DEPARTMENT
OF ECONOMICS

Essays on Household Heterogeneity in Macroeconomics

Lukas Nord

Thesis submitted for assessment with a view to
obtaining the degree of Doctor of Economics
of the European University Institute

Florence, 12 June 2023

European University Institute
Department of Economics

Essays on Household Heterogeneity in Macroeconomics

Lukas Nord

Thesis submitted for assessment with a view to
obtaining the degree of Doctor of Economics
of the European University Institute

Examining Board

Prof. Árpád Ábrahám, University of Bristol, Supervisor
Prof. Russell Cooper, EUI, Co-Supervisor
Prof. Guido Menzio, New York University
Prof. Claudio Michelacci, EIEF

Researcher declaration to accompany the submission of written work Department Economics - Doctoral Programme

I Lukas Nord certify that I am the author of the work "Essays on Household Heterogeneity in Macroeconomics" I have presented for examination for the Ph.D. at the European University Institute. I also certify that this is solely my own original work, other than where I have clearly indicated, in this declaration and in the thesis, that it is the work of others.

I warrant that I have obtained all the permissions required for using any material from other copyrighted publications.

I certify that this work complies with the Code of Ethics in Academic Research issued by the European University Institute (IUE 332/2/10 (CA 297).

I declare that this work consists of 83872 words.

Statement of inclusion of previous work:

I confirm that chapter 2 was jointly co-authored with Annika Bacher and Philipp Grübener and I contributed 1/3 of the work.
I confirm that chapter 3 was jointly co-authored with Caterina Mendicino and Marcel Peruffo and I contributed 1/3 of the work.

Signature and date:

20.03.2023

Mama und Papa
Danke, für alles.

Acknowledgements

While obtaining a Ph.D. seems like a lonely endeavor from the outside, many people have provided me with invaluable support over the last six years and have made the completion of this thesis possible.

First and foremost, I need to thank my advisors Árpád Ábrahám and Russell Cooper for years of encouragement and their constant availability to discuss my work. With their ability to always ask the right questions and challenge my thinking, they have helped me understand how to do research and have made me a better economist. I could not have asked for a better team of advisors and I can only hope to one day follow their example with my own students.

In addition to my supervisors, I have learned a lot from repeated discussions with many great economists who I had the pleasure to meet along the way, including Jesús Bueren, Edouard Challe, Giancarlo Corsetti, Alessandro Ferrari, Axelle Ferriere, Philipp Kircher, Ramon Marimon and Alex Monge. Many thanks to Guido Menzio and Claudio Michelacci for generously agreeing to be part of my thesis committee and their most valuable feedback on my work.

The parts of this thesis I enjoyed working on the most are the chapters that I have written jointly with others. I have to thank my coauthors on these chapters – Annika Bacher, Philipp Grübener, Caterina Mendicino, and Marcel Peruffo – for their friendship, for all they have taught me, and for showing me how fun it can be to do research. I am looking forward to many future collaborations.

I was lucky to consider the Economics Department at the EUI my academic home for six years and would like to thank the entire team of the department and in particular Sarah Simonsen, without whom generations of Ph.D. students would have gotten lost in the process. In addition, I thank the European Central Bank as well as the Institute for International Economic Studies (IIES) at Stockholm University for their hospitality while writing parts of this thesis.

This thesis would not have been possible without the support of many great friends inside and outside of academia. My amazing cohort at the EUI has not only helped me survive hours of work on first-year problem sets but has been a constant source of support throughout the entire Ph.D.. To the regular participants of the IEWG I owe thanks for many stimulating Friday afternoons and even more fun Friday evenings. Special thanks go to Fabian, Jannes, Ruben, Torben, and Wolfram for their unconditional friendship and for reminding me every once in a while that life exists outside of academia.

Finally, I would like to thank my entire family for their support and apologize for the long list of missed birthday parties and family gatherings while writing this thesis far away from home.

The deepest gratitude of all I owe to my parents, to whom I dedicate this thesis. They have taught me to be curious and have always encouraged me in whatever I set my mind on. I would not be who I am or where I am without them.

Abstract

This thesis contains four independent essays studying the consequences of household heterogeneity for Macroeconomics.

The first chapter studies the implications of household heterogeneity for equilibrium prices. I break with the canonical assumptions of homothetic preferences and the law of one price to show how heterogeneity in consumption baskets and search for price bargains affects posted prices. Analytical results from search theory and empirical evidence from big data on households' grocery transactions show that price distributions respond to the composition of buyers. In a quantitative heterogeneous agent model with endogenous price dispersion for multiple varieties, I find that the response of retailers to households' search effort is quantitatively important to differentiate between inequality in expenditure and consumption. It more than doubles the direct effect of paying more or less given posted prices, which has been the focus of previous literature. Furthermore, I find that household heterogeneity helps to account for the empirical cyclicality of retail prices and markups in response to aggregate shocks, and has implications for the response of prices to redistributive policies.

In the second chapter, which is joint work with Annika Bacher and Philipp Grübener, we show how households with two members can insure themselves against the job loss of a primary earner through the labor force entry of a non-participating spouse. We document empirically that this margin is predominantly used by young households. In a two-member life cycle model with endogenous arrival rates, human capital accumulation, and extensive-margin labor supply, we explore how differences in labor market opportunities and asset holdings contribute to this pattern. Our findings suggest that the age difference is predom-

inantly explained by better insurance through asset holdings for the old, while differences in arrival rates and human capital play a smaller role.

In the third chapter, which is joint work with Caterina Mendicino and Marcel Peruffo, we study differences in the exposure to bank distress along the income distribution. We develop a two-asset heterogeneous agent model with a financial sector and use this framework to show that banking sector losses disproportionately harm low-income households while rich households adjust their savings behavior to profit from fluctuations in asset prices. This is why welfare losses from bank distress are considerably more dispersed than consumption responses. We find the model-implied consumption responses to be in line with empirical evidence on the relationship between bank equity returns and consumption across households.

In the forth chapter, I study how wealth holdings can affect households' incentives to form precise expectations about future inflation rates. I document empirically how the dispersion of expectations changes along the wealth distribution and develop a consumption-savings model with costly expectation formation to study implications for the effectiveness of forward guidance policies. I show endogenous expectation formation to significantly lower the effectiveness of forward guidance policies due to selection in which households are paying attention to news about inflation.

Contents

Chapter 1

Shopping, Demand Composition, and Equilibrium Prices

Abstract This paper develops an equilibrium theory of expenditure inequality and price dispersion to study how retail prices respond to households' shopping behavior. Heterogeneity in the effort to search for prices implies that the price elasticity faced by retailers depends on the composition of demand. For a search market with price posting, I show analytically that retailers optimally charge higher markups if goods are mainly consumed by low-search-effort households. Additional predictions on the shape of posted price distributions are consistent with evidence from US supermarket scanner micro-data. I embed search for prices into an incomplete markets model with non-homothetic preferences and equilibrium price dispersion for multiple varieties. Endogenous heterogeneity in search effort allows the model to match evidence on differences in prices paid for identical goods and reduces inequality in consumption relative to expenditure. I show that the equilibrium response of posted prices across products doubles this direct effect of search on inequality. In addition, the model reconciles conflicting evidence on the cyclicality of retail markups, as aggregate shocks change the composition of demand. Finally, I find that the response of posted prices to a redistributive earnings tax compensates top earners for up to 14% of their losses.[1]

[1]Researcher(s)' own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researcher(s) and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

1.1 Introduction

Understanding inequality in households' consumption is essential to infer the welfare consequences of income and wealth inequality. A growing literature emphasizes heterogeneity in prices across households and distinguishes consumption from expenditure inequality (e.g. Aguiar and Hurst, 2005, 2007). This distinction is important because posted prices for identical products exhibit significant dispersion and poor households search for bargains to pay less for the same good. Previous work abstracts from any equilibrium effect of this shopping effort on posted prices. However, if buyers search more for cheap offers, retailers face higher competition and optimally reduce the prices they post. This response matters for distinguishing expenditure and consumption inequality because households do not buy the same basket of goods and retailers can discriminate prices across products. It also matters for understanding the impact of aggregate shocks and policies, as the adjustment of posted prices determines the full effect of changes in shopping effort on the average price of consumption across households.

This paper develops an equilibrium theory of expenditure inequality and price dispersion that accounts for the response of posted prices to the shopping behavior of heterogeneous households. First, I provide analytical results on how retailers post prices taking the level of shopping effort as given and test theoretical predictions against large US micro-data on grocery transactions. Second, I quantify the effect of retailers' price posting on the distinction between expenditure and consumption inequality in equilibrium. Finally, I highlight the implications of heterogeneous shopping effort for the cyclicality of retail prices and markups, as well as for the response of prices to redistributive earnings taxes.

The framework developed in this paper incorporates frictional goods markets in the spirit of Burdett and Judd (1983) in an Aiyagari-Bewley-Huggett economy with multiple goods. Heterogeneous households decide on their spending, savings, and shopping effort. Shopping effort is subject to a utility cost. Households allocate their total expenditure across multiple varieties of a grocery good and an outside good. Consumption baskets vary systematically across households due to non-homothetic preferences over grocery varieties. The markets for grocery varieties are subject to search frictions. For every unit of consumption they purchase, households have to search for price quotes and draw either one or two offers simultaneously from the equilibrium distribution of posted prices. Higher shopping effort increases the probability that a household observes two prices and can select the cheaper offer. The price distribution for each variety is determined endogenously as the optimal solution to retailers' price posting problem, which trades off higher margins per sale against undercutting simultaneously observed alternative offers.

The first result of this paper is that with heterogeneity in shopping effort, posted prices depend on the composition of demand. To show this analytically, I focus on retailers' price posting problem for a single variety and take households' choices as given. I derive closed-form expressions for the moments of the posted price distribution and find that demand-weighted shopping effort is a sufficient statistic for retailers to take into account rich household heterogeneity. I show that the average posted price decreases in demand-weighted effort, driven by a reduction in profit margins. A higher demand-weighted shopping effort means that the average buyer is more likely to observe two prices and substitute towards a cheaper offer. As a result, retailers face a higher average price elasticity. Therefore, if a larger share of demand comes from households exerting more shopping effort, retailers' best response is to reduce their markups and post lower prices.

In addition, the skewness of posted price distributions strictly increases in demand-weighted search effort and is independent of all other model parameters. This result provides a testable prediction that is directly linked to the mechanism generating price dispersion. For price dispersion to exist in equilibrium, retailers have to be indifferent between posting low and high prices within a distribution. To keep retailers indifferent when shopping effort increases, the distribution has to become more dense at the bottom and less dense at the top, i.e. its skewness must increase.

Empirical evidence supports the relationship between the skewness and demand-weighted search effort in micro-data on households' grocery transactions from the Nielsen Consumer Panel. I first show that high-spending, high-income, and employed households exert lower shopping effort and pay higher prices for identical barcodes. To test the relationship between households' effort and skewness, I exploit variation in demand-weighted shopping effort across products due to differences in households' consumption baskets. In line with theory, the skewness of local, barcode-level price distributions increases in the share of total expenditure for a given barcode stemming from households with higher search effort.

In equilibrium of the full model, households generate demand-weighted shopping effort for each good endogenously in response to the distributions of posted prices. I solve for the equilibrium in households' choices and posted price distributions numerically. To quantify the equilibrium effect of shopping on posted prices, I calibrate the model to evidence from the Nielsen Consumer Panel. I match differences in prices paid within and across varieties and heterogeneity in consumption baskets along the expenditure distribution, as well as price dispersion across products. In addition, the model can account for untargeted moments such as the distribution of expenditures in the data.

The calibrated model shows that the equilibrium response of posted prices more than doubles the effect of shopping effort on the difference between consumption and expenditure inequality. The literature so far has measured the effect of shopping by focusing on differences in prices paid for a given product (see e.g. Aguiar and Hurst, 2007; Arslan et al., 2021; Pytka, 2022). Under this definition, shopping reduces the cost of consumption for the bottom versus the top expenditure quintile by 2% in the model and in the data. However, because high- and low-spending households do not buy the same products, retailers target their prices to the buyers they face and post lower average margins for products in the basket of low-spending (high-search) households. In the model, I show that these differences in posted margins across varieties reduce the cost of consumption for the bottom quintile of expenditures by an additional 2.5% relative to the top quintile. Abstracting from the overall effect of shopping overstates consumption inequality between the top and bottom quintile by 5% given the same observed distribution of grocery expenditures. Half of this effect is accounted for by differences in posted margins across varieties. Focusing on welfare to account for the disutility of shopping effort, again the direct and the equilibrium effect of shopping are equally important for inequality.

