a stylized model of on-the-job search (e.g. [6]) in which each period employed workers within a team receive outside utility offers drawn from some common distribution with mean $\bar{O}$. When outside offers are independent, $\bar{O}$ is sufficiently unattractive[2] and workers have adaptive expectations over the future paths of schedule quality and wages the model yields the following structural equation describing the relationship between weekly separation rates and the level of amenities, wages and schedule quality in the team

$$\mathcal{P}(\xi, W, V^*) \equiv Pr(S_{it} = 1|\xi, W, V^*) = \frac{1}{1 + \exp(\tilde{\xi} + \tilde{\alpha}_w W + \tilde{\alpha}_v V^*)} \tag{4}$$

[1] It is also possible to generate preferences which depend on variability in hours through mean-variance or quadratic utility functions.

[2] This assumption is consistent with our worker-level data in which weekly separation rates at the individual worker level are small.

where we defined $\tilde{\xi} = \frac{\xi - \bar{O}}{1-\beta}$, $\tilde{\alpha}_v = \frac{\alpha_v}{1-\beta}$ and $\tilde{\alpha}_w = \frac{\alpha_w}{1-\beta}$ for some discount rate β. When $\alpha_v > 0$ and $\alpha_w > 0$ weekly separation rates will be decreasing in schedule quality, wages and amenities.

1.3.2 Shift Swapping and Aggregate Team Hours

We allow the realized values of schedule quality V_{it}^* which enter Equation 4 to depend on a two-stage scheduling process which incorporates both manager choices and the idiosyncratic worker shocks $\{H_{it}^*\}_{i=1}^N$. In a first-stage, before worker preference shocks have been drawn, managers use demand forecasts to determine a desired total number of hours to be worked at the team level. The distribution of total hours across workers is then determined during a second stage in which workers have access to a shift-swapping technology that enables them to adjust hours by trading their shifts with other workers. Shift-swapping tools are a standard feature in the majority of scheduling software applications.

The presence of shocks to workers' ideal hours suggests that there will be gains from trade relative to a baseline policy in which hours are distributed equally across workers. To exhaust these gains we model shift swaps by assuming that the resultant allocation of hours across workers will be equivalent to that obtained from maximizing total worker surplus subject to the constraint that total hours worked by the team coincides with the manager's desired total in the first-stage:

$$\{H_{it}\}_{i=1}^N = \underset{\{h_{it}\}_{i=1}^N}{\operatorname{argmin}} \sum_{i=1}^N (h_{it} - H_{it}^*)^2 \quad \textit{subject to} \quad \sum_{i=1}^N h_{it} = \bar{H}_t$$

where $\bar{H}_t$ is the manager's desired total team hours from the first-stage. In Appendix 1.A.2 we show that solving this program yields the following expression for worker i's allocation of hours in period t:

$$H_{it} = \frac{\bar{H}_t}{N} + \underbrace{H^*_{it} - \frac{1}{N}\sum_{m=1}^{N} H^*_{mt}}_{\pi_{it}}, \tag{5}$$

so that total hours at the worker level is increasing in both the aggregate total chosen by the manager and the net value of idiosyncratic shocks. Intuitively, the result shows that a worker's idiosyncratic preferences are more likely to be satiated in large teams whilst in smaller teams workers will be more exposed to the taste shocks of co-workers.

1.3.3 Managers and Demand Driven Scheduling

Managers have access to a linear production technology which allows the production of output Y_t as a function of customer demand A_t and total hours worked $\bar{H}_t$:

$$Y_t = A_t \bar{H}_t$$

where total hours is calculated as the sum across the N worker-level variables, H_{it}, derived in Equation 5:

$$\bar{H}_t = \sum_{i=1}^{N} H_{it}$$

and customer demand is modeled according to the ARCH(1) process given in Equation 2. Importantly, we allow for the possibility that in the cross-section the level of amenities may be endogenous to the stochastic process faced by the firm.

We assume that managers have access to a scheduling tool which allows them to adjust total scheduled hours in response to realized values of customer demand according to some policy rule $g(.)$ chosen by the manager from a class of policy rules $\mathcal{G}$

$$\bar{H}_t = g(A_t), \quad g(\cdot) \in \mathcal{G}$$

Workers are paid hourly and receive a fixed wage W_t that must be set by the manager in advance of observing A_t such that the total wage costs in period t are given by $\bar{H}_t W_t$. We model the ongoing turnover costs of replacing employees as $c \cdot Q_t$ where c is a per-worker turnover cost and Q_t is a random variable which measures the number of employees who separated in period t and whose distribution is characterized by Equation 4. The manager's problem can then be written as

$$\begin{aligned} &\max_{\{g(\cdot),\ W\}} \ \mathbb{E}\Big[\ A_t \cdot g(A_t) - W \cdot g(A_t) - c \cdot Q_t\ \Big] \\ &\textit{subject to} \qquad\qquad\qquad\qquad g(\cdot) \in \mathcal{G} \end{aligned}$$

where expectations are taken with respect to the information set in period $t-1$. Note that there are two causes of uncertainty in the model which arise from the fact that both A_t and Q_t are random variables.

Our main analysis considers policy rules in which hours-per-worker scales linearly with shocks to demand

$$\frac{g(A_t)}{N} = 1 + \phi[A_t - \mu], \quad \phi \in [0, 1] \tag{6}$$

where the parameter ϕ determines the responsiveness of allocated hours-per-worker to demand shocks and is chosen by the manager in advance of observing A_t. A value of 0 implies that employees' hours are completely fixed whilst a value of 1 implies that hours adjust one-for-one with changes to demand. Whilst restrictive, this class is representative of the policy rules used in industry settings where so-called 'labor ratios' or 'labor standards' are used to determine a target number of scheduled employee hours which scales linearly with additional units of demand.

Combining Equations 2, 5, and 6 implies that worker-level utilization is given by

$$H_{it} = 1 + \phi_t \epsilon_t + \underbrace{H^*_{it} - \frac{1}{N}\sum_{m=1}^{N} H^*_{mt}}_{\pi_{it}} \tag{7}$$

which implies that for large enough teams Q_t will follow a Poisson distribution with rate parameter given by $\lambda = N \cdot \mathcal{P}(\xi, W, \phi^2 \epsilon_t^2)$, where $\mathcal{P}(\cdot)$ is the structural function defined in Equation 4. Combining this result with our parametric assumptions on ϵ_t allows us to to derive the following closed form expression for the firm's expected turnover cost:[3]

$$\mathbb{E}[c \cdot Q_t] = \frac{cN \exp(-\tilde{\alpha}_w W_t - \tilde{\xi})}{\sqrt{1 - 2\tilde{\alpha}_v \phi_t^2 \sigma_t^2}}$$

where we have the added regularity condition $1 - \frac{2\alpha_v \phi_t^2 \sigma_t^2}{1-\beta} > 0$ to ensure existence. Intuitively, turnover costs are decreasing in wages and the level of team amenities. This follows directly from the fact that workers are less likely to receive outside offers which are preferable to their current employment when in a team which offers a high level of wages and amenities. Since variability is undesirable ($\alpha_v > 0$) we see that turnover costs are increasing in $\phi_t^2 \sigma_t^2$, which measures the amount of the conditional variance in customer demand to which the worker is exposed. This highlights the fundamental trade-off facing the firm under variable scheduling policies since marginal (expected) revenue is increasing in ϕ_t.

The closed form expression for turnover costs allows us to solve for the manager's optimal choice of W_t and ϕ_t which we collect in Proposition 1. The proof is given in Appendix 1.A.3. Related comparative statics are given by Corollary 1. Interestingly, whilst a long line

[3] Specifically, the normality assumption for e_t implies that ϵ_t^2 has a gamma distribution $G(\frac{1}{2}, 2\sigma_t^2)$.

of research examines how shocks to the conditional mean of productivity pass through to employee wages ([49], [50], [19]), our result highlights that higher order moments are relevant for the quality of worker schedules.

Proposition 1 For managers of large teams in which the average weekly separation rate is small the optimal choice of W_t and ϕ_t is given by

$$\phi_t^* = \frac{\sqrt{2\tilde{\alpha}_v \tilde{\alpha}_w^2 \sigma_t^2 + \tilde{\alpha}_v^2}}{2\tilde{\alpha}_v \tilde{\alpha}_w \sigma_t^2} - \frac{1}{2\tilde{\alpha}_w \sigma_t^2} \tag{8}$$

$$W_t^* = \frac{\ln(c\ \tilde{\alpha}_w^2 \sigma_t)}{\tilde{\alpha}_w} - \frac{\ln\left(\sqrt{2\tilde{\alpha}_v \tilde{\alpha}_w^2 \sigma_t^2 + \tilde{\alpha}_v^2} - \tilde{\alpha}_v\right)}{2\tilde{\alpha}_w} - \frac{\xi}{\tilde{\alpha}_w} \tag{9}$$

Corollary 1 (a) Wages are decreasing in the level of amenities ξ and increasing in the total amount of risk to which the worker is exposed $\sigma_t^2 \phi_t^2$. (b) The variance of observed weekly worker hours is increasing in both the conditional variance of customer demand σ_t^2 and the variance of employee preference shocks σ_H^2.

1.4 Empirical Strategy

We are interested in obtaining estimates of the following quantities

$$\textit{Parameters of Interest:} \qquad \Delta_{v^*}\mathcal{P} \equiv \frac{\partial \mathcal{P}(\cdot)}{\partial V^*}\ ,\ \tilde{\alpha_v}\ ,\ \tilde{\alpha_w}$$

where $\mathcal{P}(\cdot)$ is the structural probability that a worker separates defined in Equation 4. The parameter $\Delta_{v^*}\mathcal{P}$ measures the causal effect of schedule quality on employee turnover which is of independent interest in our setting where annual turnover rates exceed 100%. A similar parameter is studied in [7] and [8]. We will also be interested in obtaining estimates of $\frac{\alpha_v}{\alpha_w}$ which measures an employees' willingness to forgo wages in exchange for improved schedule quality. The definition given in Equation 3 indicates that V^* is unobserved and so we highlight what can be learnt when our observable measure V defined in Equation 1 is used as a proxy.

1.4.1 Endogeneity in Cross-Sectional Comparisons

Cross-sectional comparisons of separation rates between teams which fail to control for the unobserved level of amenities ξ_j which enter into Equation 4 will generally not yield consistent estimates of our target parameters. This holds even in the absence of idiosyncratic shocks to a worker's ideal number of hours. To see this, start by fixing $\sigma_H^2 = 0$ so that actual schedule quality and observed schedule quality coincide (i.e. $V_{ijt}^* = V_{ijt}$) and combine with Equation 7 to yield

$$V_{ijt} = V_{ijt}^* = \phi_{jt}^2 \epsilon_{jt}^2,$$

which indicates two mechanisms through which schedule quality and wages will be endogenous. In the first instance we will be concerned that the level of amenities in a team and the conditional variance of customer demand may be correlated such that $\mathbb{E}[\epsilon_{jt}^2 \xi_j] \neq 0$. For example, an underlying factor such as the amount of face-to-face customer interaction required to perform a role may simultaneously make a role innately undesirable and expose it to higher variability in customer demand. Recent empirical work documents a positive relationship between the productivity of firms and the level of amenities that they offer ([18], [19], [51]). These findings suggest that similar relationships may exist between amenities and empirical moments of customer demand.

Secondly, we expect that teams endowed with better quality amenities are able to offer lower quality compensation to their workers than teams with lower quality amenity endowments such that $\mathbb{E}[V_{ijt}\xi_j] \neq 0$ and $\mathbb{E}[W_{ijt}\xi_j] \neq 0$. Intuitively, a team which is conveniently located next to a public transit hub is likely to find that its weekly staff turnover rates are lower than an equivalent team offering similar quality schedules and wages but located in an inconvenient location. As a result the optimal policy of the manager will be to offer lower compensation. This result is seen formally using the model by observing the dependence of Equation 9 on ξ from which it follows that neither α_w nor α_v are identified. That neither coefficient is identified even though Equation 8 does not depend directly on ξ is a result of the fact that H_{ijt} and W_{ijt} are jointly determined by σ_{jt} which implies that $|\operatorname{Cov}(W_{ijt}, V_{ijt})| > 0$.

1.4.2 Endogeneity in Time-Series Comparisons

When there are idiosyncratic shocks to a worker's ideal number of hours, $\sigma_H^2 > 0$, comparisons of separation rates within a team over time will also fail to yield consistent estimates of target parameters. This is because worker preferences depend on both the realized level of hours worked and the idiosyncratic shocks to their ideal number of hours. Whilst we observe the former the latter are latent from the perspective of the researcher. When hours at the individual worker-level are able to adjust in response to these worker preference shocks, as is the case with a shift-swapping technology, then observed schedule quality will fail to adequately capture the actual quality experienced by the worker. Intuitively, when naively using observed schedule quality as a proxy for the quality experienced by a worker we fail to account for the fact that some of the observed variation in quality may in fact be desirable. As a result, attempts to estimate Equation 4 using V_{ijt} instead of V^*_{ijt} will generally underestimate the true value of α_v.

To show this formally, Appendix 1.A.4 shows that for large T our observed quality measure V can be mapped to actual quality V^* according to the relationship

$$V_{ijt} \rightarrow_p V^*_{ijt} + err^v_{ijt} \tag{10}$$

where err^v_{ijt} is a measurement error which depends on worker i's realized hours H_{ijt} and preference shocks H^*_{ijt}. Subsequently substituting for H_{ijt} using Equation 7 reveals that $|\operatorname{Cov}(err^v_{ijt}, V^*_{ijt})| > 0$ and highlights that the measurement error is non-classical.

1.4.3 Identification using within-firm variation and instrumental variables

Motivated by the above concerns our main identification arguments combine two approaches. Firstly, we account for the potential endogeneity of schedule quality and wages in the cross-section by leveraging the large time-dimension of our panel to focus on the variation in separation rates, schedule quality and wages observed within a team over time. To further account for the potential endogeneity which arises when using V_{ijt} as a proxy for V^*_{ijt} we use the realized variance of demand forecasts as an instrumental variable.

In Appendix 1.A.7 we use our model to show that Equation 4 can be approximated with the following quantity

$$Pr(S_{ijt} = 1|\xi, W_{ijt}, V^*_{ijt}) \approx F(\xi_j, W_{ijt}, \phi^2_{jt}\epsilon^2_{jt}) = \frac{1}{1 + \exp(\tilde{\xi} + \tilde{\alpha}_w W_{ijt} - \tilde{\alpha}_v \phi^2_{jt}\epsilon^2_t)} \tag{11}$$

which suggests a simple approach to estimation using limited information maximum likelihood ([52]). In particular we estimate the conditional distribution of separations S_{ijt} according to the following specification

$$f_S(s|X_{ijt}, \xi_j, \lambda_t) = F(X_{ijt}'\alpha + \xi_j + \lambda_t)^s[1 - F(X_{ijt}'\alpha + \xi_j + \lambda_t)]^{1-s}, \quad s \in \{0, 1\}$$

where $X_{ijt} = \left(W_{ijt}, \phi_{jt}^2\epsilon_{jt}^2\right)'$, ξ_j are team fixed-effects and λ_t are an additional set of effects controlling for calendar time.

Since ϕ_{jt} is unobserved it must also be estimated in a first-stage. Using Equation 7 we can write observed quality as

$$V_{ijt} = \mathbb{E}[\pi_{ijt}^2] + \phi_{jt}^2\epsilon_{jt}^2 + \nu_{ijt},$$

where $\nu_{ijt} = 2\pi_{ijt}\phi_{jt}\epsilon_{jt} + \pi_{ijt}^2 - \mathbb{E}[\pi_{ijt}^2]$ is an error term satisfying $\mathbb{E}[\phi_{jt}^2\epsilon_{jt}^2\nu_{ijt}] = 0$. This implies that OLS regressions will recover consistent estimates of ϕ_{jt}^2 after appropriately parameterizing ϕ_{jt}^2. We use Equation 8 to motivate the following parameterization for some flexible choice of $f(.)$:

$$\phi_{jt}^2 = f(\sigma_{jt}^2, \sigma_j^2)$$

in which ϕ_{jt} is allowed to vary as a function of the conditional and unconditional variances of team demand. We additionally include the unconditional variance of demand forecasts to account for the fact that in practice it is known that managers typically update their scheduling software configurations at irregular intervals. As a robustness we also report estimates based on parameterizations involving higher order moments.

It is well known that an incidental parameter problem can cause severe bias in nonlinear panel data models when either J or T is held fixed ([53]). Accordingly we take advantage of the length and width of our panel and rely on results showing consistency of estimators for α, $\{\xi_j\}_{j=1:J}$ and $\{\lambda_t\}_{t=1}^T$ when both J and T are allowed to grow with the sample size under an asymptotic where $T \to \infty$, $J \to \infty$ and $\frac{J}{T} \to \kappa > 0$ ([23]). Importantly, although a standard maximum likelihood estimator will be consistent, it has an asymptotic bias which can result in severe under-coverage of confidence intervals. To address this concern we utilize results from [20] and [21] who demonstrate how to construct an alternative estimator based on the split-panel Jackknife introduced by [22]. Confidence intervals constructed using this corrected estimator are shown to have significantly improved coverage properties. For average partial effects the incidental bias problem is negligible asymptotically because the order of the bias can be shown to be smaller than the rate of convergence ([21]).

1.5 Relationship to Existing Literature

We can use our model to provide an economic and econometric interpretation of identification strategies which have been proposed in the existing literature. We focus on two common strategies. The first uses the two-stage least-squares estimator and the second uses a logit specification without instrumental variables.

1.6 Estimates Using 2SLS

Strategies involving two-stage least squares use some form of the following specification described in [7]

$$S_{ijt} = \alpha_v^{2SLS} V_{ijt} + \alpha_w^{2SLS} W_{ijt} + \lambda_j + \lambda_t + e_{ijt} \tag{12}$$

$$V_{ijt} = \omega_1 Z_{ijt} + \omega_2 W_{ijt} + \kappa_j + \kappa_t + v_{ijt} \tag{13}$$

in which Z_{ijt} is an instrumental variable to be used in the first-stage described by Equation 13. The main disadvantage of the linear probability model is that it will be misspecified when the true separation rate is given by Equation 4. Although in many cases it has been shown that marginal effects from a linear probability model are similar to true marginal effects when the correct specification is known ([54]), theoretical arguments typically require fitted probabilities to be close to 0.5 in order for linear approximations to be valid. By contrast, fitted probabilities will be small in our setting. In Section 1.8 we compare estimates obtained using 2SLS specifications to those obtained from our model and do report some quantitative differences, although qualitatively our results are very similar.

In addition, we note that coefficients obtained from the two-stage least squares specification in Equation 12 cannot be interpreted directly in terms of the preference parameters from Section 1.3, however, the signs of coefficients will have an inverse relationship. For example, when $\alpha_w > 0$ and $\alpha_v > 0$, so that utility is increasing in schedule quality and wages, we expect separations to decrease when wages and quality increase which implies that $\alpha_w^{2SLS} < 0$ and $\alpha_v^{2SLS} < 0$.

Our model suggests a natural choice of instrument constructed using shocks to the demand forecasts which were rendered to managers. In particular, setting $Z_{ijt} = \epsilon_{jt}^2$ will satisfy the necessary exogeneity and relevance conditions. To see this we show in Appendix 1.A.5 how our model can be used to derive

$$\text{Cov}\Big(\epsilon_{jt}^2 \ , \ err_{ijt}^v\Big) = 0 \tag{14}$$

$$\text{Cov}\Big(\epsilon_{jt}^2 \ , \ V_{ijt}\Big) < 0 \tag{15}$$

which shows that realized demand shocks are correlated with observed schedule quality V_{ijt} but uncorrelated with the measurement error err_{ijt}^v described in Equation 10. The autoregressive conditional heteroskedasticity in the variance of customer demand implies that lagged values of ϵ_{jt}^2 may also be used as relevant instruments. The use of lagged forecasts may have additional robustness properties in cases when managers' demand forecasts make use of team outcomes such as sales or transactions.

It is useful to highlight the restrictions on economic behavior imposed by Equation 14. In particular we require that the shocks to a worker's preferred hours are independent of the demand shocks facing the team. Such an assumption may be restrictive if there are common underlying shocks which jointly determine both customer demand and worker preferences. An example of such a shock might be a holiday or an event which simultaneously increases customer demand whilst reducing an employee's desired hours. The inclusion of time effects at the weekly level across our specifications is designed to address this concern.

The economic content of this restriction turns out to be similar to that required when using the instrumental variable proposed in [7]. In that paper the authors instrument for observed schedule quality using paid days off (PDO) or unexplained absences of co-workers in the same team. Recall that Equation 5 derives worker-level shocks to utilization in period t as

$$\pi_{it} \quad = \quad H^*_{it} - \frac{1}{N_j}\sum_{m=1}^{N_j} H^*_{mt} \quad = \quad \frac{N-1}{N} H^*_{it} - \frac{1}{N_j}\sum_{m\neq i} H^*_{mt}$$

which suggests an interpretation of PDO and absences as negative shocks to the H^*_{mt} of co-workers. Using our model we see that these instruments will satisfy an exogeneity restriction when

$$\mathrm{Cov}\Big(\frac{1}{N_j}\sum_{m\neq i} H^*_{mt}\ ,\ -2\phi_t\epsilon_t H^*_{it} - 2\pi_{it}H^*_{it} + {H^*_{it}}^2\Big) = 0 \tag{16}$$

which is satisfied under our stated assumptions and similarly requires independence between employee preference shocks and team-level demand shocks. Note that in contrast to our own proposed instrument, instruments constructed from PDO and absences may not satisfy a necessary relevance condition in large teams. This follows from direct application of the WLLN to the LHS of Equation 16.

1.7 Estimates Without Instrumental Variables

Strategies using logit models without instrumental variables use some form of the following specification described in [8][4]

$$logit(S_{ijt}) = \alpha_v^{logit} V_{ijt} + \alpha_w^{logit} W_{ijt} + \lambda_j + \lambda_t \tag{17}$$

[4] In practice the authors use time-invariant team-level covariates as opposed to team fixed-effects.

In this case, directly using V_{ijt} in place of V^*_{ijt} results in misspecification. Although the resultant bias has no closed form in the logit case, it can be characterized when approximated using a linear probability model using standard formulas for omitted variable bias. In this case, we show that within-firm OLS estimates for the effect of schedule quality on separations using our observed measure of schedule quality in team j can be written as

$$\hat{\alpha}_v^{LPM} = \tilde{\alpha}_v + b \qquad \textit{where} \qquad \text{sign}(b) = \text{sign}\left(\frac{N_j - 1}{N_j}\left[\frac{1}{2} - \frac{1}{N_j^2} + \phi_{jt}^2 \frac{\sigma_{\epsilon_j}}{\sigma_H}\right]\right)$$

where N_j is the number of workers employed in team j. The result implies that separation elasticities based on naively substituting V_{ijt} for V^*_{ijt} when estimating Equation 4 will yield estimates that underestimate their true values in teams of size 2 or more. Motivated by this concern [8] include a dummy variable D_{ijt} in their specifications which indicates whether or not an employee had no input into their schedule. Whilst this may reduce the size of the resultant bias it is unclear whether it is sufficient to fully remove it.

1.8 Results

1.8.1 Main Estimates

Baseline estimates of the separation function in Equation 4 using our main observed quality measure V are reported in Table 1.3. Estimates are obtained using the limited information maximum likelihood procedure described in Section 1.4. In Column (1) we report coefficient estimates which are directly interpretable in terms of the scaled preference parameters $\tilde{\alpha}_w$ and $\tilde{\alpha}_v$ and have been bias-corrected following the procedure outlined in [21]. The corresponding average marginal effects are reported in Column (2). A similar set of estimates using our alternative V^{Daily} and $V^{Predictability}$ measures are reported in Table 1.8.

We first examine the average marginal effects which have an interpretation as the causal effects of schedule quality and wages on weekly separations. The estimated effect for log quality is -0.0057 (0.0029) and is significant at the 5% level. This result indicates that a 10% increase in schedule quality is associated with a decrease in weekly separations of approximately 0.057 percentage points. The estimated effect using log wages is -0.0050 (0.0009) and is significant at the 1% level. This result indicates that a 10% increase in wages is associated with a similar decrease in weekly separations of 0.050 percentage points. Table

	Utility Coefficients	Marginal Effects
	(1)	(2)
Log quality	0.3557^{**}	-0.0057^{**}
	(0.1815)	(0.0029)
Log wage	0.3087^{***}	-0.0050^{***}
	(0.0540)	(0.0009)
Num. obs.	2123228	2123228

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Structural estimates for the effect of schedule quality on separation rates. Standard errors in parentheses are block bootstrapped at the team level and account for first-stage estimation.

Table 1.3: Structural estimates for the effect of schedule quality on separation rates

1.8 shows that these main qualitative conclusions, that wages and schedule quality both have a negative effect on the probability of worker separations, are unchanged when estimated using our alternative quality measures, although we note that the estimate for $V^{Predictability}$ is not significant at conventional levels.

We next examine the estimated utility parameters. The estimated coefficient on schedule quality has a value of 0.36 (0.18) and is statistically significant at the 5% level which indicates that $\alpha_v > 0$ such that workers do get positive utility from better quality schedules. Similarly, the reported coefficient on wages has a value of 0.31 (0.05) and is statistically significant at the 1% level which indicates that $\alpha_w > 0$ such that, unsurprisingly, utility is also increasing in wages. Comparing the ratio of coefficients allows to interpret the value to employees of improved schedule quality when measured in terms of wages. Our estimates suggest that workers would be willing to accept a reduction in wages of 1.2% in exchange for a 1% improvement in schedule quality. Table 1.8 again shows that our qualitative conclusions are robust when estimated using our alternative quality measures although we note again that the coefficient for $V^{Predictability}$ is not significant at conventional levels. These findings are consistent with the experimental evidence reported in [35] who document a strong aversion to jobs that permit employer discretion in scheduling. In their setting job applicants are willing to take a 20 percent wage cut to avoid these jobs.

In Table 1.4 we probe the robustness of our estimates to alternative parameterizations of ϕ_{jt}. Specifically, focusing on our preferred V measure, we report estimates obtained from alternative specifications in which ϕ_{jt} is allowed to depend on higher-order moments of the unconditional distribution of team-level demand. Columns (1) and (4) report results from a specification in which ϕ_{jt} is also allowed to depend on the unconditional skewness, Columns (2) and (5) report results from a specification in which ϕ_{jt} is also allowed to depend on the unconditional skewness and kurtosis, and Columns (3) and (6) report results from a fully interacted specification involving all conditional and unconditional moments. Across all of the specifications our main conclusions are unchanged relative to our baseline estimates.

	Utility Coefficients			Marginal Effects		
	(1)	(2)	(3)	(4)	(5)	(6)
Log quality	0.2711^{**}	0.2851^{**}	0.2134^{*}	-0.0046^{**}	-0.0049^{**}	-0.0036^{*}
	(0.1555)	(0.1322)	(0.1248)	(0.0025)	(0.0021)	(0.0020)
Log wage	0.2870^{***}	0.2906^{***}	0.2722^{***}	-0.0047^{***}	-0.0048^{***}	-0.0045^{***}
	(0.0485)	(0.0474)	(0.0453)	(0.0008)	(0.0008)	(0.0007)
Moments	1	2	3	1	2	3
Num. obs.	2123228	2123228	2123228	2123228	2123228	2123228

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Structural estimates for the effect of schedule quality on separation rates. Standard errors in parentheses are block bootstrapped at the team level and account for first-stage estimation.

Table 1.4: Robustness to alternative parameterizations of ϕ_{jt}

1.8.2 2SLS Estimates

Next, we compare our main estimates to an additional set of results obtained when using the two-stage least-estimator outlined in Section 1.5. We start by first assessing the relevance of using ϵ^2_{jt} as an instrument for schedule quality. Column (1) of Table 1.5 reports coefficient estimates for the first-stage relationship described in Equation 13. Since the ARCH process for customer demand implies that lagged forecasts may also be used to construct valid instruments in Columns (2) and (3) we also include specifications estimated with instruments constructed from demand forecasts at 1 and 2 week lags respectively. Column (4) reports estimates when all instruments are included. A similar set of estimates using our alternative V^{Daily} and $V^{Predictability}$ measures are reported in Table 1.9.

Focusing on the contemporaneously displayed forecasts in Columns (1) and (4) we see that estimated coefficients are negative and statistically significant at the 1% level which is consistent with both our understanding of how managers use the scheduling software tool and the team-level relationships which were documented in Table 1.2. Intuitively, our results suggest that when shown a demand forecast which indicates that demand per worker is likely to either fall below or exceed its usual level managers respond by adjusting employee hours which causes a reduction in schedule quality. Larger deviations in demand are associated with larger decreases in schedule quality. Coefficients from specifications using only lagged demand forecasts are less precisely estimated. Table 1.9 documents similar results when using our alternative quality measures. In general, we find that the magnitude of coefficients is slightly smaller for these alternative measures.

2SLS estimates corresponding to the outcome equation given in Equation 12 are reported in Table 1.6. Column (1) reports IV estimates when contemporaneously displayed forecasts are used as a single instrumental variable. Columns (2) and (3) report estimates when additional instruments constructed using lagged demand forecasts are also included. Reported coefficients are quantitatively similar across IV and 2SLS specifications. An alternative set of estimates using our V^{Daily} and $V^{Predictability}$ measures are reported in Table 1.10

We document that whilst these findings are qualitatively consistent with our main estimates we do observe some quantitative differences. For example, we document that both improved schedule quality and wages have a significant and negative effect on the probability of worker separations. Since coefficient estimates from linear probability specifications

	(1)	(2)	(3)	(4)
ϵ_t^2	-0.132^{***}			-0.172^{***}
	(0.026)			(0.023)
ϵ_{t-1}^2		-0.030		0.032
		(0.024)		(0.021)
ϵ_{t-2}^2			0.004	0.053^{***}
			(0.023)	(0.020)
Num. obs.	2123228	2123228	2123228	2123228

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Regression estimates for the effect of shocks to the variability of customer demand on schedule quality (as measured by variability in weekly hours). Standard errors in parentheses are clustered at the team level.

Table 1.5: Effect of demand shocks on schedule quality

can be interpreted directly as marginal effects our results can be compared directly to the reported effects in Table 1.3. The magnitudes of estimated effects are slightly larger when estimated using the 2SLS specification. Specifically, the reported coefficients for schedule quality and wages in our most precisely estimated 2SLS specification are -0.02 (0.008) and -0.01 (0.002) which are 4 and 2 times larger than our baseline estimates respectively.

To further probe the robustness of our results Tables 1.11 and 1.12 report coefficient estimates obtained from specifications using instruments constructed from lagged demand forecasts only. Table 1.11 reports IV estimates for our preferred V measure when forecasts from the previous week are used as a single instrumental variable. Table 1.12 reports results when additional instruments constructed using demand forecasts from higher-order lags are also included. Estimated coefficients are quantitatively and qualitatively similar to those from the specification using contemporaneously displayed demand forecasts.

1.8.3 Comparison to Fixed-Effects Estimators

Section 1.4 predicted that fixed-effects estimators which do not instrument for schedule quality may be upward biased on account of the fact that some of the observed variability in schedule quality may in fact be desirable from the perspective of the worker. To evaluate this prediction Table 1.13 reports coefficient estimates when estimating Equation 17 using

	(1)	(2)	(3)
Log quality	-0.0310^{***}	-0.0239^{***}	-0.0237^{***}
	(0.0109)	(0.0090)	(0.0084)
Log wage	-0.0127^{***}	-0.0109^{***}	-0.0109^{***}
	(0.0030)	(0.0025)	(0.0024)
Additional Lags	None	1	2
Num. obs.	2123228	2123228	2123228

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Coefficient estimates from 2SLS specifications. Standard errors in parentheses are clustered at the team level.

Table 1.6: Estimates using 2SLS for effect of schedule quality on separation rates

observed quality without instruments. Across all of our measures of schedule quality the reported marginal effects are higher as predicted by the theory. The magnitude of the differences between both sets of reported estimates is consistent with similar findings in [7] who document similar differences between their OLS and IV estimates.

1.8.4 Heterogeneity by gender

A growing literature documents noticeable heterogeneity by gender in the extent to which employees value having flexibility and control over their schedule ([35], [55], [37]). In our current setting such results suggest that the relationship between separation rates and schedule quality may also vary when estimated separately by gender. As a simple way to test this idea we examine the implications of interacting schedule quality in a 2SLS specification with a dummy variable corresponding to employee gender. Results are reported in Table 1.7. Across all of our measures of schedule quality we are unable to reject the null that gender has no effect on our reported coefficients.

Whilst this specification is parsimonious, simple interactions of schedule quality and gender implicitly impose additional restrictions on preferences. For example, a single interaction imposes the assumption that there are no differences by gender in the extent to which employees value non-quality attributes such as wages or amenities. Given this concern we also investigated alternative specifications in which we either included additional interaction terms or re-estimated our baseline specification separately for male and female sub-samples. Our qualitative conclusions were similar across each of these specifications.

The model in Section 1.3 suggests a simple mechanism which may be used to reconcile these findings with the heterogeneous effects reported in the existing literature. First, recall that our definition of schedule quality V^* uses the deviation in a worker's weekly hours from some ideal level. One possibility is that even if there is no heterogeneity in the underlying preferences for V^*, there may still be systematic differences in the distributions characterizing

	(1)	(2)	(3)
Log quality	-0.0311^{***}	-0.0382^{**}	-0.0893^{**}
	(0.0108)	(0.0149)	(0.0348)
Log quality x male	0.0002	-0.0008^{*}	0.0004
	(0.0004)	(0.0005)	(0.0002)
Log wage	-0.0127^{***}	-0.0248^{***}	-0.0003
	(0.0030)	(0.0079)	(0.0020)
Quality Measure	V	V^{Daily}	$V^{Predictability}$
Num. obs.	2123228	2123228	2123228

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Key coefficient estimates from 2SLS specifications for each of the proposed measures of schedule quality interacted with gender. Standard errors in parentheses are clustered at the team level.