In addition, accounting for heterogeneity in shopping effort has implications for the cyclicality of prices and markups in response to aggregate shocks. I implement an aggregate shock based on the decline in net worth and losses in labor earnings during the Great Recession. The model generates a 0.7% decline in average prices paid upon impact. 0.6 percentage points are accounted for by changes in posted prices as retailers respond with lower markups to an increase in demand-weighted shopping effort. Only 0.1 percentage points can be attributed to a decline in the average price paid relative to the average posted price. This finding shows how focusing on prices paid relative to prices posted understates the effect of shopping on the cost of consumption over the business cycle.

The change in posted prices reported above is almost entirely driven by the decline in wealth. Losses in earnings have little impact on retailers' price posting despite accounting for a similar loss in disposable resources. This result arises because wealth losses are relatively more concentrated at the top of the income distribution and earnings losses at the bottom. In response to a loss in her earnings or wealth, any household increases her search effort and reduces consumption. If low-income households reduce their consumption, the composition of demand shifts in favor of high-income households with low search effort, incentivizing retailers to raise prices. Therefore, in response to earnings losses at the bottom of the distribution, this shift in demand composition offsets the increase in individual search effort. In response to a decline in wealth at the top, the increase in individual effort and the composition effect go in the same direction and unambiguously reduce prices and markups. This result reconciles seemingly conflicting empirical evidence, suggesting procyclical price and markup responses to house price shocks (Stroebel and Vavra, 2019) and acyclical responses to unemployment fluctuations (Anderson et al., 2020). Overall, composition effects reduce the on-impact response of posted prices to the combined shock by one third.

Finally, I show that the response of posted prices to shifts in demand composition partially compensates net contributors to redistributive policies for the decline in their income. To do so, I introduce a flat tax on labor earnings and rebate the proceeds lump-sum to all households. As this policy redistributes resources towards low-income (high-shopping-effort) households, it increases their share in aggregate demand and hence increases demand-weighted shopping effort. In an economy with a higher level of redistribution, retailers therefore optimally choose to reduce their markups and post lower prices. This channel compensates net contributors in the top quintile of expenditures for 5-14% of the decline in their after-tax earnings.

The paper is structured as follows: Section 1.1.1 discusses related literature. Section 1.2 presents analytical results on the response of posted prices to shopping effort. Section 1.3 provides empirical evidence on shopping effort and price distributions. Section 1.4 outlines the equilibrium model and its calibration. Section 1.5 studies the implications for inequality. Section 1.6 presents the results on cyclicality and policies. Section 1.7 concludes.

1.1.1 Related Literature

Search Frictions. Seminal contributions on price search in the goods market include Butters (1977), Varian (1980), and Burdett and Judd (1983). I build on the latter, which has been widely applied in macroeconomic research.[2] I extend the previous work by providing analytical results on how the moments of posted price distributions respond to the distribution of shopping effort in a Burdett-Judd market.

In shopping economies with rich heterogeneity in income and wealth, Arslan et al. (2021) take posted prices as given and Pytka (2022) endogenizes the price distribution for a single good. Both papers focus on the direct effect of shopping on prices paid for life-cycle inequality and the response to idiosyncratic income shocks in a stationary economy. The equilibrium search framework presented in

[2]See e.g. Albrecht et al. (2021), Burdett and Menzio (2018), and Menzio (2021). Additional work on the macroeconomics of goods market frictions and households' shopping behavior includes e.g. Angelini and Brès (2022), Bai et al. (2019), Coibion et al. (2015), Gaballo and Paciello (2021), Kryvtsov and Vincent (2021), Petrosky-Nadeau and Wasmer (2015), and Sara-Zaror (2022).

this paper is the first with rich household heterogeneity and endogenous price distributions for multiple varieties. I employ it to study how the equilibrium response of posted prices to shopping affects inequality.

Equilibrium effects of shopping effort on posted prices allow Alessandria (2009) to explain movements of relative prices across countries and Kaplan and Menzio (2016) to generate self-fulfilling unemployment fluctuations. Both setups feature price dispersion for a single good and stylized heterogeneity with finite types of shoppers. I show how rich heterogeneity can affect the cyclicality of retail prices due to shifts in demand composition when accounting for the incidence of aggregate shocks.[3]

Expenditure Inequality. The paper also relates to the empirical literature on expenditure inequality (e.g. Aguiar and Bils, 2015; Attanasio and Pistaferri, 2016; Coibion et al., 2021). Most closely related are the seminal contributions on the direct effect of shopping effort on prices paid by Aguiar and Hurst (2005, 2007) and subsequent work (e.g. Aguiar et al., 2013; Broda et al., 2009; Griffith et al., 2009; Nevo and Wong, 2019; Pisano et al., 2022; Pytka, 2022). My findings suggest that the equilibrium response of posted prices more than doubles the direct effect of shopping on inequality studied previously.

Non-Homotheticities. The literature on non-homotheticities dates back to Engel's Law in 1857. Most closely related is the recent work focusing on CES-preferences at the barcode level (Argente and Lee, 2021; Auer et al., 2022; Faber and Fally, 2022; Handbury, 2021; Jaravel, 2019). Non-homotheticities at this disaggregated level are often interpreted as substitution along a quality margin (Bils and Klenow, 2001; Bisgaard Larsen and Weissert, 2020; Ferraro and Valaitis, 2022; Jaimovich et al., 2019). Mongey and Waugh (2022) generate non-homotheticities from logit-preferences in an incomplete-markets economy. None of the previous work considers interactions with shopping effort and price dispersion.

Retail Prices and Markups. The paper also extends the empirical literature on retail prices and markups. Seminal work by Kaplan and Menzio (2015) and

[3]Huo and Ríos-Rull (2015) develop a framework with heterogeneous households and directed search for quantities and show how shifts in demand composition can affect productivity.

Kaplan et al. (2019) provides evidence on the structure of price distributions but does not consider their co-movement with demand composition across products. Stroebel and Vavra (2019) find retail prices and markups to respond procyclically to local variations in house prices and attribute this pattern to empirically observed changes in shopping behavior. Anderson et al. (2020) find markups paid to co-vary positively with proxies for local income, driven by differences in products bought. Their findings are in line with the theory of this paper.

Closely related is the complementary work of Sangani (2022), providing evidence on higher markups for goods bought by high-income households. He rationalizes his findings by combining the single-variety model of Burdett and Judd (1983) with stylized household heterogeneity and studies the implications of increasing income inequality for the rise in aggregate markups. In contrast, I provide direct empirical evidence on the mechanism, testing predictions on the relationship between shopping and the shape of price distributions. My focus is the feedback between equilibrium prices and inequality. Hence, I develop a model with rich household heterogeneity in the tradition of Bewley (1977) and Aiyagari (1994), featuring non-homothetic preferences and endogenous price distributions for multiple varieties.

1.2 The Mechanism: Price Posting with Search Frictions

To study how household heterogeneity can affect posted price distributions, I analyze retailers' price posting problem in a frictional product market. Throughout this section, I take the distribution of households, their shopping effort, and consumption choices as given and focus on the distribution of posted prices. I build on the price posting problem of a single-variety retailer in a market with consumer search as introduced by Burdett and Judd (1983) and Pytka (2022). Within this framework, I characterize analytically how moments of the posted price distribution respond to the distribution of households.

1.2.1 Retailers' Problem and Posted Price Distributions

Consider the market for a single variety j, which is produced at homogeneous marginal cost κ_j, and for which all consumers have identical maximum willingness to pay $\bar{p}_j > \kappa_j$. The variety is sold by a continuum of homogeneous retailers of measure one. The demand side of the market consists of a continuum of households indexed by their type i, with λ_i being the distribution over types. A type i household is characterized by her disposable resources x_i and consumes a quantity $c_j(x_i) \geq 0$ of variety j, which she splits into a measure $c_j(x_i)$ of infinitesimal purchases. The market for the variety is subject to incomplete information. For each purchase she makes, a household observes either one or two price postings, drawn at random from the equilibrium distribution of posted prices $F_j(p)$. The probability of observing two price draws for any given purchase is determined by the household's shopping effort $s(x_i) \in [0, 1]$, i.e. shopping effort is the intensity with which households search for a second price observation. For purchases with a single price observation the household buys the good if the observed price is below the maximum willingness to pay $\bar{p}_j$. Purchases with two simultaneous price observations are made at the lowest offer below $\bar{p}_j$.

Retailers' Problem. Retailers commit to a price for variety j before meeting any buyers. They post prices to maximize their profits, taking expectations over the type of household they will meet in the market and how likely any type is to see a second price offer simultaneously. The total profits of a retailer posting price p are given by

$$\pi_j(p) = \underbrace{C_j}_{\substack{\text{demand} \\ \text{per retailer} \\ \text{(market size)}}} \underbrace{\left[\int \frac{\lambda^i c_j(x_i)}{C_j} \left[(1 - s(x_i)) + s(x_i) 2(1 - F_j(p))\right] di\right]}_{\substack{\text{sales per demand} \\ \text{(market share)}}} \underbrace{(p - \kappa_j)}_{\substack{\text{profit} \\ \text{per sale} \\ \text{(margin)}}},$$

where $C_j = \int \lambda_i c_j(x_i) di$ is total demand for variety j and $\frac{\lambda_i c_j(x_i)}{C_j}$ the fraction of demand accounted for by households of type i. In words, profits are given as the margin per sale $(p - \kappa_j)$ times total demand per retailer (C_j) times the market share. To determine her market share, the retailer considers the likelihood with which any buyer she meets in the market observes a second price quote simultaneously: With probability $\frac{\lambda_i c_j(x_i)}{C_j}$ she meets a type i household, and with probability $s(x_i)$ this household has a simultaneous second price observation conditional on being type i. In this case the retailer only makes a sale if her price offer is lower than the second quote, which conditional on posting price p occurs with probability $(1-F_j(p))$.[4] The problem can be simplified to

$$\pi_j(p) = C_j\left[(1-\bar{s}_j)+\bar{s}_j 2(1-F_j(p))\right](p-\kappa_j), \tag{1.1}$$

where

$$\bar{s}_j = \int \frac{\lambda_i c_j(x_i)}{C_j} s(x_i) di \tag{1.2}$$

is the *demand-weighted average search effort* in the market. Deciding on the price to post in this market, retailers trade off between margins per sale and their market share. A higher price increases the margin earned per sale $(p-k_j)$ but increases the probability to be undercut by a competitor $F_j(p)$ and hence decreases demand at the extensive margin. Taking into account demand-weighted search effort $\bar{s}_j$ is key for the second effect as it determines the ex-ante likelihood that the average buyer observes a second price and therefore the ex-ante likelihood any retailer has to compete for a purchase. In this sense, $\bar{s}_j$ determines the price elasticity of demand across retailers.

As retailers are homogeneous, a non-degenerate equilibrium price distribution requires them to be indifferent between posting a range of prices. For any price increase on the support of the posted distribution, the benefit of earning a higher margin on the current number of sales has to be exactly offset by the cost of a loss in market share. Formally, this requires $\frac{\partial \pi(p)}{\partial p} = 0$ which yields

$$\underbrace{C_j\left[(1-\bar{s}_j)+\bar{s}_j 2(1-F_j(p))\right]}_{\text{current sales}} = \underbrace{C_j\left[\bar{s}_j 2 f_j(p)\right]}_{\text{loss in sales}}\underbrace{(p-\kappa_j)}_{\substack{\text{current}\\ \text{margin}}}. \tag{1.3}$$

The market size C_j cancels from the expression as retailers are infinitesimal and only compete over their share in a total number of sales they take as given. Demand-weighted average shopping effort $\bar{s}_j$ summarizes all relevant infor-

[4]The multiplication of the second term by 2 captures that the retailer can be either the first or second of two price observations.

mation about the distribution of households and is a sufficient statistic for the retailer to post a price.

Posted Price Distribution. For given κ_j, $\bar{p}_j$, and $0 < \bar{s}_j < 1$, Burdett and Judd (1983) and Pytka (2022) show that a unique and continuous equilibrium distribution of posted prices $F_j(p)$ exists with compact support $[\underline{p}_j, \bar{p}_j]$, where

$$F_j(p) = \begin{cases} 0 & \text{if } p < \underline{p}_j \\ 1 - \frac{1-\bar{s}_j}{2\bar{s}_j}\frac{\bar{p}_j - p}{p - \kappa_j} & \text{if } p \in [\underline{p}_j, \bar{p}_j] \\ 1 & \text{if } p > \bar{p}_j \end{cases} \tag{1.4}$$

and

$$\underline{p}_j = \kappa_j + (\bar{p}_j - \kappa_j)\frac{1-\bar{s}_j}{1+\bar{s}_j}.$$

Retailers play a mixed strategy, randomizing prices over the interval $[\underline{p}_j, \bar{p}_j]$ according to the density $f_j(p)$ associated with $F_j(p)$. The distribution of posted prices depends on the marginal cost κ_j and households' maximum willingness to pay $\bar{p}_j$, as well as demand-weighted shopping effort $\bar{s}_j$, but is independent of total demand per retailer C_j. As marginal cost are constant across retailers, $F_j(p)$ is a distribution of markups.