Table 1.7: Heterogeneity in 2SLS estimates

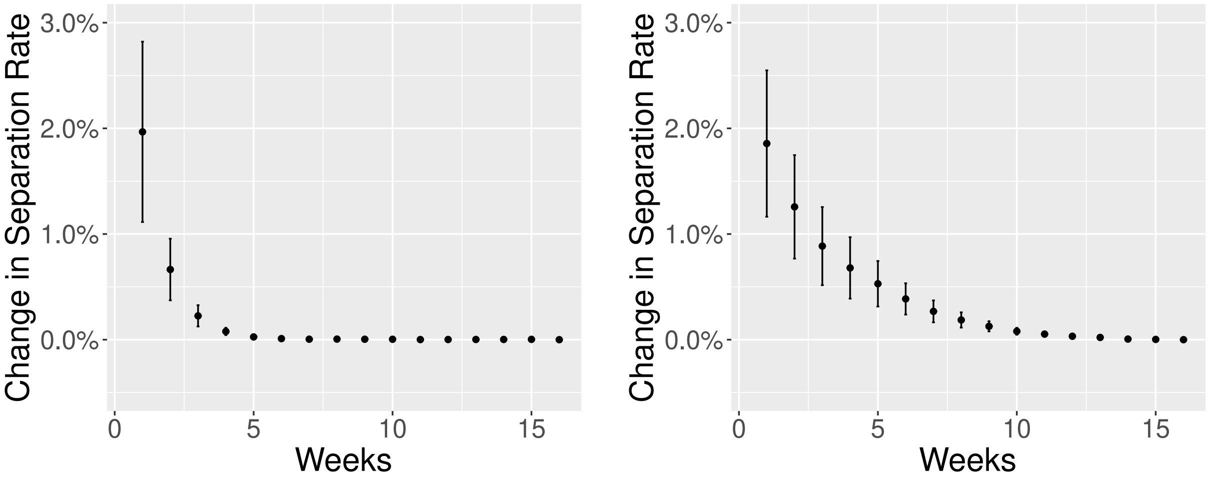
the worker-level shocks to this ideal level. When expressed in terms of our model parameters this suggests that σ_H^2 may be heterogeneous even though α_v is not.

1.9 Impulse Responses

To better contextualize the economic significance of our findings we also report impulse responses implied by the model's estimates. Specifically, we simulate the effect of a 90th percentile demand shock at time t on separation rates over the subsequent 16 weeks. Figure 1.3 estimates the impact of this demand shock to average separation rates when fixing the unconditional variance of customer demand in teams at its long-run average. In order to incorporate the effects of different amenities across teams we simulate responses for each of the observed teams in our sample and report the subsequent averages. We divide the figure into two parts to illustrate the differences which arise on account of the stochastic process which drives demand. Specifically, in the left panel we report the average effect of the shock to a firm with an ARCH coefficient of 0.33. In the right panel we report the average effect of the shock to a firm with an ARCH coefficient of 0.66. Both panels include 95% confidence intervals obtained from a block-bootstrap of the two-stage LIML procedure described in Section 1.4.

The figures reveal clear differences with regards to the extent to which the effects of demand shocks propagate forward to subsequent periods. Specifically, we see that when the ARCH coefficient is given by 0.33 the firm experiences an immediate increase in it's separation rate of almost 2%. This effect quickly dissipates and after 4 weeks separations have returned to their steady state averages. Integrating under the curve implies that the total number of separations over the 16 week period increases by 0.25%. By contrast, when the ARCH coefficient is given by 0.66 we see that separation rates remain elevated for a prolonged period. Our results indicate that persistence in the variability of demand can have significant implications for cumulative separation rates. Integrating under the curve implies that the total number of separations over the 16 week period now increases by almost 1%.

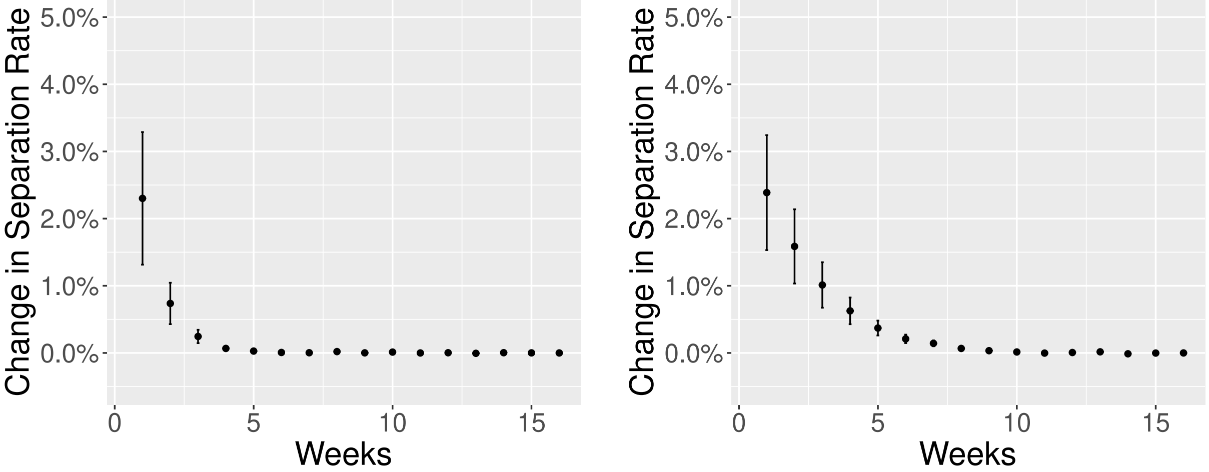
Our impulse responses can also be used to highlight the effects of non-linearities in Equation 4 on average separation rates. In Figure 1.4 we repeat the analysis but instead report the effects of a 90th percentile demand shock at time t on the median separation rate over the subsequent 16 weeks. We see that when the ARCH coefficient is given by 0.33 the dynamics are similar between the median and mean separation rates. By contrast, when the ARCH coefficient is given by 0.66 we see that the mean separation rate adjusts more slowly to its steady-state value than the median.



(a) ARCH = 0.33

(b) ARCH = 0.66

Figure 1.3: IRFs - Average Separation Rates



(a) ARCH = 0.33

(b) ARCH = 0.66

Figure 1.4: IRFs - Median Separation Rates

1.10 Conclusion

The goals of this paper were to understand the mechanisms through which scheduling policies impact the quality of worker schedules, to provide causal estimates of the effects of manager scheduling policies on employee turnover and to obtain quantitative estimates of worker valuations for unpredictable and variable schedules. To this end we developed a model of workforce scheduling at the weekly level which highlighted the role of both worker preference shocks and team-level demand shocks in explaining how the realized quality of worker schedules is determined. Using matched employer-employee panel data obtained from a global provider of workforce management software we documented meaningful heterogeneity in the quality of schedules across teams that is consistent with our model predictions.

Our preferred empirical estimates suggest that improved schedule quality has a significant and negative effect on the number of employee separations. We document that a 1% increase in schedule quality is associated with subsequent decreases in worker separation rates of 0.01 percentage points and find that workers would be willing to accept a reduction in wages of between 1.2% for a 1% improvement in schedule quality. We highlight that the economic significance of these effects over prolonged periods depends on the nature of the stochastic process for customer demand at the team-level. Our results add to a growing literature studying the importance to workers of flexibility and control over one's schedule ([35], [37], [38]).

We outline a few caveats and extensions of our study. First, when microfounding the turnover costs which are internalized by managers when making scheduling decisions our analysis assumes that workers form adaptive expectations over the future path of realized schedule quality. Although we believe this assumption to be a reasonable approximation of actual worker behavior it may be inappropriate when used to evaluate counterfactual worker responses to proposed policy changes. Second, since our model is primarily one of worker utilization it abstracts away from firm-level decisions related to optimal team size. Whilst relaxing both of these restrictions would be interesting it presents challenges for identification on account of the need to incorporate both non-linear dynamics and agent expectations thereof. Finally, in future work it will be interesting to extend the model to incorporate a distribution of teams to better understand how the cross-sectional distribution of schedule quality and wages arises in a general equilibrium context.

1.A Proofs and Derivations

1.A.1 Deriving the Separation Equation of Workers

At the start of each period t a worker i chooses between continued employment with their current team and an outside offer yielding lifetime utility O_{it}. If they choose to remain employed at j they receive their flow utility. If instead they choose to accept their outside offer they are assumed to quit immediately and do not receive their flow utility. We assume that O_{it} can be written as

$$\bar{O}_{it} = \bar{O} + \nu_{Oit}$$

where $\bar{O}$ captures the average value of outside offers and ν_{Oit} is an additional idiosyncratic shock that also follows a Markov process with independent innovations across individuals which are drawn from a logistic distribution. Additional restrictions are required if we wish to allow $\bar{O}$ to vary systematically across workers or teams, as such differences will not be separately identified from amenities.[5]

The value function for worker i at the start of period t is given by:

$$V(\xi, W_t, V_t^*, \nu_{it}) = \max\Big\{u_{it}(W_t, V_t^*, \xi) + \beta\mathbb{E}[V(\xi, W_{t+1}, V_{t+1}^*, \nu_{i,t+1})] \ , \ \bar{O} + \nu_{Oit}\Big\} \tag{18}$$

so that the optimal policy is to accept any outside offers which exceed the current flow utility and expected continuation utility from continued employment at j. Importantly, the model highlights that worker behavior depends on how they form expectations about the path of future utilization. In turn, this implies that the optimal behavior of firms will depend

[5] [6] interpret ν_{0it} as representing outside job offers which requires additional modelling for the dependence of ν_{0it} on W_t.

on their beliefs as to how workers form expectations. Our baseline analysis assumes that workers have adaptive expectations over the future path of schedule quality and wages:

$$\mathbb{E}\left[V_{t+1}^{*}|V_{t}^{*}\right] = V_{t}^{*}$$

$$\mathbb{E}\left[W_{t+1}|W_{t}\right] = W_{t}$$

To make progress let $\bar{V}(\xi, W_t, V_t^*)$ denote the expected value function defined as

$$\bar{V}(\xi, W_t, V_t^*) \equiv \int V(\xi, W_t, V_t^*, \nu_{it}) g(\nu_{it}) d\nu_{it}, \tag{19}$$

where $g(.)$ denotes the density of taste innovations. When the distribution of outside offers is sufficiently unattractive (which is implied in our current setting where weekly separation rates are c.2%) then this expected value function will be approximately linear and of the following form:

$$\bar{V}(\xi, W_t, V_t^*) \approx \frac{1}{1-\beta}\xi + \frac{\alpha_w}{1-\beta}W_t + \frac{\alpha_v}{1-\beta}V_t^* \tag{20}$$

To see this, start by guessing that the expected value function has the following form:

$$\bar{V}(\xi, W_t, V_t^*) = \gamma_0 + \gamma_1\xi + \gamma_2 W_t + \gamma_3 V_t^*$$

Combining (18) with the definition of the expected value function yields

$$
\begin{aligned}
&\gamma_0 + \gamma_1 \xi + \gamma_2 W_t + \gamma_3 V_t^* \\
&= \int \Big(\max \Big\{ u_{it}(W_t, V_t^*, \xi) + \beta \mathbb{E}[\gamma_0 + \gamma_1 \xi + \gamma_2 W_{t+1} + \gamma_3 V_{t+1}^*] \ , \ \bar{O} + \nu_{Oit} \Big\} \Big) g(\nu_{it}) d\nu_{it} \\
&= \int \Big(\max \Big\{ u_{it}(W_t, V_t^*, \xi) + \beta [\gamma_0 + \gamma_1 \xi + \gamma_2 W_t + \gamma_3 V_t^*] \ , \ \bar{O} + \nu_{Oit} \Big\} \Big) g(\nu_{it}) d\nu_{it} \\
&= \ln \Big(\exp \left(\alpha_w W_t + \alpha_v V_t^* + \xi + \beta [\gamma_0 + \gamma_1 \xi + \gamma_2 W_t + \gamma_3 V_t^*] \right) + \exp \left(\bar{O} \right) \Big)
\end{aligned}
$$

where the first line follows from the definition of the expected value function, the second line follows from the linearity of expectation and our assumption of adaptive expectations, and the final line follows from standard results about the expected maximum of iid Type I Extreme value random variables. Exponentiating both sides yields

$$
\begin{aligned}
&\exp \left(\gamma_0 + \gamma_1 \xi + \gamma_2 W_t + \gamma_3 V_t^* \right) \\
&= \exp \left(\alpha_w W_t + \alpha_v V_t^* + \xi + \beta [\gamma_0 + \gamma_1 \xi + \gamma_2 W_t + \gamma_3 V_t^*] \right) + \exp \left(\bar{O} \right) \\
&\approx \exp \left(\alpha_w W + \alpha_v V_t^* + \xi + \beta [\gamma_0 + \gamma_1 \xi + \gamma_2 W_t + \gamma_3 V_t^*] \right)
\end{aligned}
$$

where the last line holds for sufficiently unattractive values of $\bar{O}$. Matching coefficients yields $\gamma_0 = 0$, $\gamma_1 = \frac{1}{1-\beta}$, $\gamma_2 = \frac{\alpha_w}{1-\beta}$ and $\gamma_3 = \frac{\alpha_v}{1-\beta}$

1.A.2 Deriving the Shift Swap Allocation

Here we derive worker-level utilization rates under a shift-swapping technology that maximizes total worker surplus. Specifically, we assume that the shift swapping allocation in period t is equivalent to the allocation obtained from solving the following program:

$$\min_{\{H_{it}\}_{i=1}^{N}} \sum_{i=1}^{N} (H_{it} - H_{it}^{*})^{2} \quad \textit{subject to} \quad \sum_{i=1}^{N} H_{it} = \bar{H}_{t}$$

where $\bar{H}_t$ is the total number hours that must be allocated. The corresponding Lagrangian is given by

$$\mathcal{L} = \sum_{i=1}^{N} (H_{it} - H_{it}^{*})^{2} - \lambda\Big(\sum_{i=1}^{N} H_{it} - \bar{H}_{t}\Big)$$

which yields the following first-order condition which holds for any pair of workers i and m

$$H_{it} - H_{it}^{*} = H_{mt} - H_{mt}^{*}$$

Re-arranging for H_{it} and substituting into the constraint yields

$$\begin{aligned}\bar{H}_{t} &= H_{it} + \sum_{m \neq i} \Big(H_{it} - H_{it}^{*} + H_{mt}^{*}\Big) \\ &= NH_{it} - (N-1)H_{it}^{*} + \sum_{m \neq i} H_{mt}^{*} \\ &= NH_{it} - NH_{it}^{*} + \sum_{m=1}^{N} H_{mt}^{*}\end{aligned}$$

which can be re-arranged to obtain the following expression for worker i's utilization in period t:

$$\begin{aligned} H_{it} &= \frac{\bar{H}_t}{N} + H_{it}^{*} - \frac{1}{N}\sum_{m=1}^{N} H_{mt}^{*} && (21) \\ &= \frac{\bar{H}_t}{N} + \pi_{it} && (22) \end{aligned}$$

1.A.3 Proof of Proposition 1

Using Equation 6 the manager's problem can be written as

$$\max_{\{\phi,\ W\}} \ \mathbb{E}\Big[\ NA_t + N\phi_t A_t^2 - N\phi_t A_t \mu - NW_t - NW_t\phi_t(A_t - \mu) - c \cdot Q_t\ \Big]$$

where for large teams Q_t follows a Poisson distribution with rate parameter given by $\lambda = N \cdot \mathcal{P}(\xi, W_t, \phi_t^2 \epsilon_t^2)$, where $\mathcal{P}(\cdot)$ is the structural function defined in Equation 4. Applying the law of iterated expectations with the result given in Equation 4 yields

$$\begin{aligned}
\mathbb{E}[Q_t] &= \mathbb{E}\Big[\mathbb{E}\left[Q_t|\epsilon_t\right]\Big] \\
&= N\mathbb{E}\Big[\mathcal{P}(\xi, W, \phi^2\epsilon_t^2)] \\
&\approx \frac{N\exp(-\tilde{\alpha}_w W_t - \tilde{\xi})}{\sqrt{1 - 2\tilde{\alpha}_v \phi_t^2 \sigma_t^2}}
\end{aligned}$$

where the last line follows by combining (i) the approximation that for large x we have that $\exp(-x) \approx \frac{1}{1+\exp(x)}$ and (ii) the closed form expression for the moment generating function of the gamma distribution.[6]

The manager's problem can then be written as

$$\max_{\{\phi_t,\ W_t\}} \quad N_t\mu + N_t\phi_t\sigma_t^2 - N_tW_t - \frac{cN_t\exp(-\tilde{\alpha}_w W_t - \tilde{\xi})}{\sqrt{1 - 2\tilde{\alpha}_v \phi_t^2 \sigma_t^2}}$$

[6] when ϵ_t is normally distributed then ϵ_t^2 has a chi-squared distribution which is a special case of the gamma distribution.

The first-order conditions are given by

$$\frac{c\tilde{\alpha}_w \exp(-\tilde{\alpha}_w W_t - \tilde{\xi})}{\sqrt{1 - 2\tilde{\alpha}_v \phi_t^2 \sigma_t^2}} - 1 = 0 \tag{23}$$

$$\frac{2c\tilde{\alpha}_v \phi \exp(-\tilde{\alpha}_w W_t - \tilde{\xi})}{(1 - 2\tilde{\alpha}_v \phi_t^2 \sigma_t^2)^{1.5}} + 1 = 0 \tag{24}$$

which can be combined to yield the desired result.

1.A.4 Deriving the Measurement Error Expression

We use the model from Section 1.3 to show that observed quality V_{ijt} can be written as

$$V_{ijt} = V^*_{ijt} + err^v_{ijt} + o_p(1) \tag{25}$$

where the definition of observed quality corresponds to V from Section 1.2

$$V_{ijt} \equiv -(H_{ijt} - \bar{H}_{ij})^2$$

To start, we need to account for the observed heterogeneity in the average hours worked by employees within a team. To do so, we modify our parametric assumption describing the shocks to a worker's preferred hours H^*_{ijt} to allow for an employee-specific mean $\bar{H}^*_{ij}$

$$H^*_{ijt} \sim \mathcal{N}(\bar{H}^*_{ij}, \sigma^2_H)$$

We impose that $\frac{1}{N_j}\sum_{i=1}^{N_j} \bar{H}^*_{ij} = h_j$ which requires that the average preferred hours in a team equals some constant h_j. Setting $h_j = 1$ maintains consistency with our presentation of the model in Section 1.3, although our results are not dependent on any particular choice of h_j. Making the necessary adjustments to Equation 7 to account for this change yields

$$H_{ijt} = \bar{H}^*_{ij} + \phi_{jt}\epsilon_{jt} + (H^*_{ijt} - \bar{H}^*_{ij}) - \frac{1}{N_j}\sum_{m=1}^{N}(H^*_{mjt} - \bar{H}^*_{mj}) \tag{26}$$

which implies that $\bar{H}_{ij} \equiv \frac{1}{T}\sum_{t=1}^{T} H_{ijt}$ is an unbiased and consistent estimator of $\bar{H}^*_{ij}$.

Second, use the definition of V^*_{ijt} to write

$$\begin{aligned} V^*_{ijt} &= -(H_{ijt} - H^*_{ijt})^2 \\ &= -\Big(H_{ijt} - \bar{H}^*_{ij} - (H^*_{ijt} - \bar{H}^*_{ij})\Big)^2 \\ &= -\Big(H_{ijt} - \bar{H}^*_{ij}\Big)^2 + 2(H_{ijt} - \bar{H}^*_{ij})(H^*_{ijt} - \bar{H}^*_{ij}) \; - \; (H^*_{ijt} - \bar{H}^*_{ij})^2 \end{aligned}$$

Combining the above results then yields the desired result

$$\begin{aligned} V_{ijt} \; &= \; V^*_{ijt} \; - \; 2(H_{ijt} - \bar{H}^*_{ij})(H^*_{ijt} - \bar{H}^*_{ij}) \; + \; (H^*_{ijt} - \bar{H}^*_{ij})^2 \; + \; o_p(1) \\ &= V^*_{ijt} + err^v_{ijt} + o_p(1) \end{aligned}$$

1.A.5 Deriving the Orthogonality Condition

In this section we use results from Appendix 1.A.4 to show that conditional on a given team j the following exogeneity condition holds when $H^*_{ijt} \perp \epsilon_{jt}$

$$\mathrm{Cov}_j\Big(\epsilon^2_{jt}\ ,\ err^v_{ijt}\Big) = 0$$

By combining the definition of err^v_{ijt} from Appendix 1.A.4 with Equation 26 we see that the desired result is equivalent to

$$\mathrm{Cov}_j\Big(\epsilon^2_{jt}\ ,\ -2\phi_{jt}\epsilon_{jt}(H^*_{ijt} - \bar{H}^*_{ij}) - 2\pi_{ijt}(H^*_{ijt} - \bar{H}^*_{ij}) + (H^*_{ijt} - \bar{H}^*_{ij})^2\Big) = 0$$

where π_{ijt} is given by $(H^*_{ijt} - \bar{H}^*_{ij}) - \frac{1}{N_j}\sum_{m=1}^{N}(H^*_{mjt} - \bar{H}^*_{mj})$. Using linearity of the covariance operator gives

$$\begin{aligned}\mathrm{Cov}_j\Big(\epsilon^2_{jt}\ ,\ -2\phi_{jt}\epsilon_{jt}(H^*_{ijt} - \bar{H}^*_{ij}) - 2\pi_{ijt}(H^*_{ijt} - \bar{H}^*_{ij}) + (H^*_{ijt} - \bar{H}^*_{ij})^2\Big) &= \underbrace{-2\,\mathrm{Cov}_j\Big(\epsilon^2_{jt}\ ,\ \phi_{jt}\epsilon_{jt}(H^*_{ijt} - \bar{H}^*_{ij})\Big)}_{A} \\ &\quad - \underbrace{2\,\mathrm{Cov}_j\Big(\epsilon^2_{jt}\ ,\pi_{ijt}(H^*_{ijt} - \bar{H}^*_{ij})\Big)}_{B} \\ &\quad + \underbrace{\mathrm{Cov}_j\Big(\epsilon^2_{jt}\ ,(H^*_{ijt} - \bar{H}^*_{ij})^2\Big)}_{C}\end{aligned}$$

Combining the definition of π_{ijt} with our assumption that $H^*_{ijt} \perp \epsilon_{jt}$ implies that the terms B and C are equal to zero. To see that the term A is also zero observe that

$$\begin{aligned}\mathrm{Cov}_j\Big(\epsilon^2_{jt}\ ,\ \phi_{jt}\epsilon_{jt}(H^*_{it} - \bar{H}^*_{ij})\Big) &= \mathbb{E}_j[\phi_{jt}\epsilon^3_t(H^*_{it} - \bar{H}^*_{ij})] - \mathbb{E}_j[\epsilon^2_t]\mathbb{E}_j[\phi_{jt}\epsilon_t(H^*_{it} - \bar{H}^*_{ij})] \\ &= \mathbb{E}_j[\phi_{jt}\epsilon^3_{jt}]\mathbb{E}_j[(H^*_{ijt} - \bar{H}^*_{ij})] - \mathbb{E}_j[\epsilon^2_{jt}]\mathbb{E}_j[\phi_{jt}\epsilon_{jt}]\mathbb{E}_j[(H^*_{ijt} - \bar{H}^*_{ij})] \\ &= 0\end{aligned}$$

1.A.6 Characterizing Endogeneity from Measurement Error

Consider the following structural outcome equation for the effect of wages and schedule quality on separations conditional on a fixed team j

$$S_{ijt} = \alpha_0 + \alpha_w W_{ijt} - \alpha_v \Big(H_{ijt} - H^*_{ijt} \Big)^2 + \nu_{ijt}$$

where we assume that $\mathbb{E}_j[\nu_{ijt} W_{ijt}] = 0$ and $\mathbb{E}_j \left[\nu_{ijt} \left(H_{ijt} - H^*_{ijt} \right)^2 \right] = 0$. We want to characterize the endogeneity bias when estimating α_v using only the observable components of schedule quality. It is helpful to first introduce the following variable which describes deviations in the shocks to preferred hours from their means

$$\tilde{H}^*_{ijt} = H^*_{ijt} - \bar{H}^*_{ij}$$

Next, combine Equations 25 and 26 to re-write the outcome equation

$$\begin{aligned} S_{ijt} = \tilde{\alpha}_0 + \alpha_w W_{ijt} - \alpha_v & \underbrace{\left(\phi^2_{jt} \epsilon^2_{jt} + 2\phi_{jt} \epsilon_{jt} \Big(\tilde{H}^*_{ijt} - \frac{1}{N_j} \sum_{m=1}^{N_j} \tilde{H}^*_{mjt} \Big) + \Big(\tilde{H}^*_{ijt} - \frac{1}{N_j} \sum_{m=1}^{N_j} \tilde{H}^*_{mjt} \Big) \right)}_{V_{ijt}} \\ + \nu_{ijt} - & \underbrace{\alpha_v \Big(-2\phi_{jt} \epsilon_{jt} \tilde{H}^*_{ijt} + \frac{1 - N_j}{N_j} ((\tilde{H}^*_{ijt})^2 - \mathbb{E}[(\tilde{H}^*_{ijt})^2]) + \frac{1}{N_j} \sum_{m \neq i} \tilde{H}^*_{ijt} \tilde{H}^*_{mjt} \Big)}_{err^v_{ijt}} \end{aligned}$$

where $\tilde{\alpha}_0 = \alpha_0 + \frac{1-N_j}{N_j} \mathbb{E}[(\tilde{H}^*_{ijt})^2]$ is the new structural intercept. Whether or not we can recover α_v depends on the extent to which our observable data is related to the unobservables which enter into the expression for the structural errors. Standard application of the formulas for omitted variable bias imply that we have

$$\alpha_v^{OLS} = \alpha_v + \frac{\text{Cov}(V_{ijt}, err^v_{ijt})}{\text{Var}(V_{ijt})}$$

which we calculated by utilizing the independence of ϵ_{jt} and H^*_{ijt} along with the properties of moments of the normal distribution.

1.A.7 Deriving Approximation Used in Limited Information Likelihood

Here we show how to obtain the limited information maximum likelihood estimator when team sizes are large. Recall that our definition of schedule quality is given by

$$V^*_{ijt} \equiv -\left(H_{ijt} - H^*_{ijt}\right)^2$$

and that Equation 26 gives the following expression for worker-level utilization rates

$$\begin{aligned} H_{ijt} &= \bar{H}^*_{ij} + \phi_{jt}\epsilon_{jt} + \pi_{ijt} \\ \pi_{ijt} &= (H^*_{ijt} - \bar{H}^*_{ij}) - \frac{1}{N}\sum_{m \neq i}(H^*_{mjt} - \bar{H}^*_{mj}) \end{aligned}$$

where $\bar{H}^*_{ij}$ is the mean value of the shocks to worker i's preferred weekly hours. We see that for large teams $\pi_{ijt} \rightarrow_p (H^*_{ijt} - \bar{H}^*_{ij})$ which implies that $V^*_{ijt} \rightarrow_p -\phi^2_{jt}\epsilon^2_{jt}$

1.A.8 Discussion on Endogenizing Team Size

Here we discuss how the size of a team might be endogenized within the model. A natural approach is to consider a firm which solves for its optimal size when taking the manager's behavior as given:

$$\max_{N} \ \mathbb{E}_{firm}\Big[\ A_t \cdot g^*(A_t) - W_t^* \cdot g^*(A_t) - cQ_t \ \Big]$$

where $\{g(\cdot)^*, W_t^*\}$ denotes the manager's optimal choice of policy rule and wages. Note that $g(\cdot)$ and Q_t are both implicit functions of N. The firm's expectation is taken with respect to the unconditional distribution of random variables. Directly applying the same assumptions from Section 1.3 reveals that the problem is approximately linear in N so that the model needs to be adjusted to ensure that everything remains well-behaved.

A natural approach replaces the constant per-worker replacement cost c with some strictly convex function $C(q)$. Closed form solutions for the manager's revised problem are available for certain choices of $C(\cdot)$. A natural specification uses $C(q) = c_q(q^2 - q)$ where c_q is some additional cost parameter (by properties of the Poisson distribution the inclusion of a linear term ensures that expected costs depend only on q^2). In general obtaining a closed form solution for the firm's subsequent outer problem is challenging on account of the fact that the manager's optimal policies have a non-linear dependence on ϵ_{t-1}^2.

1.A.9 Discussion on Infrequent Adjustment

Suppose that firms only update their choices of W_{jt} and ϕ_{jt} at irregular intervals. As a result, our parameterization of ϕ_{jt} will also need to control for higher-order moments in the distribution of customer demand. To see this, consider the following expression which uses Equation 11 to derive expected turnover costs when $W_{jt} = W$ and $\phi_{jt} = \phi$

$$\begin{aligned} \mathbb{E}\Big[c \cdot Q_t\Big] &= \mathbb{E}\Big[\mathbb{E}[c \cdot Q_t|\epsilon_t]\Big] \\ &= \mathbb{E}\Big[c \cdot \mathcal{P}(\xi, W, \phi^2\epsilon_t^2)\Big] \end{aligned}$$

A second-order Taylor approximation then yields

$$\mathbb{E}\Big[c \cdot Q_t\Big] \approx c \cdot \mathcal{P}(\xi, W, \phi^2\sigma_\epsilon^2) + \frac{c \cdot \mathcal{P}''(\xi, W, \phi^2\sigma_\epsilon^2)}{2} \underbrace{\mathbb{E}\Big[\epsilon_t^4 - 2\sigma^2\epsilon_t^2 + \sigma^4\Big]}_{K}$$

which highlights the dependence on expectations of higher-order moments of ϵ_t.

To see why these higher-order moments don't appear in our baseline model, we can use the normality assumption for the standardized residual in Equation 2 to derive

$$\mathbb{E}_{t-1}\Big[\epsilon_t^4 - 2\sigma_t^2\epsilon_t^2 + \sigma_t^4\Big] = 2\sigma_t^4$$

from which it follows that when choosing wages and pass-throughs each period the optimal policy will only depend on the conditional variance. However, when making longer-run decisions the manager will need to account for higher-order moments.

1.B Additional Tables and Figures

	Utility Coefficients		Marginal Effects	
	(1)	(2)	(3)	(4)
Log quality	1.5121***	0.2932	−0.0245***	−0.0050
	(0.4893)	(1.1263)	(0.0079)	(0.0183)
Log wage	1.0001***	0.2024***	−0.0162***	−0.0033***
	(0.2538)	(0.0629)	(0.0041)	(0.0010)
Quality Measure	V^{Daily}	$V^{Predictability}$	V^{Daily}	$V^{Predictability}$
Num. obs.	2123228	2123228	2123228	2123228

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Structural estimates for the effect of schedule quality on separation rates. Standard errors in parentheses are block bootstrapped at the team level and account for first-stage estimation.

Table 1.8: Structural estimates for the effect of schedule quality on separation rates

	(1)	(2)	(3)	(4)
ϵ^2_{jt}	−0.107***	−0.099***	−0.046***	−0.048***
	(0.027)	(0.023)	(0.011)	(0.009)
$\epsilon^2_{j,t-1}$		−0.000		0.008
		(0.022)		(0.008)
$\epsilon^2_{j,t-2}$		−0.018		−0.006
		(0.022)		(0.008)
Quality Measure	V^{Daily}	V^{Daily}	$V^{Predictability}$	$V^{Predictability}$
Num. obs.	2123228	2123228	2123228	2123228

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Regression estimates for the effect of shocks to the variability of customer demand on schedule quality. Standard errors in parentheses are clustered at the team level.

Table 1.9: Effect of demand shocks on schedule quality

	(1)	(2)	(3)	(4)
Log quality	-0.0384^{***}	-0.0370^{**}	-0.0894^{**}	-0.0818^{**}
	(0.0149)	(0.0148)	(0.0349)	(0.0328)
Log wage	-0.0247^{***}	-0.0239^{***}	-0.0002	-0.0006
	(0.0079)	(0.0078)	(0.0020)	(0.0019)
Quality Measure	V^{Daily}	V^{Daily}	$V^{Predictability}$	$V^{Predictability}$
Additional Lags	No	Yes	No	Yes
Num. obs.	2123228	2123228	2123228	2123228

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Key coefficient estimates from 2SLS specifications for each of the proposed measures of schedule quality. Standard errors in parentheses are clustered at the team level.

Table 1.10: Estimates using 2SLS with alternative quality measures

	(1)	(2)	(3)
Log quality	-0.1103	-0.0537^{*}	-0.1632
	(0.0968)	(0.0292)	(0.1025)
Log wage	-0.0331	-0.0326^{**}	0.0036
	(0.0250)	(0.0153)	(0.0054)
Quality Measure	V	V^{Daily}	$V^{Predictability}$
Additional Lags	No	No	No
Num. obs.	2123228	2123228	2123228

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Key coefficient estimates from 2SLS specifications for each of the proposed measures of schedule quality. Specification uses single lagged instrument only. Standard errors in parentheses are clustered at the team level.

Table 1.11: Estimates using 2SLS with only lagged instruments

	(2)	(4)	(6)
Log quality	-0.0775^{*}	-0.0411^{*}	-0.1120
	(0.0461)	(0.0238)	(0.0728)
Log wage	-0.0246^{**}	-0.0261^{**}	0.0010
	(0.0120)	(0.0125)	(0.0038)
Quality Measure	V	V^{Daily}	$V^{Predictability}$
Additional Lags	Yes	Yes	Yes
Num. obs.	2123228	2123228	2123228

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Key coefficient estimates from 2SLS specifications for each of the proposed measures of schedule quality. Specification uses multiple lagged instruments. Standard errors in parentheses are clustered at the team level.

Table 1.12: Estimates using 2SLS with only lagged instruments

	(1)	(2)	(3)
Log quality	0.0003^{***}	-0.0001	0.0005^{***}
	(0.0000)	(0.0002)	(0.0000)
Log wage	-0.0047^{***}	-0.0048^{***}	-0.0045^{***}
	(0.0007)	(0.0007)	(0.0007)
Quality Measure	V	V^{Daily}	$V^{Predictability}$
Num. obs.	2123228	2123228	2123228

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Key coefficient estimates from OLS specifications for each of the proposed measures of schedule quality. Standard errors in parentheses are clustered at the team level.

Table 1.13: Estimates using OLS for effect of schedule quality on separation rates

CHAPTER 2

HETEROGENEITY OF CONSUMPTION RESPONSES TO INCOME SHOCKS IN THE PRESENCE OF NONLINEAR PERSISTENCE

2.1 Introduction

The empirical analysis of consumption and income dynamics has an important place in a number of key areas of economic research and policy design. A large literature aims at understanding income persistence, income inequality and income volatility, see [56], [57] and references in [58]. A parallel literature studies how income shocks impact consumption and savings decisions, see [59] and [60] among many other references. In this paper our goal is to empirically document the nature of consumption responses, with a particular focus on household heterogeneity and nonlinear persistence.

Economic models inform the empirical analysis of consumption and income. In a standard incomplete markets model of the life cycle, how much a household consumes in a given period is determined by the level of assets, the stage of the life cycle, as well as the income stream, see [61] for a comprehensive review. Changes to income components with different degrees of persistence lead to different consumption responses. In addition, the shape of the consumption function may differ among households for a variety of reasons, such as heterogeneity in preferences or discounting, household-specific returns to assets, or heterogeneous access to other sources of insurance.