1.2.2 The Effect of Heterogeneous Shopping on Posted Prices

How does the distribution of households affect posted prices? As demand-weighted effort $\bar{s}_j$ is a sufficient statistic for retailers' pricing decision, an answer to this question can be split into two steps: (i) How does $\bar{s}_j$ change with the distribution of households? and (ii) How does the distribution of posted prices respond to changes in $\bar{s}_j$?

Demand-Weighted Shopping Effort. Focus first on how the distribution of households determines $\bar{s}_j$. Equation (1.2) implies that a retailer takes into account type i households' shopping effort according to their share in total demand $\frac{\lambda_i c_j(x_i)}{C_j}$. From here on out I will refer to the vector of these shares as *demand composition*. Differences in demand composition shift the weights attached to each household's idiosyncratic search behavior. In this way, heterogeneity in individual effort $s(x_i)$ creates a role for demand composition to affect $\bar{s}_j$

and through it posted prices. $\bar{s}_j$ is higher if a larger share of demand is accounted for by households with a higher shopping effort. Taking (shifts in) demand composition into account is important to fully capture how changes in the distribution of disposable income affect $\bar{s}_j$. Consider an increase in type i's disposable resources x_i holding the resources of all other households constant. The derivative of $\bar{s}_j$ w.r.t. type i's disposable resources is given by

$$\frac{\partial \bar{s}_j}{\partial x_i} = \frac{\lambda_i c_j(x_i)}{C_j}\frac{\partial s(x_i)}{\partial x_i} + \frac{\lambda_i}{C_j}(s(x_i) - \bar{s}_j)\frac{\partial c_j(x_i)}{\partial x_i}. \tag{1.5}$$

The first term is the change in type i's shopping effort, which is weighted by her share in demand. The second term captures shifts in demand composition. Whether a change in type i's income increases or decreases $\bar{s}_j$ through its effects on shopping behavior and on demand composition depends on the properties of $s(x_i)$ and $c(x_i)$. The sign of the direct effect on shopping is pinned down by the slope of the shopping policy function $\frac{\partial s(x_i)}{\partial x_i}$. The demand composition effect depends on a household's position in the distribution of shopping effort interacted with the change in her consumption policy. If j is a normal good $\left(\frac{\partial c_j(x_i)}{\partial x_i} > 0\right)$ shifts in demand composition increase $\bar{s}_j$ in response to increases in the disposable income of high shopping households with $s(x_i) > \bar{s}_j$ and decrease $\bar{s}_j$ in response to increases in x_i for low shopping households $(s(x_i) < \bar{s}_j)$. If j is an inferior good $\left(\frac{\partial c_j(x_i)}{\partial x_i} < 0\right)$ the argument is reversed.

Moments of the Posted Price Distribution. Even without disciplining households' behavior we can assess how changes in $\bar{s}_j$ affect posted prices. Given $\bar{s}_j$, equation (1.4) determines the ensuing posted price distribution. Taking the derivative of (1.4) with respect to $\bar{s}_j$ yields $\frac{\partial F_j(p)}{\partial \bar{s}_j} \geq 0$, i.e. a distribution with lower demand-weighted shopping effort $\bar{s}_j$ has first-order stochastic dominance over any distribution with higher $\bar{s}_j$ and hence a greater probability to observe high posted prices. Figure 1.1 highlights this result graphically. It shows that for given κ_j and $\bar{p}_j$ a lower level of $\bar{s}_j$ shifts mass of the posted density towards the maximum willingness to pay, away from the marginal cost.

Given the analytical characterization of $F_j(p)$, the problem yields closed form solutions for the moments of the distribution and their relation with $\bar{s}_j$. Expressions for the first three central moments are presented in Proposition 1.

Proposition 1. *The mean μ_j^F, standard deviation σ_j^F, and skewness γ_j^F of the posted price distribution $F_j(p)$ for given κ_j, $\bar{p}_j$, and $0 < \bar{s}_j < 1$ can be derived as*

(i)

$$\mu_j^F = \kappa_j + \underbrace{(\bar{p}_j - \kappa_j)\frac{1-\bar{s}_j}{2\bar{s}_j} log\left(\frac{1+\bar{s}_j}{1-\bar{s}_j}\right)}_{\textit{average posted margin}},$$

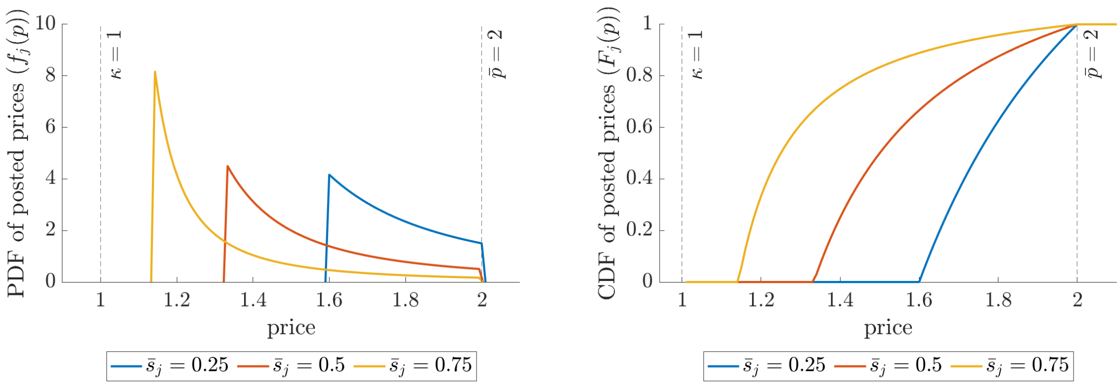
(ii)

$$\sigma_j^F = \sqrt{(\bar{p}_j - \kappa_j)^2 \left(\frac{1-\bar{s}_j}{1+\bar{s}_j} - \left(\frac{1-\bar{s}_j}{2\bar{s}_j}\right)^2 log\left(\frac{1+\bar{s}_j}{1-\bar{s}_j}\right)^2\right)},$$

(iii)

$$\gamma_j^F = \frac{\frac{1-\bar{s}_j}{4\bar{s}_j}\left(1-\left(\frac{1-\bar{s}_j}{1+\bar{s}_j}\right)^2\right) - 3\frac{(1-\bar{s}_j)^2}{2\bar{s}_j+2\bar{s}_j^2} log\left(\frac{1+\bar{s}_j}{1-\bar{s}_j}\right) + 2\left(\frac{1-\bar{s}_j}{2\bar{s}_j}\right)^3 log\left(\frac{1+\bar{s}_j}{1-\bar{s}_j}\right)^3}{\left(\frac{1-\bar{s}_j}{1+\bar{s}_j} - \left(\frac{1-\bar{s}_j}{2\bar{s}_j}\right)^2 log\left(\frac{1+\bar{s}_j}{1-\bar{s}_j}\right)^2\right)^{\frac{3}{2}}}.$$

Proof. *Follows from equation* (1.4) *and the standard formulas for the first three central moments of any continuous distribution.* ■



(a) Density **(b)** Cumulative Density

Figure 1.1: Posted Price Distributions

Note: Price distributions derived from retailers' optimal price posting problem for marginal cost $\kappa_j = 1$, maximum willingness to pay $\bar{p}_j = 2$, and three levels of demand-weighted shopping effort $\bar{s}_j$.

Proposition 2 implies that the average price posted is increasing in marginal cost κ_j and maximum willingness to pay $\bar{p}_j$, but decreasing in demand-weighted shopping effort $\bar{s}_j$. Figure 1.2a illustrates this result graphically.

Proposition 2. *The mean of the posted price distribution μ_j^F is strictly increasing in marginal cost κ_j and maximum willingness to pay $\bar{p}_j$, but strictly decreasing in demand-weighted search effort $\bar{s}_j$ for $0 < \bar{s}_j < 1$, i.e.*

$$\textit{(i)}\ \frac{\partial \mu_j^F}{\partial \kappa_j} > 0,\ \textit{(ii)}\ \frac{\partial \mu_j^F}{\partial \bar{p}_j} > 0,\ \textit{(iii)}\ \frac{\partial \mu_j^F}{\partial \bar{s}_j} < 0.$$

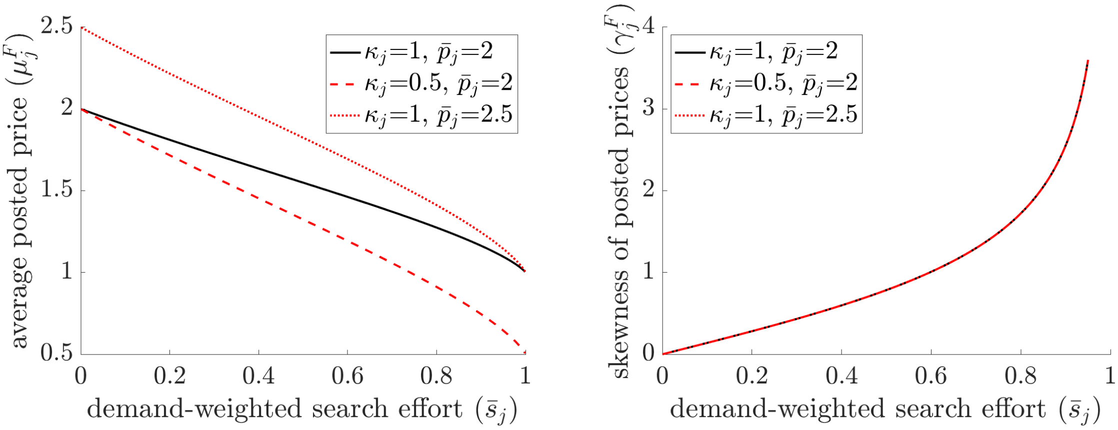
Proof. *Follows from taking first derivatives of μ_j^F.* ■

The effect of shopping on the average price posted operates through changes in profit margins over marginal cost κ_j, which are strictly decreasing in equilibrium shopping effort $\left(\frac{\partial(\mu_j^F - \kappa_j)}{\partial \bar{s}_j} < 0\right)$. Higher demand-weighted shopping effort increases the price elasticity a seller faces. An increase in $\bar{s}_j$ makes it more likely that the average buyer observes a second price, and hence tilts sellers' tradeoff between higher margins and retaining market share in favor of the latter.

In the limit, the setup approaches two well known special cases: If all buyers observe two prices simultaneously ($\bar{s}_j = 1$), retailers solve a Bertrand competition problem and post marginal cost ($\mu_j^F = \kappa_j$). If no buyer observes two prices simultaneously ($\bar{s}_j = 0$), all retailers have a monopoly for any buyer they meet and extract buyers maximum willingness to pay ($\mu_j^F = \bar{p}_j$). Households' shopping effort determines a market's position between these two extremes by regulating the price elasticity retailers face.

Together with the effect of (changes in) demand composition on $\bar{s}_j$ described above, the relationship between the average price posted μ_j^F and demand-weighted effort captures the mechanism at the heart of this paper: If a larger share of demand is accounted for by low-search households, retailers face a lower average price elasticity (lower $\bar{s}_j$) and optimally post higher prices (markups). This is how taking into account retailers' optimal response to changes in demand composition yields equilibrium effects of heterogeneity in shopping effort on posted prices.

The theoretical relation between demand-weighted shopping effort $\bar{s}_j$ and the skewness of the price distribution provides a sharp, empirically testable prediction of the response of posted prices to households' shopping effort. As shown in Proposition 3 below and highlighted graphically in Figure 1.2b, the skewness



(a) Mean

(b) Skewness

Figure 1.2: Moments of the Posted Price Distribution

Note: Theoretical moments of the posted price distribution $F_j(p)$ as a function of demand-weighted shopping effort $\bar{s}_j$, for different values of marginal cost κ_j and maximum willingness to pay $\bar{p}_j$.

of the posted price distribution is a function only of demand-weighted shopping effort $\bar{s}_j$ and independent of parameters. Furthermore, it is strictly increasing in $\bar{s}_j$.

Proposition 3. *The skewness of the posted price distribution γ_j^F is strictly increasing in demand-weighted search effort $\bar{s}_j$ for $0 < \bar{s}_j < 1$, but independent of marginal cost κ_j and maximum willingness to pay $\bar{p}_j$, i.e.*

$$(i)\ \frac{\partial \gamma_j^F}{\partial \kappa_j} = 0, \ \ (ii)\ \frac{\partial \gamma_j^F}{\partial \bar{p}_j} = 0, \ \ (iii)\ \frac{\partial \gamma_j^F}{\partial \bar{s}_j} > 0.$$

Proof. *Follows from taking first derivatives of γ_j^F.* ■

The intuition for this finding goes back to retailers' indifference condition in equation (1.3). Given a distribution of posted prices, an increase in demand-weighted search effort increases sales of retailers with low prices and decreases them for retailers with high prices, as households on average buy more at cheaper offers. This increases the benefit of raising prices at the bottom and decreases the benefit at the top. To offset this effect and keep retailers indifferent between posting low and high prices, the loss in market share when raising prices has to increase at the bottom and decrease at the top. This requires the distribution of posted prices to be more dense at the bottom and less dense at the top. A more (less) dense distribution increases (decreases) the number of competitors that additionally undercut a retailer when raising prices marginally.

A distribution that is more dense at the bottom and less dense at the top exhibits higher skewness.

Robustness. Appendix A.1 shows that the distribution of posted prices remains unchanged when introducing free entry and fixed cost of operating and that under reasonable calibrations heterogeneity in marginal cost leaves average prices decreasing and the skewness increasing in shopping effort.[5]

1.3 Evidence on Shopping Effort and Price Distributions

This section provides empirical evidence on shopping effort across households and price distributions across goods. First, I focus on how shopping effort changes with households' expenditure to provide motivation for the quantitative model below. Second, I test Proposition 3 empirically. I exploit differences in demand composition across goods and evidence on households' shopping effort to show that the skewness of price distributions indeed increases in demand-weighted shopping effort.

Data. For all empirical results I rely on data from the Nielsen Consumer Panel for 2007-2019. The dataset provides detailed information on the grocery purchases of approximately 60,000 US households per year, recording both quantities purchased and prices paid for every store visit at the barcode level. In addition, the data contains annual information on households' demographic characteristics such as income, household composition, employment, and the place of residence.[6]

1.3.1 Shopping Effort across Households

Studying how shopping behavior changes across households requires a measure of search effort. The Nielsen dataset does not provide direct information on the time spent searching for prices. I therefore rely on two proxies for households'

[5] I focus on homogeneous marginal cost as they should be interpreted as wholesale cost and wholesale price differentiation among retailers within a geographic area is prohibited in the US under the federal Robinson-Patman Act and more commonly applied state legislations (e.g. Nakamura (2008)).

[6] Further information on the dataset is provided in Appendix A.2.1.

shopping effort. First, I focus on the outcome of the search process – heterogeneity in the prices paid for identical barcodes – and construct household level price indices in the spirit of Aguiar and Hurst (2007). Second, I consider the number of stores households visit. Kaplan and Menzio (2015) show that an effective way to reduce prices paid is to visit more stores or the same store more often, controlling for the number of purchases.

Price Index. The total cost of household i's consumption bundle in year t across all barcodes j is

$$X_{it} = \sum_j p_{jit} c_{jit},$$

where p_{jit} is the quantity-weighted average price paid by household i in year t for barcode j and c_{jit} is the respective quantity consumed. I further compute counterfactual cost $\tilde{X}_{it}$ assuming the household pays the national, quantity-weighted average price across all households $\tilde{p}_{jt}$ for each transaction of barcode j such that[7]

$$\tilde{X}_{it} = \sum_j \tilde{p}_{jt} c_{jit}.$$

A household's price index is defined as the ratio between true and counterfactual cost

$$P_{it} = \left(\frac{X_{it}}{\tilde{X}_{it}} - 1 \right) * 100.$$

It can be interpreted as the percentage difference in the cost of household i's consumption bundle in year t relative to paying average prices for each of the barcodes purchased. A high index value indicates relatively high prices paid within barcodes and therefore low shopping effort. The relationship between this price index and households' overall expenditure levels is not trivially positive: While higher prices on a given basket of goods necessarily increase expenditure, high-spending households could in principle be buying a larger basket but be paying less for each individual product. The size of the basket is controlled for by normalizing the actual cost of consumption X_{it} by counterfactual spending $\tilde{X}_{it}$. I regress annual individual price indices on household (η_i) as well as

[7] I define $\tilde{p}_{jt}$ anually at the national level. An alternative, more restrictive definition that has been commonly used in the literature defines average prices at the local and quarterly level. I show in Appendix A.2.2 that this definition is subject to a small sample bias attenuating results. Nevertheless, findings based on local average prices are qualitatively similar.

year-state (α_{st}) fixed effects and controls (Z_{it})

$$P_{it} = \eta_i + \alpha_{st} + \gamma Z_{it} + \epsilon_{it}. \tag{1.6}$$

The fixed effects control for all constant unobserved characteristics and for local economic conditions at the year-state level. The vector Z_{it} contains a set of time-varying observed characteristics: Most importantly, I include the logarithm of annual grocery spending to measure how shopping effort changes along the expenditure distribution. To control for other variables commonly associated with shopping effort, I include dummies for households' taxable income between \$30k-60k, \$60k-100k, or above \$100k (omitting the category below \$30k as baseline), the number of non-employed household heads (baseline is no non-employed head), whether the (male) household head is of working age (25-65), and the square root of household size.[8] Nielsen-provided household-level sampling weights are applied throughout. Standard errors are clustered at the household level.

Table 1.1 reports selected results of estimating equation (1.6).[9] It shows that a higher level of household expenditures is associated with paying higher prices for identical goods. In addition, prices paid increase in income and decrease in the number of non-employed household heads. The findings suggest that households with higher spending or income and fewer non-employed household heads exert lower shopping effort. The results for income and employment status are in line with the findings reported in the literature.[10] The result on expenditure is novel.

The relation between expenditure levels and prices paid is sizeable, especially relative to the relation between prices and income. A move from the lowest to the highest income bin increases prices paid by 0.326 percent. Each income bin accounts for roughly 25% of households, so a move from the lowest to the highest bin is approximately equivalent to moving from the lowest to the highest

[8]Grocery expenditures are equivalence scale adjusted by dividing by the square root of household size and deflated to 2019 USD with the urban CPI. Income is reported in the Nielsen dataset as a binned variable and refers to the tax base of the previous year, i.e. household income two years prior to the survey wave. To the extent that household income is persistent at a two-year horizon, it can be seen as an approximation of households' current income but is a noisy measure of the true value.

[9]Full results are reported in column (1) of Table A.1 in Appendix A.2.2.

[10]See e.g. Aguiar and Hurst (2007), Kaplan and Menzio (2016) and Pytka (2022).

Table 1.1: Shopping Effort across Households

	price index	trips per purchase
	(1)	(2)
log(expenditure)	0.706***	−0.042***
	(0.078)	(0.001)
income 30k-60k	0.080*	−0.001*
	(0.046)	(0.001)
income 60k-100k	0.178***	−0.002**
	(0.057)	(0.001)
income >100k	0.326***	−0.002**
	(0.070)	(0.001)
1 non-employed household head	−0.236***	0.002***
	(0.037)	(0.000)
2 non-employed household heads	−0.422***	0.004***
	(0.068)	(0.001)
mean		0.15
FE year-state	X	X
FE household	X	X
Observations	801,398	801,398

Note: Regression of shopping effort on household characteristics. Column (1) index of individual prices paid vs. national annual average price. Column (2) annual number of shopping trips divided by number of purchases. Data from Nielsen Consumer Panel waves 2007-2019. Observations weighted with Nielsen provided sample weights. Standard errors clustered at the household level. *p<0.1; **p<0.05; ***p<0.01.

income quartile. In comparision, doubling a household's grocery expenditures is associated with a 0.706 percent increase in prices for the same barcode. The top expenditure quintile spends roughly five times as much as the bottom quintile, translating the coefficient into a 2.8% difference in prices paid between the bottom and top of the expenditure distribution.

Trips per Purchase. I define a shopping trip as a visit to a unique store at a unique day. To control for the size of the consumption basket, I divide the number of annual trips a household undertakes by the number of her purchases, where a purchase is defined as a transaction involving a unique barcode in a given store on a given day. Column (2) of Table 1.1 shows that the results obtained via the price index also hold when measuring shopping effort as the number of trips per purchase: Households with higher expenditure or income make fewer trips per purchase while households with more non-employed members make more trips. Again the coefficient on households' expenditure is sizeable: The average number of trips per purchase is 0.15 (a household makes on average 6.67 purchases per shopping trip). Doubling a household's expenditures reduces it by 0.042.

The strong relation between households' shopping effort and expenditure levels is well in line with a mechanism introduced in Pytka (2022). As households with higher expenditure make more purchases, they have to search more often to achieve the same average reduction in prices. Hence, reducing the average price paid becomes more costly as the size of a household's basket increases. I take the strong relation between spending levels and households' shopping behavior as motivation for developing a quantitative model centered around households' expenditure below.

1.3.2 Demand Composition and Price Distributions

Having identified dimensions of heterogeneity in shopping effort, I move on to an empirical test of the relationship between shopping and posted price distributions. Testing for a reduction in retailers margins in response to higher shopping effort would require data on markups at the seller-good level. However, for a test of the relationship between the skewness of price distributions and effort outlined in Proposition 3, it is sufficient to observe price distributions and demand-weighted shopping effort. Both are available in the Nielsen dataset.

Demand-Weighted Shopping Effort. According to the theory outline in Section 1.2, retailers should consider the shopping effort of households weighted by their share in overall demand for the variety they sell. I exploit variation in demand composition across products and compute for each barcode the national, annual expenditure shares stemming from different groups of households, sorted by their shopping effort.[11] Building on the results above, I consider separately

[11] I use annual and national shares as Nielsen is representative at this level. Using aggregate rather than local shares is justified by the evidence on uniform price setting of large retail chains across locations (see e.g. DellaVigna and Gentzkow, 2019), making national rather than local demand composition the relevant statistic for their price setting.

the five quintiles of the expenditure distribution, four bins of household income, as well as the number of non-employed household heads. To be in line with the predictions from Section 1.2 and the results on households' shopping behavior above, the skewness of price distributions should be decreasing in the expenditure share coming from high-spending or high-income households, but increasing in the share of demand from households with more non-employed heads.

Price Distributions. A price distribution consists of all transactions observed for a barcode j, within a region r and time period t. In line with Kaplan and Menzio (2015), I define a region as a Scantrack Market Area (SMA) and the time period to be a quarter.[12] The price associated with a transaction is defined as the total amount paid less of coupon values, divided by the quantity purchased. To control for outliers, I drop all transactions for which the reported amount paid less of coupons is zero or negative. For the baseline analysis, I consider all price distributions containing at least 25 transactions and compute the skewness of each distribution weighting individual price observations with household weights and quantities purchased.

Estimation. To test for the relationship between skewness and shopping derived from theory, I regress the skewness of a price distribution (j, r, t) on the national expenditure shares of each household group g, for variety j in the respective year $y(t)$. I run separate regressions defining groups based on expenditure quintiles, income bins, and the number of non-employed household heads, excluding the lowest expenditure quintile, the lowest income bin and households with no non-employed head respectively as a baseline. The specification is given in equation (1.7).

$$skew_{j,r,t} = \theta_m + \mu_{r,t} + \sum_{g=2}^{G} \beta_g share_{j,g,y(t)} + \varepsilon_{j,r,t} \tag{1.7}$$

[12] The choice for what definition of a region and which time period to consider trades off between two forces: A narrow definition ensures that any variation in prices can be confidently allocated to (and exploited by) search frictions, while it also reduces the number of price observations per distribution and hence makes the analysis more noisy. For the ensuing analysis to be valid it is not necessary that households have access to every price within a region, but only that the distribution of prices is identical for any subregion. As Scantrack Markets are defined by industry professionals as target regions for marketing purposes, retailers pricing can be assumed to be sufficiently similar within such regions to ensure identical price distributions throughout.

To control for local economic conditions and product characteristics, I include time-region fixed effects ($\mu_{r,t}$) as well as fixed effects for Nielsen-defined product modules (θ_m). I do not control for barcode fixed effects to exploit variation in expenditure shares across different barcodes.[13] The included fixed effects demean the skewness by product category and by region at a given point in time. Therefore, the coefficients of interest β_g are identified by the covariation of demand composition and differences in the skewness of distributions among closely substitutable barcodes within a given region and period. All regressions are weighted by the total amount of expenditures contained in the respective price distributions. Standard errors are clustered at the barcode-year level.