Our starting point is the nonlinear panel data framework proposed by [1] (ABB hereafter) which involves a Markovian permanent-transitory model of income, and a flexible age-dependent nonlinear consumption rule that is a function of assets, permanent income and transitory income. ABB found that individual income dynamics feature nonlinearities that matter for economic decisions. Specifically, they found evidence that the persistence of past earnings varies substantially with the sign and magnitude of shocks across the past earnings distribution. Thus, ex ante identical individuals may have experienced a very different propagation of a past shock into their income depending on their history of subsequent shocks. Using a balanced panel from the PSID, from 1999 to 2009, ABB showed how nonlinear income dynamics lead to nonlinear responses of consumption to income shocks.[7]

Given this background we make three main contributions. First, we exploit the important extension to the set of consumption goods in the recent waves of the PSID to produce new estimates of the degree of nonlinear persistence and consumption insurance. The improved panel survey redesign in the 1999 PSID was further enhanced in 2005 and, in addition to food at home and food away from home, includes health expenditures, utilities, gasoline, car maintenance, transportation, education, clothing, and leisure activities, see [64]. We bring this together with the detailed data on earnings, family income, and financial and real estate assets. Using the 2005-2017 PSID panel survey waves, we estimate the nonlinear nature of income shocks and the consumption implications of the insurance to income shocks. In addition, unlike ABB we do not restrict the sample to be balanced. This leads us to consider a larger and more comprehensive sample, more than 2000 households compared to approximately 800 in ABB.

Our second main contribution is to empirically document household heterogeneity in consumption responses. To do so, we move away from the partial insurance consumption growth framework of [60] and estimate a dynamic model where we specify the entire conditional distribution of consumption given assets, age, and the income components. This modeling approach contrasts with that adopted in ABB, who specified the link between consumption and its determinants using a nonlinear mean model with separable heterogeneity. Allowing for non-separabilities, we show how to estimate the joint distribution of latent and observed variables, and to consistently estimate log-derivatives of the consumption function as a result.[8]

[7] See [62] and [63] for recent applications of the nonlinear dynamic approach introduced in ABB.

[8] As we will explain below, our approach exploits the weak exogeneity of the observed state variables (i.e., assets and income components), conditional on a latent time-invariant type, to identify average response functions, see [65] for a review of identification results in models with non-separable heterogeneity. Relaxing exogeneity would require valid instruments and appropriate structure on the first stage ([66]). Also, while

The average log-derivatives of the consumption function that we focus on are nonlinear coefficients quantifying how well insured households are, at different points of the life cycle and depending on their level of assets. Importantly, we model the consumption function as heterogeneous across households, by indexing consumption on a latent time-invariant continuous type. This unobserved consumer type may reflect heterogeneity in economic primitives, and leads to different consumption derivative responses for two households that are at the same point of the life cycle, face the same income stream, and own the same level of assets. We show this heterogeneity to be a salient feature of the PSID.

To study a larger sample using a more complex model, we modify the computational techniques that ABB relied on. The use of new computational tools represents our third main contribution. Specifically, we examine improved sequential computational methods for the estimation of the nonlinear latent/hidden quantile Markov model. The Markovian structure for latent earnings components allows us to make use of Sequential Monte-Carlo (SMC) methods to improve the Markov Chain Monte Carlo algorithm, see [68] for a review. SMC methods can be used to generate efficient proposals within a Particle Markov Chain Monte Carlo (PMCMC) algorithm, as proposed by [69]. We develop an implementation in the latent Markov setting of this paper. The PMCMC approach allows us to produce numerically robust estimates of derivatives of log-consumption with respect to the latent income components, in a nonlinear quantile model that allows for unobserved types.

Empirically, we confirm the nonlinear income dynamics found in ABB while documenting new patterns in consumption responses. The estimated quantile Markovian permanent-transitory model of income reveals asymmetric persistence of earnings and income shocks. We show the use of enhanced computational techniques leads to essentially the same results

the distribution of consumption responses is generally not identified beyond its mean, partial information about this distribution can be obtained by using a result from [67]. We will apply this strategy to compute a lower bound on the variance of responses.

as ABB in their balanced sample. However, estimates based on SMC techniques are more stable numerically. The use of sequential Monte Carlo methods allows us to draw robust conclusions in our larger unbalanced sample, and to document nonlinear patterns in the dynamics of income.

Our main results concern the nature of consumption responses to income shocks. We find that older and wealthier households adjust their consumption less as a response to an income shock than younger and less wealthy households. For our main sample of dual earners the average derivative of log-consumption to the persistent income component is 0.33 on overall average, yet it can be much higher for younger households with low levels of assets and, conversely, as low as 0.10 for older and wealthy households. These findings are qualitatively consistent with the implications of standard life-cycle models of consumption and saving behavior. We show that accounting for latent income components with varying degrees of persistence, and for unobserved heterogeneity in consumption, are both important to accurately document these patterns quantitatively. Heterogeneity in consumer responses to income shocks matters for understanding the impact not only of fiscal policies but also of monetary policies which, as [70] notes, can create large redistribution in favor of high MPC agents and be expansionary over and beyond the effect on real interest rates.

Our key finding is that consumption responses vary substantially with unobserved types. Our results clearly separate lower consumption types, who appear to follow the life-cycle patterns in consumption responses implied by standard models, from higher types, whose consumption responses to income shocks vary little with either assets levels or the stage of the life cycle. High-type households consistently have higher consumption levels, and relative to low-type households they have slightly higher incomes and levels of assets. For the younger low types, consumption responses to persistent income shocks exceed 0.50 while for older low types this falls below 0.20. Moreover, based on bootstrapped confidence intervals we conclude the difference between the two coefficients is significant at conventional levels. For the higher types, consumption responses are flatter across age and assets, and differences across age and assets are insignificant. These findings shed new light on the presence of heterogeneity in consumption behavior across households, on which there has been extensive micro- and macroeconomic research, see [71], [72], and references therein.

We examine several mechanisms that could lead to such heterogeneous consumption responses. First, the fact that high types consume more and hold more assets is difficult to reconcile with an explanation based on heterogeneity in preferences or discounting. Second, we estimate a specification that allows for latent heterogeneity in asset accumulation and find that the heterogeneity in consumption responses is virtually unaffected. Lastly, we link a subset of household heads in our sample (33%) to their parents, using the intergenerational linkages that the PSID provides. We find that high-type household heads have on average parents with higher consumption and income levels, suggesting that the heterogeneous responses that we find might in part reflect heterogeneity in access to other sources of insurance such as parental insurance.

We show the main results are robust to a number of specification changes. In particular, while we use disposable income in most of the analysis, we find similar patterns when using pre-tax labor income, with some quantitative differences. In addition, we find that including households where one member may not be working does not lead to major changes in our results. Lastly, we probe the robustness of our scalar individual effect modeling approach by allowing for a separate effect of education on consumption responses, in addition to the latent type. While the heterogeneity results remain qualitatively similar, the findings based on this specification allow us to discuss some limitations of our scalar individual effect modeling approach and to motivate future work.

The outline of the paper is as follows. In Section 2.2 we describe the sample and present motivating evidence on the nature of consumption responses. In Section 2.3 we provide a general description of the model, and in Section 2.4 we discuss implementation and present the computational methods we use. We then show our main empirical results in Section 2.5. In Section 2.6 we study possible mechanisms for those results. In Section 2.7 we show results based on extensions of our main model. We conclude in Section 2.8. An appendix describes implementation and provides additional empirical results.

2.2 Data

In this section we describe the PSID sample, and we provide preliminary motivating evidence about how consumption responds to income changes.

2.2.1 The PSID sample

We rely on the newly redesigned PSID, from 2005 to 2017. Since 1999, the PSID presents a unique combination of longitudinal data on income, consumption, and assets holdings for the US. Unlike the annual information available every year before 1997, after 1999 a new wave is only available every other year. Since 2005, the consumption information has been enhanced, with additional categories, see [73]. The recent waves include food at home and away from home, gasoline, health, transportation, utilities, clothing, and leisure activities. [64] provide a detailed analysis of the post-2005 data and assess the new methodology developed by the PSID for collecting household expenditure data. The new survey methodology allows unfolding brackets as well as choice of time-frame for different consumption categories. They show that since 2005 the PSID has captured almost all expenditures measured in the cross-sectional Consumer Expenditure Survey (CE) and suggest the new measurement design is likely to improve on the accuracy of the expenditure data. For this reason, we expect the post-2005 PSID to provide more accurate information about household consumption patterns than the earlier period used in ABB.

Another difference with ABB is that we do not restrict the panel to be balanced. Following [74], we focus on a sample of household heads that participate in the labor market and are between 25 and 60 years old. Since we do not model labor supply, either at the extensive or intensive margin, in our baseline sample we focus on households where both adult members are working and present in at least two waves, and we keep their first spell of non-zero income observations. We refer to this baseline as the "dual earners" sample. However, in Section 2.7 we will also present results based on a broader sample that includes households where only one member is employed.

Our final dual earners sample contains 2,113 households and seven biennial waves from 2005-2017. In Table 2.1 we report some descriptive statistics about this sample. Food consumption, which was the only consumption item available in the PSID prior to the redesign of the data set, accounts for approximately one fourth of total non-durables consumption. Net disposable income is approximately 30% lower than pre-tax labor income. Since it is disposable income and not pre-tax income that should affect consumption decisions, we will focus on disposable income in most of the analysis. In Section 2.7 we will also present results using pre-tax labor income.

Table 2.1 also shows that total wealth tends to decrease around the 2008 recession, whereas income and especially consumption seem more stable over the period. See [75] for an analysis of consumption, income and wealth using the PSID with a focus on the great recession. In our analysis we will not focus on business cycle fluctuations, and we will attempt to remove calendar time effects in a prior partialling-out estimation step.

In Appendix Table 2.4 we show additional statistics in order to describe the unbalanced structure of the panel sample. In the first column of that table we report statistics for households who are only observed for one wave, although we do not include these households in our main sample due to our focus on unobserved heterogeneity. More than half of households in our main sample are observed for at most three waves. For this reason, it will be important to account for the unbalancedness of the PSID in the modeling of income and consumption dynamics.

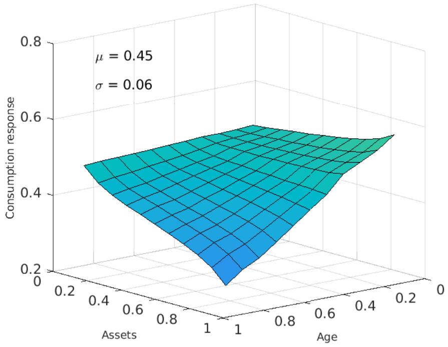
Table 2.1: Descriptive Statistics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	2005	2007	2009	2011	2013	2015	2017
Food	10,681.46	10,652.44	10,356.33	10,516.91	10,778.89	11,287.65	11,916.79
	(5,280.66)	(5,497.57)	(5,035.15)	(5,107.21)	(5,744.91)	(5,385.16)	(5,673.31)
Non-durables (excl. food)	28,476.06	29,563.67	28,264.68	28,694.76	30,310.30	29,906.71	28,432.69
	(19,445.13)	(19,881.54)	(19,295.93)	(18,331.37)	(18,247.37)	(17,265.61)	(14,547.69)
Total Non-durables	39,179.31	40,233.90	38,669.21	39,265.89	41,129.95	41,246.63	40,383.30
	(22,220.87)	(22,516.17)	(21,678.39)	(21,154.18)	(20,962.80)	(19,845.41)	(17,547.23)
Home equity	161560.91	169580.40	137089.26	121021.37	111956.54	113269.94	130350.80
	(216942.00)	(229763.44)	(197997.93)	(166538.89)	(154874.43)	(143419.48)	(144146.96)
Negative Equity Dummy	0.01	0.01	0.03	0.03	0.02	0.01	0.01
	(0.08)	(0.10)	(0.16)	(0.16)	(0.15)	(0.09)	(0.10)
Wealth (excl. home)	206679.75	278971.16	269420.39	247951.44	231130.23	256813.63	333757.83
	(709285.07)	(1.00e+06)	(933414.69)	(536086.47)	(516957.59)	(566105.75)	(1.06e+06)
Total wealth	446917.54	512678.86	448989.83	388763.07	349033.92	370083.56	448654.75
	(970857.51)	(1.25e+06)	(1.14e+06)	(656915.67)	(621844.77)	(636801.00)	(1.07e+06)
Labor income	126181.76	127847.66	133105.34	129458.55	128366.66	124779.30	131051.39
	(143916.08)	(148500.93)	(194142.24)	(129247.51)	(128479.97)	(72,585.03)	(69,355.95)
Net income	95,598.70	97,089.32	100204.10	99,234.77	98,238.57	95,004.23	99,192.91
	(86,212.45)	(89,857.83)	(116281.39)	(78,750.29)	(77,931.32)	(46,552.59)	(45,252.48)
Observations	1288	1544	1400	1149	1023	948	755

Notes: PSID, 2005-2017. Means of variables, standard deviations in parentheses. Our baseline measure of consumption includes the following categories: food at home, food delivery, eating out, food stamps, clothing, gasoline, utilities, telephone bills, automobile insurance, parking, transport, education, childcare, institutional medical services, doctor services, prescriptions, health insurance, and trips and other recreation.

Following a common practice in the previous literature on income dynamics, we will work with residuals of log-disposable income on a set of demographics and time indicators. This partialling-out is meant to make household demographics as comparable to each other as possible, and to control for aggregate time effects. Specifically, we net out household size, year of birth, state indicators, number of kids, race of both adults, a higher education indicator for both adults interacted with age indicators, and a full set of age indicators interacted with year indicators. We similarly construct residuals of log-consumption and log-assets net of the

Figure 2.1: Average derivative of log-consumption with respect to log-income



Notes: The graph shows averages of the derivative of log-consumption with respect to log-income, conditional on log-income, age and log-assets. Estimates are based on a linear regression of log-consumption on a second-order polynomial in log-income, age, and log-assets. The two horizontal axes show age and assets percentiles. μ and σ denote the mean and standard deviation of the average derivatives, respectively.

same set of controls. Working with logarithms requires removing observations with zero or negative assets, which reduces the number of observations by approximately 200 households per year. In Appendix Table 2.5 we report additional statistics for a sample which includes households with negative asset balances.

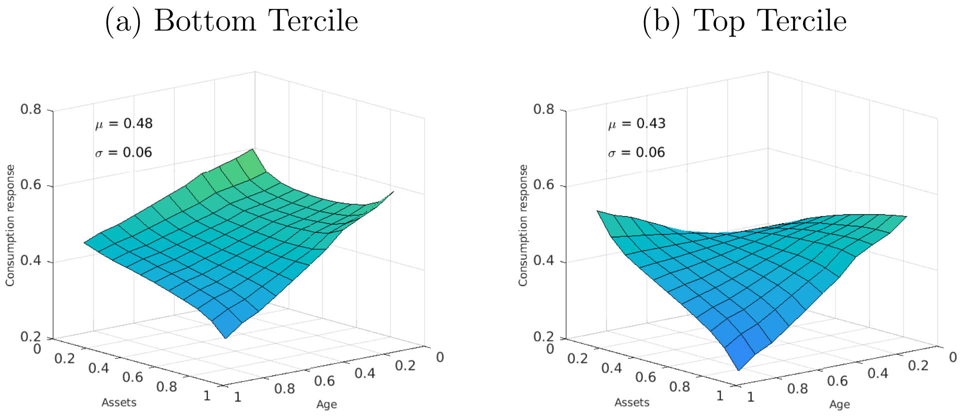
2.2.2 A first look at consumption responses

We will analyze the PSID sample using a dynamic model of income, consumption, and assets holdings. The model is flexibly parameterized and it features various latent variables. Before describing how we specify the model and estimate it, here we provide preliminary motivating evidence about consumption and income, only using observed covariates and simple econometric methods. We highlight two features of the data in turn.

In Figure 2.1, we show average derivatives of log-consumption with respect to log-income, controlling for age and log-assets.[9] The derivative effect is 0.45 on average, with a standard

[9] Here and in the following we simply refer to log-income residuals in a regression on demographics and time indicators as "log-income", and we similarly refer to log-consumption residuals and log-assets residuals

Figure 2.2: Quantile derivatives of log-consumption with respect to log-income



(a) Bottom Tercile

(b) Top Tercile

Notes: The graphs show averages of the derivatives of quantile functions of log-consumption with respect to log-income, conditional on income, age and assets. In the left graph we report results for the bottom tercile (averaged over a fine grid of percentiles), in the right graph we report results for the top tercile. Estimates are based on quantile regressions of log-consumption on a second-order polynomial in log-income, age, and log-assets. The two horizontal axes show age and assets percentiles.

deviation of 0.07. In particular, wealthier and older households have a lower derivative (i.e., lower than 0.30), suggesting that they are relatively well insured against income shocks. In contrast, younger and less wealthy households have a higher derivative (i.e., higher than 0.50), suggesting less ability to insure.

In Figure 2.2, we show quantile derivatives of log-consumption with respect to log-income. In the left graph, we average quantile derivatives over the bottom tercile, while in the right graph we report an average over the top tercile. We see that these quantile derivative coefficients tend to be somewhat higher at the bottom of the consumption distribution (0.48 on average) than at the top (0.43 on average). The main difference between the two graphs concerns the younger and less wealthy households, for whom the derivative drops from 0.60 to 0.40 when moving from the bottom tercile to the top tercile.[10]

as "log-consumption" and "log-assets", respectively.

[10] In Appendix Figure 2.10 we show bootstrapped confidence bands corresponding to both Figures 2.1 and 2.2.

This evidence is suggestive of the presence of heterogeneity in consumption responses and insurance. However, there are several reasons why it may be incomplete and quantitatively inaccurate. Standard consumption models imply that income components with varying degrees of persistence have a different impact on consumption. Hence, while in Figure 2.1 we report derivatives with respect to observed income, in a model where log-income is the sum of a persistent and a transitory component, economically-relevant consumption derivatives should be computed with respect to the latent components of income. To do so, a dynamic model with latent variables is needed. The heterogeneity suggested by Figure 2.2 is similarly ambiguous. Indeed, consumption quantiles are likely to reflect a combination of time-invariant household heterogeneity and time-varying shocks. Distinguishing the two requires estimating a dynamic panel data model that features latent heterogeneity explicitly. In the next two sections we describe such a model, and we explain how we estimate it using the PSID.

2.3 Overview of the model

2.3.1 Consumption behaviour

Our primary interest is to understand how shocks to income translate into consumption for different types of consumers. Consumers are allowed to differ along a number of dimensions, specifically according to their assets, the stage in their life cycle, observable characteristics, and unobserved heterogeneity. Our underlying framework is one where households act as single agents with access to a single risk-free asset. They receive income shocks each period and make consumption decisions subject to a period-to-period budget constraint. We assume all distributions are known to households, and there is no aggregate uncertainty.

In modeling the dynamic responses of consumption to earnings shocks, one strategy is to specify the functional form of the utility function and the distributions of the shocks, and to calibrate or estimate the model's parameters by comparing the model's predictions with the data, see [76] and references therein. Another strategy is to follow the partial insurance approach of [60] and linearize the Euler equation, with the help of the budget constraint. The approach we follow in this paper builds on the framework introduced in ABB. It differs from the earlier strategies as we directly estimate the consumption rule that comes from the optimization problem. In this approach the level of consumption is modeled as a function of beginning of period assets, income components, consumer characteristics and individual heterogeneity. The framework we develop here is a generalization of the main specification in ABB to allow for individual unobserved heterogeneity and a more flexible policy rule. The shape of the consumption function and its derivatives will depend on the distributions of beliefs about future incomes and characteristics. We are therefore able to document a rich set of derivative effects but, as our model does not separate the role of preferences from expectations, we cannot recover counterfactuals that involve a change in the income process.

In our approach, the income process is modeled using the framework of ABB which allows for nonlinear persistence. In this framework, log-income is decomposed into a predetermined life-cycle component and two latent stochastic factors that represent the level of persistent income and the level of transitory income. We consider an unbalanced panel of households, $i = 1, ..., N$, in which household i is observed T_i consecutive time periods. For any household i at time t we denote the persistent income component as η_{it} and assume it follows a nonlinear first-order Markov process. The transitory income component ε_{it} is assumed to be distributed independently across time and independent of the $\eta's$. Log-income residuals are then $y_{it} = \eta_{it} + \varepsilon_{it}$. The details of the income specification are developed in the next subsection.

Given beginning-of-period-t assets a_{it}, and the realizations of the persistent and transitory income components η_{it} and ε_{it}, consumers make their consumption choices according to the policy rule

$$c_{it} = g_t \left(a_{it}, \eta_{it}, \varepsilon_{it}, age_{it}, \xi_i, \nu_{it} \right), \quad i = 1, ..., N, \quad t = t_i, ..., t_i + T_i - 1, \tag{27}$$

where t_i denotes the period when i enters the panel, c_{it} is log-consumption for household i in period t, a_{it} is log-assets, age_{it} is the age of the household head in period t, and unobserved heterogeneity is given by the "fixed effect" ξ_i.[11] As mentioned above, both c_{it} and a_{it} are net of common effects of age and other demographics, and of time indicators. We also allow consumption choices to depend on transitory preference shocks ν_{it}, with arbitrary dimension.

Our main goal is to estimate the empirical consumption response parameters

$$\phi(age_{it}, a_{it}, \eta_{it}, \varepsilon_{it}, \xi_i) = \mathbb{E}_{\nu_{it}} \left[\frac{\partial g_t \left(a_{it}, \eta_{it}, \varepsilon_{it}, age_{it}, \xi_i, \nu_{it} \right)}{\partial \eta} \right]. \tag{28}$$

Average derivative effects such as (28) can be identified without restricting the dimensionality of ν_{it}, see [65] and references therein.[12] Reporting features of estimates of the individual transmission parameters

$$\phi_{it} = \phi(age_{it}, a_{it}, \eta_{it}, \varepsilon_{it}, \xi_i)$$

in the PSID will shed light on how much variation there is in consumption responses and insurance, over the life cycle and as a function of assets and income. Importantly, the dependence of the consumption function on the latent type ξ_i will allow us to document individual heterogeneity in consumption responses. Exploring the relationship between ϕ_{it}

[11] Below we will postulate that ξ_i follows a certain distribution (albeit a rather flexible one) conditional on cohort, education and income. An alternative description of ξ_i would thus be as a "correlated random effect".

[12] However, in our setting, some of the arguments of the structural function g_t (i.e., η_{it}, ε_{it}, and ξ_i) are latent. Identification of average derivatives thus requires showing that the distribution of $(c_{it}, a_{it}, \eta_{it}, \varepsilon_{it}, age_{it}, \xi_i)$ is identified.

and ξ_i is a main objective of this paper.

In order to estimate the consumption function g_t in (27), one needs to recover the persistent and transitory income components η_{it} and ε_{it}, and the time-invariant consumption type ξ_i, all of which are unobserved to the econometrician. For this purpose, we will estimate a dynamic model of income and consumption with latent variables, following ABB.

Asset accumulation. Estimation of the consumption function g_t requires taking a stand on the accumulation of assets. A simple case is when current assets only depend on lagged assets, income, and consumption, but not on the latent income components and heterogeneity separately. This would hold in a textbook asset accumulation rule with a constant risk-free interest rate, for example. Under the assumption that asset accumulation does not depend on the latent variables, one can estimate the consumption function consistently without having to model the assets process, in the spirit of partial likelihood estimation. We will use this approach in our main results. More generally, our approach can allow the latent income components and type heterogeneity to affect current assets, and we will report results based on such a specification as well, see Subsection 2.6.3.

Dispersion of consumption derivatives. Lastly, while we focus on recovering the average response parameters ϕ_{it}, the distribution of the consumption derivatives

$$\frac{\partial c_{it}}{\partial \eta} = \frac{\partial g_t\left(a_{it}, \eta_{it}, \varepsilon_{it}, age_{it}, \xi_i, \nu_{it}\right)}{\partial \eta},$$

conditional on $(a_{it}, \eta_{it}, \varepsilon_{it}, age_{it}, \xi_i)$, is generally not identified unless ν_{it} is scalar and has a monotone effect on g_t. Yet, using an insight from [67], one can compute a lower bound on the variance of the consumption derivatives $\frac{\partial c_{it}}{\partial \eta}$, even though the variance itself is not identified. We make this point formally in Appendix 2.D, and we will report empirical estimates of bounds on variances as a complement to our main average coefficients.

2.3.2 Income and consumption

Our modeling of the income process closely follows ABB, with the main difference that we extend the model to an unbalanced panel. Specifically, let y_{it} be the log-disposable income of household i in year t, net of common effects of age and other demographics, and time indicators. We specify the following persistent-transitory model

$$y_{it} = \eta_{it} + \varepsilon_{it}, \qquad i = 1, ..., N, \; t = t_i, ..., t_i + T_i - 1, \tag{29}$$

where the persistent and transitory components η_{it} and ε_{it}, respectively, are zero-mean continuous latent variables given age.

We model the processes η_{it} and ε_{it} using their quantile representations. Let $Q_A(B, v)$ be a generic notation for the conditional quantile of A given B, evaluated at the percentile v in the unit interval. The quantile representation of A given B implies that $A = Q_A(B, V)$, where V is standard uniform independent of B.[13]

The persistent income component η_{it} follows a nonlinear first-order Markov process with age-specific transitions; that is,[14]

$$\eta_{it} = Q_\eta(\eta_{i,t-1}, age_{it}, u^\eta_{it}), \qquad (u^\eta_{it} \mid \eta_{i,t-1}, age_{it}) \sim iid\, \text{Uniform}\,(0,1)\,, \qquad t > t_i. \tag{30}$$

In order to model entry in the panel, we let the initial persistent latent component η_{i,t_i} depend on years of education and birth cohort of the household head, and on age at entry

[13] For example, $Q_A(B, 0.50)$ is the conditional median of A given B, and $Q_A(B, 0.90)$ is the conditional 90th percentile of A given B. The fact that $A = Q_A(B, V)$, where V is standard uniform independent of B, is referred to as the Skorohod representation in the literature, see, e.g., [77].

[14] In our sequential model, we assume that $u^\eta_{it} \mid \eta^{t-1}_i, age^t_i$ is standard uniform, where η^{t-1}_i and age^t_i denote sequences of lags of η and age. For conciseness we leave the full conditioning implicit in the notation.

in the sample:

$$\eta_{i,t_i} = Q_{\eta_1}(cohort_i, educ_i, age_{i,t_i}, u_i^{\eta_1}), \quad (u_i^{\eta_1} \mid cohort_i, educ_i, age_{i,t_i}) \sim iid\, \text{Uniform}\,(0,1)\,. \tag{31}$$

In turn, the transitory component ε_{it} is assumed to be independent over time and independent of η_{is} for all s with an age-specific distribution,

$$\varepsilon_{it} = Q_{\varepsilon}(age_{it}, u_{it}^{\varepsilon}), \quad (u_{it}^{\varepsilon} \mid age_{it}) \sim iid\, \text{Uniform}\,(0,1)\,. \tag{32}$$

Note that the income process is common across households. In this paper we do not attempt to model latent time-invariant heterogeneity in the income process beyond heterogeneity in initial conditions. However, we allow for an unobserved type that affects consumption and may be correlated with income.

Turning to consumption, we let the unobserved heterogeneity variable ξ_i be correlated with birth cohort, education, and income; that is, we specify

$$\xi_i = Q_{\xi}(cohort_i, educ_i, income_i, u_i^{\xi}), \quad \left(u_i^{\xi} \mid cohort_i, educ_i, income_i\right) \sim iid\, \text{Uniform}\,(0,1)\,. \tag{33}$$

Here $income_i$ is a measure of the household's "normal" income. In our baseline specification we will take $income_i$ to be the average log-income over the period of observation. In addition, note that the age at entry in the panel does not affect ξ_i given cohort, education, and income. Hence, ξ_i is a time-invariant household characteristic that does not depend on when the household starts being recorded in the PSID, whereas the value of the initial persistent latent component in (31) depends on the stage of the life cycle the household was at when she entered the panel.

We then specify the log-consumption function as

$$c_{it} = Q_c(a_{it}, \eta_{it}, \varepsilon_{it}, age_{it}, \xi_i, u^c_{it}), \qquad (u^c_{it} \mid a_{it}, \eta_{it}, \varepsilon_{it}, age_{it}, \xi_i) \sim iid\,\text{Uniform}\,(0,1)\,. \tag{34}$$

For the purpose of documenting consumption responses, it is important to know under which conditions estimating (34) allows one to learn about features of the household's consumption function g_t in (27). Suppose that the transitory preference shocks ν_{it} in (27) are i.i.d., independent of past assets and income components, age, and latent type ξ_i. If in addition ν_{it} are scalar and have a monotone impact on the consumption function g_t, then g_t will be identified based on (34), up to a nonlinear transformation of its last argument. Moreover, when the economic primitives are such that ν_{it} are multidimensional or have a non-monotone impact on consumption, the conditional mean function of log-consumption implied by (27) will still be identified based on (34), even though the individual consumption function g_t will not be identified in general. Indeed, under our assumptions we have

$$\phi_{it} = \mathbb{E}_{\nu_{it}}\left[\frac{\partial g_t\left(a_{it}, \eta_{it}, \varepsilon_{it}, age_{it}, \xi_i, \nu_{it}\right)}{\partial \eta}\right] = \mathbb{E}_{u^c_{it}}\left[\frac{\partial Q_c(a_{it}, \eta_{it}, \varepsilon_{it}, age_{it}, \xi_i, u^c_{it})}{\partial \eta}\right].$$

In other words, using quantile methods to flexibly estimate the function Q_c in (34), we will be able to consistently estimate our main target parameters, which are the average derivative quantities ϕ_{it}.

Note that, under mild assumptions, the consumption response parameters in (28) are equal to the derivatives of the conditional mean of consumption given the state variables,

$$\phi_{it} = \frac{\partial}{\partial \eta}\mathbb{E}\left[c_{it} \,\middle|\, a_{it}, \eta_{it}, \varepsilon_{it}, age_{it}, \xi_i\right].$$

However, η_{it}, ε_{it} and ξ_i are unobserved in the data, so it is not enough to model the conditional mean $\mathbb{E}\left[c_{it} \mid a_{it}, \eta_{it}, \varepsilon_{it}, age_{it}, \xi_i\right]$ to recover our key parameters ϕ_{it}. ABB specified a nonlinear mean model with separable heterogeneity. A concern with their specification is that it might be too restrictive as a model of the conditional distribution of c_{it} given $(a_{it}, \eta_{it}, \varepsilon_{it}, age_{it}, \xi_i)$. In contrast, in this paper we employ a quantile specification to achieve a more flexible modeling of that conditional distribution.

In our baseline model where assets do not depend on the latent variables η_{it}, ε_{it}, and ξ_i directly, a specification of the assets process is not needed. However, assuming that asset accumulation does not depend on the latent variables might be restrictive if, for example, assets returns are heterogeneous and the assets process is not independent of ξ_i. For this reason, we will also estimate a model where we specify a reduced-form assets process as

$$a_{i,t+1} = Q_a(a_{it}, \eta_{it}, \varepsilon_{it}, c_{it}, age_{it}, \xi_i, u^a_{i,t+1}), \qquad \left(u^a_{i,t+1} \mid a_{it}, \eta_{it}, \varepsilon_{it}, c_{it}, age_{it}, \xi_i\right) \sim iid \, \text{Uniform}\,(0,1)\,, \tag{35}$$

where in addition $u^a_{i,t+1}$ and $u^c_{i,t+1}$ are independent. In this model, we will specify initial assets holdings as

$$\begin{aligned} a_{i,t_i} &= Q_{a_1}(\eta_{i,t_i}, age_{i,t_i}, cohort_i, educ_i, \xi_i, u^{a_1}_{i,t_i}), \\ &\left(u^{a_1}_{i,t_i} \mid \eta_{i,t_i}, age_{i,t_i}, cohort_i, educ_i, \xi_i\right) \sim iid \, \text{Uniform}\,(0,1)\,. \end{aligned} \tag{36}$$

To summarize the framework laid out in this section, we have described a model with three latent components. The time-invariant type ξ_i is intended to capture household pre-sample-period observed and unobserved heterogeneity. The other two latent components enter the income process. The persistent component η_{it} captures household heterogeneity that results from the accumulation of persistent shocks over time. Finally, independent transitory shocks ε_{it} with an age-specific distribution combine with the persistent component and its profile to produce observed labor income.

The presence of the latent type ξ_i as an argument of the consumption function may potentially reflect several mechanisms. For example, ξ_i may indicate preference or discounting heterogeneity. Alternatively, it may capture heterogeneity in returns to assets. Yet another possible interpretation of ξ_i is as additional resources that are available to the household but not observed in the data, such as consumption insurance provided by parents. We will examine the plausibility of these various mechanisms empirically in Section 2.6. We let the latent type ξ_i correlate with income through the conditioning on $income_i$ in (33). In addition, although here we will use our most parsimonious specification as a baseline when reporting results, in an extension we will let ξ_i enter asset accumulation directly, see equations (35)-(36).

The model thus features two levels of heterogeneity: (a) demographics and time effects, which we partial out linearly in an initial step, and (b) the latent type ξ_i, which we include as part of our nonlinear model. We will study the possibility of an additional nonlinear impact of demographic heterogeneity in Subsection 2.7.4.

2.4 Estimation methodology and implementation

To specify and estimate the model, we closely follow ABB, with some differences. While in this section we focus on estimation and practical implementation, we note that given the similarity of the model's structure to that of ABB, nonparametric identification can be shown using the arguments they provide. Those arguments rely on insights from the literature on nonparametric instrumental variable models and nonlinear models with latent variables (see, among others, [78], [79], and [80]).

2.4.1 Specification

Following ABB, we model all conditional quantile functions using linear quantile specifications at a grid of percentiles. As an example, we model the conditional quantile function of the persistent latent component of income in (30) as

$$Q_\eta(\eta_{i,t-1}, age_{it}, \tau) = \sum_{k=0}^{K} a_k^\eta(\tau)\varphi_k(\eta_{i,t-1}, age_{it}), \tag{37}$$

where φ_k are low-order products of Hermite polynomials in age and the lagged persistent latent component of income, and $a_k^\eta(\tau)$ are piecewise-linear polynomial functions of τ. In practice we use a grid of 11 equidistant percentiles. In addition, following ABB we augment the model by specifying $a_k^\eta(\tau)$ using an exponential modeling of the tails of the intercept coefficients. We use similar specifications for all the other equations (32)-(36). We provide details in Appendix 2.A.

A difference with ABB is that, while they modeled the nonlinear mean of log-consumption and assumed separable errors, here we flexibly estimate the entire conditional quantile function of log-consumption in (34) without imposing separability between u_{it}^c and the other determinants of consumption. This is important for estimating the average consumption derivative parameters ϕ_{it} in the presence of latent variables, in a way which is robust to the presence of non-separabilities implied by the economic model.

Fully nonlinear estimation of consumption quantiles has implications for the econometric specification of the model, given that the type ξ_i is a latent variable. Indeed, note that ξ_i and the conditional quantile function Q_c are not separately nonparametrically identified, since it is always possible to take a transformation of ξ_i, and to undo it in Q_c.[15] In a general quantile model such as (34), we impose the following restriction:

$$\mathbb{E}[c_{it} \,|\, a_{it} = \overline{a}, \eta_{it} = \overline{\eta}, \varepsilon_{it} = \overline{\varepsilon}, age_{it} = \overline{age}, \xi_i = \xi] = \int_0^1 Q_c(\overline{a}, \overline{\eta}, \overline{\varepsilon}, \overline{age}, \xi, \tau)d\tau = \xi, \quad \text{for all } \xi, \tag{38}$$

[15] For example, for any invertible function ψ we can write $Q_c(\xi) = (Q_c \circ \psi^{-1})(\psi(\xi))$.

where $\overline{a}, \overline{\eta}, \overline{\varepsilon}, \overline{age}$ are some fixed reference values for log-assets, persistent and transitory income components, and age. Imposing this restriction resolves the indeterminacy.[16] In this way, ξ_i is measured in consumption units, which is meaningful when studying its distribution. In the implementation we set $\overline{a}, \overline{\eta}, \overline{age}$ to be the unconditional sample averages of log-assets, log-income and age, respectively, and we set $\overline{\varepsilon}$ to zero.

2.4.2 Estimation

To estimate the model we adapt the multi-step approach proposed by ABB to our setting. *In a first step*, we compute regression residuals of log-income, log-consumption, and log-assets on a set of controls, which includes demographics and time indicators, see Section 2.2 for the full list of controls. This allows us to construct the residualized variables y_{it}, c_{it}, and a_{it}.

In a second step, we estimate the income process. To this end, we use a stochastic EM algorithm ([81]), which alternates between draws of the latent income components η_{it} and ε_{it}, and parameter updates based on the latent draws. The updates are performed using quantile regressions, similarly to ABB. For example, to estimate the parameters $a_k^{\eta}(\tau)$ at a grid of τ values in (37), we run multiple quantile regressions.[17]

To generate the latent draws, we depart from ABB who relied on Metropolis Hastings, and use a Sequential Monte Carlo sampling method. We describe this method in the next subsection. The reason for using a different sampler compared to ABB is numerical stability. Indeed, the performance of Metropolis Hastings tends to deteriorate as the length of the panel and the number of households increase. In the longer and larger panel sample we use in this paper, Sequential Monte Carlo methods tend to be more robust to numerical issues

[16] If Q_c is linear, (38) selects a form of the fixed effect that is inclusive of all the intercept components. See [79] and the subsequent literature for related assumptions.

[17] Before every update step, we compute an empirical counterpart of the left-hand side in (38) by regressing log-consumption on the draws of η, ε, ξ, log-assets, and age, and we set ξ_i to be the corresponding predicted value. See Appendix 2.A for details.

such as initialization and seeding than Metropolis Hastings in our experience. A feature of Sequential Monte Carlo methods is that they take advantage of the Markovian structure of the model to improve performance relative to naive importance sampling.

In a third step, we estimate the consumption function, for given values of the parameters governing the income process. We perform this step using a similar strategy to the one we use for income. In this case also, we depart from ABB in the sampling step of the stochastic EM algorithm. However, the presence of the latent type ξ_i further complicates implementation, since one needs to repeatedly draw ξ_i together with the sequences of persistent and transitory components. To generate valid draws, we rely on the pseudo-marginal Markov Chain Monte Carlo algorithm proposed by [69], which itself makes use of Sequential Monte Carlo sampling. We describe our implementation in the next subsection.

Quantile monotonicity. Given our quantile modeling, the parameters satisfy monotonicity restrictions (e.g., [82]). For example, in (37) the mapping $\tau \mapsto a_k^{\eta}(\tau)\varphi_k(\eta_{i,t-1}, age_{it})$ is non-decreasing. In practice we do not enforce monotonicity in estimation. However, in each expectation step of the stochastic EM algorithm we draw from the likelihood implied by the estimated parameters. This ensures that we obtain posterior draws from a valid distribution of η's and ξ's, irrespective of the lack of monotonicity of the quantile parameter estimates. To provide intuition in a simple setup, note that to draw η_{it} according to model (37) one can compute, as in [83],

$$\widetilde{\eta}_{it} = \sum_{k=0}^{K} \widehat{a}_k^{\eta}(u_{it}^{\eta})\varphi_k(\widetilde{\eta}_{i,t-1}, age_{it}) \text{ for } t > t_i, \quad \widetilde{\eta}_{i,t_i} = \eta_{i,t_i},$$

where u_{it}^{η} are i.i.d. standard uniform. Although the estimates $\widehat{a}_k^{\eta}(\tau)$ may not satisfy monotonicity restrictions, this approach produces $\widetilde{\eta}_{it}$ draws from a valid distribution function. In our setting we use this strategy to generate *posterior* draws of η's and ξ's, see Appendix 2.A for details.

Asymptotic distribution and inference. Under the assumption that the parametric model is correctly specified,[18] averages of parameter draws are consistent and asymptotically normal with an asymptotic variance that can be estimated by bootstrap or analytical approximations, see [84] and ABB for details. We will report confidence bands computed using two versions of the bootstrap: a parametric bootstrap that relies on the model's structure for simulations, and a nonparametric bootstrap clustered at the household level.

2.4.3 Computational sampling techniques

Here we describe how we draw latent variables in every step of the stochastic EM algorithm. We present, in turn, the methods we use for the latent income components $\eta_{it}, \varepsilon_{it}$, and for the latent consumption type ξ_i. In practice we run these simulation steps in parallel across households, which makes it easy to estimate the model on an unbalanced panel.

Income components: Sequential Monte Carlo. Estimating the income process requires solving a nonlinear filtering problem, where $\eta_{i,t_i}, ..., \eta_{i,t_i+T_i-1}$ are latent variables. To draw from their posterior distribution given the income data we use a Sequential Monte Carlo (SMC) approach, see [68] and [85] for surveys.

To describe the SMC approach, we focus on the problem of sampling $\eta_{i,t_i}, ..., \eta_{i,t_i+T_i-1}$ for a single household i from the posterior distribution $f(\eta_{i,t_i}, ..., \eta_{i,t_i+T_i-1}|y_{i,t_i}, ..., y_{i,t_i+T_i-1})$. In practice we sample in parallel across households. With importance sampling, one might first sample directly from some proposal distribution $\pi(\eta_{i,t_i}, ..., \eta_{i,t_i+T_i-1})$, and then re-sample

[18] One may also view the parametric model as a sieve approximation to a nonparametric distribution, where the size of the grid of τ values, and hence the number of parameters, would grow with the sample size at an appropriate rate. The theoretical justification we mention here is for a well-specified parametric model.

using importance sampling weights

$$w_i \propto \frac{f(\eta_{i,t_i}, ..., \eta_{i,t_i+T_i-1}|y_{i,t_i}, ..., y_{i,t_i+T_i-1})}{\pi(\eta_{i,t_i}, ..., \eta_{i,t_i+T_i-1})},$$

where $\propto$ is a proportionality symbol. However, finding a suitable proposal distribution in our flexible nonlinear model is challenging. Instead, we try and generate draws (also called "particles") sequentially.

At $t = t_i$, we initialize S particles $\eta_{i,t_i}^{(s)}$ from a suitable proposal distribution $\pi(\eta_{i,t_i})$. Re-sampling with weights

$$w_{i,t_i}^{(s)} \propto \frac{f\left(\eta_{i,t_i}^{(s)}|y_{i,t_i}\right)}{\pi\left(\eta_{i,t_i}^{(s)}\right)}$$

gives S particles approximately distributed according to $f(\eta_{i,t_i}|y_{i,t_i})$.

At $t = t_i + 1$, we now aim to approximate

$$f(\eta_{i,t_i}, \eta_{i,t_i+1}|y_{i,t_i}, y_{i,t_i+1}) = \frac{f(y_{i,t_i+1}|\eta_{i,t_i+1})f(\eta_{i,t_i+1}|\eta_{i,t_i})}{f(y_{i,t_i+1}|y_{i,t_i})}f(\eta_{i,t_i}|y_{i,t_i}).$$

Since we already have S particles approximately distributed according to $f(\eta_{i,t_i}|y_{i,t_i})$, we can simply use a second proposal distribution $\pi(\eta_{i,t_i+1}|\eta_{i,t_i})$ to extend these existing particles. Re-sampling with weights

$$w_{i,t_i+1}^{(s)} \propto \frac{f\left(\eta_{i,t_i+1}^{(s)}|y_{i,t_i+1}, \eta_{i,t_i}^{(s)}\right)}{\pi\left(\eta_{i,t_i+1}^{(s)}|\eta_{i,t_i}^{(s)}\right)}$$

gives S particles approximately distributed according to $f(\eta_{i,t_i}, \eta_{i,t_i+1}|y_{i,t_i}, y_{i,t_i+1})$. The process continues until we obtain S particles approximately distributed as[19]

$$f(\eta_{i,t_i}, ..., \eta_{i,t_i+T_i-1} | y_{i,t_i}, ..., y_{i,t_i+T_i-1})$$

The choice of proposal distributions π is important for numerical performance. We found that a simple generalization of a linear permanent-transitory earnings model with Gaussian errors performed well. Specifically, we postulate the following model:

$$y_{it} = \eta_{it} + \varepsilon_{it}, \quad \varepsilon_{it} \sim iid\, \mathcal{N}(0, \sigma_\varepsilon^2), \tag{39}$$

$$\eta_{it} = m(\eta_{i,t-1}, age_{it}) + v_{it}^\eta, \quad v_{it}^\eta \sim iid\, \mathcal{N}(0, \sigma_v^2), \tag{40}$$

where ε_{it} and v_{it}^η are independent at all lags, and m is a Hermite polynomial. We re-estimate this model at each iteration of the stochastic EM algorithm, and then set $\pi(\eta_{it}|\eta_{i,t-1})$ to be the posterior distribution based on it. We provide details about the implementation of the SMC sampler in Appendix 2.A. In addition, we provide a comparison of the SMC and Metropolis Hastings sampling methods in the ABB sample in Appendix 2.C. We find that, while our SMC algorithm recovers similar estimates of nonlinear persistence to those reported in ABB, the SMC method is less sensitive to numerical instability than Metropolis Hastings.

Unobserved type in consumption: Particle Markov Chain Monte Carlo. In order to incorporate unobserved heterogeneity ξ_i, we embed the SMC sampler into a Particle Markov Chain Monte Carlo (PMCMC) algorithm, following [69]. We use this method to estimate the parameters of the consumption process, after having estimated the parameters

[19] In practice, re-sampling at every time increment can result in degeneracy among the available particles. For this reason, we instead use an adaptive rule which avoids degeneracy (see Creal, 2012).

of the income process.

To outline the PMCMC approach, suppose we wish to sample $\xi_i, \eta_{i,t_i}, ..., \eta_{i,t_i+T_i-1}$ from the posterior distribution $f(\xi_i, \eta_{i,t_i}, ..., \eta_{i,t_i+T_i-1} \,|\, w_{i,t_i}, ..., w_{i,t_i+T_i-1})$, where $w_{it} = (y_{it}, c_{it}, a_{it})$ is a vector of household i's observed income, consumption and assets at time t. In the PMCMC approach, SMC algorithms are used to generate efficient proposals to be used within a Metropolis Hastings algorithms. An important feature of these methods is that they only rely upon the availability of unbiased estimates of the marginal likelihood $f(w_{i,t_i}, ..., w_{i,t_i+T_i-1}|\xi_i)$, which are readily available as a by-product of the SMC algorithm. The use of unbiased estimates of a target distribution within a Metropolis Hastings algorithm can be viewed more generally as an example of a pseudo-marginal approach in which the resulting algorithms can be presented as bona fide Metropolis Hastings samplers whose marginal distribution is the target distribution of interest. We provide details about the implementation of the PMCMC sampler in Appendix 2.A.

2.5 Main results

In this section we present the main empirical results on income and consumption, obtained using our baseline nonlinear model with unobserved heterogeneity.

2.5.1 Income persistence

We start by reporting the results on nonlinear income persistence. In the left graph of Figure 2.3 we show the derivative of the conditional quantile function of log-income given lagged log-income and age, with respect to lagged log-income. Formally, we compute an estimate of

$$\rho_y(y, age, \tau) = \frac{\partial Q_y(y, age, \tau)}{\partial y}, \qquad \text{for } \tau \in (0,1),$$

where Q_y is the conditional quantile function of log-income given lagged log-income and age, and average it with respect to age. The nonlinear persistence parameters $\rho_y(y, age, \tau)$ can be interpreted as heterogeneous autoregressive coefficients, which may depend on both the income level y and the income shock τ.[20] We plot the derivative as a function of lagged log-income (which we refer to as "initial income") and of the innovation in the quantile model (which we refer to as "income shock").

The results show that most households, for most shocks, have current disposable incomes that are quite persistent, with a derivative coefficient that is above 0.80. However, households with low initial income and high income shocks have incomes that are substantially less persistent, with a coefficient as low as 0.40. Likewise, persistence is also low for households with high initial income and low income shocks, with a coefficient of a similar magnitude. These nonlinear persistence estimates are closely related to those found by ABB on a smaller balanced sample drawn from the earlier pre-recession years of the PSID.

In the right graph of Figure 2.3 we show nonlinear income persistence, but now for the persistent latent component η_{it}. That is, we show

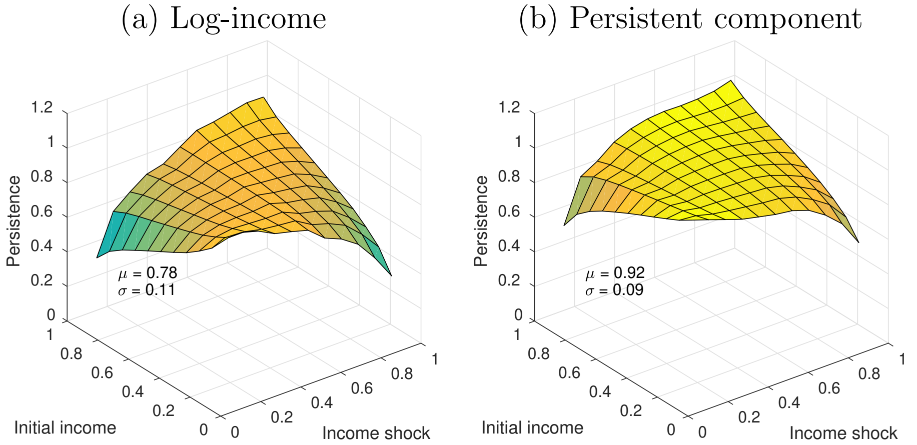
$$\rho_\eta(\eta, age, \tau) = \frac{\partial Q_\eta(\eta, age, \tau)}{\partial \eta}, \qquad \text{for } \tau \in (0,1),$$

where Q_η is the conditional quantile function of η_{it} given $\eta_{i,t-1}$ and age, see (30). We plot the derivative as a function of $\eta_{i,t-1}$ ("initial income") and the innovation in the quantile model ("income shock").[21] We see that average persistence is higher than for the case of log disposable income — it is 0.92 in the right graph, versus 0.78 in the left graph — due

[20] In Appendix Figure 2.11 we show a different projection of the same three-dimensional surfaces, to ease visualization.

[21] To produce the plot, we use posterior draws computed from the model. We proceed similarly when plotting all subsequent results involving latent variables.

Figure 2.3: Nonlinear income persistence


(a) Log-income

(b) Persistent component

Notes: PSID, 2005-2017 sample, disposable income, dual earners. The left graph shows quantile derivatives of log-income with respect to lagged log-income, $\rho_y(y, age, \tau)$ averaged over y. The right graph shows quantile derivatives of the persistent latent component η_{it} with respect to η_{it-1}, $\rho_\eta(\eta, age, \tau)$ averaged over η, in a model estimated using sequential Monte Carlo with a stochastic EM algorithm. In this case, the two horizontal axes show percentiles of η_{it-1} ("initial income") and conditional percentiles of η_{it} given η_{it-1} ("income shock"), respectively.

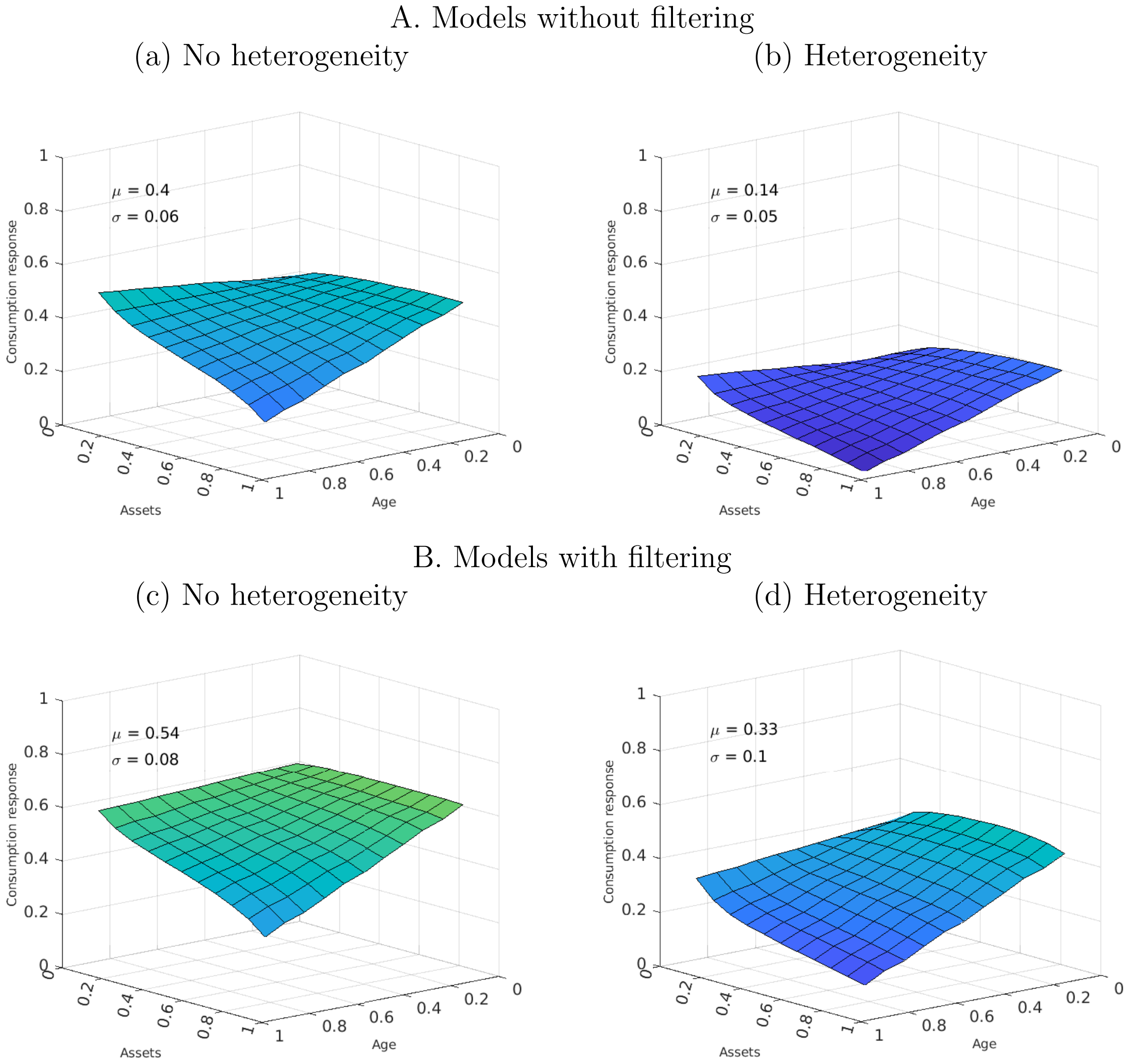
to the removal of the transitory income component. For households with high values of initial persistent income and high shocks, persistence is close to unity, and similarly for households with low initial persistent income and low shocks.[22] The nonlinear pattern for the persistent latent component η_{it} is qualitatively similar to the one for log-income, although it is quantitatively less pronounced.

These nonlinear persistence patterns are rather precisely estimated, see the parametric bootstrap 95% confidence bands in Appendix Figure 2.12 and the nonparametric bootstrap 95% confidence bands in Appendix Figure 2.13. In addition, comparing Figure 2.3 to Appendix Figure 2.8, we see that, while nonlinearities are somewhat more salient in our larger and more recent sample compared to the balanced sample used in ABB, the persistence patterns in both cases are comparable.[23]

[22] Note that it is possible for the nonlinear income persistence measure to exceed one.

[23] In Figure 2.3 we average the persistence measure across age values. In contrast, the main nonlinear

Figure 2.4: Average consumption responses


A. Models without filtering

(a) No heterogeneity

(b) Heterogeneity

B. Models with filtering

(c) No heterogeneity

(d) Heterogeneity

Notes: PSID, 2005-2017 sample, dual earners. The graphs show the average derivative of log-consumption with respect to log-income (in the top panel) and the persistent latent component η_{it} (in the bottom panel). The left graphs correspond to a model without unobserved heterogeneity ξ_i in consumption, whereas the right graphs correspond to a model with unobserved heterogeneity ξ_i. The two horizontal axes show age and assets percentiles, respectively.

2.5.2 Average consumption responses to income shocks

The main goal of the paper is to study heterogeneity in consumption responses to unexpected changes in income. That is, the way income shocks are transmitted into consumption which underpins the degree of "partial insurance" achieved by the household. In this subsection, and the next, we document several key features of household partial insurance, which we measure using the household-and-time-varying transmission coefficients

$$\phi_{it} = \phi(age_{it}, a_{it}, \eta_{it}, \varepsilon_{it}, \xi_i)$$

given by the average derivative effects (28) introduced in Subsection 2.3.1. The transmission coefficient ϕ_{it} quantifies the change in consumption induced by an exogenous marginal change in the persistent latent component of income.

In Figure 2.4 we start by showing how the mean of the estimated transmission parameters ϕ_{it} varies with assets levels and over the life cycle. We compare four specifications. The "models without filtering" in the upper panel correspond to specifications without transitory component ε_{it}, so the derivative on the right-hand side of (28) is taken with respect to log current disposable income y_{it} instead of the persistent latent component η_{it}. The "models with filtering" in the lower panel allow for a separate role of η_{it} and ε_{it}. For both models with and without filtering, we distinguish two specifications with and without unobserved heterogeneity ξ_i, in the left and right columns, respectively.

Figure 2.4 shows that all specifications agree quite well qualitatively. In particular, the association between consumption and income or its persistent latent component is weaker for older and wealthier households. At the same time, there are important quantitative dif-

persistence figures in ABB are evaluated at a reference age value. The analog of Figure 3(a) in ABB is Figure S3 in their supplemental appendix.

ferences between the four specifications. We find that allowing for unobserved heterogeneity ξ_i tends to dampen the consumption impacts of income shocks, the difference being particularly noticeable for the models without filtering where average responses decrease from 0.40 to 0.14. The impact of heterogeneity can be explained by the fact that, according to our estimates, ξ_i is positively correlated with income, see Section 2.6. In contrast, allowing for a transitory income component tends to increase consumption responses to income shocks, as is typically the case in estimates that correct for measurement error bias. As a result, in our main model with unobserved heterogeneity and a transitory component, the lower right hand graph shows an estimated average response parameter of 0.33. There are strong differences by assets and age too, with the estimated average transmission coefficient dropping toward 0.10 for older and wealthier households, while for younger households the estimated mean transmission rises to around 0.40.[24]

Comparison with ABB. It is informative to compare the average responses in Figure 2.4 to the results obtained by ABB. In a model without heterogeneity but with a transitory component, ABB found an average transmission coefficient of 0.38. This is lower than the responses in Figure 2.4 (c), which are 0.54 on average.[25] As we previously noted, the period of observation, the sample of households, and the income measure used in ABB all differ from the ones we focus on in the current paper. In particular, ABB focus on labor income as opposed to disposable income. Our estimates of consumption responses based on labor

[24] The relative magnitudes of the nonlinear estimates in Figure 2.4 are reminiscent of the situation in a linear model with a mismeasured persistent regressor and fixed effects, where the (positive) fixed effects bias and the (negative) measurement error bias tend to offset each other, while only accounting for fixed effects exacerbates the measurement error bias ([86]).

[25] The consumption responses in a model without heterogeneity in ABB can be found in their Figure 5(c). In addition, ABB also reported average responses based on a model with unobserved heterogeneity, albeit using a different specification for the consumption rule. They found lower responses in this case, amounting to 0.32 on average, see Figure S24(b) in the supplementary appendix of ABB.

income are substantially lower than the responses based on disposable income shown in Figure 2.4, see Section 2.7.

Test of homogeneity. By comparing average response coefficients in models with and without household-specific heterogeneity, one can assess whether the data supports an homogeneous model without latent types. To do this, in Appendix Figure 2.14 we report confidence bands based on the nonparametric bootstrap clustered at the household level for the average responses depicted in the lower panel of Figure 2.4. We find a 95% confidence interval for the mean across these responses of $[0.50, 0.59]$ in the model without heterogeneity, and of $[0.21, 0.44]$ in the model with heterogeneity. The fact that the two confidence intervals do not overlap represents a formal rejection, at the 5% level, of the null hypothesis of homogeneity. The same conclusion holds when we use the parametric bootstrap to produce confidence intervals, see Appendix Figure 2.15.

2.5.3 Heterogeneity in consumption responses to income shocks

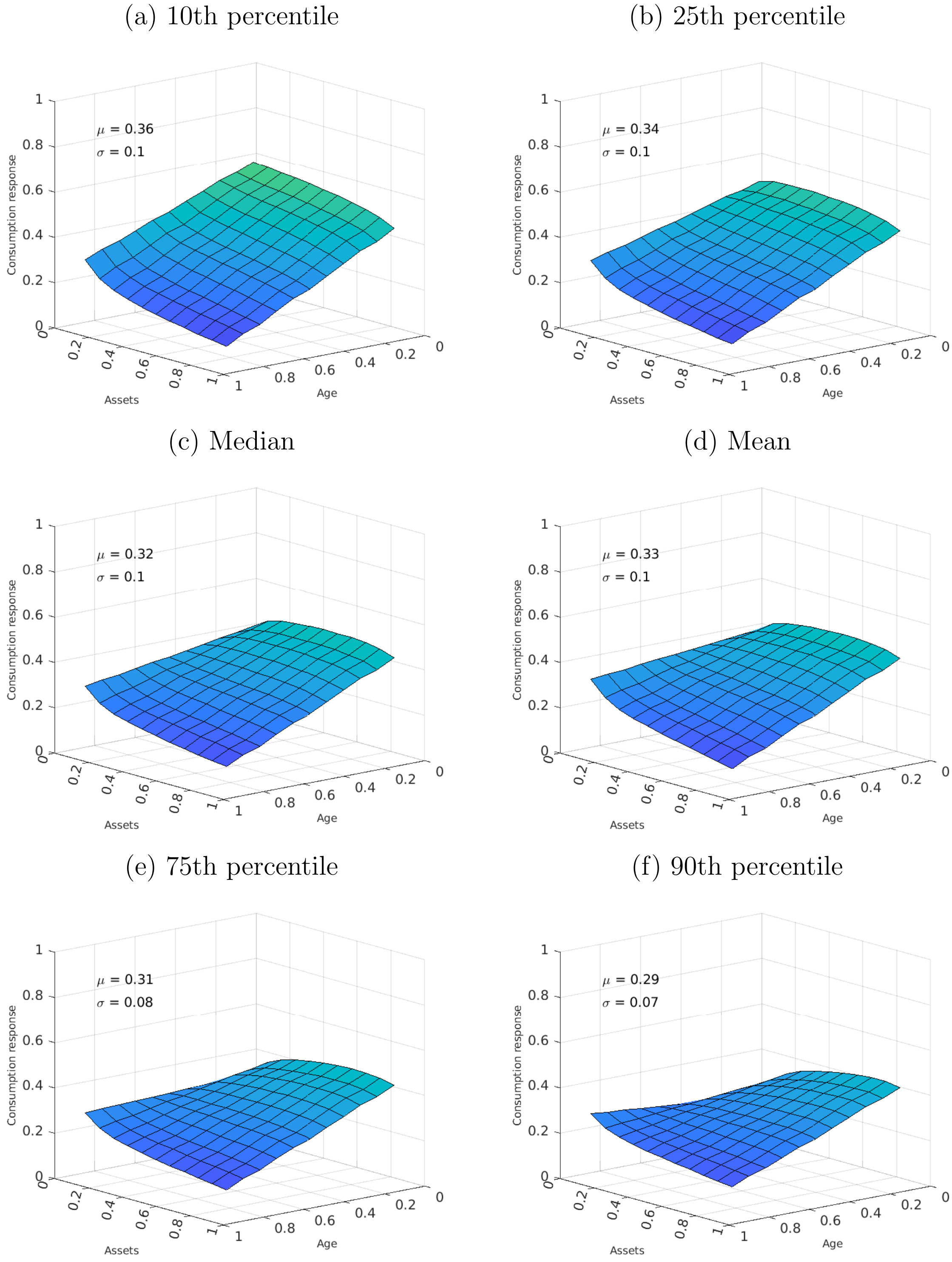
We have already seen that the introduction of unobserved heterogeneity has a systematic effect on the estimated average response of consumption to changes in income. We hypothesize that there are also systematic differences in responses across consumers that differ according to unobserved heterogeneity. To examine this, we study how consumption responses differ among households that are at the same point in the life cycle and have the same level of assets. For this purpose, we show how the transmission coefficients ϕ_{it} vary by quantiles of the unobserved type ξ_i, in addition to showing how they vary with assets levels and over the life cycle.

In Figure 2.5 we show transmission parameters as a function of assets and age, for five different percentiles of ξ_i, and we also show the average across ξ_i values. The results show clear evidence of household heterogeneity in consumption responses to income shocks. Consider the 10th percentile of ξ_i, in the top left graph. For these "low consumption type" households, average transmission is 0.36, yet the magnitude of the transmission coefficient varies substantially with age and assets. Indeed, while younger and less wealthy households have transmission coefficients of close to 0.60, the coefficient is as low as 0.10 for older and wealthier households. This pattern is qualitatively consistent with the implications of a standard life-cycle model of consumption and saving behavior in which persistent shocks are harder to self-insure for young consumers and for those consumers with low levels of net assets.

This "life-cycle consistent" pattern of responses is maintained through to the median type, albeit less pronounced. As we move to the higher consumer types, a pattern that is much less sensitive to assets and age appears. Consider the 90th percentile of ξ_i, in the bottom right graph of Figure 2.5. For these high-type households, the transmission coefficients are 0.29 on average, hence lower than the coefficients of the low-type households. In addition, the variation of the transmission coefficients with assets and age is less pronounced than for the low types. Indeed, while coefficients are approximately 0.15 for the older and wealthier households, the young and less wealthy households have coefficients that do not exceed 0.40. These patterns for the high-types are less in accordance with the forces at play in conventional life-cycle models of the individual household.

In order to provide measures of uncertainty associated with our main results, we rely on the bootstrap. We report results based on a parametric bootstrap approach, where we use the model to simulate bootstrapped data sets given parameter estimates. In Appendix Figure 2.16 we report pointwise 95% bands for the transmission parameters of Figure 2.5. We see that our estimates are rather precise. As a complement to the parametric bootstrap, in Appendix Figure 2.17 we report pointwise 95% bands based on the nonparametric bootstrap clustered at the household level. Precision is lower in this case, which is is not surprising, since, relative to the clustered nonparametric bootstrap, the parametric bootstrap exploit

Figure 2.5: Heterogeneity in consumption responses


(a) 10th percentile

(b) 25th percentile

(c) Median

(d) Mean

(e) 75th percentile

(f) 90th percentile

Notes: See the notes to Figure 2.4. Here we report the results by percentiles of heterogeneity ξ_i in consumption.

Table 2.2: Summarizing heterogeneity across types, parametric bootstrap

A. 90th vs 10th percentile of ξ

	Young, low assets	Old, high assets	Δ
High ξ	0.31 [0.20,0.39]	0.22 [0.12,0.34]	0.09 [-0.03,0.19]
Low ξ	0.48 [0.40,0.62]	0.21 [0.13,0.33]	0.27 [0.16,0.38]
Δ	-0.17 [-0.36,-0.06]	0.01 [-0.15, 0.13]	-0.18 [-0.34,-0.06]

B. 75th vs 25th percentile of ξ

	Young, low assets	Old, high assets	Δ
High ξ	0.36 [0.28,0.42]	0.21 [0.14,0.31]	0.15 [0.04,0.20]
Low ξ	0.45 [0.38,0.55]	0.21 [0.15,0.31]	0.24 [0.14,0.31]
Δ	-0.09 [-0.17,-0.03]	0.00 [-0.08, 0.06]	-0.09 [-0.17,-0.03]

Notes: See the notes to Figure 2.4. Here we report average consumption responses for young and low assets households compared to old and high assets households, for different percentiles of heterogeneity ξ_i in consumption. Values are calculated by evaluating the average consumption response for households at a fixed percentile of ξ_i when assets and age are fixed at the τth percentile. Reported values for young and low assets households are then shown by averaging over $\tau \in (0, 0.5)$. Reported values for old and high assets households are then shown by averaging over $\tau \in (0.5, 1)$. Parametric bootstrap 95% confidence intervals based on 200 replications are shown in brackets.

our modeling of the time-series dependence.

As a summary measure of the salient dimensions of heterogeneity that we find, in the top panel of Table 2.2 we report estimates of average transmission parameters for various categories of households: high and low types, corresponding to ξ_i being at the 90th percentile or the 10th percentile, young/low assets for whom age and assets are below the median, and old/high assets for whom age and assets are above the median. In the bottom panel we repeat the exercise for high types corresponding to ξ_i being at the 75th percentile and low types corresponding to ξ_i being at the 25th percentile. Alongside point estimates, we report 95% confidence intervals based on the parametric bootstrap.