Table 1.2: Demand Composition and the Skewness of Price Distributions

	by expenditures			by income		by employment
	all	working age		working age		all
	(1)	(2)		(3)		(4)
expenditure	-1.638^{***}	-1.467^{***}	income	-0.136	1 non-employed	0.864^{***}
quintile 2	(0.242)	(0.206)	30k-60k	(0.133)	household head	(0.115)
expenditure	-2.309^{***}	-2.076^{***}	income	-0.824^{***}	2 non-employed	1.011^{***}
quintile 3	(0.256)	(0.221)	60k-100k	(0.155)	household heads	(0.210)
expenditure	-3.067^{***}	-2.582^{***}	income	-0.820^{***}		
quintile 4	(0.258)	(0.219)	>100k	(0.139)		
expenditure	-3.412^{***}	-3.007^{***}				
quintile 5	(0.253)	(0.224)				
FE product module	X	X		X		X
FE quarter-SMA	X	X		X		X
Observations	3,026,551	3,026,404		3,026,404		3,026,551

Note: Regression of the skewness of price distributions on demand shares by household groups. Price distributions defined as all transactions of a barcode within a Scantrack Market Region and quarter. Demand shares defined as the share of national annual spending on a barcode by each group of households. Data from Nielsen Consumer Panel waves 2007-2019. Observations weighted by total sales in given price distribution. Standard errors clustered at the barcode-year level. *p<0.1; **p<0.05; ***p<0.01.

Table 1.2 reports the results. The skewness of price distributions is monotonically decreasing in the share of expenditure stemming from higher spending households (column (1)). The coefficients should be interpreted as the relative skewness compared to the omitted baseline group. For column (1): If a barcode is bought entirely by households in the fifth quintile of the expenditure distribu-

[13] Nielsen-defined product modules are the first level of aggregation above barcodes and capture product characteristics at a granular level. Examples of product modules in Nielsen are e.g. "fresh apples" or "fresh oranges" for different categories of fresh fruits.

tion, the skewness of its price distribution decreases by 3.4 relative to a barcode bought entirely by the first quintile. All differences w.r.t. the baseline group are statistically significant at the 1%-level. The finding is robust to measuring expenditure shares conditional on the (male) household head being between age 25-65 to account for spending patterns of student and retiree households (column (2)). Similar findings pertain by income group, again conditioning on working age households (column (3)). In addition, the skewness is monotonically increasing in the number of non-employed household heads (column (4)). All specifications suggest one conclusion: The skewness of price distributions decreases in the share of expenditure stemming from low-effort households. This is well in line with Proposition 3 and provides strong evidence in favor of the theoretical relationship between search effort and posted prices.

Robustness. In Appendix A.2.3, I report further robustness with respect to how the skewness of price distributions is measured for the specification of column (1) of Table 1.2. Table A.2 in Appendix A.2.3 reports results without using weights in the regression, computing the skewness based on unweighted price observations, or based on household weights only. All findings are robust to using alternative weighting schemes. The decrease of skewness in expenditure is also robust when using Kelly's measure of skewness, which is less sensitive to outliers.[14] As Table A.3 in Appendix A.2.3 shows, results become quantitatively stronger if considering only price distributions with at least 50 or 100 transactions. The robustness tests alleviate potential concerns that price distributions are constructed based on transaction data sampled from households. As the findings are robust to not weighting by quantities purchased and focusing on distributions with many transactions (where each posted price has a better chance of entering the sample) it is unlikely that households' purchase behavior is driving the results.[15]

[14]The units of the coefficients are not comparable for Kelly's measure of skewness, so no statements can be made about the relative magnitude of the results in column (2) of Table A.2 in Appendix A.2.3.

[15]In future work, I plan to extend the analysis to the Nielsen Retail Scanner dataset providing information on posted prices sampled directly from stores.

1.4 A Theory of Inequality and Price Dispersion

In Section 1.2 I have taken households' choices as given to derive analytical results on how posted prices respond to the choices of households. In equilibrium, households' shopping effort and consumption choices are themselves a function of the distributions of posted prices. To account for this feedback between households' choices and retailers' price posting, this section develops an equilibrium theory of expenditure inequality and price dispersion and disciplines it against evidence from the Nielsen Consumer Panel.

1.4.1 Households with Non-Homothetic Preferences and Shopping

Households are infinitely lived and heterogeneous in their labor earnings zw. w is the common wage rate per unit of labor and z households' idiosyncratic labor productivity, evolving exogenously according to a first order Markov process. Households supply z efficiency units of labor inelastically. In addition, they earn a return r per unit of beginning of period assets a. Households decide jointly on their future asset holdings a', quantities consumed of each variety $j \in J$ of a grocery good $\{c_j\}_{j=1}^J$ and an outside (non-grocery) good c_O, and shopping effort s.

Households' decision problem can be split into two stages. In a first stage a household divides her resources between savings a' and total expenditure e to solve

$$\begin{aligned} V(z,a) = \max_{e,a' \geq 0} U(e) + \beta \mathbb{E}_{z'|z} V(z',a') \\ \text{s.t. } e + a' \leq (1+r)a + zw. \end{aligned} \tag{1.8}$$

The utility of expenditure $U(e)$ summarizes the second stage in which households decide on their allocation of consumption across grocery varieties and the outside good as well as their choice for shopping effort, conditional on expenditure. They solve

$$
\begin{aligned}
U(e) = & \max_{s\in[0,1],\{c_j\}_{j=1}^{J},c_O} u(\mathcal{C}) - v(s,\mathcal{C}) \\
\text{s.t.} \quad & \mathcal{C} = (c_G)^{\alpha}(c_O)^{1-\alpha} \\
& c_G = \left[\sum_{j=1}^{J} (\mathcal{C})^{\frac{q_j}{\sigma}} (c_j)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} \\
& c_O + \sum_{j=1}^{J} p_j(s)c_j \leq e.
\end{aligned} \tag{1.9}
$$

The outside good is taken to be the numeraire and its price is normalized to 1. $u(\cdot)$ are households' preferences over the consumption aggregator $\mathcal{C}$ and $v(s,\mathcal{C})$ is the disutility of exerting shopping effort. I assume that the disutility of effort depends on the level of consumption $\mathcal{C}$ to capture in reduced form that households have to search more often for prices if they have a larger consumption basket.[16]

Due to the two stage setup, the distribution of expenditures fully determines the distribution of shopping effort and consumption baskets across households. The structure allows me to focus on data moments of the expenditure distribution when disciplining households' shopping and consumption policies below.

Consumption Allocation. The aggregator $\mathcal{C}$ is a Cobb-Douglas function defined over grocery and non-grocery consumption. Grocery consumption c_G is itself a non-homothetic CES aggregator over varieties $j \in J$ in the spirit of Comin et al. (2021) and Handbury (2021). For given total consumption $\mathcal{C}$ and shopping effort s, it defines a demand system across varieties that can be characterized in terms of expenditure shares ω_j, where the optimal allocation satisfies

$$
\frac{\omega_j}{\omega_k} = \mathcal{C}^{q_j - q_k} \left(\frac{p_j(s)}{p_k(s)}\right)^{1-\sigma}.
$$

Varieties should be considered close substitutes and can be thought of as different barcodes within a Nielsen defined product module. I focus on this low level of product differentiation as a significant degree of non-homotheticities occurs

[16]This mechanism is micro-founded in Pytka (2022) who also provides evidence that conditional on employment high-income households spend more time making purchases and rules out that this is due to shopping as a leisure activity.

at this granular definition of a variety.[17] The parameters $\{q_j\}_{j=1}^{J}$ govern the expenditure elasticity of demand: With $\mathcal{C}$ increasing in expenditures, the relative expenditure share of variety j vs. variety k $\left(\frac{\omega_j}{\omega_k}\right)$ is increasing in total spending e iff $q_j > q_k$. In line with the literature's interpretation of more expensive varieties among close substitutes as higher quality products, I will refer to varieties with a high q_j as *high-quality*.[18] Under this interpretation, the non-homotheticities considered arise because high-spending households have a stronger taste for quality.

The price of the optimal grocery consumption bundle (of one unit c_G) is given as

$$p_G(\mathcal{C}, s) = \left(\sum_{j=1}^{J} \mathcal{C}^{q_j} \left(p_j(s)\right)^{1-\sigma} \right)^{\frac{1}{1-\sigma}}.$$

The Cobb-Douglas aggregator $\mathcal{C}$ implies optimal shares of expenditures on groceries e_G and the outside good e_O given by

$$\omega_G = \frac{e_G}{e} = \frac{p_G(\mathcal{C}, s) c_G}{e} = \alpha \quad \text{and} \quad \omega_O = \frac{e_O}{e} = \frac{c_O}{e} = 1 - \alpha.$$

For given shopping effort s and expenditure level e the consumption aggregator is a solution to the non-linear equation

$$\mathcal{C} = \frac{e}{P(\mathcal{C}, s)}, \tag{1.10}$$

where the price index associated with $\mathcal{C}$ is given as

$$P(\mathcal{C}, s) = \left(\frac{p_G(\mathcal{C}, s)}{\alpha}\right)^{\alpha} \left(\frac{1}{1-\alpha}\right)^{1-\alpha}.$$

Shopping Effort. The price a household pays for any variety j of the grocery good is a function of her shopping effort. Households' optimal choice of shopping effort equates the marginal benefits of shopping with the marginal disutility of exerting effort such that

[17] In Appendix A.2.4 I compared consumption baskets along the expenditure distribution in the Nielsen data and show that significant non-homotheticities arise when defining a product as a barcode as compared to aggregating goods at the product module level. This is in line with e.g. Jaravel (2019).

[18] See e.g. Bils and Klenow (2001), Bisgaard Larsen and Weissert (2020), or Argente and Lee (2021).

$$\underbrace{-v_s(s,\mathcal{C})}_{\substack{\text{marginal disutility}\\ \text{of shopping}}} = \underbrace{\frac{\partial p_G(\mathcal{C},s)}{\partial s} c_G}_{\substack{\text{change}\\ \text{in BC}}} \underbrace{(1-\alpha)\frac{\mathcal{C}}{c_O}\left[u'(\mathcal{C}) - v_{\mathcal{C}}(s,\mathcal{C})\right]}_{\substack{\text{marginal benefit}\\ \text{of relaxing BC}}}. \tag{1.11}$$

The benefit of shopping is the change in the budget constraint times the marginal utility of additional available resources, here expressed as the marginal utility of consuming one more unit of c_O.[19] Increasing shopping effort reduces the cost of consumption and relaxes the budget by $\frac{\partial p_G(\mathcal{C},s)}{\partial s} c_G$, where the change in the price index depends on households' consumption basket and the return to shopping for each variety. It is given by

$$\frac{\partial p_G(\mathcal{C},s)}{\partial s} = p_G(\mathcal{C},s)^{\sigma} \sum_{j=1}^{J} \mathcal{C}^{q_j} \left(p_j(s)\right)^{-\sigma} \underbrace{\frac{\partial p_j(s)}{\partial s}}_{\substack{\text{return to}\\ \text{shopping}}}.$$

The relationship between prices and shopping effort, i.e. the return to shopping effort $\frac{\partial p_j(s)}{\partial s}$, is an equilibrium object and depends on the distribution of posted prices.

1.4.2 Equilibrium in the Goods Market and Return to Shopping

Production. All grocery varieties and the outside good are produced and sold to retailers at marginal cost by fully competitive production firms. Producers operate linear technologies with labor N as the single input factor. I assume production functions $y_O = N_O$ and $y_j = \frac{1}{\kappa_j} N_j$, so that with the outside good as the numeraire the equilibrium wage is determined as $w = 1$ and the marginal cost of producing a unit of variety j is given by κ_j. In line with Kaplan and Menzio (2016), I assume that households can transform the outside good into groceries of variety j at a rate $\frac{1}{\bar{p}_j}$ implying a maximum willingness to pay $\bar{p}_j$ to purchase variety j from a retailer. Households' assets are invested in a risk-free bond at exogenous interest rate r. Under these assumptions, the model outcomes can be interpreted as the equilibrium of a small open economy or as the equilibrium in a subregion (state) of a large economy like the US.

[19] At the optimal solution, marginal utility could also be expressed in terms of spending the additional unit of available resources on any of the grocery varieties or the composite grocery good c_G.

Market Structure. The outside good is traded in a perfectly competitive market and its price is independent of shopping effort, whereas all varieties of grocery consumption are sold in markets that are subject to search frictions. There is a separate search market for each grocery variety j. The distribution of posted prices for each variety $F_j(p)$ is determined by the optimal price posting of a mass of single-variety retailers. As before, retailers post prices for variety j given equilibrium search effort $\bar{s}_j$, marginal cost κ_j and maximum willingness to pay $\bar{p}_j$. I deviate from the setup laid out in Section 1.2 and assume that a retailer selling variety j is subject to a per period fixed cost of operation K_j and the mass M_j of active retailers selling variety j is determined by free entry. Profits of posting price p in the market for variety j are given by

$$\tilde{\pi}_j(p) = \frac{C_j}{M_j}\left[(1-\bar{s}_j)+\bar{s}_j 2(1-F_j(p))\right](p-\kappa_j) - K_j = \frac{\pi_j(p)}{M_j} - K_j,$$

where $\pi_j(p)$ is retailers' profits of posting p in the version of the model without fixed cost of operating and a given mass one of retailers as outlined in Section 1.2. Appendix A.1.1 proves that this setup yields a distribution of posted prices equivalent to the one derived in Section 1.2, while the mass of entrants M_j ensures that $\tilde{\pi}_j(p) = 0$, i.e. retailers make zero profits in equilibrium. Beyond ensuring zero profits, M_j and K_j do not influence equilibrium allocations and only their product is uniquely determined by $\pi_j(p) = M_j K_j$ for any p on the support of $F_j(p)$.