We find that, while for high consumption types at the 90th percentile the transmission of income shocks is only 0.09 higher for young/low assets households and insignificant at conventional levels, for low types at the 10th percentile the average response coefficient is 0.27 higher for the young and low assets and significant at the 5% level. This supports our main conclusion regarding the fact that the behavior of low types appears to be consistent with a standard life-cycle model of consumption and saving, yet the behavior of high types appears less consistent with the mechanisms of the model. In addition, the cross-type difference $0.09 - 0.27 = -0.18$ between these two estimates, which is akin to a difference-in-differences estimate, is significant at the 5% level.[26]

The results in this section, based on a dynamic model with latent income components and unobserved heterogeneity, provide evidence for the presence of heterogeneous types of consumers, confirming what Figure 2.2 suggested. In the next section, we develop the implications of these results for life-cycle patterns of consumption and savings, and we examine

[26] In Appendix Table 2.6 we report confidence intervals based on the nonparametric bootstrap clustered at the household level. In this case, for low types below the 10th percentile the average response coefficient remains significantly higher for the young and low assets. However, the cross-type difference is insignificant at the 5% level.

various possible mechanisms for the patterns in transmission parameters displayed in Figure 2.5.

Dispersion of consumption responses around their means ϕ_{it}. While our main focus is on the average consumption response parameters ϕ_{it}, there may be dispersion around those averages. In Appendix 2.D we show how to compute an upper bound on the share of variance in responses $\frac{\partial c_{it}}{\partial \eta}$ explained by the means ϕ_{it}, obtained by calculating a lower bound on the variance of $\frac{\partial c_{it}}{\partial \eta}$ conditional on $(a_{it}, \eta_{it}, \varepsilon_{it}, age_{it}, \xi_i)$. The reason why only bounds are available is because transitory preference shocks ν_{it}, which may generate additional heterogeneity in responses beyond the mean transmission parameters ϕ_{it}, may be multi-dimensional. We report estimates of the upper bounds on the variance shares in Appendix Figure 2.18. We find high variance shares, in many cases higher than 80%, suggesting that the ϕ_{it} parameters capture a large part of the heterogeneity in responses (although we note that, since those are upper bounds, this evidence does not strictly speaking rule out the presence of substantial additional heterogeneity).

2.6 Candidate mechanisms to explain the heterogeneity

In this section we study various mechanisms that might potentially explain the type heterogeneity that we find.

2.6.1 Three candidate mechanisms

Informed by standard models of consumption and saving decisions, which guide our empirical analysis, we can outline three candidate mechanisms to explain the heterogeneous types that we document.

A first possible explanation is heterogeneity in preferences and discounting. There is a long history of incorporating discount rate heterogeneity to help explain lifetime wealth accumulation, for example [87] and [88]. Everything else equal, individuals with higher marginal utility of consumption will consume more, and hold fewer assets. Individuals with higher discount factors will delay consumption relative to those with lower discount factors, and hold more assets. This type of heterogeneity should lead to high-type households consuming more and holding fewer assets. We examine this hypothesis by showing how consumption and assets profiles depend on the latent type.

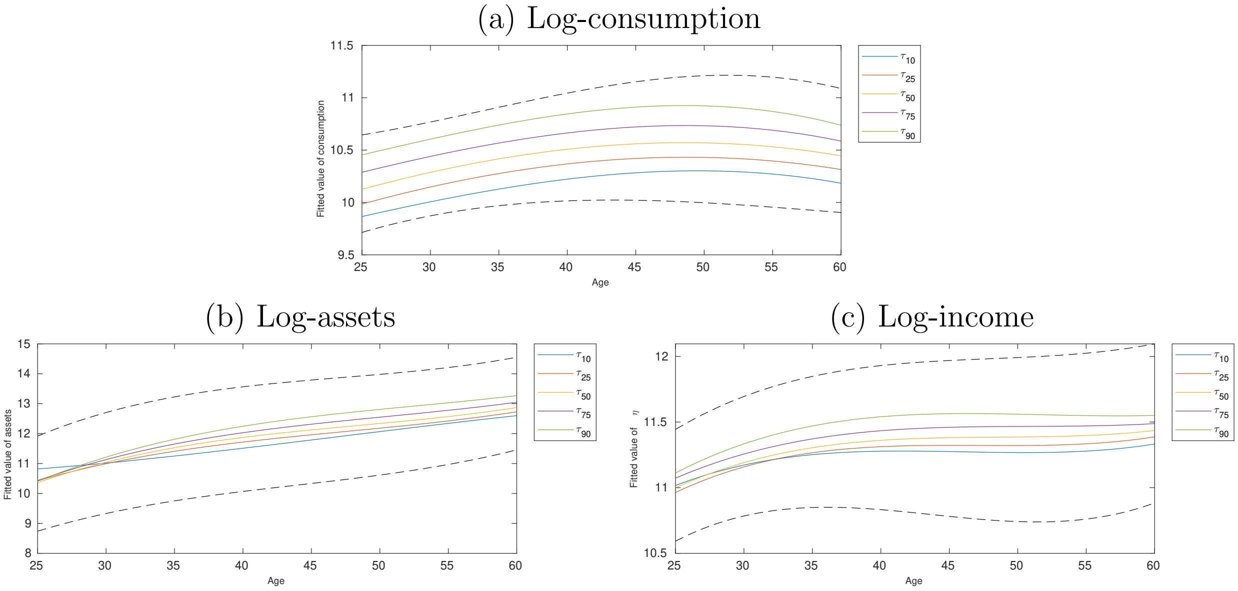
A second candidate explanation is heterogeneity in returns to assets. The rate of return is a key determinant of consumption choice in standard models, so the types we find might in fact reflect heterogeneity in those returns across households. [89] find evidence of individual heterogeneity in returns to wealth using administrative records from Norway. We examine this heterogeneity in the PSID by estimating an extension of the model with heterogeneity in the asset accumulation rule (see equations (35)-(36)), and by empirically documenting the form of this rule.

A third candidate explanation is heterogeneity in access to external resources, such as parental insurance. Individuals with access to other forms of insurance would be expected to consume more, for a comparable level of income and assets. [90], [91] and, more recently, [92] and [93], use the generational links in the PSID to document a significant role for parents and family networks in providing additional insurance. To probe this hypothesis, we link the household heads in the PSID to their parents, and study how the latent types relate to parental income, wealth, and consumption.

2.6.2 Life-cycle profiles

As a step towards examining the plausibility of a preference and discounting channel, we show the life-cycle profiles implied by our dynamic model, for various percentiles of the unobserved heterogeneity ξ_i. In the top panel of Figure 2.6 we show consumption profiles, in logs.[27] We see that consumption levels are monotone in the types. This is partly a result of our restriction in (38), which implies monotonicity at the reference age. In addition, comparing the dispersion of the solid lines (which correspond to the ξ_i percentiles) with the dashed lines (which correspond to 10th and 90th unconditional percentiles of log-consumption), we see that type heterogeneity explains a large part of the overall variation in log-consumption. Our results imply that ξ_i accounts for 25% of the variance of log-consumption.[28]

Figure 2.6: Life-cycle profile



(a) Log-consumption

(b) Log-assets

(c) Log-income

Notes: Average non-residualized log-consumption in graph (a), log-assets in graph (b), and persistent latent component of log-income in graph (c), for different ages and percentiles of ξ_i (10%, 25%, Median, 75%, 90%). The dashed lines show the age-specific unconditional 10th and 90th percentiles for each outcome measure.

[27] To draw these profiles we proceed by simulation, using a similar strategy to ABB. In addition, in the graphs we show non-residualized variables; that is, we add back the predictions of the first-stage regressions to the residuals of log-consumption, log-assets, and log-income. Note these predictions include the effects of calendar time in addition to those of demographics.

[28] In Appendix Figure 2.19 we plot the median and 10th and 90th percentile of log-consumption, over the life cycle, for three percentiles of ξ_i (10th, median, and 90th). This confirms that the between-ξ_i dispersion of consumption is substantial, even though there is large within-ξ_i variation as well.

In the bottom panel of Figure 2.6 we show the profiles of assets and income, in logs. In the left graph we see that, similarly to consumption, assets are monotone in types. This suggests that, while high-type households consume more than low types, they also hold more assets. However, the variation in types explains a relatively small share of the overall variation in log-assets. Note that, while the restriction in (38) imposes that log-consumption increases with the type ξ_i at particular covariates values, nothing in our approach restricts log-assets to be monotone in the type. Quantitatively, we find that ξ_i accounts for 3% of the variance of log-assets. In the right graph we show the results for the persistent latent component of income. We see the same monotone behavior in the type as for consumption and assets. Our results imply that ξ_i accounts for 4% of the variance of the persistent latent component of log-income.[29] We have already seen that the correlation between the latent type and income is sufficient to generate sizable differences between specifications with and without latent heterogeneity, see Figure 2.4.

Overall, our results show that high-type households consume more, hold more assets, and have higher income. Quantitatively, individual types mainly differ in their consumption profiles. While these findings do not rule out that differences in preferences and discounting may be present in the data, they are difficult to reconcile with this channel being the main driver of the heterogeneity in consumption responses that we find.

2.6.3 Heterogeneity in consumption and assets

We next assess the role of heterogeneity in assets returns as an explanation for type heterogeneity. For this purpose, we estimate a specification where asset accumulation depends on the latent type ξ_i, see equations (35)-(36). The results based on this specification are

[29] In Appendix Figure 2.20 we plot the median and 10th and 90th percentile of log-assets and the persistent latent component of log-income, over the life cycle, for three percentiles of ξ_i. The results confirm that most of the dispersion in assets and income is within-ξ_i.

similar to the baseline ones for both income and consumption. In Appendix Figure 2.21 we show the type heterogeneity in consumption responses to variation in the persistent latent component of income, and find overall very similar responses to the ones based on a specification without assets heterogeneity. In Appendix Figures 2.22 and 2.23 we report estimates of assets responses, by type, in this generalized specification that allows the latent type to enter the asset accumulation rule. We find that the association between lagged assets and current assets conditional on lagged income and consumption increases with the latent type, and that assets responses are higher for the young, decrease with the level of lagged assets, and increase with the type ξ_i, especially for older households.

Overall, the results based on the extended specification with latent heterogeneity in assets and consumption suggest that returns to assets are indeed heterogeneous across households in the data. However, allowing the heterogeneity to enter asset accumulation does not materially affect the conclusions regarding the heterogeneity in consumption responses.

2.6.4 Heterogeneity in parental insurance

A third candidate mechanism is heterogeneity in access to other forms of insurance, such as parental insurance. In order to examine the plausibility of this mechanism, we take advantage of the inter-generational linkages available in the PSID to match households to their parents. This aspect makes the PSID uniquely suited to study income and consumption dynamics in the presence of links across generations. Specifically, we start by matching the heads of each household to those households headed by a parent of the head. If matches to the household head are not available, we alternatively try and match the spouse of each household to those households headed by a parent of the spouse. In our baseline sample we are able to successfully match approximately 33% of households.

Given this matched panel dataset, we then regress posterior means of the types ξ_i on various parental outcomes, such as consumption, income, and assets. In Table 2.3 we report

Table 2.3: Heterogeneity and parental outcomes

A. All households

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Parent consumption	0.05	0.07					0.04	0.05
	(0.02)	(0.02)					(0.02)	(0.02)
Parent income			0.03	0.04			0.02	0.02
			(0.01)	(0.01)			(0.02)	(0.02)
Parent assets					0.01	0.01	0.00	0.00
					(0.01)	(0.01)	(0.01)	(0.01)
Controls	No	Yes	No	Yes	No	Yes	No	Yes

B. Young adults only

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Parent consumption	0.05	0.06					0.03	0.04
	(0.02)	(0.02)					(0.02)	(0.02)
Parent income			0.03	0.04			0.02	0.02
			(0.01)	(0.01)			(0.02)	(0.02)
Parent assets					0.01	0.01	0.00	0.00
					(0.01)	(0.01)	(0.01)	(0.01)
Controls	No	Yes	No	Yes	No	Yes	No	Yes

Notes: PSID, 2005-2017 sample, household heads aged 25-60 (top panel) and 25-45 (bottom panel). Regressions of posterior ξ_i draws on parental outcomes. Parental links are obtained for approximately 33% of panel. Parental outcomes are obtained as average residuals net of cohort and year effects. Results are based on 10 posterior draws per household. Controls include an education dummy for the household head and a quadratic specification for first period age. Standard errors clustered at the household level do not account for the uncertainty in the posterior parameter estimates.

the results of various specifications with different sets of controls. For robustness, in addition to the results for all households (in the top panel) we also report results for household heads who are less than 45 years old (in the bottom panel). We find that parental income and consumption correlate positively with the mean type, although the correlation with assets is insignificant from zero at conventional levels. When including all parental variables together, parental consumption remains significantly positively correlated with the type. This suggests that, indeed, the latent type ξ_i may partly reflect heterogeneous access to parental insurance. This interpretation is further supported by the monotonicity of assets in the type documented in Figure 2.6.

However, these results are purely indicative and we leave it to future work to assess whether this channel is quantitatively important.

2.7 Other results and extensions

In this section we report results based on extensions of the model and other robustness checks.

2.7.1 Impulse responses

We start by reporting impulse responses implied by the model's estimates. In Figure 2.7 we estimate the impact of a shock to the persistent latent component of income, η_{it}, at age 34. The figure is divided into three parts. In the upper part, we report the difference between the average persistent latent component of income for households hit by the shock and the average persistent latent component of income for households hit by a "median" shock, i.e., corresponding to the 50th percentile of η_{it} conditional on $\eta_{i,t-1}$. To highlight the heterogeneity in impulse responses, we show results for various percentiles of the latent type distribution. In the middle and bottom parts of the figure we proceed similarly for log-consumption and log-assets, respectively, instead of the income component.

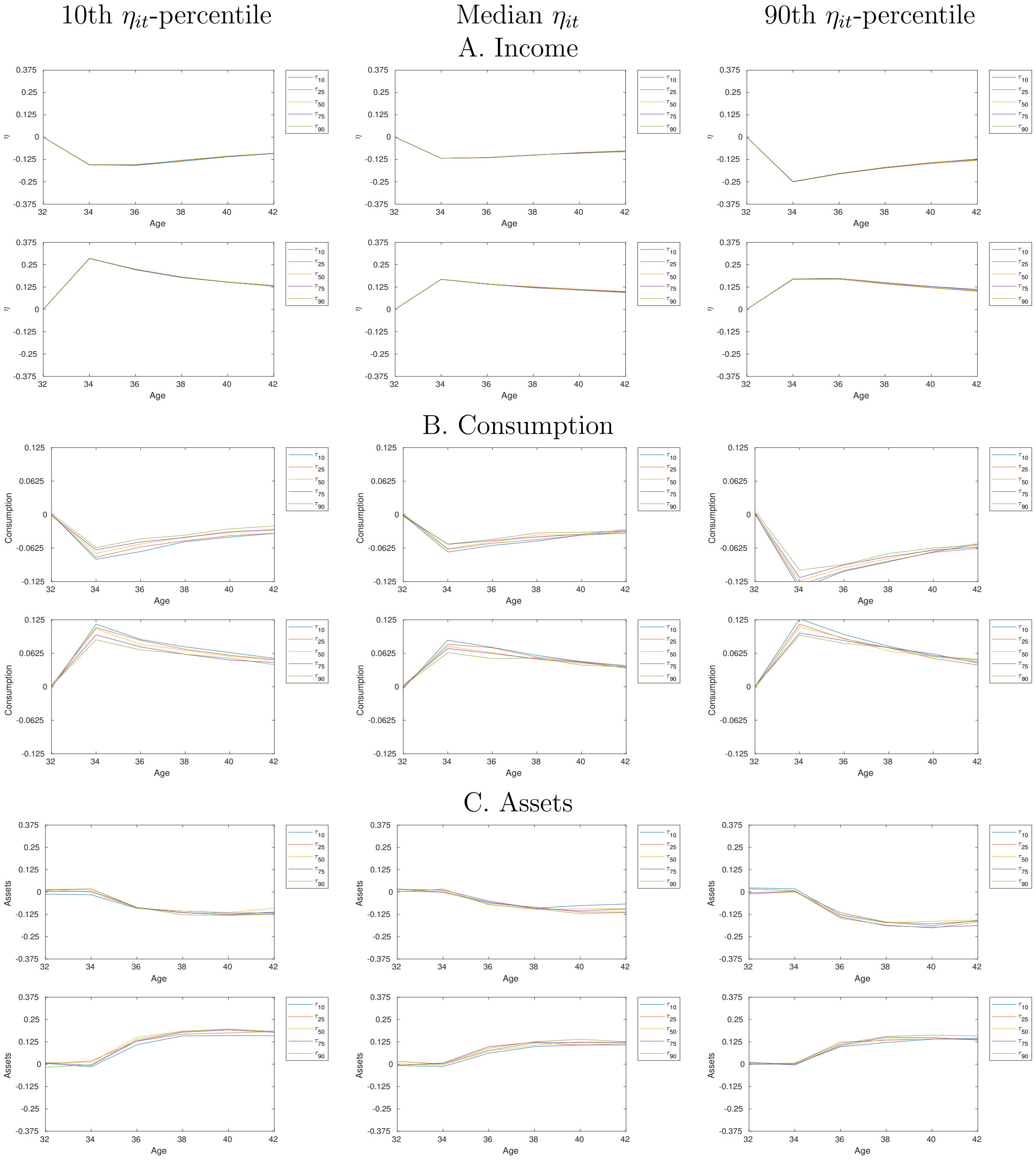
Within each part of the figure, we show impulse responses for various values of initial income and the shock. In the left, middle and right columns we consider households who are at the 10th, 50th and 90th percentile of the distribution of the persistent income component at age 32, respectively. In the top (respectively, bottom) subpanels, we show the results for a shock at the 10th (respectively, 90th) percentile of the distribution of shocks. Hence, top subpanels correspond to negative income shocks, whereas bottom subpanels correspond to positive income shocks.

Focusing first on the upper part of Figure 2.7, and moving across columns, we observe that negative shocks tend to have a stronger impact for those on higher income, and that positive shocks have a stronger impact for those on lower income. This illustrates the nonlinear persistence in the income process documented in ABB. In addition, the fact that all lines corresponding to different values of the latent type ξ_i reflects our assumption that the income process does not depend on ξ_i.

Moving then to the middle part of Figure 2.7, we also observe nonlinearities in consumption responses, although those are stronger for the negative income shocks than for the positive ones. In addition, the differences between lines reflect the heterogeneity between types. In particular, low types with higher income tend to respond more strongly to negative shocks than other types. To further illustrate this heterogeneity, in Appendix Figure 2.24 we show how consumption levels evolve, on impact, after an income shock.

Lastly, focusing on the bottom part of Figure 2.7 we see only moderate differences in assets evolution after a shock depending on the initial income level. In Appendix Figure 2.25 we show impulse responses based on the model that allows for heterogeneity in both assets and consumption, see equations (35)-(36) and the results discussed in Subsection 2.6.3. The responses to a shock to the persistent latent component of income are overall similar to the ones based on the model without heterogeneity in the asset accumulation rule.

Figure 2.7: Heterogeneity in impulse responses



10th η_{it}-percentile

Median η_{it}

90th η_{it}-percentile

A. Income

B. Consumption

C. Assets

Notes: Impulse responses shown for shocks at the 10th (top subpanels) and 90th (bottom subpanels) percentiles, relative to median. See the text for a description. The different lines correspond to different percentiles of ξ_i.

2.7.2 Robustness to the complexity of the quantile model used in estimation

The complexity of our empirical specification is controlled in part by the number of knots at which we evaluate the quantiles of the variables in the model (i.e., the income components, consumption, and the latent type). Our estimates of the functions, such as $a_k^{\eta}(\tau)$ in (37), interpolate between those τ values. Hence, a large number of knots can approximate any continuous quantile function well, while a small number of knots may provide a worse approximation. However, in estimation one faces the usual bias/variance trade-off, and the impact of the number of knots on the estimates is a priori unclear. To probe the sensitivity of our main results to the number of knots, we report average consumption derivatives based on 19 knots in Appendix Figure 2.26. By comparison, our baseline results were obtained using 11 knots (see Figure 2.5). Overall the two sets of estimates agree very well.

2.7.3 Robustness to income definition and sample restriction

Next, we probe the robustness of our results to changes in income definition and sample restriction. While our main results rely on using disposable, post-tax income, in Appendix Figure 2.27 we report results on nonlinear income persistence based on pre-tax labor income. In Appendix Figure 2.28 we report the corresponding results for heterogeneity in consumption responses. The findings suggest a higher degree of nonlinearity in income persistence, and a higher degree of consumption insurance, compared to the results based on disposable income. This is not surprising, as the non-proportionality in the tax system can be interpreted as a source of insurance to households. Moreover, since the results in ABB were based on labor income, these findings help explain the quantitative differences between the results in ABB and the ones we report in this paper when relying on disposable income.

Another important feature of our sample is the restriction to dual earner households. While this restriction is motivated by the goal to abstract from extensive labor supply decisions, it also results in a smaller and potentially more insured sample. We have estimated our model on a larger sample that also includes single earners, where the second member of the household is not working.[30] In Appendix Figure 2.29 we report the results for income persistence, and in Appendix Figure 2.30 we reproduce our main results on heterogeneity in consumption responses to income shocks. Our findings are qualitatively unchanged relative to our baseline sample of dual earners.

2.7.4 Additional dimensions of heterogeneity

Our specification of the consumption function flexibly allows for heterogeneity in income, assets, age effects, and the effect of the latent type, see equation (27). However, it is possible that the effects of additional observed and unobserved factors might matter for consumption insurance. For example, differences in education and birth cohorts might be associated with different consumption responses to income shocks. In Appendix Figure 2.32 we show that neither education nor cohort are strongly associated with the latent type ξ_i. Yet, it is theoretically possible that they enter the consumption function, and interact with income components in meaningful ways, even though our modeling approach rules out this possibility.

To tentatively explore this question, in Appendix Figure 2.33 we report consumption responses to income shocks, by type ξ_i, in a specification that also controls for a fully interacted education indicator. Since we do not re-estimate the model with latent variables, we view this exercise as indicative. We see that the consumption responses across types are qualitatively similar to the baseline ones, yet those responses appear somewhat muted. This motivates future work extending our framework to allow for multiple observed and unobserved sources of heterogeneity in consumption insurance and income processes across

[30] In Appendix Table 2.7 we show descriptive statistics for this broader sample.

households.

2.8 Conclusion

The motivation for this research has been to better understand nonlinear income dynamics and heterogeneous consumption responses to changes in income. In this paper we have developed methods that build on and extend [1], and we have applied them to a larger and more comprehensive sample from the PSID which includes a richer set of consumption categories. We have developed computational tools to better handle larger and more complex models, including in settings with unbalanced panels, within a nonlinear quantile-based latent variables framework. These new data and tools allow us to go beyond confirming the presence of nonlinear income and consumption dynamics, and to document rich heterogeneity in consumption responses across households.

Our results point to consumption responses to income shocks that vary substantially with unobserved types. We distinguish lower types, who appear to follow the life-cycle patterns in consumption responses implied by standard models, from higher types, whose consumption responses to income shocks vary little with either assets levels or the stage of the life cycle. High-type households consistently have higher consumption levels and, relative to low-type households, have slightly higher incomes and levels of assets. For the younger low types, consumption responses to persistent income shocks are close to 0.60 while for older low types this falls to 0.10. For the higher types, consumption responses are flatter across age and assets.

We examined alternative mechanisms that could lead to such heterogeneous consumption responses. The fact that high types both consume more and hold more assets is difficult to reconcile with an explanation based on heterogeneity in preferences or discounting. We also argue that it is difficult to align with a specification that allows for latent heterogeneity in asset accumulation, finding that the heterogeneity in consumption responses is virtually unaffected by this extension. To explore a third mechanism, parental insurance, we used the inter-generational linkages in the PSID to link a subset of household heads in our sample to their parents. We found that high-type household heads have on average parents with higher consumption and income levels, suggesting that the heterogeneous responses might in part reflect heterogeneity in access to other sources of insurance such as parental insurance.

Our findings motivate further work on two fronts. First, whilst we have examined several mechanisms and found a correlation between the latent types and parental consumption, we lack a quantitative understanding of how these and other factors shape the household differences in consumption responses and insurance. Second, although we have leveraged a single-latent-factor model to maintain tractability in the presence of heterogeneous responses, generalizing the model to account for other sources of heterogeneity is an important next step. In particular, it would be valuable to extend the model to allow for time-invariant heterogeneity in income, in addition to the latent consumption type.

2.A Modeling and estimation details

2.A.1 Empirical specification

Earnings components. Let φ_k, for $k = 0, 1, ...$, denote a dictionary of functions, with $\varphi_0 = 1$. In practice we use low-order products of Hermite polynomials for φ_k. We specify, for $t \in \{t_i + 1, ..., t_i + T_i - 1\}$, the conditional quantile function of η_{it} given $\eta_{i,t-1}$ and age_{it} as in (37). We specify the quantile function of ε_{it} (for $t = 1, ..., T$) given age_{it}, and that of

η_{i1} given age at the start of the period age_{i1}, in a similar way. Specifically, we set

$$\begin{aligned} Q_{\varepsilon}(age_{it},\tau) &= \sum_{k=0}^{K} a_k^{\varepsilon}(\tau)\varphi_k(age_{it}), \\ Q_{\eta_1}(cohort_i, educ_i, age_{i,t_i},\tau) &= \sum_{k=0}^{K} a_k^{\eta_1}(\tau)\varphi_k(cohort_i, educ_i, age_{i,t_i}), \end{aligned}$$

with outcome-specific choices for K and φ_k.

Consumption type. To specify the latent type we set

$$Q_{\xi}(cohort_i, educ_i, income_i,\tau) = \sum_{k=0}^{K} a_k^{\xi}(\tau)\varphi_k(cohort_i, educ_i, income_i).$$

Consumption rule. To specify the consumption process we set

$$Q_c(a_{it},\eta_{it},\varepsilon_{it}, age_{it},\xi_i,\tau) = \sum_{k=1}^{K} a_k^{c}(\tau)\varphi_k(a_{it},\eta_{it},\varepsilon_{it}, age_{it},\xi_i). \tag{41}$$

To fix the scale of the function we impose that

$$\int_0^1 Q_c(\overline{a},\overline{\eta},\overline{\varepsilon},\overline{age},\xi,\tau)d\tau = \xi,$$

which translates into linear restrictions on the parameters $\int_0^1 a_k^c(\tau)d\tau$.

Assets evolution. For initial assets we set

$$Q_{a_1}(\eta_{i,t_i}, age_{i,t_i}, cohort_i, educ_i,\xi_i,\tau) = \sum_{k=0}^{K} a_k^{a_1}(\tau)\varphi_k(\eta_{i,t_i}, age_{i,t_i}, cohort_i, educ_i,\xi_i). \tag{42}$$

For assets evolution we set

$$Q_a(a_{it}, \eta_{it}, \varepsilon_{it}, c_{it}, age_{it}, \xi_i, \tau) = \sum_{k=0}^{K} a_k^a(\tau)\varphi_k(a_{it}, \eta_{it}, \varepsilon_{it}, c_{it}, age_{it}, \xi_i, \tau). \tag{43}$$

Implementation. We base our implementation on ABB, and model the functions $a_k(\tau)$ as piecewise-linear interpolating splines on a grid $[\tau_1, \tau_2]$, $[\tau_2, \tau_3]$, ... , $[\tau_{L-1}, \tau_L]$, contained in the unit interval. We extend the specification of the intercept coefficient a_0 on $(0, \tau_1]$ and $[\tau_L, 1)$ using a Laplace model indexed by λ_- (for the left tail) and λ_+ (for the right tail). All a_k for $k \geq 1$ are constant on $[0, \tau_1]$ and $[\tau_L, 1]$, respectively. We denote $a_{k\ell} = a_k(\tau_\ell)$. In practice, we take $L = 11$ and $\tau_\ell = \ell/(L+1)$. We use tensor products of Hermite polynomials for φ_k, each component of the product taking as argument a standardized variable.

2.B Estimation

Overview of the estimation strategy. We start by estimating the earnings parameters. Next, we recover estimates of the consumption, assets, and type parameters, given the previous earnings estimates.

Parameters. We collect all parameters governing the income process into a vector θ, given by

$$\theta = \left(a^\eta, \lambda^\eta, a^\varepsilon, \lambda^\varepsilon, a^{\eta_1}, \lambda^{\eta_1}\right).$$

Likewise, we collect all parameters governing the consumption process into a vector μ, given by

$$\mu = \left(a^\xi, \lambda^\xi, a^c, \lambda^c, a^{a_1}, \lambda^{a_1}, a^a, \lambda^a\right).$$

We estimate θ and μ sequentially.

Model's restrictions Let $\rho_\tau(u) = u(\tau - \mathbf{1}\{u \leq 0\})$ denote the "check" function of quantile regression. Consider the parameters of Q_η; that is, the $a_{k\ell}^\eta$ and the corresponding Laplace parameters λ^η. The true values of $a_{k\ell}^\eta$ maximize

$$E\left[\sum_{t=t_i+1}^{t_i+T_i-1} \int \rho_{\tau_\ell}\left(\eta_t - \sum_{k=0}^{K} a_{k\ell}^\eta \varphi_k(\eta_{t-1}, age_{it})\right) f_i(\eta) d\eta\right] = 0,$$

where f_i is the posterior distribution of the $(\eta_{i,t_i}, ..., \eta_{i,t_i+T_i-1})$ given the data and the true parameter values. In turn, the true values of λ^η satisfy

$$\overline{\lambda}_-^\eta = \\ -\frac{E\left[\sum_{t=t_i+1}^{t_i+T_i-1} \int \mathbf{1}\left\{\eta_t \leq \sum_{k=0}^{K} \overline{a}_{k1}^\eta \varphi_k(\eta_{t-1}, age_{it})\right\} f_i(\eta) d\eta\right]}{E\left[\sum_{t=t_i+1}^{t_i+T_i-1} \int \left(\eta_t - \sum_{k=0}^{K} \overline{a}_{k1}^\eta \varphi_k(\eta_{t-1}, age_{it})\right) \mathbf{1}\left\{\eta_t \leq \sum_{k=0}^{K} \overline{a}_{k1}^\eta \varphi_k(\eta_{t-1}, age_{it})\right\} f_i(\eta) d\eta\right]},$$

with an analogous formula for the upper tail parameter λ_+^η. The model implies related restrictions on all the other quantile and tail parameters in θ and μ.

Likelihood function. The likelihood function is, letting $z_i = (cohort_i, educ_i)$ and $\mathcal{T}_i = \{t_i, ..., t_i + T_i - 1\}$,

$$\begin{aligned}
& f(y_i^{\mathcal{T}_i}, c_i^{\mathcal{T}_i}, a_i^{\mathcal{T}_i}, \eta_i^{\mathcal{T}_i}, \xi_i \,|\, age_i^{\mathcal{T}_i}, z_i; \theta, \mu) \\
& = \prod_{t \in \mathcal{T}_i} f(c_{it} | a_{it}, \eta_{it}, y_{it}, \xi_i, age_{it}; \mu) \\
& \quad \times \prod_{t \in \mathcal{T}_i, t > t_i} f(a_{it} | a_{i,t-1}, y_{i,t-1}, c_{i,t-1}, \eta_{i,t-1}, \xi_i, age_{it}; \mu) \\
& \quad\quad \times \prod_{t \in \mathcal{T}_i} f(y_{it} | \eta_{it}, age_{it}; \theta) \prod_{t \in \mathcal{T}_i, t > t_i} f(\eta_{it} | \eta_{i,t-1}, age_{it}; \theta) \\
& \quad\quad\quad \times f(a_{i,t_i} | \eta_{i,t_i}, age_{i,t_i} z_i, \xi_i; \mu) f(\eta_{i,t_i} \,|\, z_i, age_{i,t_i}; \theta) f(\xi_i \,|\, z_i, income_i; \mu),
\end{aligned}$$

where notice we have imposed the assumption that ξ_i is independent of $(y_i^{\mathcal{T}_i}, \eta_i^{\mathcal{T}_i})$ given $(z_i, income_i)$.

Similarly to ABB, the likelihood function is available in closed form. For example, we have

$$\begin{aligned}
f(y_{it} | \eta_{it}, age_{it}; \theta) \;=\; & \mathbf{1}\{y_{it} - \eta_{it} < A_{it}^{\varepsilon}(1)\} \, \tau_1 \lambda_-^{\varepsilon} \exp\left[\lambda_-^{\varepsilon}\left(y_{it} - \eta_{it} - A_{it}^{\varepsilon}(1)\right)\right] \\
& + \sum_{\ell=1}^{L-1} \mathbf{1}\{A_{it}^{\varepsilon}(\ell) \leq y_{it} - \eta_{it} < A_{it}^{\varepsilon}(\ell+1)\} \frac{\tau_{\ell+1} - \tau_\ell}{A_{it}^{\varepsilon}(\ell+1) - A_{it}^{\varepsilon}(\ell)} \\
& + \mathbf{1}\{A_{it}^{\varepsilon}(L) \leq y_{it} - \eta_{it}\} (1 - \tau_L) \lambda_+^{\varepsilon} \exp\left[-\lambda_+^{\varepsilon}\left(y_{it} - \eta_{it} - A_{it}^{\varepsilon}(L)\right)\right],
\end{aligned}$$

where

$$A_{it}^{\varepsilon}(\ell) \equiv \sum_{k=0}^{K} a_{k\ell}^{\varepsilon} \varphi_k(age_{it}) \text{ for all } (i, t, \ell).$$

Note that the likelihood function is non-negative by construction. In particular, drawing from the posterior density of η automatically produces rearrangement of the various quantile curves ([82]).

Estimation algorithm. Like in ABB, starting from initial parameter values, we iterate between two steps.

In the stochastic E-step, we draw M values $\eta_i^{(m)} = (\eta_{i,t_i}^{(m)}, ..., \eta_{i,t_i+T_i-1}^{(m)})$ and $\xi_i^{(m)}$ from their posterior distribution. In practice we take $M = 1$.

In the M-step, we estimate parameters by solving empirical counterparts of the population restrictions. This involves running multiple quantile regressions in order to estimate the $a_{k\ell}$ parameters, and estimating the λ parameters which are available in closed form.

Solving the indeterminacy in consumption. To impose the restriction (38), which solves the indeterminacy in the relationship between consumption and the latent type, we proceed as follows. At the start of every M-step, given draws $\eta_i^{(m)}$ and $\xi_i^{(m)}$, we regress c_{it} on polynomials in a_{it}, $\eta_{it}^{(m)}$, $\varepsilon_{it}^{(m)} = y_{it} - \eta_{it}^{(m)}$, age_{it}, and $\xi_i^{(m)}$, using the same polynomial specification as in the quantile model for log-consumption. Letting $\widehat{c}_{it}$ denote the predicted value at $(\overline{a}, \overline{\eta}, \overline{\varepsilon}, \overline{age}, \xi_i^{(m)})$, we then reset $\widehat{c}_{it} \mapsto \xi_i^{(m)}$.