Return to Search. To determine the relationship $p_j(s)$ between prices paid for variety j and households' shopping effort, I follow Pytka (2022). The distribution of effective prices for a single purchase of variety j when exerting shopping effort s is given as

$$G_j(p|s) = (1-s)F_j(p) + s(1-(1-F_j(p))^2).$$

The assumptions that households split their total demand for each variety into a continuum of purchases and that every price observation is an i.i.d. random draw from $F_j(p)$ ensure that there is no uncertainty about the total cost of purchasing a quantity c_j. Pytka (2022) shows that the average price paid per purchase of variety j is

$$p_j(s^i) = \mathbb{E}_j^G[p|s^i] = \mu_j^F - s^i \left(\mu_j^F - \mathbb{E}_j^F[min\{p',p''\}]\right), \tag{1.12}$$

such that

$$\frac{\partial p_j(s^i)}{\partial s^i} = -\left(\mu_j^F - \mathbb{E}_j^F[min\{p',p''\}]\right) = \text{const.} < 0.$$

The two constants μ_j^F and $\frac{\partial p_j(s^i)}{\partial s^i} < 0$ are sufficient statistics to capture the impact of the price distribution of variety j on households' behavior, i.e. all that households need to know to decide on their demand for each variety and their shopping effort. This feature simplifies the computational solution of the model significantly as households do not need to keep track of the entire price distribution. It also suggests that matching the average price and a measure of price dispersion across varieties is sufficient to discipline $p_j(s^i)$, as the dispersion of prices is closely related to the equilibrium object $\frac{\partial p_j(s^i)}{\partial s^i}$.

Equilibrium. A stationary equilibrium in the economy consists of households' value function $V(z,a)$, consumption policy functions $\left\{c_O(z,a), \{c_j(z,a)\}_{j=1}^J\right\}$, shopping policy $s(z,a)$, expenditure policy $e(z,a)$ and savings policy $a'(z,a)$, the induced distribution of households across states $\lambda(z,a)$, aggregated demand $\{C_j\}_{j=1}^J$ and demand-weighted shopping effort $\{\bar{s}_j\}_{j=1}^J$ for each variety, posted price distributions $\{F_j(p)\}_{j=1}^J$ and implied pricing functions $\{p_j(s)\}_{j=1}^J$, where

(i) Given $\{p_j(s)\}_{j=1}^J$, households' value and policy functions solve (1.8) and (1.9).

(ii) The distribution of households is a stationary solution to the law of motion

$$\lambda(z',a') = \int\int \lambda(z,a) Pr(z'|z) \mathbb{1}_{a'=a(z,a)} \, dz \, da.$$

(iii) Aggregated demand for variety j is given by

$$C_j = \int\int \lambda(z,a) c_j(z,a) \, dz \, da.$$

(iv) Demand weighted shopping effort for variety j is given by

$$\bar{s}_j = \int\int \frac{\lambda(z,a)c_j(z,a)}{C_j} s(z,a)\, dz\, da.$$

(v) Given $\{\bar{s}_j\}_{j=1}^{J}$, the posted price distributions $\{F_j(p)\}_{j=1}^{J}$ solve (1.4).

(vi) Given $\{F_j(p)\}_{j=1}^{J}$, the pricing functions $\{p_j(s)\}_{j=1}^{J}$ satisfy (1.12).

1.4.3 Calibration

I calibrate the model at annual frequency. The calibration proceeds in three steps: I first calibrate the income process outside of the model, describe functional forms and set some parameters exogenously, and finally calibrate all remaining parameters to match targets on expenditure composition, price dispersion, and macro aggregates.

Income Process. The process for idiosyncratic productivity is the same as in Ferriere et al. (2022) and Mendicino et al. (2022). I assume an AR(1) with innovations from a Gaussian mixture, to capture higher moments of income risk as reported e.g. in Guvenen et al. (2021b). I target the cross-sectional variance of income, as well as moments of the distribution of income changes. Data targets are obtained from De Nardi et al. (2020). Details on the calibration of the income process are delegated to Appendix A.3.1.

Functional Forms and External Parameters. I assume CRRA preferences for $u(\cdot)$ and a disutility of shopping effort as a function of the consumption aggregator $\mathcal{C}$ such that

$$u(\mathcal{C}) = \frac{\mathcal{C}^{1-\phi}-1}{1-\phi} \quad \text{and} \quad v(s,\mathcal{C}) = \psi_1 \mathcal{C}^{\psi_2}\frac{s^2}{1-s}.$$

The term $\frac{s^2}{1-s}$ ensures that households will prefer an interior solution for s.[20] I restrict $\psi_1 > 0$ and $\psi_2 > 0$ to obtain a disutility of effort increasing in $\mathcal{C}$ and capture the need to search more often for prices when making more purchases. In this spirit, ψ_2 determines the economies of scale in shopping effort.

[20] It yields $v(0,\mathcal{C}) = 0$, $v(1,\mathcal{C}) = \infty$, $v_s(0,\mathcal{C}) = 0$, $v_s(1,\mathcal{C}) = \infty$. Under these assumptions households optimally choose $0 < s < 1$ iff $\frac{\partial p_j(s^i)}{\partial s^i} < 0$. An interior solution for s facilitates the computational solution of the model.

The calibrated version of the model features three varieties (levels of quality), i.e. $J = 3$.[21] In line with the evidence on low-level elasticities of substitution sampled in Jaravel and Olivi (2021), I set $\sigma = 2$. Furthermore, I normalize the medium quality to $q_2 = 0$ and marginal cost of the lowest quality to the outside good, i.e. $\kappa_1 = 1$. The CRRA parameter is set to $\phi = 2$ and the annual real interest rate to $r = 0.02$.

Based on Broda and Parker (2014), I set $\alpha = 0.35$ to the share of non-durable and services consumption covered by the Nielsen dataset. This is a conservative choice, as search frictions and price dispersion can be expected to matter beyond the products covered by Nielsen. Without data for additional product categories, I restrict the search friction to the share of goods covered in Nielsen. Under this assumption, I interpret all results on overall consumption $\mathcal{C}$ and welfare as a lower bound on the consequences of shopping frictions for inequality and report results based on grocery consumption c_G separately as an approximation for an economy with $\alpha = 1$.

Internal Parameters. The remaining parameters to be calibrated are $(\psi_1, \psi_2, \kappa_2, \kappa_3,\ \bar{p}_1, \bar{p}_2, \bar{p}_3, \beta, q_1, q_3)$. As they do not influence allocations, I do not need to account for the fixed cost K_j in the calibration. I impose $q_1 = -q_3$ and $\bar{p}_j = a + b(\kappa_j - \kappa_1)$. This leaves eight parameters for which I target eight moments, divided into three groups.

At the aggregate level, I target a wealth-to-income ratio of 3. While all parameters can influence all moments, the one most closely linked to the wealth to income ratio is β. Furthermore, I target an average retail markup of 1.39, computed based on the US Census' Annual Retail Trade Survey.[22] This value lies within the set of results reported for retail markups in the literature, ranging from 1.31 (Sangani, 2022) to 1.45 (Hall, 2018). The aggregate markup is closely related to ψ_1, which governs average shopping effort and hence the average price elasticity of demand across retailers.

[21] Considering versions of the model with 4 or 5 varieties does not alter the conclusions drawn below.

[22] I use data for 2007-2019 and take sales divided by purchases net of the change in inventories for *food and beverage stores*, *health and personal care stores*, and *general merchandise stores* as the categories most closely reflecting the retailers covered in Nielsen. I weight markups across categories by total sales.

Table 1.3: Calibration Targets and Model Fit

Target	Data	Model	Source
basket overlap (Q1 vs. Q5)	63.28%	63.76%	Nielsen (2007-2019)
Δp across varieties (Q1 vs. Q5)	7.2%	7.2%	Nielsen (2007-2019)
Δp within varieties (Q1 vs. Q5)	2%	2%	Nielsen (2007-2019)
mean(CoV_j)	0.1920	0.1915	Nielsen (2007-2019)
$CoV_2 - CoV_1$	-0.0120	-0.0078	Nielsen (2007-2019)
$CoV_3 - CoV_1$	-0.0203	-0.0211	Nielsen (2007-2019)
average markup	1.39	1.39	ARTS (2007-2019)
wealth/income	3	3	

Note: Results of the internal calibration of $(\psi_1, \psi_2, \kappa_2, \kappa_3, \bar{p}_1, \bar{p}_2, \bar{p}_3, \beta, q_1, q_3)$.

A second set of moments targets price dispersion across varieties. These moments are closely related to the return to search. Capturing the right returns to search across varieties (and therefore across consumption baskets) is important for the correct identification of the elasticity of shopping to expenditures. Targets for price dispersion are computed from the Nielsen Consumer Panel based on the same definition of a price distribution as in Section 1.3.2, i.e. pooling transactions for a given barcode within a Scantrack region and a quarter. To account for differences in the average price across barcodes in the data, I focus on the coefficient of variation (CoV).[23] I target the expenditure-weighted average CoV across all price distributions. In addition, I run a regression of the CoV on demand composition by barcode as specified in equation (1.7) for the skewness, including on the right-hand side the quintiles of the expenditure distribution. I target the implied differences in the CoV across varieties based on the endogenous demand composition (spending shares across quintiles) in the model. Targets for price dispersion interact most closely with the values for $\bar{p}_j$ relative to κ_j.

The final set of moments contains targets on expenditure composition across households, again measured from the Nielsen data. This set of targets is particularly important as it identifies the elasticities of consumption baskets and shopping effort to households' expenditure and with them the main mechanism of this paper. To discipline how consumption baskets change across households, I target the (dis)similarity in expenditure shares ω_j at the barcode level between

[23] Targeting the coefficient of variation is equivalent to normalizing all prices by the mean and computing the standard deviation of normalized prices as e.g. in Kaplan and Menzio (2015).

Table 1.4: Calibrated Parameter Values

Parameter	Value	Notes
J	3	
σ	2.0	Jaravel and Olivi (2021)
α	0.35	Broda and Parker (2014)
ϕ	2.0	
r	0.02	
q_j	$[-0.67\ 0\ 0.67]$	
ψ_1	0.0023	
ψ_2	0.21	
κ_j	$[1\ 1.06\ 1.219]$	
$\bar{p}_j$	$[2.55\ 2.65\ 2.90]$	$2.55 + 1.6(\kappa_j - \kappa_1)$
β	0.9232	

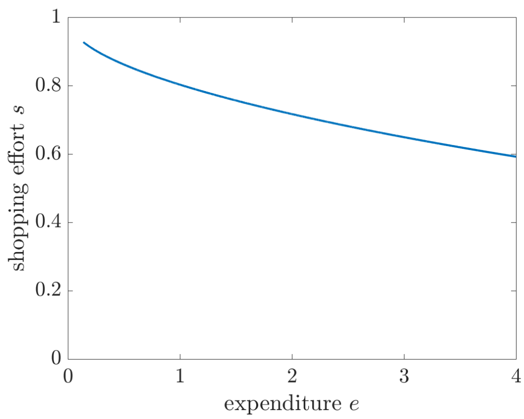
Note: Summary of calibrated parameter values. $(J, \alpha, \sigma, \phi, r)$ are set externally. $(\psi_1, \psi_2, \kappa_2, \kappa_3, \bar{p}_1, \bar{p}_2, \bar{p}_3, \beta, q_1, q_3)$ are calibrated internally.

the first and the fifth quintile of the expenditure distribution. For this purpose, I interpret the vector of expenditure shares for a group of households as a discrete distribution over the universe of available varieties (barcodes) and measure the similarity between two such distributions as the histogram overlap. Full details on the construction of this target are provided in Appendix A.2.4. The barcode-level overlap between consumption baskets at the bottom and top of the expenditure distribution is about 63%. In addition, I target the annual savings as a share of respective grocery expenditure of households at the bottom quintile of expenditures relative to the top quintile, due to (i) buying similar varieties that are cheaper on average and (ii) paying less for identical varieties. I measure (i) as the difference between the per-unit average prices of a barcode and the per-unit average price of all barcodes within a Nielsen product module. (ii) is measured as the difference in the price a household pays for a barcode to the average price for this barcode across all households. More details on the construction of these targets are provided in Section 1.5.1. Price differences across varieties reduce expenditures of the bottom relative to the top quintile by 7.2% and price differences within varieties by 2.0% of annual spending. Given the returns to search across varieties (and consumption baskets) identified by the moments on price dispersion, the difference in prices paid for identical varieties identifies ψ_2, which governs the shape of households' shopping policy along the expenditure distribution. The overlap in consumption baskets and price differences across varieties interact closely with relative expenditure elasticities q_j as well as κ_j and $\bar{p}_j$ across varieties.