Stochastic E-step (income estimation). The target for a given household i is the posterior distribution

$$f(\eta_{i,t_i}, ..., \eta_{i,t_i+T_i-1} | y_{i,t_i}, ..., y_{i,t_i+T_i-1}).$$

At $t = t_i$, we initialize S particles $\eta_{i,t_i}^{(s)}$ from the following proposal distribution π:

$$\begin{aligned}
&\eta_{i,t_i} \sim \mathcal{N}(\mu_i, \sigma^2),\\
&\mu_i = \left(1 - \frac{\sigma_{\eta_1}^2}{\sigma_{\eta_1}^2 + \sigma_\varepsilon^2}\right) \sum_{k=0}^{K} \beta_k^\varepsilon \varphi_k(cohort_i, educ_i, age_{i,t_i}) + \frac{\sigma_{\eta_1}^2}{\sigma_{\eta_1}^2 + \sigma_\varepsilon^2} y_{i,t_i},\\
&\sigma^2 = \frac{c}{\frac{1}{\sigma_{\eta_1}^2} + \frac{1}{\sigma_\varepsilon^2}},
\end{aligned}$$

where the β_k^ε, $\sigma_{\eta_1}^2$ and σ_ε^2 are parameters estimated by running OLS counterparts to the M-step quantile regressions (in the previous stochastic EM iteration), and $c \geq 1$ is a constant (we take $c = 2$). Time $t = t_i$ re-sampling weights are then given by

$$w_{i,t_i}^{(s)} \propto \frac{f(\eta_{i,t_i}^{(s)}|y_{i,t_i})}{\pi(\eta_{i,t_i}^{(s)})},$$

where π is the normal density with mean μ_i and variance σ^2. These weights, which are available in closed form, are used to re-sample particles with replacement from the set of particles $\eta_{i,t_i}^{(s)}$, if the effective sample size $\frac{1}{\sum_{s=1}^{S}(w_{i,t_i}^{(s)})^2}$ exceeds some threshold (see below). This simple adaptive rule avoids degeneracy of the particles. After re-sampling we reset $w_{i,t_i}^{(s)} = \frac{1}{S}$. Otherwise we keep all the existing particles and weights.

At $t = t_i + r > t_i$, we use the following proposal distribution, again denoted as π, to generate new draws to append to the existing set of particles:

$$\begin{aligned}
\eta_{i,t_i+r} \,|\, \eta_{i,t_i+r-1} &\sim \mathcal{N}(\widetilde{\mu}_{i,r}, \widetilde{\sigma}^2), \\
\widetilde{\mu}_{i,r} &= \left(1 - \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2}\right) \sum_{k=0}^{K} \beta_k^\varepsilon \varphi_k(\eta_{i,t+r-1}, age_{i,t_i+r}) + \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2} y_{i,t_i+r}, \\
\widetilde{\sigma}^2 &= \frac{c}{\frac{1}{\sigma_\eta^2} + \frac{1}{\sigma_\varepsilon^2}},
\end{aligned}$$

where again the β_k^ε, σ_η^2 and σ_ε^2 are parameters estimated by running OLS counterparts to the M-step quantile regressions. The re-sampling weights are given by

$$w_{i,t_i+r}^{(s)} \propto w_{i,t_i+r-1}^{(s)} \frac{f(\eta_{i,t_i+r}^{(s)}|y_{i,t_i+r}, \eta_{i,t_i+r-1}^{(s)})}{\pi(\eta_{i,t_i+r}^{(s)}|\eta_{i,t_i+r-1}^{(s)})},$$

which are used to re-sample particles if the effective sample size $\frac{1}{\sum_{s=1}^{S}\left(w_{i,t_i+r}^{(s)}\right)^2}$ exceeds the threshold.

Stochastic E-step (consumption estimation). The target for a given household is the posterior distribution

$$f(\xi_i, \eta_{i,t_i}, ..., \eta_{i,t_i+T_i-1} \,|\, x_{i,t_i}, ..., x_{i,t_i+T_i-1}),$$

where $x_{it} = (y_{it}, c_{it}, a_{it}, age_{it})$ is a vector of household i's observed income, consumption, assets and age at time t. Algorithm 1 below provides a pseudo-code for the implementation. The SMC sampling steps (used to generate efficient proposals within a Metropolis Hastings algorithm) are identical to those outlined above with the exception that re-sampling weights at times $t = t_i$ and $t > t_i$ are now given by

$$w_{i,t_i}^{(s)} \propto \frac{f\left(\eta_{i,t_i}^{(s)} | \xi_i^*, x_{i,t_i}\right)}{\pi\left(\eta_{i,t_i}^{(s)}\right)},$$

and

$$w_{i,t_i+r}^{(s)} \propto \frac{f\left(\eta_{i,t_i+r}^{(s)} | \xi_i^*, x_{i,t_i+r}, \eta_{i,t_i+r-1}^{(s)}\right)}{\pi\left(\eta_{i,t_i+1}^{(s)} | \eta_{i,t_i}^{(s)}\right)},$$

respectively, where ξ_i^* is a draw from a random walk proposal. We make use of the same proposal distributions π as in the income estimation.

In the very first iteration of the stochastic EM algorithm we initialize the Metropolis Hastings chains using random draws from the following proposal:

$$\xi_i^* \sim \mathcal{N}(\nu_i, \omega^2),$$

where $\nu_i = \sum_{k=0}^{K} \beta_k^{\xi} \varphi_k(cohort_i, educ_i, income_i)$. The parameters β_k^{ξ} and ω^2 are estimated by running OLS counterparts to the corresponding M-step quantile regressions. At subsequent iterations of the stochastic EM we initialize the Metropolis Hastings chains using draws from the previous iteration. After initialization we use a Gaussian random walk proposal with variance $3.5\omega^2$.

Whilst running the SMC samplers we obtain unbiased estimates of the marginal likelihood which can be calculated recursively as $\hat{p}(x_{i,t_i+r},...|\xi_i^*) = \sum_{s=1}^{S} \hat{p}(x_{i,t_i+r-1},...|\xi_i^*) w_{i,t_i+r}^{(s)}$. The unbiasedness of these marginal likelihood estimates implies that the resulting algorithm can be represented as a bona fide Metropolis Hastings algorithm yielding the desired target as its marginal.

Pseudo-code of the stochastic EM algorithm. A short pseudo-code for the algorithm we use is presented in Algorithm 1.

Algorithm 1. *(Stochastic EM)*

```
1: for $\ell$=1:L do
2:     Stochastic E-Step:
3:     Set $\xi_i^0$ and $(\eta^0_{i,t_i},...,\eta^0_{i,t_i+T_i-1})$ to some starting values.[31]
4:     for k=1:K do
5:         Sample $\xi_i^* \sim q(.|\xi_i^{k-1})$, where q is a proposal distribution.[32]
6:         Run an SMC algorithm targeting $p(\eta_{i,t_i},...,\eta_{i,t_i+T_i-1}|\xi_i^*, w_{i,t_i},...,w_{i,t_i+T_i-1})$.
7:         Store the marginal likelihood estimate, $\hat{p}(\xi_i^*) = p(w_{i,t_i},...,w_{i,t_i+T_i-1}|\xi_i^*)$, and the
           resulting particles $\eta^*_{i,t_i},...,\eta^*_{i,t_i+T_i-1}$, both of which are available as output of the SMC
           algorithm in line 6.
8:         Let f denote the density of $\xi_i$, whose expression is given in Appendix 2.A. With
           probability $\min\left(1, \frac{\hat{p}(\xi_i^*)f(\xi_i^*)q(\xi_i^{k-1}|\xi_i^*)}{\hat{p}(\xi_i^{k-1})f(\xi_i^{k-1})q(\xi_i^*|\xi_i^{k-1})}\right)$ set $\xi_i^k = \xi_i^*$ and
           $(\eta^k_{i,t_i},...,\eta^k_{i,t_i+T_i-1}) = (\eta^*_{i,t_i},...,\eta^*_{i,t_i+T_i-1})$; otherwise set $\xi_i^k = \xi_i^{k-1}$ and $(\eta^k_{i,t_i},...,\eta^k_{i,t_i+T_i-1}) =$
```

[31] When $\ell > 1$ we simply take ξ_i^0 to be the ξ_i draw from the previous $(\ell - 1)$ step. When $\ell = 1$ we always accept the first proposal. In both cases, we run an SMC algorithm (see line 6 in the pseudo-code) based on ξ_i^0 to generate a draw $(\eta^0_{i,t_i},...,\eta^0_{i,t_i+T_i-1})$.

[32] In practice, we use a random walk proposal. We tune the variance of the proposal so that the acceptance rate is approximately 30%.

```
        $(\eta_{i,t_i}^{k-1}, ..., \eta_{i,t_i+T_i-1}^{k-1})$.
    end for
    Keep the last values $\xi_i^K$ and $(\eta_{i,t_i}^{K}, ..., \eta_{i,t_i+T_i-1}^{K})$.
    M-Step:
    Estimate the quantile parameters by quantile regressions given the draws $\xi_i^K$ and
        $(\eta_{i,t_i}^{K}, ..., \eta_{i,t_i+T_i-1}^{K})$, as explained in Appendix 2.A. Estimate the Laplace tail parameters.
    Update the parameters of the proposal distribution, as explained in Appendix 2.A.
end for
```

Practical issues: number of particles and threshold for effective sample size. In practice, we set an i-specific number of particles equal to $S_i = 50T_i$, where T_i is the number of observations of household i. We set the threshold for effective sample size to $S_i/2$.

Practical issues: specification. In practice we set the following polynomial degrees K for our baseline specification, chosen after some experimentation:

- Q_η: $K^\eta = 3$, $K^{age} = 2$.
- Q_{η_1}: $K^{educ} = 1$, $K^{cohort} = 1$, $K^{age} = 2$.
- Q_ε: $K^{age} = 2$.
- Q_c: $K^{age} = 1$, $K^a = 2$, $K^\eta = 2$, $K^\varepsilon = 1$, $K^\xi = 1$.
- Q_a: $K^{age} = 1$, $K^a = 2$, $K^y = 1$, $K^c = 1$.
- Q_{a_1}: $K^{age} = 1$, $K^\eta = 1$, $K^\xi = 1$, $K^{education} = 1$, $K^{cohort} = 1$.
- Q_ξ: $K^{income} = 1$, $K^{educ} = 1$, $K^{cohort} = 1$.

Practical issues: starting values. In practice, we start the algorithm from different parameter values. For example, for the initial values of the quantile parameters in η_{it}, we run quantile regressions of log-earnings on lagged log-earnings and age. We proceed similarly to set other starting parameter values, including those for the proposal distributions. In addition, we use latent draws from the income model as initial draws when estimating the consumption model. We experimented with a number of other choices.

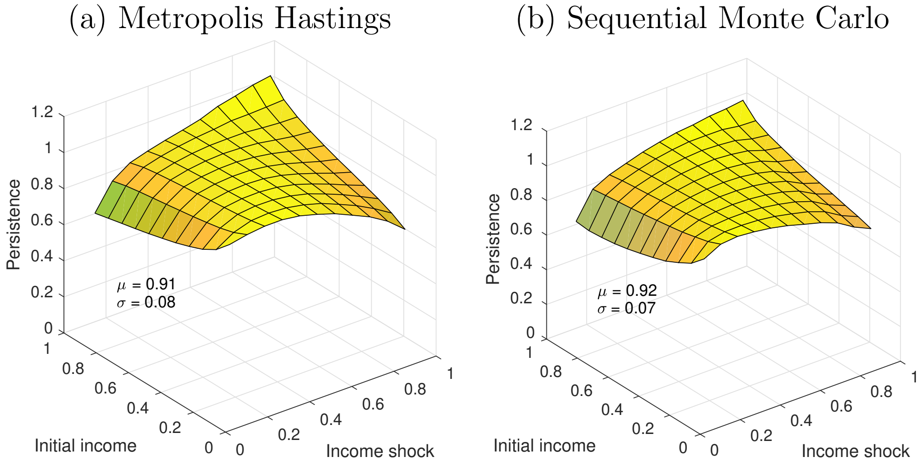
Practical issues: numerical performance. Our aim is to ensure that the stochastic EM parameter Markov chains mix well. Among the factors that influence mixing (as measured by the decay rate of auto-correlations along the parameter Markov chains), we found three key ones to be the number of particles, the length of the Metropolis chains, and the number of iterations in the overall EM algorithm. Given our experiments, we found that setting moderate numbers for the first two (we set $S_i = 50T_i$ particles, as indicated above, and we run each Metropolis chain for 50 iterations), and relatively large numbers for the third (we run the stochastic EM for 2000 iterations), gave best performance given computation constraints in our short panel data setting.

2.C Numerical comparison with ABB

The SMC approach differs from the Metropolis Hastings method that was used in ABB. Here we compare the income persistence implied by SMC and Metropolis Hastings, when using the original 6-wave balanced panel from ABB.

In Figure 2.8 we show the nonlinear income persistence predicted by the algorithm using SMC, and compare it to the estimates based on the Metropolis Hastings algorithm from ABB. We see that the results are little affected by the change in method. In particular, we see that, for households with a low persistent income component, high shocks are associated with less income persistence, and for households with a high persistent income component,

Figure 2.8: Comparing Metropolis Hastings and Sequential Monte Carlo in the balanced panel used in [1]



(a) Metropolis Hastings

(b) Sequential Monte Carlo

Notes: 6-wave balanced sample from the PSID used in ABB, 1999-2009. The graphs show the quantile derivatives of the persistent income component η_{it} with respect to η_{it-1}, averaged over ages in the sample. In the left graph we show the result obtained using a Metropolis Hastings, using the codes from ABB. In the right graph we show the results obtained using the Sequential Monte Carlo algorithm. The two horizontal axes show percentiles of η_{it-1} ("initial income") and conditional percentiles of η_{it} given η_{it-1} ("income shock"), respectively.

low shocks are associated with more income persistence. These patterns differ from the implications of a linear process such as a random walk, where income persistence would be flat, independent of both the income level and the income shock. Formally, the income persistence measure proposed by ABB is, in the case of the persistent income component η_{it},

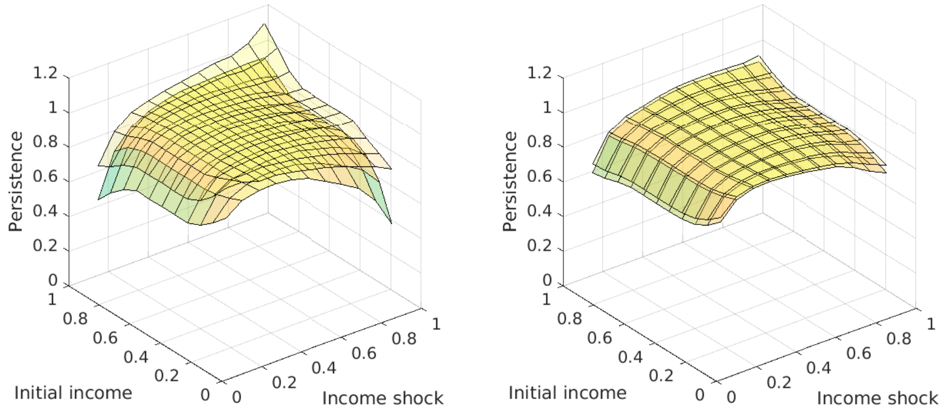
$$\rho(\eta, age, \tau) = \frac{\partial Q_\eta(\eta, age, \tau)}{\partial \eta}, \quad \tau \in (0,1), \tag{44}$$

where Q_η is the quantile function appearing in (30).[33]

The income persistence results reported in ABB are based on comparing various estimation runs, and selecting the one that provides the highest value of the likelihood. However, compared to Metropolis Hastings used in ABB, we found the SMC approach to be more effec-

[33] Note that $\rho(\eta, age, \tau)$ also depends on age, which we average out in Figure 2.8.

Figure 2.9: Pointwise numerical stability bands of nonlinear persistence estimates



(a) Metropolis Hastings

(b) Sequential Markov Chain

Notes: 6-wave balanced sample from the PSID used in ABB, 1999-2009. The graphs show the quantile derivatives of the persistent income component η_{it} with respect to η_{it-1}, averaged over ages in the sample, and evaluated pointwise at the 2.5th and 97.5th percentiles over 200 runs of the stochastic EM algorithm, using different seeds every time. In the left graph we show the result obtained using a Metropolis Hastings sampler, using the codes from ABB. In the right graph we show the results obtained using the Sequential Monte Carlo algorithm. The two horizontal axes show percentiles of η_{it-1} ("initial income") and conditional percentiles of η_{it} given η_{it-1} ("income shock"), respectively.

tive at reducing the numerical instability across estimation runs. To illustrate this, in Figure 2.9 we report numerical stability bands that indicate the variability of income persistence estimates obtained from 200 runs of our estimation algorithm using different seeds, based on the two different sampling methods. In the left graph of the figure we report results based on Metropolis Hastings. In the right graph we report results based on the SMC algorithm we rely on in this paper. The SMC results show substantially less numerical variability.

Lastly, although reported estimates in ABB appear reliable in the shorter balanced sample, in our experience increasing the number of households and the length of the panel makes it more challenging to rely on Metropolis Hastings for sampling. In contrast, we found our SMC implementation to remain numerically stable in such cases.

2.D Which features of the consumption policy rule can be identified?

Consider a structural policy rule of the form

$$C = g(X, \nu),$$

where ν, of unrestricted dimension, is independent of X. To simplify the presentation we assume that X is scalar. In this paper, C denotes consumption, and X contains all state variables, including the income components. Denote the conditional quantile function of C given X as $Q(X, \tau)$. Hence, for U uniform independent of X, we can write

$$C = Q(X, U).$$

We are interested in moments of the marginal effects

$$\Delta_x C = \frac{\partial g(x,\nu)}{\partial x}.$$

The key challenge is that, while Q is identified from data on (C, X), g is generally not.

Average responses. We have, under standard conditions,

$$\mathbb{E}\left[\Delta_x C\right] = \frac{\partial}{\partial x}\mathbb{E}\left[g(x,\nu)\right],$$

hence

$$\mathbb{E}\left[\Delta_x C\right] = \frac{\partial}{\partial x}\mathbb{E}\left[C \,|\, X = x\right],$$

or, equivalently,

$$\mathbb{E}\left[\Delta_x C\right] = \frac{\partial}{\partial x}\mathbb{E}\left[Q(x,U)\right],$$

that is,

$$\mathbb{E}\left[\Delta_x C\right] = \mathbb{E}\left[\frac{\partial}{\partial x}Q(x,U)\right].$$

Hence, average marginal effects are identified, irrespective of the dimensionality of ν and the monotonicity properties of g.

Variance of responses. By Theorem 2.1 in [67] we have

$$\mathbb{E}\left[\Delta_x C \,|\, X = x, C = Q(x,\tau)\right] = \frac{\partial}{\partial x}Q(x,\tau),$$

for all τ and x. We thus can write

$$\frac{\partial}{\partial x}Q(x,U) = \mathbb{E}\left[\Delta_x C\right] + V,$$

where

$$V = \frac{\partial}{\partial x} Q(x, U) - \mathbb{E}\left[\Delta_x C\right].$$

Now, V has mean zero, and variance

$$\begin{aligned} \operatorname{Var}\left(\frac{\partial}{\partial x} Q(x, U)\right) &= \operatorname{Var}\left(\mathbb{E}\left[\Delta_x C \mid X = x, C = Q(x, U)\right]\right) \\ &= R^2 \operatorname{Var}\left(\Delta_x C\right), \end{aligned}$$

where R^2 corresponds to the nonparametric regression of $\Delta_x C$ on C and X. Hence, the variance of $\frac{\partial}{\partial x} Q(x, U)$ underestimates the variance of $\Delta_x C$, by an amount that depends on how well C and X explain $\Delta_x C$.

For example, if ν is scalar and has a monotone effect on g, then $R^2 = 1$ and the variances are equal. In that case, $Q = g$, and g is identified. More generally, even though g is may not be identified, the mean of $\frac{\partial g(x,\nu)}{\partial x}$ is identified and one can compute a lower bound on the variance of $\frac{\partial g(x,\nu)}{\partial x}$.

2.E Additional tables and figures

2.E.1 Tables and figures for Section 2.2

Table 2.4: Additional descriptive statistics about the unbalanced panel

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Waves 1	Waves 2	Waves 3	Waves 4	Waves 5	Waves 6	Waves 7
Age	38.29	39.66	40.58	40.95	40.90	40.08	38.25
	(10.51)	(10.70)	(10.13)	(9.55)	(9.23)	(8.52)	(6.71)
Education	4.88	4.95	5.00	5.06	5.05	5.16	5.14
	(1.09)	(1.10)	(1.10)	(0.99)	(1.05)	(0.95)	(0.98)
Kids	1.22	1.05	1.08	0.90	1.02	1.17	1.35
	(1.16)	(1.15)	(1.17)	(1.00)	(1.22)	(1.08)	(1.01)
Food	10,224.82	10,297.24	10,231.33	10,417.36	10,618.86	10,873.81	10,339.80
	(5,618.54)	(4,871.19)	(4,884.57)	(5,322.02)	(5,205.25)	(5,295.29)	(5,566.94)
Non-durables (excl. food)	24,446.69	25,271.07	27,640.10	26,705.96	27,597.78	29,553.66	27,365.30
	(23,423.94)	(14,975.07)	(20,170.04)	(19,519.81)	(18,453.14)	(18,044.51)	(18,732.42)
Total Non-durables	34,818.81	35,657.00	37,929.68	37,137.98	38,269.81	40,427.48	37,731.48
	(26,171.72)	(17,197.60)	(22,674.04)	(22,345.36)	(21,752.17)	(20,778.80)	(21,432.31)
Home equity	94,353.18	93,634.64	134445.57	142168.95	146854.98	144322.37	145431.99
	(221908.96)	(157549.09)	(218194.70)	(196533.44)	(231684.48)	(171917.47)	(182450.48)
Negative Equity Dummy	0.03	0.01	0.02	0.02	0.03	0.02	0.01
	(0.16)	(0.12)	(0.13)	(0.12)	(0.16)	(0.15)	(0.10)
Wealth (excl. home)	236379.23	151718.99	192237.00	207947.79	245846.12	149537.06	144836.15
	(1.85e+06)	(452508.81)	(480574.21)	(1.03e+06)	(713417.52)	(437249.71)	(607971.51)
Total wealth	369397.05	283854.13	387068.75	414928.35	464594.00	349645.50	352285.71
	(2.26e+06)	(648961.96)	(714758.54)	(1.30e+06)	(1.00e+06)	(604613.97)	(791042.37)
Labor income	105504.37	106842.60	121094.14	134196.63	136728.34	118852.50	117218.18
	(131690.40)	(90,625.35)	(131051.85)	(226750.06)	(132471.38)	(65,851.87)	(53,500.79)
Net income	83,800.29	84,063.03	92,061.60	100974.16	101662.12	90,869.67	90,045.96
	(80,287.10)	(57,270.53)	(80,307.77)	(132837.00)	(79,228.69)	(42,442.62)	(34,698.92)
Observations	1002	668	484	263	223	177	299

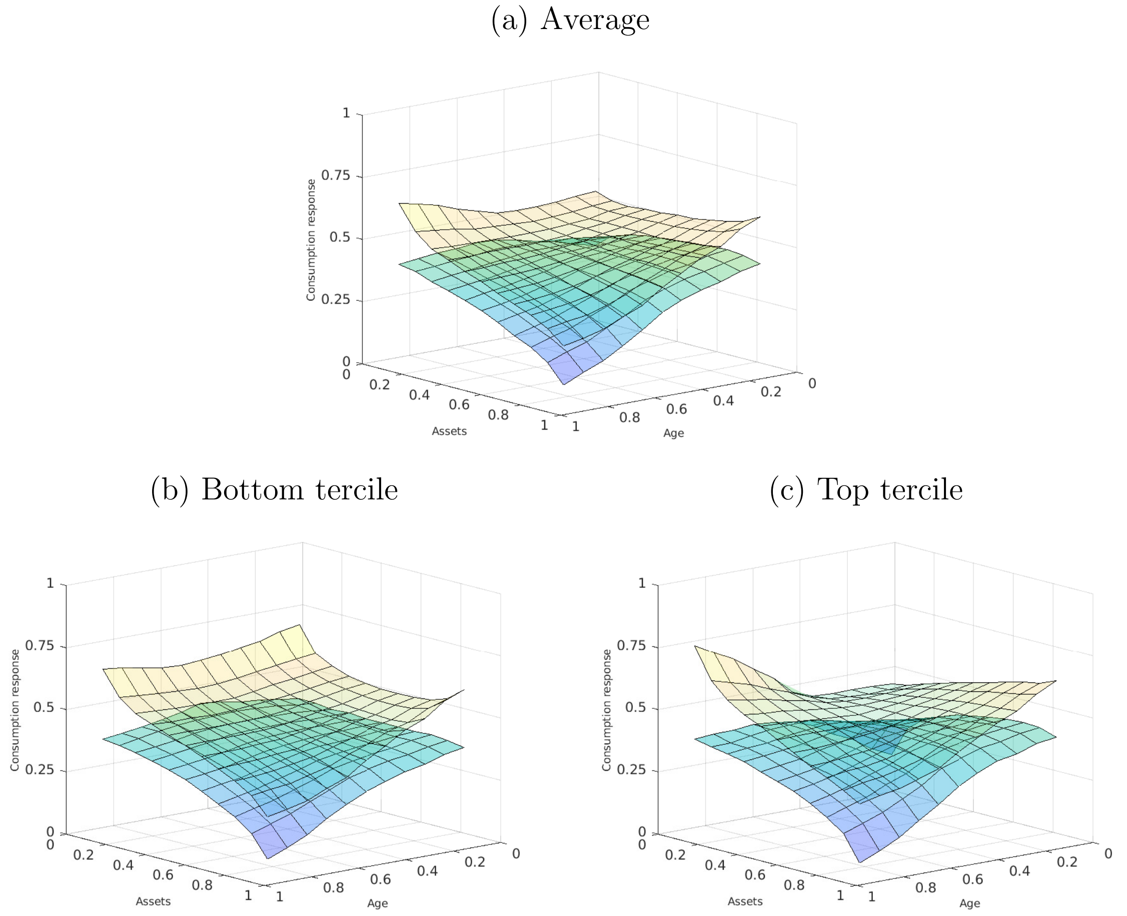
Notes: PSID, 2005-2017. Means of variables, standard deviations in parentheses.

Table 2.5: Additional descriptive statistics about the main sample, including negative assets

	(1) 2005	(2) 2007	(3) 2009	(4) 2011	(5) 2013	(6) 2015	(7) 2017
Food	10,632.46	10,584.68	10,231.23	10,517.05	10,701.87	11,154.70	11,761.09
	(5,299.94)	(5,480.66)	(4,985.32)	(5,039.51)	(5,575.75)	(5,262.43)	(5,514.86)
Non-durables (excl. food)	28,005.27	29,138.06	27,784.64	28,336.56	30,089.00	29,597.75	28,312.56
	(18,936.74)	(19,416.39)	(18,768.90)	(17,696.35)	(17,860.16)	(17,018.40)	(14,572.47)
Total Non-durables	38,669.31	39,750.28	38,081.61	38,921.75	40,869.60	40,824.46	40,119.38
	(21,699.98)	(22,033.74)	(21,113.38)	(20,391.50)	(20,440.32)	(19,538.10)	(17,414.63)
Home equity	150404.41	156582.18	117029.77	97,240.09	91,229.90	94,851.48	108298.61
	(212201.08)	(224409.31)	(192280.79)	(161856.05)	(146908.40)	(135356.38)	(135913.11)
Negative Equity Dummy	0.01	0.01	0.07	0.08	0.06	0.02	0.02
	(0.12)	(0.12)	(0.26)	(0.28)	(0.24)	(0.14)	(0.12)
Wealth (excl. home)	188962.86	255179.00	230841.97	201148.10	183919.95	203580.64	272524.94
	(683870.01)	(964936.86)	(874673.16)	(497471.78)	(476877.85)	(519691.15)	(1.01e+06)
Total wealth	411875.17	470628.74	384224.26	314392.55	279919.36	298432.12	368142.89
	(940132.05)	(1.20e+06)	(1.07e+06)	(617956.79)	(578419.27)	(590056.80)	(1.01e+06)
Labor income	122972.70	124391.48	126510.00	123237.46	121745.86	120544.04	125475.14
	(139187.13)	(143195.31)	(182296.90)	(119741.17)	(118132.57)	(72,546.62)	(66,226.60)
Net income	93,504.28	94,804.55	95,893.52	95,289.90	94,087.25	92,224.02	95,572.56
	(83,501.16)	(86,771.32)	(109386.52)	(73,204.74)	(71,919.39)	(46,205.59)	(43,329.49)
Observations	1397	1684	1616	1399	1269	1192	968

Notes: PSID, 2005-2017. Means of variables, standard deviations in parentheses.

Figure 2.10: Consumption responses at various quantiles, confidence bands



(a) Average

(b) Bottom tercile

(c) Top tercile

Notes: See the notes to Figures 2.1 and 2.2. Bootstrapped pointwise 95% confidence bands clustered at the household level.

2.E.2 Tables and figures for Section 2.5

Table 2.6: Summarizing heterogeneity across types, nonparametric bootstrap

A. 90th vs 10th percentile of ξ

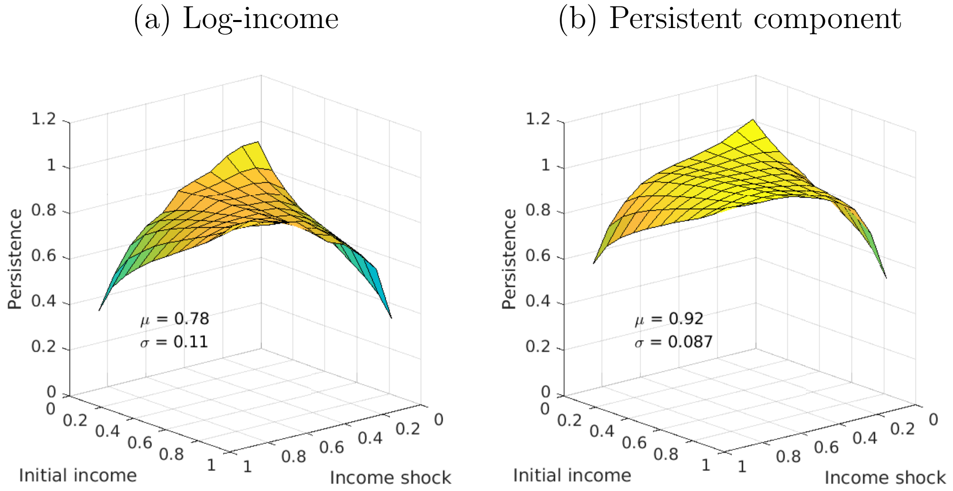
	Young, low assets	Old, high assets	Δ
High ξ	0.31	0.22	0.09
	[0.08,0.48]	[0.04,0.44]	[-0.15,0.22]
Low ξ	0.48	0.21	0.27
	[0.27,0.68]	[0.04,0.42]	[0.01,0.43]
Δ	-0.17	0.01	-0.18
	[-0.56,0.08]	[-0.27,0.28]	[-0.54,0.08]

B. 75th vs 25th percentile of ξ

	Young, low assets	Old, high assets	Δ
High ξ	0.36	0.21	0.15
	[0.19,0.46]	[0.08,0.40]	[-0.02,0.24]
Low ξ	0.45	0.21	0.24
	[0.27,0.58]	[0.09,0.39]	[0.02,0.34]
Δ	-0.09	0.00	-0.09
	[-0.27,0.04]	[-0.13,0.14]	[-0.25,0.04]

Notes: See the notes to Figure 2.4. Here we report average consumption responses for young and low assets households compared to old and high assets households, for different percentiles of heterogeneity ξ_i in consumption. Values are calculated by evaluating the average consumption response for households at a fixed percentile of ξ_i when assets and age are fixed at the τth percentile. Reported values for young and low assets households are then shown by averaging over $\tau \in (0, 0.5)$. Reported values for old and high assets households are then shown by averaging over $\tau \in (0.5, 1)$. Nonparametric bootstrap 95% confidence intervals clustered at the household level based on 200 replications are shown in brackets.

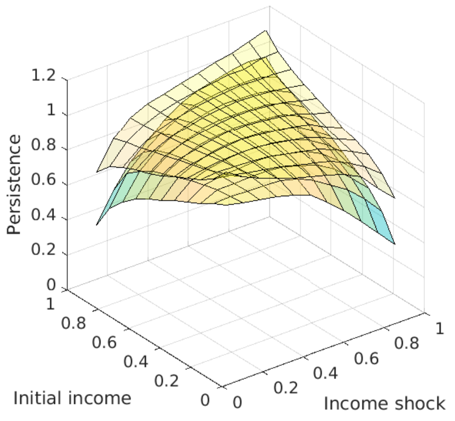
Figure 2.11: Nonlinear income persistence


(a) Log-income

(b) Persistent component

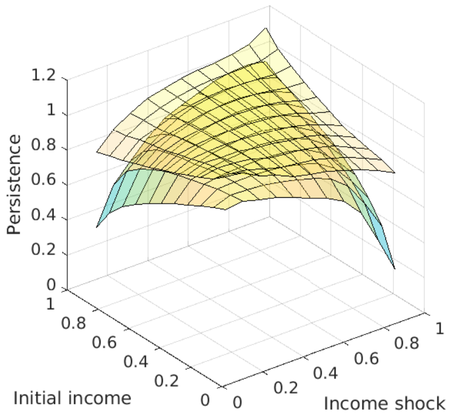
Notes: PSID, 2005-2017 sample, disposable income, dual earners from an alternative perspective. The left graph shows quantile derivatives of log-income with respect to lagged log-income. The right graph shows quantile derivatives of the persistent latent component η_{it} with respect to η_{it-1}, model estimated using sequential Monte Carlo with a stochastic EM algorithm. The two horizontal axes show percentiles of η_{it-1} ("initial income") and conditional percentiles of η_{it} given η_{it-1} ("income shock"), respectively.