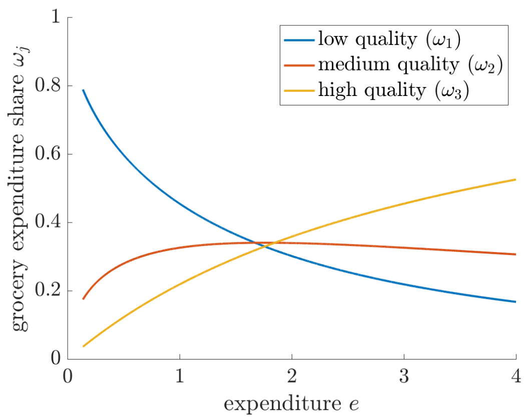
A description of the algorithm applied to solve the model is delegated to Appendix A.3.2. Table 1.3 summarizes all targets and shows that the model is able to match the moments considered. Table 1.4 reports the calibrated parameter values. Noteworthy is the fact that the calibration yields $\psi_2 < 1$. This value implies increasing returns to scale in shopping effort as the utility cost of exerting a given effort s increases less than one-for-one with the consumption aggregator $\mathcal{C}$.

1.4.4 Model Properties and Validation

Policy Functions. The policy functions implied by the first stage spending-savings problem are standard, expenditures and future asset holdings are increasing in disposable resources $(zw+(1+r)a)$ and the expenditure policy is concave. Figure 1.3 displays selected policy functions for the second stage problem of choosing shopping effort and allocating spending across goods. It shows that shopping effort is decreasing and mildly convex in households' total expenditure e and that households with higher expenditures allocate a larger share of their consumption basket to varieties with higher quality (higher elasticity q_j) due to their non-homothetic preferences.[24]



(a) Shopping Policy



(b) Grocery Expenditure Shares

Figure 1.3: Shopping Effort and Consumption Baskets

Note: Model implied shopping effort s and expenditure shares ω_j across grocery varieties j as a function of total household expenditure e. Expenditure shares are computed as total spending of a household on variety j divided by her total grocery expenditures $e^G = \alpha e$.

[24]The policy function for non-grocery consumption follows trivially from $c_O = (1-\alpha)e$.

Demand Composition. Table 1.5 shows that the non-homotheticity in households' preferences leads to significant differences in demand composition across grocery varieties. While households in the bottom quintile of the expenditure distribution account for 11.38% of the spending on the low quality variety, they account for only 2.49% of spending on the high quality variety. On the other hand, households in the top quintile of the expenditure distribution account for 51.45% of spending on the highest quality variety but only 26.42% on the lowest quality variety. Intuitively, each household is more important for the demand of varieties in her own consumption basket.[25]

Table 1.5: Demand Composition of Grocery Varieties

	quintile of expenditures				
	Q1	Q2	Q3	Q4	Q5
low quality (q_1)	0.1138	0.1752	0.2107	0.2360	0.2642
medium quality (q_2)	0.0555	0.1228	0.1838	0.2521	0.3859
high quality (q_3)	0.0249	0.0770	0.1432	0.2405	0.5145

Note: Model implied demand shares by varieties of the grocery good and expenditure quintile. Demand shares are computed as total sales of variety j to quintile g divided by total sales of variety j.

Price Distributions. Table 1.6 provides summary statistics for the model generated price distributions. In line with households' shopping policy and the demand composition across varieties, demand-weighted shopping effort $\bar{s}_j$ is decreasing in quality q_j. The average price paid μ_j^G is lower than the average price posted μ_j^F for each variety due to households' search for cheaper offers. The standard deviation of posted prices σ_j^F increases with q_j and so does the return to search (i.e. $\frac{\partial p_j(s^i)}{\partial s^i}$ is more negative). This implies that high-spending households exert lower search effort despite a higher return to search for their consumption basket.

The skewness of price distributions decreases in the demand share of high spending households, qualitatively in line with the empirical results in Section 1.3.2. The differences in skewness generated by the model can account for one third of the differences predicted by the empirical results based on the model implied demand composition. The lower magnitude is of little concern for

[25]Tabel A.7 in Appendix A.3.5 shows this both for the model and the data. For a more formal discussion of how to measure this in the data refer to Appendix A.2.4.

Table 1.6: Price Distributions of Grocery Varieties

		quality of grocery variety		
		low (q_1)	medium (q_2)	high (q_3)
demand-weighted shopping effort	$\bar{s}_j$	0.76	0.73	0.71
average price posted	μ_j^F	1.49	1.60	1.83
average price paid	μ_j^G	1.38	1.48	1.71
standard deviation of posted prices	σ_j^F	0.30	0.31	0.33
return to search	$\frac{\partial p_j(s^i)}{\partial s^i}$	-0.15	-0.16	-0.17
skewness of posted prices	γ_j^F	1.51	1.42	1.33

Note: Summary statistics of model implied price distributions for each grocery variety j.

the results below, as by construction of the pricing function $p_j(s)$ households' choices are determined by the first two moments of a distribution.

Inequality. Table 1.7 reports the distribution of total disposable income in the model $(zw+ra)$ as well as data on total household income after taxes and transfers from the Congressional Budget Office (CBO). The model generates realistic inequality in income.

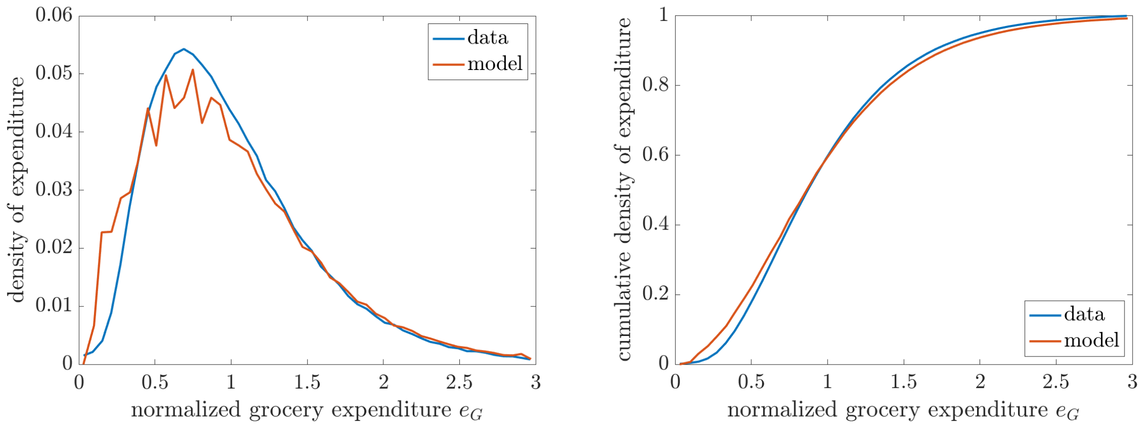
Table 1.7: Income Distribution – Model vs. Data

	quintile of post-tax income				
	Q1	Q2	Q3	Q4	Q5
Model	5.02%	10.50%	15.78%	23.89%	44.79%
Data	6.44%	10.91%	14.70%	20.31%	47.65%

Note: Fit of the model implied income distribution. In the model, income is measured as labor and financial income $(zw+ra)$. Data moments for household income after taxes and transfers from Congressional Budget Office (CBO) for 2007-2018.

Due to the two-stage setup of the household problem, shopping effort and the allocation of consumption across varieties are determined conditional on expenditure e. Matching the distribution of expenditures therefore ensures a realistic distribution of shopping and consumption policies in the model. Figure 1.4 plots the model implied distribution of grocery expenditures along with its equivalent from the Nielsen dataset.

While the calibration targets include moments of the labor earnings process and match the empirical overlap in consumption baskets, as well as price differences within and across varieties, the dispersion in households' expenditure is not included in the calibration. Nevertheless, the model does remarkably well in capturing the empirically observed distribution of grocery expenditures. This



(a) Density **(b)** Cumulative Density

Figure 1.4: Distribution of Grocery Expenditures – Model vs. Data

Note: Fit of the model implied distribution of grocery expenditures $e_G = \alpha e$. Model and data distributions normalized to mean 1 and sorted into 50 equally spaced bins between 0 and the maximum expenditure level. Data from Nielsen Consumer Panel 2007-2019. Household-level grocery expenditures are equivalence scale adjusted and deflated with the Urban CPI to 2019. I winsorize the top 1% of data.

finding provides confirmation that the model is a suitable framework for studying the relationship between expenditure inequality and posted prices.

1.5 Implications for Expenditure Inequality

1.5.1 Price Differences across Households

Before studying the consequences of shopping effort for inequality, I quantify price differences across and within varieties along the expenditure distribution. To make the model and data comparable along this dimension, I focus on the contribution of price differences relative to expenditure and provide a novel decomposition of households' grocery spending.

The first step of a price-based decomposition of expenditure inequality in the data is to make per-unit prices comparable across products. Therefore, I sort all barcodes into groups of similar products, defining a group as all items within a Nielsen product module measured in the same unit, and normalize prices and quantities by the size of a product.[26] Total annual grocery expenditure e_i^G for a

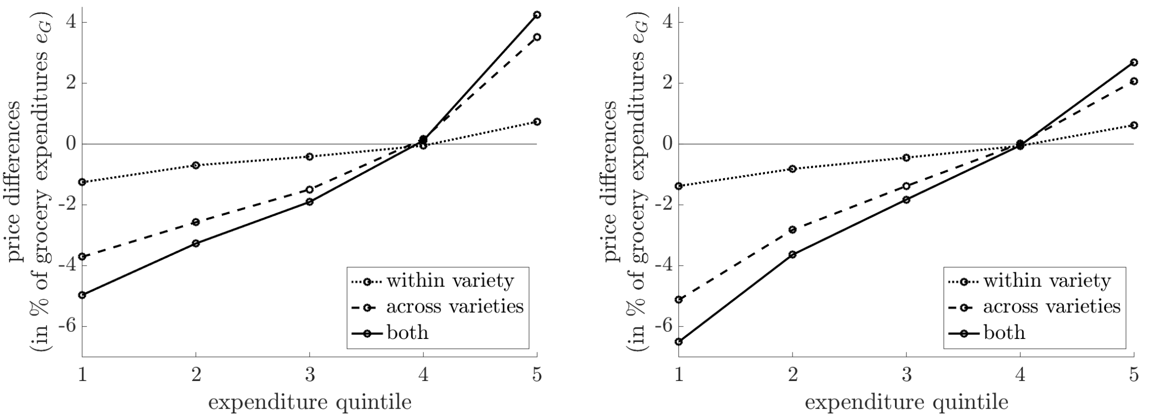
[26] E.g. I group together all barcodes in the module "fresh apples" that are measured in pieces and divide the price of each barcode by the number of individual apples included to get a price per apple. Most product modules have one dominant unit of measurement and there is no systematic difference in purchases across households in the unit dimension as Appendix A.2.4 shows.

given household i is the sum of spending e^G_{ijk} over all barcodes j in all groups k. Spending per barcode is the quantity-weighted average per-unit price paid by the household for barcode j in group k p_{ijk} times the units consumed c_{ijk}. Further, define $\hat{p}_{jk}$ as the quantity-weighted average price paid for variety j across all households and $\tilde{p}_k$ as the average of $\hat{p}_{jk}$ within group k. For example, for the module "fresh apples" measured in pieces $\tilde{p}_k$ is the average price per apple across all households and all barcodes of apples, $\hat{p}_{jk}$ the average price for one specific type of apple across households, and p_{ijk} the average price one specific household pays for one specific type of apple. In the model, I consider a single (representative) group k and all three varieties j as close substitutes of identical size within that group. Decompose e^G_i as

$$
\begin{aligned}
e^G_i &= \sum_k \sum_{j \in J_k} e^G_{ijk} = \sum_k \sum_{j \in J_k} p_{ijk} c_{ijk} \\
&= \underbrace{\sum_k \sum_{j \in J_k} (p_{ijk} - \hat{p}_{jk}) c_{ijk}}_{\substack{\text{within varieties} \\ \text{(direct effect of shopping)}}} + \underbrace{\sum_k \sum_{j \in J_k} (\hat{p}_{jk} - \tilde{p}_k) c_{ijk}}_{\text{across varieties}} + \underbrace{\sum_k \sum_{j \in J_k} \tilde{p}_k c_{ijk}}_{\text{quantities}} .
\end{aligned}
$$

The first term captures price differences within identical products between what an individual household pays relative to other households. In the model, this is driven by differences in the direct effect of shopping on reducing the average price per unit given the posted price distribution. The second term captures differences across products, due to heterogeneity in consumption baskets. In line with the interpretation of demand elasticities q_j, the literature has referred to these differences in the average price across close substitutes as quality (e.g. Argente and Lee, 2021; Bisgaard Larsen and Weissert, 2020; Jaimovich et al., 2019). The last term summarizes households' counterfactual expenditure absent any price differences within or across products, where all variation of expenditures within a group k is due to differences in the quantity consumed.