Figure 2.12: Nonlinear persistence in η_{it}, 95% pointwise confidence bands (parametric bootstrap)



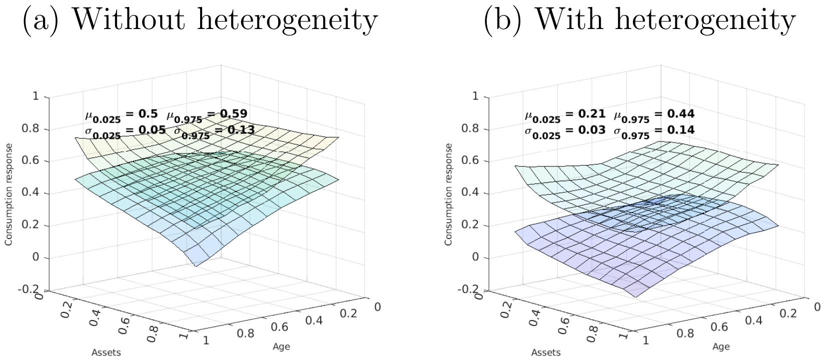
Notes: Pointwise 95% confidence bands based on the parametric bootstrap. 200 replications.

Figure 2.13: Nonlinear persistence in η_{it}, 95% pointwise confidence bands (nonparametric bootstrap)



Notes: Pointwise 95% confidence bands based on nonparametric bootstrap. 200 replications. Bootstrap is clustered at the household level.

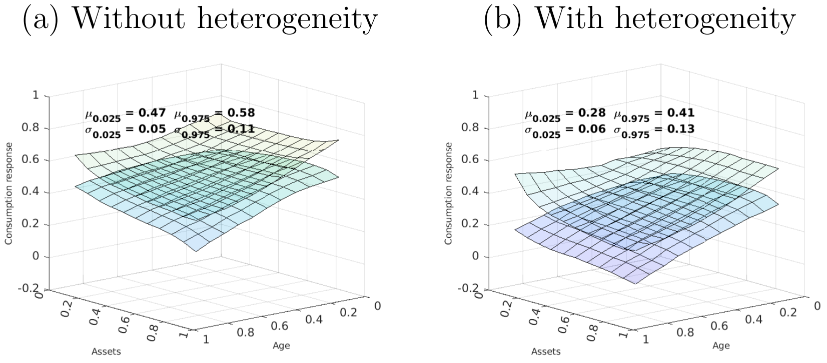
Figure 2.14: Average insurance in model with and without heterogeneity 95% pointwise confidence bands (nonparametric bootstrap)


(a) Without heterogeneity

(b) With heterogeneity

Notes: Pointwise 95% confidence bands based on nonparametric bootstrap. 200 replications. Bootstrap is clustered at the household level.

Figure 2.15: Average insurance in model with and without heterogeneity 95% pointwise confidence bands (parametric bootstrap)

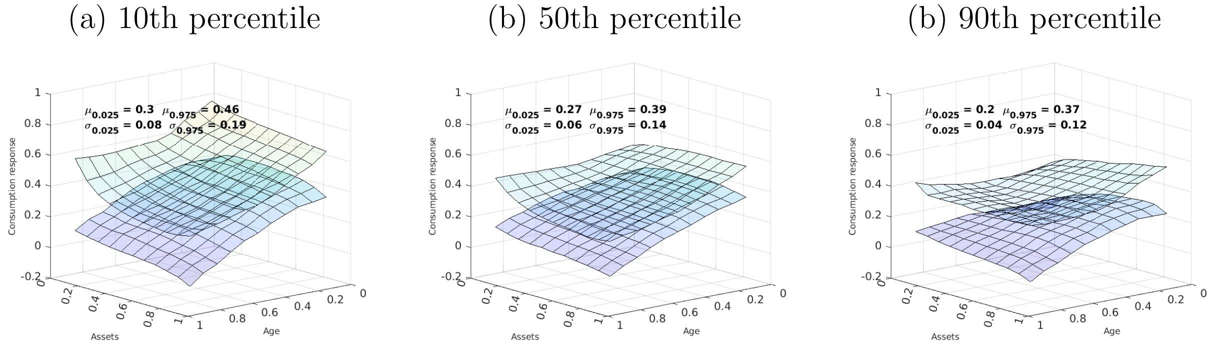


(a) Without heterogeneity

(b) With heterogeneity

Notes: Pointwise 95% confidence bands based on parametric bootstrap.

Figure 2.16: Heterogeneity in consumption responses, 95% pointwise confidence bands (parametric bootstrap)



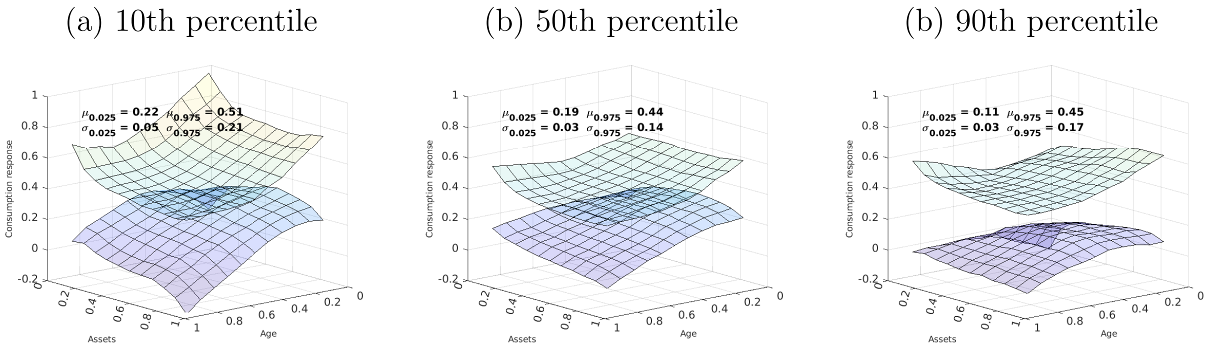
(a) 10th percentile

(b) 50th percentile

(b) 90th percentile

Notes: Pointwise 95% confidence bands based on parametric bootstrap. 200 replications.

Figure 2.17: Heterogeneity in consumption responses, 95% pointwise confidence bands (nonparametric bootstrap)

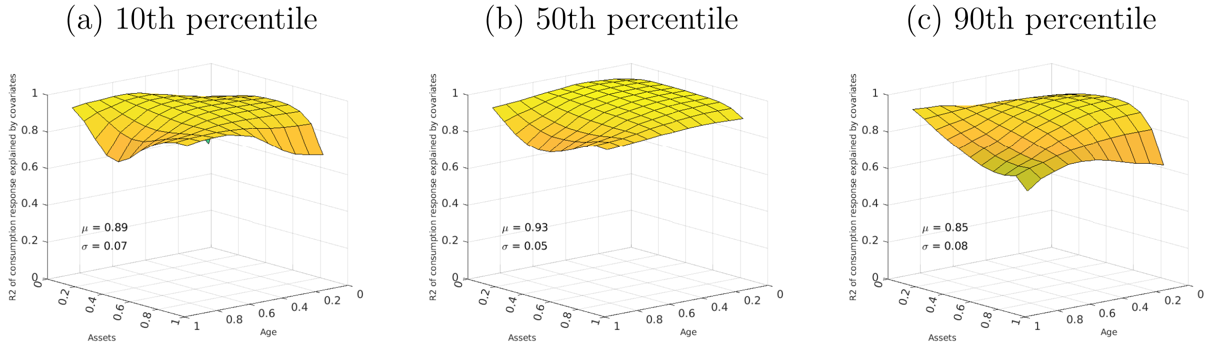

(a) 10th percentile

(b) 50th percentile

(b) 90th percentile

Notes: Pointwise 95% confidence bands based on nonparametric bootstrap. 200 replications. Bootstrap is clustered at the household level.

Figure 2.18: Heterogeneity in residual variation of consumption responses


(a) 10th percentile

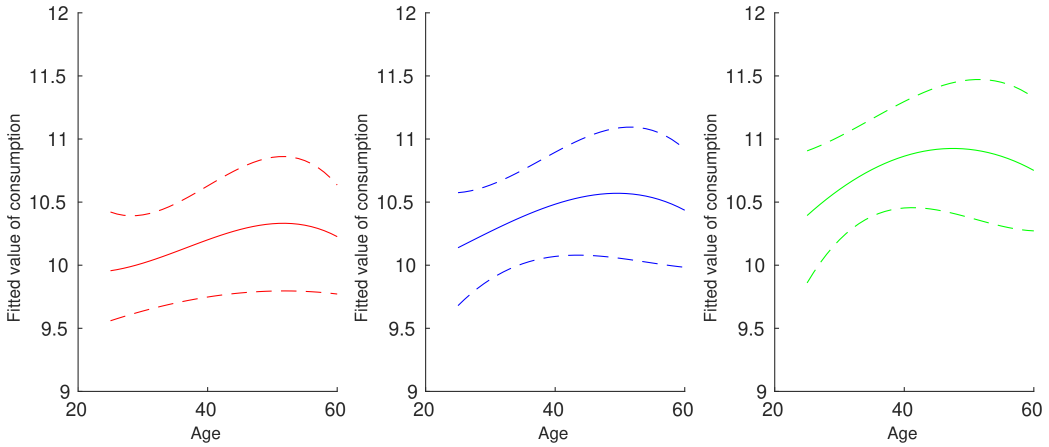
(b) 50th percentile

(c) 90th percentile

Notes: See the notes to Figure 2.5. The figure shows an upper bound on the proportion of the variation in consumption responses to η_{it} explained by the average consumption response, conditional on age and assets, see Section 2.D of the appendix. The various graphs corresponds to different percentiles of ξ_i.

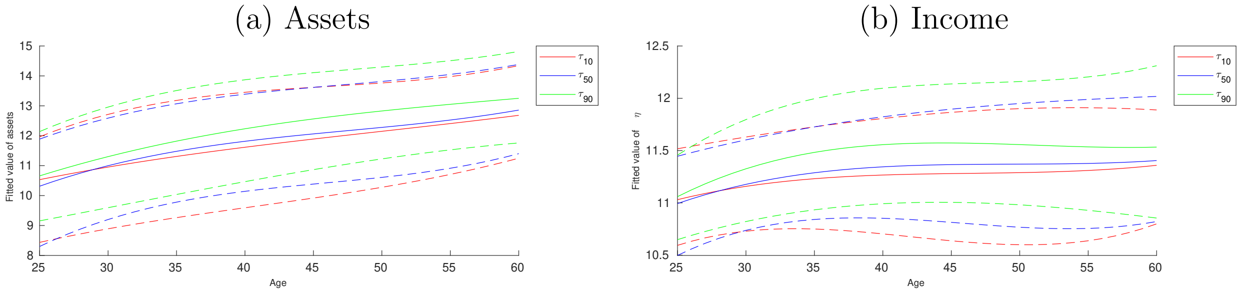
2.E.3 Tables and figures for Section 2.6

Figure 2.19: Life-cycle profile of log-consumption, for different percentiles of unobserved types



Notes: Average non-residualized log-consumption, for different ages and percentiles of ξ_i (10%, Median, 90%). The dashed lines show the age-specific and ξ_i-specific 10th and 90th percentiles of log-consumption.

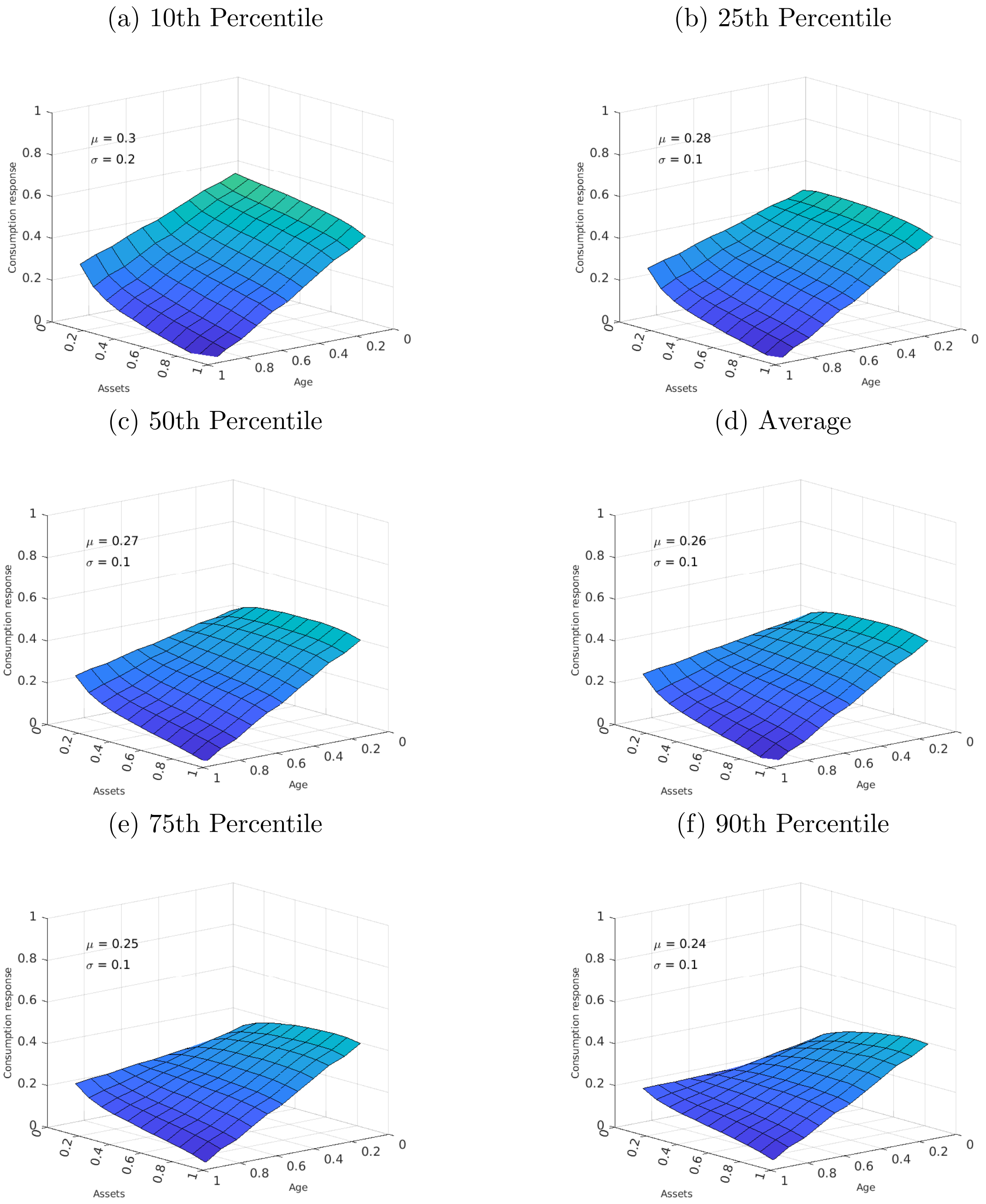
Figure 2.20: Life-cycle profiles of log-assets and log-income, for different percentiles of unobserved types


(a) Assets

(b) Income

Notes: Average non-residualized log-assets and persistent latent component of log-income, for different ages and percentiles of ξ_i (10%, Median, 90%). The dashed lines show the age-specific and ξ_i-specific 10th and 90th percentiles for each outcome measure.

Figure 2.21: Heterogeneity in consumption responses, model with heterogeneity in assets



(a) 10th Percentile

(b) 25th Percentile

(c) 50th Percentile

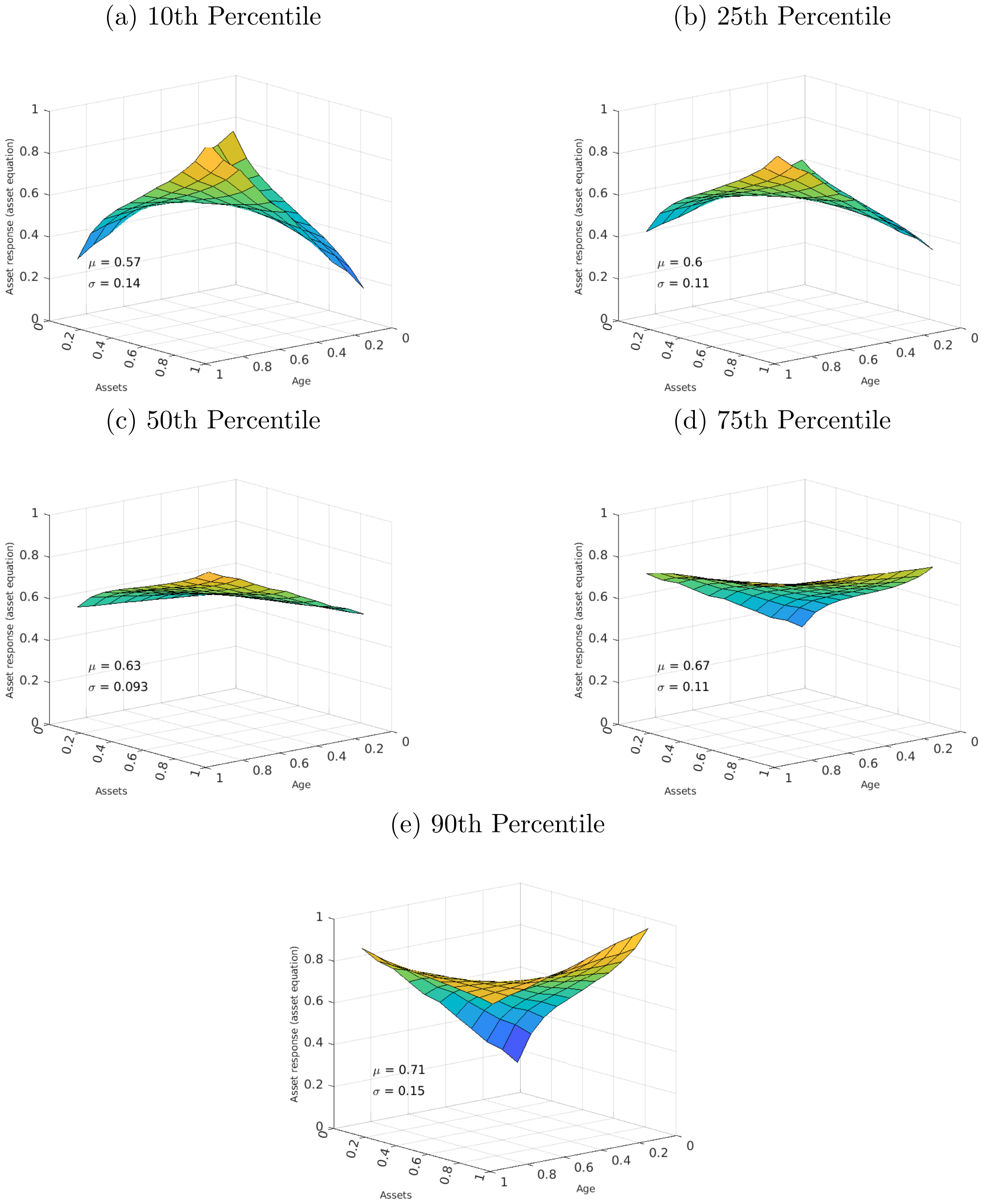
(d) Average

(e) 75th Percentile

(f) 90th Percentile

Notes: See the notes to Figure 2.5. The results are based on a model with latent heterogeneity ξ_i in consumption and assets. Here we report the results by percentiles of heterogeneity ξ_i.

Figure 2.22: Heterogeneity in assets dynamics, model with heterogeneity in assets


(a) 10th Percentile

(b) 25th Percentile

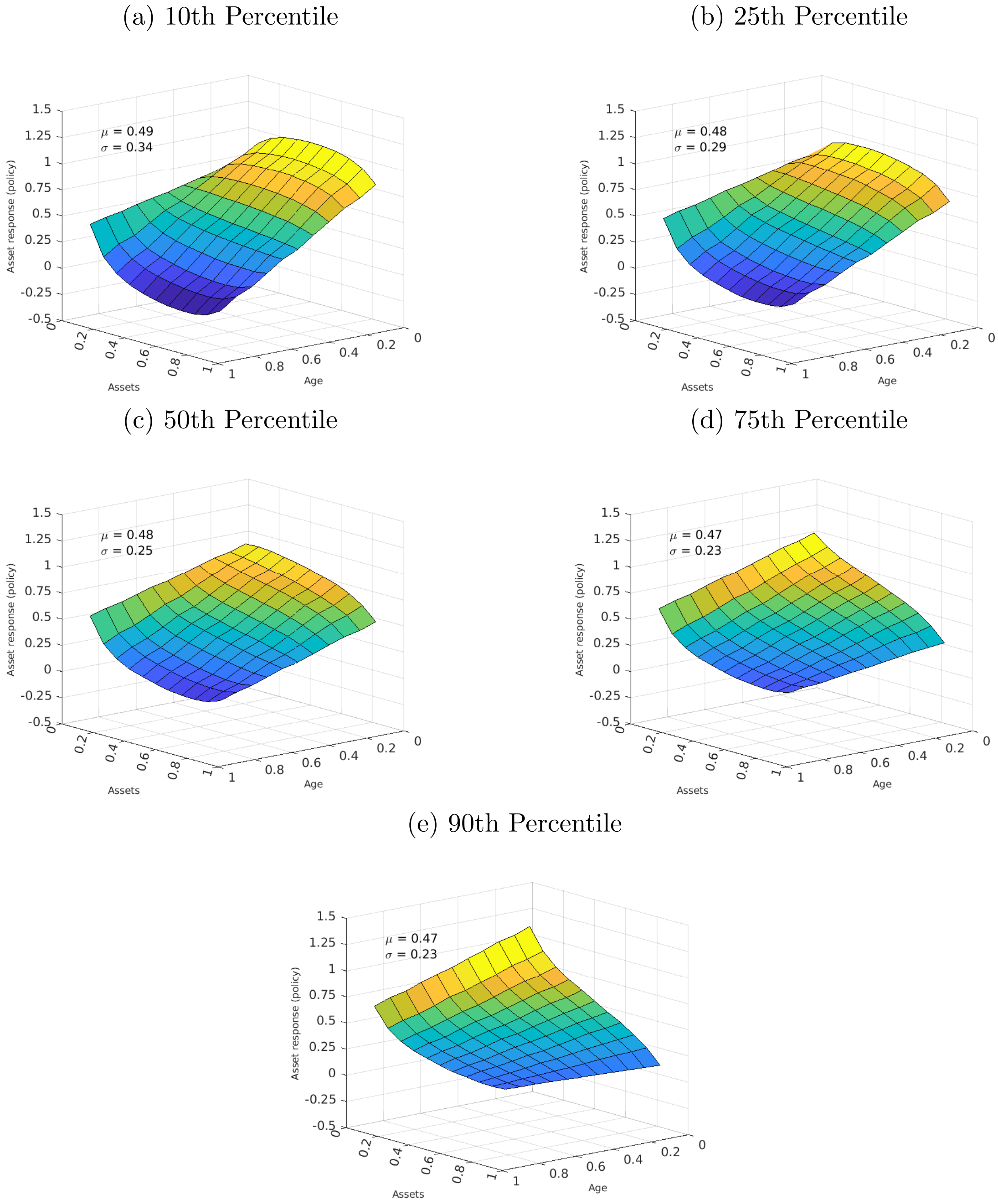
(c) 50th Percentile

(d) 75th Percentile

(e) 90th Percentile

Notes: The figure shows the average total derivative of log-assets with respect to lagged log-assets, conditional on lags of log-assets, income components, log-consumption, age, and the latent type. Here we report the results by percentiles of heterogeneity ξ_i.

Figure 2.23: Heterogeneity in assets responses, model with heterogeneity in assets



(a) 10th Percentile

(b) 25th Percentile

(c) 50th Percentile

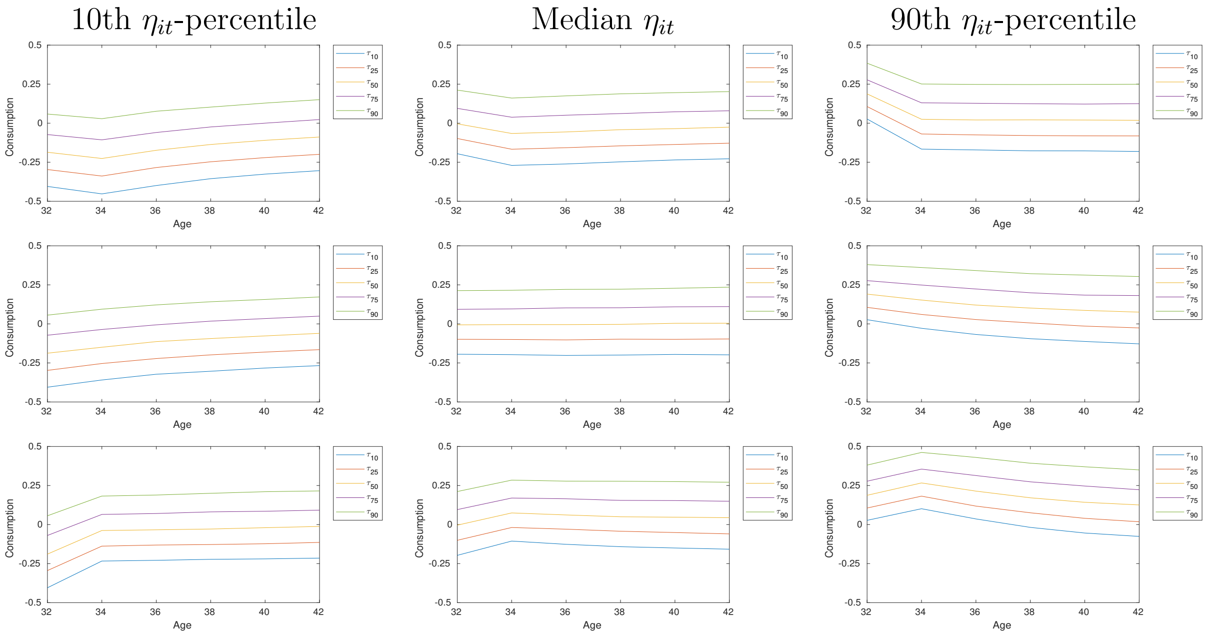
(d) 75th Percentile

(e) 90th Percentile

Notes: The figure shows the average total derivative of log-assets with respect to lagged η, conditional on lags of log-assets, income components, age and the latent type. Here we report the results by percentiles of heterogeneity ξ_i.

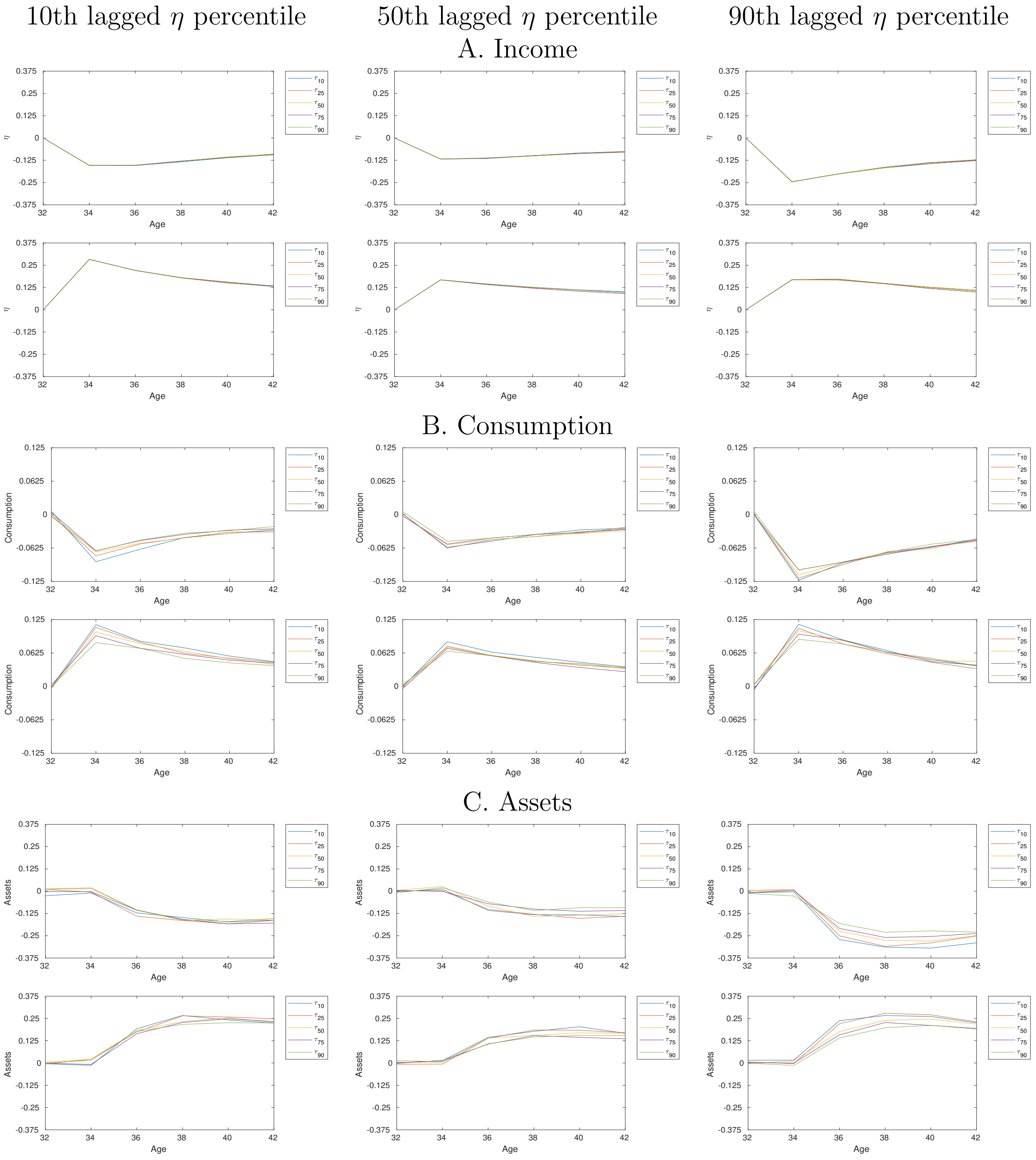
2.E.4 Tables and figures for Section 2.7

Figure 2.24: Heterogeneity in impulse responses: consumption trajectories



Notes: Trajectories shown for shocks at the 10th (top subpanel), 50th (middle subpanel) and 90th (bottom subpanel) percentiles.

Figure 2.25: Heterogeneity in impulse responses, model with heterogeneity in assets



10th lagged η percentile

50th lagged η percentile

90th lagged η percentile

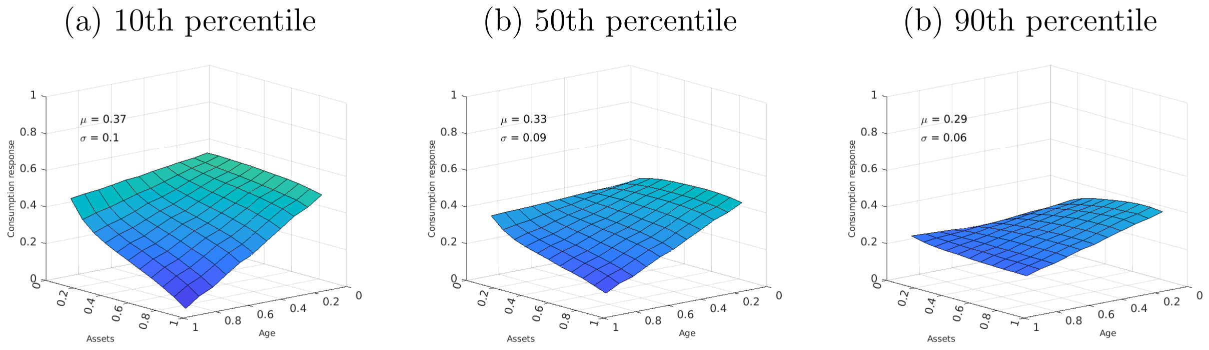
A. Income

B. Consumption

C. Assets

Notes: Impulse responses shown for shocks at the 10th (top subpanels) and 90th (bottom subpanels) percentiles, relative to median.

Figure 2.26: Heterogeneity in consumption responses based on 19 knots

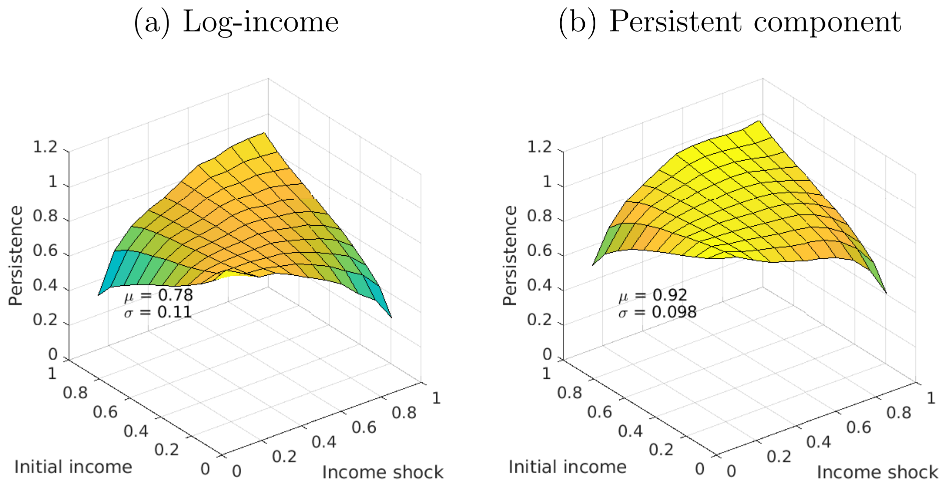

(a) 10th percentile

(b) 50th percentile

(b) 90th percentile

Notes: See the notes to Figure 2.5. In this figure we use 19 knots in estimation. For our baseline results in Figure 2.5 we used 11 knots.

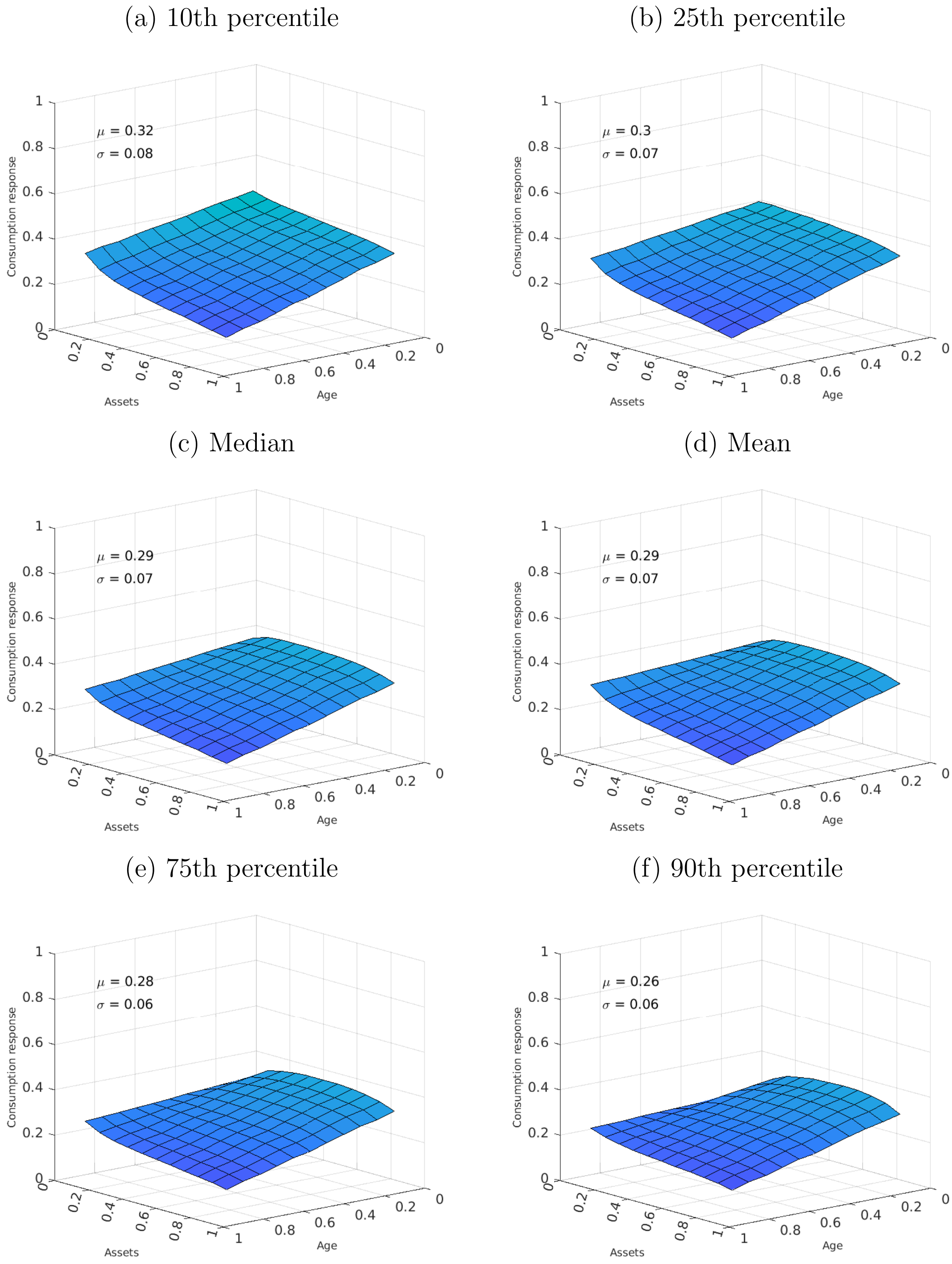
Figure 2.27: Nonlinear income persistence, labor income


(a) Log-income

(b) Persistent component

Notes: PSID, 2005-2017 sample, household labor income. The left graph shows quantile derivatives of log-income with respect to lagged log-income. The right graph shows quantile derivatives of the persistent latent component η_{it} with respect to η_{it-1}, model estimated using sequential Monte Carlo with a stochastic EM algorithm. The two horizontal axes show percentiles of η_{it-1} ("initial income") and conditional percentiles of η_{it} given η_{it-1} ("income shock"), respectively.

Figure 2.28: Heterogeneity in consumption responses, labor income



(a) 10th percentile

(b) 25th percentile

(c) Median

(d) Mean

(e) 75th percentile

(f) 90th percentile

Notes: See the notes to Figure 2.5. Household labor income. Here we report the results by percentiles of heterogeneity ξ_i in consumption.

Table 2.7: Descriptive statistics about the main sample without dual earners restriction

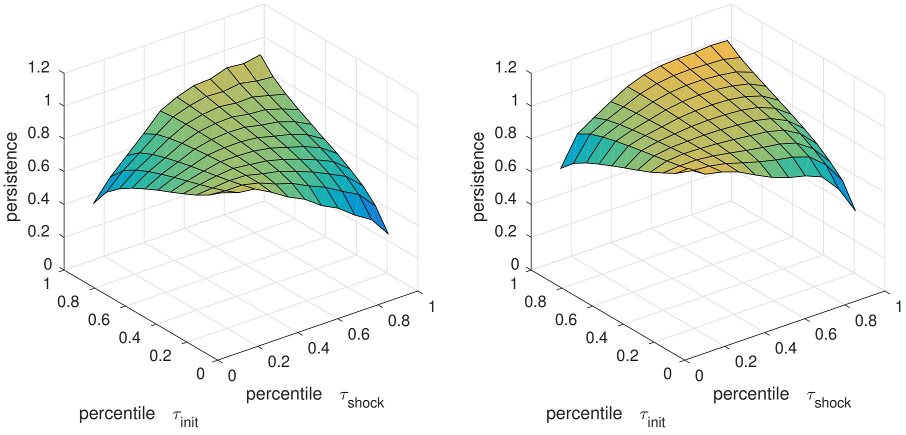
	(1) 2005	(2) 2007	(3) 2009	(4) 2011	(5) 2013	(6) 2015	(7) 2017
Food	10,739.58	10,629.51	10,294.05	10,523.17	10,728.10	11,195.42	12,049.05
	(5,602.43)	(5,617.21)	(5,131.22)	(5,066.30)	(5,701.24)	(5,290.79)	(5,872.98)
Non-durables (excl. food)	27,847.42	28,588.68	27,339.21	27,883.75	29,368.71	29,549.63	27,907.17
	(23,625.00)	(20,214.59)	(19,243.79)	(19,340.37)	(19,382.75)	(19,794.72)	(16,322.99)
Total Non-durables	38,625.09	39,265.87	37,731.22	38,532.60	40,205.65	40,843.42	40,002.57
	(26,482.07)	(23,195.91)	(21,701.55)	(21,932.92)	(22,194.81)	(22,563.15)	(19,525.89)
Home equity	168358.82	176300.55	136154.76	121783.91	116463.18	118089.14	133596.76
	(262246.00)	(283429.69)	(207398.72)	(175957.85)	(165962.97)	(152785.21)	(150451.14)
Negative Equity Dummy	0.01	0.01	0.03	0.03	0.03	0.01	0.01
	(0.08)	(0.10)	(0.17)	(0.17)	(0.16)	(0.10)	(0.10)
Wealth (excl. home)	211547.79	279544.52	278268.96	268297.79	260584.33	291511.77	346692.78
	(1.09e+06)	(1.16e+06)	(1.02e+06)	(704058.90)	(656770.57)	(765195.01)	(1.07e+06)
Total wealth	461075.98	521015.10	457730.90	411004.15	383583.34	409600.91	464296.87
	(1.51e+06)	(1.51e+06)	(1.23e+06)	(841641.87)	(762537.13)	(834428.29)	(1.09e+06)
Labor income	121962.17	120618.90	124276.87	121469.61	127809.57	124560.33	129948.36
	(155403.02)	(143009.67)	(181097.03)	(129296.65)	(241344.84)	(172615.50)	(115383.29)
Net income	93,333.70	93,262.22	95,306.98	95,476.95	98,924.10	95,790.47	99,431.91
	(92,700.09)	(86,962.24)	(108935.54)	(82,869.14)	(145844.22)	(98,144.09)	(69,882.14)
Observations	1730	2004	1843	1578	1436	1321	1090

Notes: PSID, 2005-2017. Means of variables, standard deviations in parentheses.

Figure 2.29: Nonlinear income persistence, no dual earners restriction

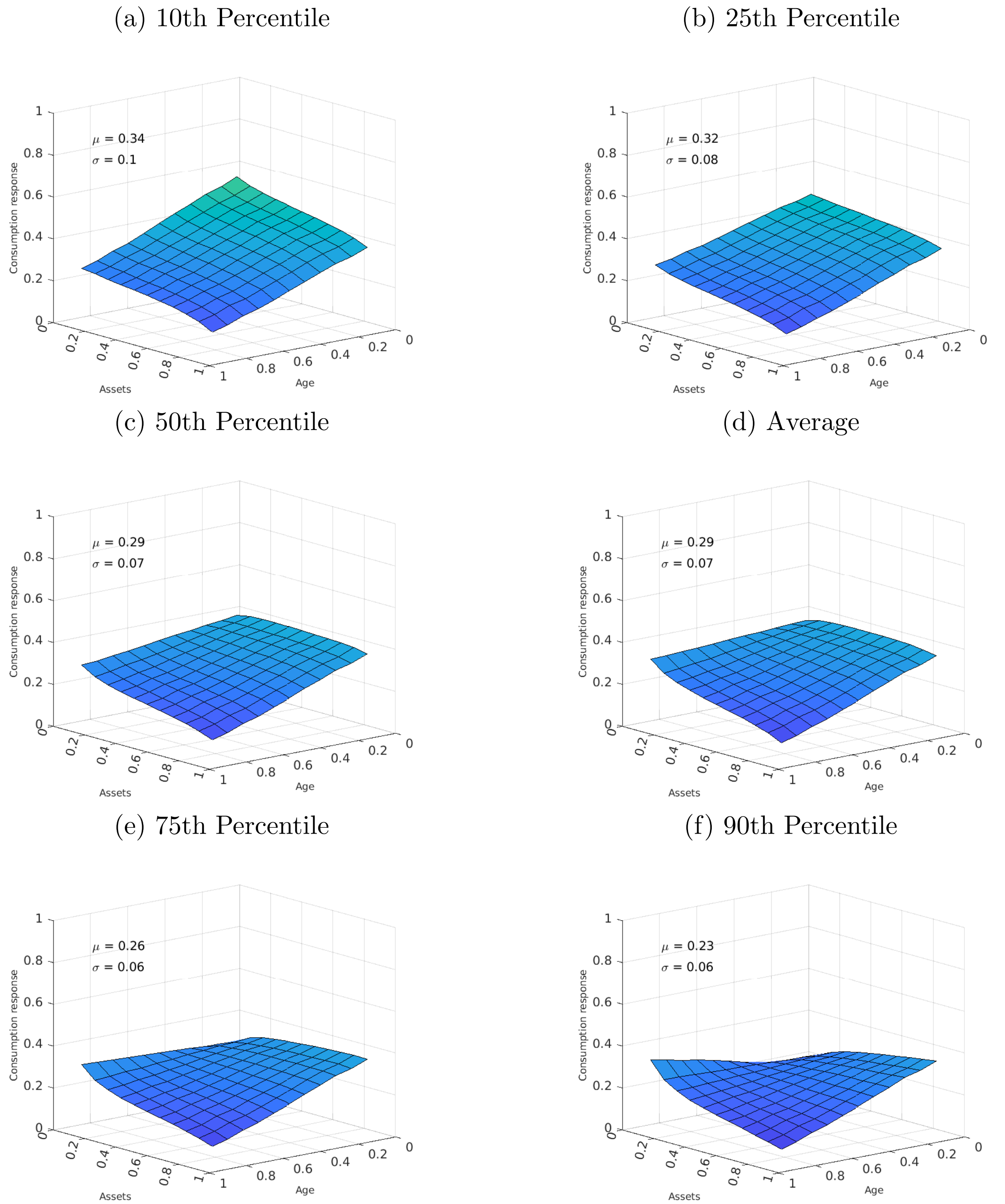
(a) Log-income

(b) Persistent component



Notes: PSID sample, no dual earners restriction. See the notes to Figure 2.3.

Figure 2.30: Heterogeneity in consumption responses, no dual earners restriction



(a) 10th Percentile

(b) 25th Percentile

(c) 50th Percentile

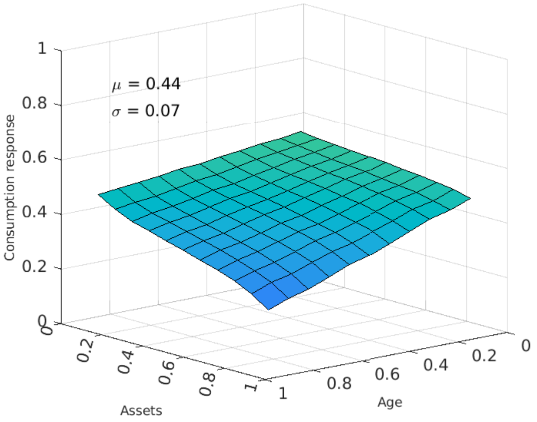
(d) Average

(e) 75th Percentile

(f) 90th Percentile

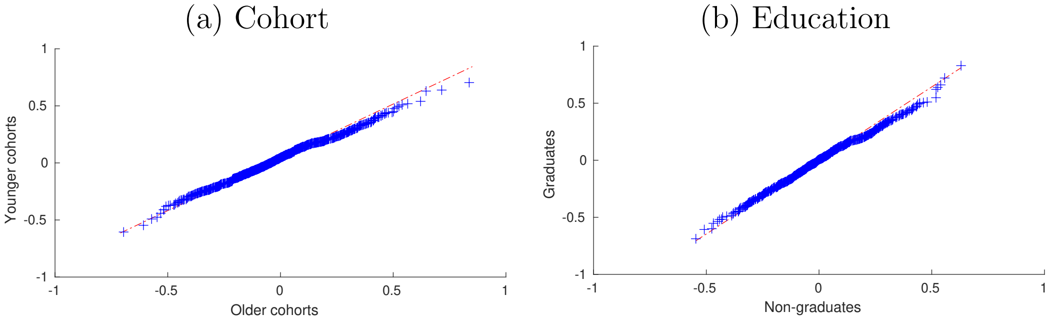
Notes: See the notes to Figure 2.4. No dual earners restriction. Here we report the results by percentiles of heterogeneity ξ_i in consumption.

Figure 2.31: Average consumption responses to labor income



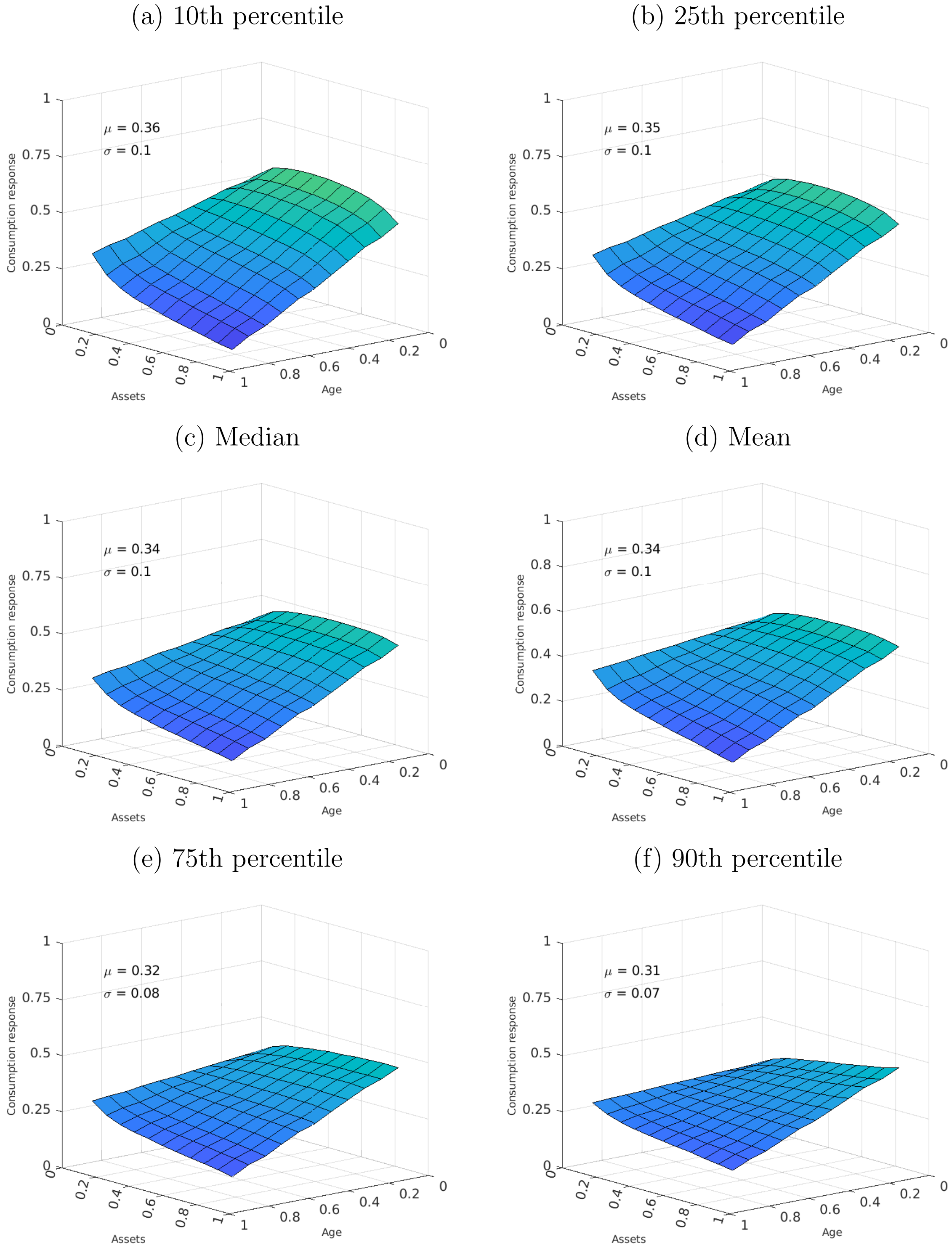
Notes: PSID, 2005-2017 sample, dual earners, labor income. The graph shows the average derivative of log-consumption with respect to the persistent latent component η_{it} in a model without unobserved heterogeneity ξ_i in consumption. The two horizontal axes show age and assets percentiles, respectively.

Figure 2.32: Quantile-quantile plots for ξ_i by observables



Notes: Quantile-quantile plots shown for (a) graduates and non-graduates (b) born before 1969 and born after 1969. The graphs show the quantiles of ξ_i indicated on the x-axis against the quantiles of ξ_i indicated on the y-axis.

Figure 2.33: Heterogeneity in consumption responses controlling for education



(a) 10th percentile

(b) 25th percentile

(c) Median

(d) Mean

(e) 75th percentile

(f) 90th percentile

Notes: See the notes to Figure 2.4. We report average derivatives in a regression that includes a full set of interactions with a binary higher education indicator. Here we report the results by percentiles of heterogeneity ξ_i in consumption.

REFERENCES

[1] M. Arellano, R. Blundell, and S. Bonhomme. Earnings and consumption dynamics: A nonlinear panel data framework. *Econometrica*, 85:693–734, 2017. v, 65, 112, 124

[2] Lonnie Golden. Irregular work scheduling and its consequences. *Economic Policy Institute Briefing Paper*, (394), 2015. 2

[3] Elaine D McCrate et al. *Unstable and on-call work schedules in the United States and Canada*. International Labour Office, Inclusive Labour Markets, Labour Relations and . . . , 2018. 2

[4] Susan J Lambert, Julia R Henly, and Jaeseung Kim. Precarious work schedules as a source of economic insecurity and institutional distrust. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, 5(4):218–257, 2019. 2

[5] Susan J Lambert and Anna Haley. Implementing work scheduling regulation: Compliance and enforcement challenges at the local level. *ILR Review*, 74(5):1231–1257, 2021. 2

[6] Kenneth Burdett and Dale T Mortensen. Wage differentials, employer size, and unemployment. *International Economic Review*, pages 257–273, 1998. 3, 19, 46

[7] Alon Bergman, Guy David, and Hummy Song. 'i quit': Schedule volatility as a driver of voluntary employee turnover. *Available at SSRN 3910077*, 2022. 4, 6, 10, 25, 30, 31, 39

[8] Joshua Choper, Daniel Schneider, and Kristen Harknett. Uncertain time: Precarious schedules and job turnover in the us service sector. *ILR Review*, 75(5):1099–1132, 2022. 4, 6, 9, 25, 32, 33

[9] Alan Manning. A generalised model of monopsony. *The Economic Journal*, 116(508): 84–100, 2006. 4

[10] Boyan Jovanovic. Job matching and the theory of turnover. *Journal of political economy*, 87(5, Part 1):972–990, 1979. 4

[11] Zeynep Ton and Robert S Huckman. Managing the impact of employee turnover on performance: The role of process conformance. *Organization Science*, 19(1):56–68, 2008. 4

[12] John P Hausknecht, Charlie O Trevor, and Michael J Howard. Unit-level voluntary turnover rates and customer service quality: implications of group cohesiveness, newcomer concentration, and size. *Journal of Applied Psychology*, 94(4):1068, 2009. 4

[13] Qin Li, Ben Lourie, Alexander Nekrasov, and Terry Shevlin. Employee turnover and firm performance: Large-sample archival evidence. *Management Science*, 68(8):5667–5683, 2022. 4

[14] Adam Smith. *The wealth of nations [1776]*, volume 11937. na, 1776. 4

[15] Hae-shin Hwang, Dale T Mortensen, and W Robert Reed. Hedonic wages and labor market search. *Journal of Labor Economics*, 16(4):815–847, 1998. 4

[16] Stéphane Bonhomme and Gregory Jolivet. The pervasive absence of compensating differentials. *Journal of Applied Econometrics*, 24(5):763–795, 2009. 4

[17] Timothy J Gronberg and W Robert Reed. Estimating workers' marginal willingness to pay for job attributes using duration data. *Journal of Human Resources*, pages 911–931, 1994. 4

[18] Jason Sockin. Show me the amenity: Are higher-paying firms better all around? 2022. 5, 26

[19] Thibaut Lamadon, Magne Mogstad, and Bradley Setzler. Imperfect competition, compensating differentials, and rent sharing in the us labor market. *American Economic Review*, 112(1):169–212, 2022. 5, 24, 26

[20] Iván Fernández-Val and Martin Weidner. Fixed effects estimation of large-t panel data models. *Annual Review of Economics*, 10:109–138, 2018. 5, 29

[21] Iván Fernández-Val and Martin Weidner. Individual and time effects in nonlinear panel models with large n, t. *Journal of Econometrics*, 192(1):291–312, 2016. 5, 29, 34

[22] Geert Dhaene and Koen Jochmans. Split-panel jackknife estimation of fixed-effect models. *The Review of Economic Studies*, 82(3):991–1030, 2015. 5, 29

[23] Peter CB Phillips and Hyungsik R Moon. Linear regression limit theory for nonstationary panel data. *Econometrica*, 67(5):1057–1111, 1999. 5, 29

[24] Julia R Henly, H Luke Shaefer, and Elaine Waxman. Nonstandard work schedules: Employer-and employee-driven flexibility in retail jobs. *Social service review*, 80(4): 609–634, 2006. 6, 10

[25] Susan J Lambert, Peter J Fugiel, and Julia R Henly. Schedule unpredictability among early career workers in the us labor market: A national snapshot. *Chicago, IL: Employment Instability, Family Well-being, and Social Policy Network, University of Chicago*, 2014. 6, 10, 12, 14

[26] Daniel Schneider and Kristen Harknett. Consequences of routine work-schedule instability for worker health and well-being. *American Sociological Review*, 84(1):82–114, 2019. 6, 10, 12, 14

[27] Lonnie Golden and Alison Dickson. Precarious times at work: Detrimental hours and scheduling in illinois and how fair workweek policies will improve workers' well-being. *Available at SSRN 3795584*, 2020. 6, 10

[28] Elizabeth Ananat, Anna Gassman-Pines, and John Fitz-Henley II. The effects of the emeryville fair workweek ordinance on the daily lives of low-wage workers and their families. Technical report, National Bureau of Economic Research, 2022. 6, 10

[29] Kristen Harknett, Daniel Schneider, and Sigrid Luhr. Who cares if parents have unpredictable work schedules?: Just-in-time work schedules and child care arrangements. *Social Problems*, 69(1):164–183, 2022. 6

[30] Samantha A Conroy, Dorothea Roumpi, John E Delery, and Nina Gupta. Pay volatility and employee turnover in the trucking industry. *Journal of Management*, 48(3):605–629, 2022. 6

[31] Richard Thaler and Sherwin Rosen. The value of saving a life: evidence from the labor market. In *Household production and consumption*, pages 265–302. NBER, 1976. 6

[32] Charles Brown. Equalizing differences in the labor market. *The Quarterly Journal of Economics*, 94(1):113–134, 1980. 6

[33] Greg J Duncan and Bertil Holmlund. Was adam smith right after all? another test of the theory of compensating wage differentials. *Journal of Labor Economics*, 1(4): 366–379, 1983. 6

[34] Thomas J Kniesner, W Kip Viscusi, Christopher Woock, and James P Ziliak. The value of a statistical life: Evidence from panel data. *Review of Economics and Statistics*, 94 (1):74–87, 2012. 6

[35] Alexandre Mas and Amanda Pallais. Valuing alternative work arrangements. *American Economic Review*, 107(12):3722–59, 2017. 6, 35, 39, 44

[36] M Keith Chen, Peter E Rossi, Judith A Chevalier, and Emily Oehlsen. The value of flexible work: Evidence from uber drivers. *Journal of political economy*, 127(6):2735–2794, 2019. 6

[37] Kuan-Ming Chen, Claire Ding, John A List, and Magne Mogstad. Reservation wages and workers' valuation of job flexibility: Evidence from a natural field experiment. Technical report, National Bureau of Economic Research, 2020. 6, 39, 44

[38] Haoran He, David Neumark, and Qian Weng. Do workers value flexible jobs? a field experiment. *Journal of Labor Economics*, 39(3):709–738, 2021. 6, 44

[39] Marshall Fisher, Santiago Gallino, and Serguei Netessine. Setting retail staffing levels: A methodology validated with implementation. *Manufacturing & Service Operations Management*, 23(6):1562–1579, 2021. 7

[40] Serguei Netessine, Marshall Fisher, and Jayanth Krishnan. Labor planning, execution, and retail store performance: An exploratory investigation. *Execution, and Retail Store Performance: An Exploratory Investigation (January 3, 2010)*, 2010. 7

[41] Andres Musalem, Marcelo Olivares, and Ariel Schilkrut. Retail in high definition: Monitoring customer assistance through video analytics. *Manufacturing & Service Operations Management*, 23(5):1025–1042, 2021. 7

[42] Caleb Kwon and Ananth Raman. Lateness and absenteeism in retail stores. *Available at SSRN 4189723*, 2022. 7, 10

[43] Marshall Fisher, Jayanth Krishnan, and Serguei Netessine. Retail store execution: An empirical study. *Available at SSRN 2319839*, 2006. 7

[44] Caleb Kwon, Ananth Raman, and Jorge Tamayo. Human-computer interactions in demand forecasting and labor scheduling decisions. *Available at SSRN 4296344*, 2022. 7, 10

[45] Alon Bergman, Hummy Song, Guy David, Joanne Spetz, and Molly Candon. The role of schedule volatility in home health nursing turnover. *Medical Care Research and Review*, page 10775587211034310, 2021. 10

[46] US BLS. Bls jolts report 2022. 2022. 12

[47] Julia R Henly and Susan J Lambert. Unpredictable work timing in retail jobs: Implications for employee work–life conflict. *Ilr Review*, 67(3):986–1016, 2014. 14, 19

[48] Arne L Kalleberg. Precarious work, insecure workers: Employment relations in transition. *American sociological review*, 74(1):1–22, 2009. 19

[49] Luigi Guiso, Luigi Pistaferri, and Fabiano Schivardi. Insurance within the firm. *Journal of Political Economy*, 113(5):1054–1087, 2005. 24

[50] David Card, Ana Rute Cardoso, Joerg Heining, and Patrick Kline. Firms and labor market inequality: Evidence and some theory. *Journal of Labor Economics*, 36(S1): S13–S70, 2018. 24

[51] Dale Mortensen. *Wage dispersion: why are similar workers paid differently?* MIT press, 2005. 26

[52] Douglas Rivers and Quang H Vuong. Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of econometrics*, 39(3):347–366, 1988. 28

[53] Jerzy Neyman and Elizabeth L Scott. Consistent estimates based on partially consistent observations. *Econometrica: Journal of the Econometric Society*, pages 1–32, 1948. 28

[54] Joshua D Angrist and Jörn-Steffen Pischke. *Mostly harmless econometrics: An empiricist's companion.* Princeton university press, 2009. 30

[55] Matthew Wiswall and Basit Zafar. Preference for the workplace, investment in human capital, and gender. *The Quarterly Journal of Economics*, 133(1):457–507, 2018. 39

[56] R. Moffitt and P. Gottschalk. Trends in the covariance structure of earnings in the united states: 1969-1987. *University of Wisconsin Institute for Research on Poverty*, (1001-93), 1993. 65

[57] Michael Baker and Gary Solon. Earnings dynamics and inequality among canadian men, 1976-1992: Evidence from longitudinal income tax records. *Journal of Labor Economics*, 21(2):267–288, 2003. 65

[58] C. Meghir and L. Pistaferri. Earnings, consumption and life cycle choices. 2011. 65

[59] R. Hall and F. Mishkin. The sensitivity of consumption to transitory income: Estimates from panel data of households. *Econometrica*, 50(2):261–81, 1982. 65

[60] Richard Blundell, Luigi Pistaferri, and Ian Preston. Consumption inequality and partial insurance. *American Economic Review*, 98(5):1887–1921, 2008. 65, 66, 76

[61] A. Deaton. *Understanding consumption.* Oxford University Press, 1992. 65

[62] M. De Nardi, G. Fella, and G. Paz-Pardo. Nonlinear household earnings dynamics, self-insurance, and welfare. *Journal of the European Economic Association*, 18(2):890–926, 2020. 66

[63] B. Anghel, H. Basso, O. Bover, J. M. Casado, L. Hospido, M. Izquierdo, and E. Vozmediano. Income, consumption and wealth inequality in spain. *SERIEs*, 9(4):351–387, 2018. 66

[64] P. Andreski, L. Geng, M. Z. Samancioglu, and R. Schoeni. Estimates of annual consumption expenditures and its major components in the psid in comparison to the ce. *American Economic Review: Papers & Proceedings*, 104(5):132–135, 2014. 66, 70

[65] R. L. Matzkin. Nonparametric identification in structural economic models. *Annual Review of Economics*, 5(1):457–486, 2013. 66, 77

[66] G. W. Imbens and W. K. Newey. Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica*, 77(5):1481–1512, 2009. 66

[67] S. Hoderlein and E. Mammen. Identification of marginal effects in nonseparable models without monotonicity. *Econometrica*, 75(5):1513–1518, 2007. 67, 78, 127

[68] D. Creal. A survey of sequential monte carlo methods for economics and finance. *Econometric reviews*, 31(3):245–296, 2012. 67, 87

[69] C. Andrieu, A. Doucet, and R. Holenstein. Particle markov chain monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(3):269–342, 2010. 67, 86, 89

[70] A. Auclert. Monetary policy and the redistribution channel. *American Economic Review*, 109(6):2333–2367, 2019. 68

[71] S. Alan, M. Browning, and M. Ejrnæs. Income and consumption: A micro semistructural analysis with pervasive heterogeneity. *Journal of Political Economy*, 126(5):1827–1864, 2018. 69

[72] E. Crawley and A. Kuchler. Consumption heterogeneity: Micro drivers and macro implications. 2020. 69

[73] G. Li, R. F. Schoeni, S. Danziger, and K. K. Charles. New expenditure data in the psid: Comparisons with the ce. *Monthly Lab. Rev.*, 133:29, 2010. 70

[74] R. Blundell, L. Pistaferri, and I. Saporta-Eksten. Consumption smoothing and family labor supply. *American Economic Review*, 106(2):387–435, 2016. 71

[75] D. Krueger, K. Mitman, and F. Perri. Macroeconomics and heterogeneity, including inequality. 2015. 71

[76] G. Kaplan and G. L. Violante. How much consumption insurance beyond self-insurance? *American Economic Journal: Macroeconomics*, 2:53–87, 2010. 76

[77] V. Chernozhukov and C. Hansen. An iv model of quantile treatment effects. *Econometrica*, 73(1):245–261, 2005. 79

[78] W. K. Newey and J. L. Powell. Instrumental variable estimation of nonparametric models. *Econometrica*, 71(5):1565–1578, 2003. 83

[79] Y. Hu and S. M. Schennach. Instrumental variable treatment of nonclassical measurement error models. *Econometrica*, 76(1):195–216, 2008. 83, 85

[80] D. Wilhelm. Identification and estimation of nonparametric panel data regressions with measurement error. *Cemmap working paper*, (CWP34/15), 2015. 83

[81] S. F. Nielsen. The stochastic em algorithm: estimation and asymptotic results. *Bernoulli*, pages 457–489, 2000. 85

[82] V. Chernozhukov, I. Fernández-Val, and A. Galichon. Quantile and probability curves without crossing. *Econometrica*, 78(3):1093–1125, 2010. 86, 117

[83] J. A. Machado and J. Mata. Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics*, 20(4):445–465, 2005. 86

[84] M. Arellano and S. Bonhomme. Nonlinear panel data estimation via quantile regressions. *The Econometrics Journal*, 19(3):C61–C94, 2016. 87

[85] N. Kantas, A. Doucet, S. S. Singh, J. Maciejowski, and N. Chopin. On particle methods for parameter estimation in state-space models. *Statistical science*, 30(3):328–351, 2015. 87

[86] Z. Griliches and J. A. Hausman. Errors in variables in panel data. *Journal of Econometrics*, 31:93–118, 1986. 95

[87] P. Krusell and A. Smith. Income and wealth heterogeneity in the macroeconomy. *Journal of Political Economy*, 2(1):245–272, 1998. 102

[88] L. Hendricks. How important is discount rate heterogeneity for wealth inequality. *Journal of Economic Dynamics and Control*, 31(9):3042–3068, 2007. 102

[89] A. Fagereng, L. Guiso, D. Malacrino, and L. Pistaferri. Heterogeneity in returns to wealth and the measurement of wealth inequality. *American Economic Review*, 106(5): 651–655, 2016. 102

[90] J. Altonji, F. Hayashi, and L. Kotlikoff. Is the extended family altruistically linked? direct tests using micro data. *American Economic Review*, 82(5):1177–1198, 1992. 102

[91] F. Hayashi, J. Altonji, and L. Kotlikoff. Risk-sharing between and within families. *Econometrica*, 64(2):261–294, 1996. 102

[92] K. K. Charles, S. Danziger, G. Li, and R. Schoeni. The intergenerational correlation of consumption expenditures. *American Economic Review*, 104(5):136–140, 2014. 102

[93] O. Attanasio, C. Meghir, and C. Mommaerts. Insurance in extended family networks. *NBER Working Paper*, (21059), 2019. 102