Figure 1.5a shows the results of the empirical decomposition by quintile of the expenditure distribution, expressed as a fraction of grocery expenditure. Households at the bottom of the distribution have about 5.5% lower expenditure due to deviations from the average price $\tilde{p}_k$ within module-unit bins. 1.5pp. are due to lower prices within products, i.e. driven by the direct effect of shopping effort. 4pp. are due to lower prices across the products bought. At the top of the distribution, price differences increase total spending by 4%, 3.5pp. of which due to differences across products. In between, the contribution of price differences within and across products is monotonically increasing in expenditure. The reported magnitudes are well in line with the findings of e.g. Aguiar and Hurst (2007) and Bisgaard Larsen and Weissert (2020) under alternative approaches.



(a) Data

(b) Model

Figure 1.5: Price Differences along the Expenditure Distribution

Note: Price differences by expenditure quintile in the model and data, as a share of households' grocery spending. "Within variety" refers to differences between the price paid by a given household and the average price for a given product (direct effect of shopping), while "across varieties" is the difference in average prices across different products. Data from the Nielsen Consumer Panel 2007-2019. In the data, the price within products is computed by barcode and the price across products is computed across barcodes within a product module (by unit of measurement).

We can interpret these findings in terms of their contribution to expenditure inequality. Inequality is often measured as a ratio between the top and the bottom of the distribution (e.g. Aguiar and Bils, 2015). Define $\tilde{e}_i^G = \sum_k \sum_{j \in J_k} \tilde{p}_k c_{ijk}$ as a household's counterfactual grocery spending absent price differences within and across products. The results of Figure 1.5a imply that price differences within and across varieties can account for 10% of expenditure inequality between the top and bottom quintile. I.e. expenditure inequality measured as the ratio of counterfactual expenditure $\tilde{e}^G$ between the top to bottom quintile of the expenditure distribution is about 10% lower compared to true expenditure e^G.[27]

Figure 1.5b shows that the model is able to reproduce the empirical patterns. It is important to note that only the differences within and across varieties between the lowest and highest quintile are included in the set of targeted moments. The

[27]Figure A.5 in Appendix A.2.5 shows that this number increases to 12% when considering deciles.

model does well at reproducing the levels and slope of both margins along the entire expenditure distribution.

Robustness. Table A.4 in Appendix A.2.5 provides estimates from a regression of the contribution of price differences within and across barcodes on households' expenditure and controls to show that their relationship is robust to other household characteristics.[28]

1.5.2 Equilibrium Effects of Shopping on Inequality

While the previous section provides descriptive evidence on how prices contribute to expenditure inequality, interpreting their contribution to real inequality requires one further step. Price differences across households can be either due to differences in marginal costs for the products bought or due to differences in the margins paid. It is important to distinguish between the two. While a higher marginal cost cannot be avoided to purchase a preferred good, a higher margin implies that it would be feasible to achieve the same consumption allocation with lower expenditures. Higher margins reflect the cost imperfect competition in the product market imposes on households. While I cannot distinguish between marginal cost and margins in the data, the model allows me to disentangle them.

Direct and Equilibrium Effect of Shopping. In the previous decomposition, differences in prices within varieties are entirely due to margin differences. They capture the direct effect of differences in households' shopping effort, paying less for an identical product given the posted price distribution. However, differences in average (posted) prices across varieties can be due to either marginal costs or average margins posted. As Proposition 1 has shown, differences in average posted margins across varieties can be attributed to the equilibrium effect of search frictions and shopping behavior on posted prices. This effect operates through retailers' response to demand-weighted shopping effort $\bar{s}_j$, which changes across varieties due to the differences in demand composition induced by non-homotheticities in households' preferences. I therefore attribute

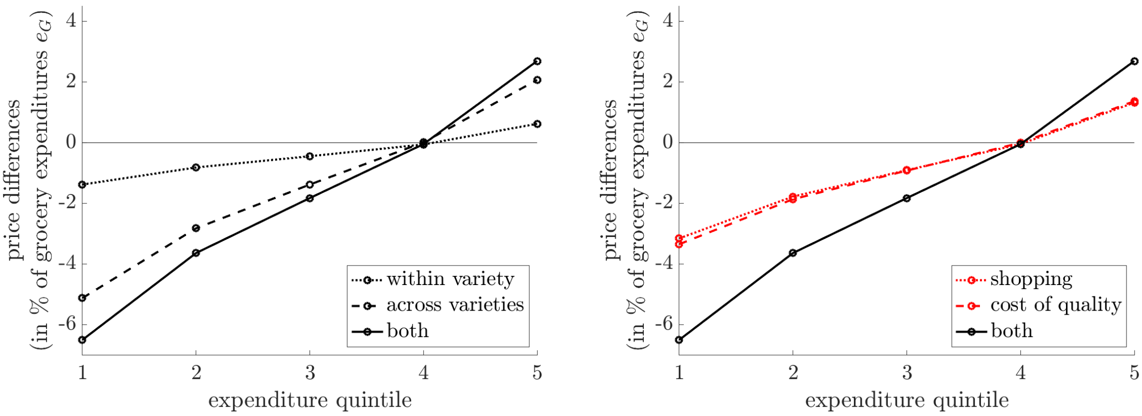
[28] I include income, employment status, age, household size, household and year-state fixed effects. To relate price differences within products to households' shopping effort, I report again the evidence on trips-per-purchase from Section 1.3.1. Running a regression of the price differences within and across varieties on log-expenditure in the model slightly underpredicts the estimates from the data (0.68 vs. 0.95 for differences within and 2.5 vs. 3.4 across) but is consistent with the relative magnitudes.

differences in average margins across varieties ($\hat{p}_{jk} - \kappa_{jk}$) to the equilibrium effect of shopping. This leaves differences in marginal cost of variety j relative to the average marginal cost in group k ($\kappa_{jk} - \tilde{\kappa}_k$) to explain the remaining price gap across products. In line with the interpretation of higher q_j as quality, I will refer to this term as *cost of quality*. I adjust the decomposition accordingly.

$$
\begin{aligned}
e_i^G &= \underbrace{\sum_k \sum_{j \in J_k} (p_{ijk} - \hat{p}_{jk}) c_{ijk}}_{\substack{\text{within varieties} \\ \text{(direct effect of shopping)}}} + \underbrace{\sum_k \sum_{j \in J_k} (\hat{p}_{jk} - \tilde{p}_k) c_{ijk}}_{\text{across varieties}} + \sum_k \sum_{j \in J_k} \tilde{p}_k c_{ijk} \\
&= \underbrace{\sum_k \sum_{j \in J_k} (p_{ijk} - \hat{p}_{jk}) c_{ijk} + ((\hat{p}_{jk} - \kappa_{jk}) - (\tilde{p}_k - \tilde{\kappa}_k)) c_{ijk}}_{\text{shopping (direct+equilibrium)}} + \underbrace{\sum_k \sum_{j \in J_k} (\kappa_{jk} - \tilde{\kappa}_k) c_{ijk}}_{\text{cost of quality}} + \sum_k \sum_{j \in J_k} \tilde{p}_k c_{ijk}
\end{aligned}
$$

Figure 1.6 presents the results for the adjusted decomposition alongside the original results from Figure 1.5b. It shows that accounting for the equilibrium effect of search frictions and differences in $\bar{s}_j$ on margins across varieties more than doubles the contribution of shopping to expenditure inequality, relative to the direct effect on price differences within products. This makes the overall effect of shopping equally important as differences in marginal costs across varieties. According to this decomposition, shopping alone through its direct and equilibrium effects can account for 5% of inequality in household expenditures between the bottom and top quintile of expenditures. This contribution is entirely due to the consequences of imperfect competition and should therefore be seen as a reduction in real consumption inequality relative to expenditure inequality.

Consumption Inequality. The effect of heterogeneous shopping behavior on the relationship between expenditure and consumption inequality can be best understood by the following thought experiment: Fix the distribution of expenditures and allow households to optimally choose their consumption bundles assuming (i) all households pay the average price $\hat{p}_{jk}$ for each variety or (ii) all households pay the marginal cost plus the average margin $\kappa_j + (\tilde{p}_k - \tilde{\kappa}_j)$ for each variety. The former eliminates the direct effect of shopping on inequality. The latter additionally shuts down its equilibrium effect and is the sum of the quantity term and the cost of quality above.



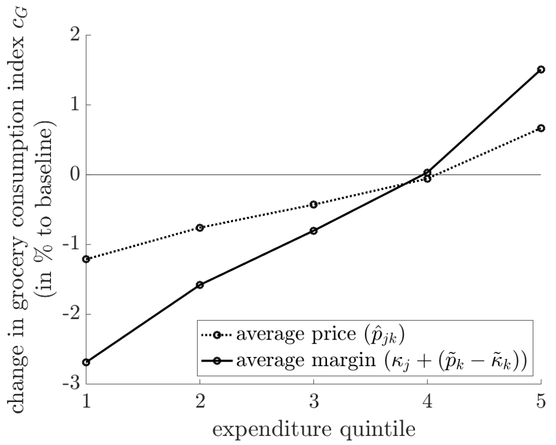
(a) Model (Original)

(b) Model (Adjusted)

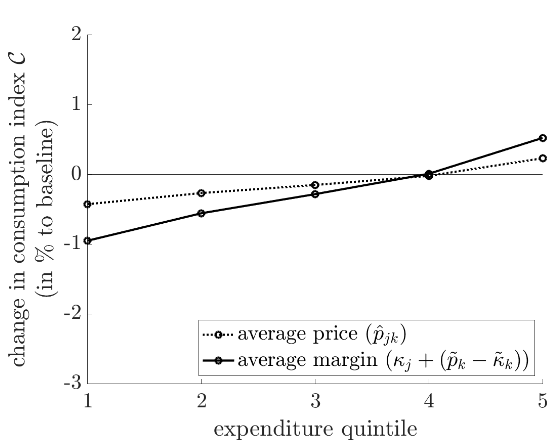
Figure 1.6: Price Differences along the Expenditure Distribution - The Effect of Shopping
Note: Price differences by expenditure quintile in the model, as a share of households' grocery spending. In the *original* decomposition, "within variety" refers to differences between the price paid by a given household and the average price for a given product (direct effect of shopping), while "across varieties" is the difference in average prices across different products. In the *adjusted* decomposition, "shopping" refers to price differences within and margin differences across products (direct and equilibrium effect of shopping) and "cost of quality" to differences in marginal costs across products.

Figure 1.7a shows by how much grocery consumption would change on average within each expenditure quintile under the alternative prices. Without the direct and indirect effect of shopping, grocery consumption c_G for the bottom quintile would be approximately 3% lower at identical expenditures and 2% higher for the top quintile. About half of the difference can be attributed to the direct and equilibrium effect respectively. We can again measure inequality as the ratio of top to bottom quintiles' consumption. The findings confirm that with identical inequality in grocery expenditures, inequality in consumption would be approximately 5% higher without the direct and equilibrium effect of shopping. The effect is muted when focusing on total consumption $\mathcal{C}$ in Figure 1.7b, as the outside good is not subject to any search frictions. The result on $\mathcal{C}$ sets the lower bound for shopping's overall effect on consumption inequality to 1.5%. With similar search frictions for all consumption, the overall effect would be close to the results reported for c_G.

Welfare. The findings on consumption inequality do not take into consideration the full consequence for households' welfare, as they abstract from the disutility of shopping effort. To interpret the equilibrium effect of shopping in terms of welfare, I again fix the expenditure distribution and compute the one-period



(a) Grocery Consumption



(b) Total Consumption

Figure 1.7: Shopping Effort and Consumption Inequality

Note: Change in households' grocery (c_G) and total consumption ($\mathcal{C}$) absent shopping effort. Counterfactuals fix the expenditure distribution and let households choose a basket assuming they have to pay (i) the average price within each grocery variety (shutting down the direct effect of shopping) or (ii) marginal cost plus the average margin across varieties (shutting down the direct and equilibrium effect).

utility for each household under the assumption that no household exerts any search effort and all households (i) pay the average prices $\hat{p}_{jk}$ or (ii) pay the marginal cost plus average profit margins $\kappa_{jk} + (\tilde{p}_k - \tilde{\kappa}_k)$ for each variety. I then ask by how much consumption $\mathcal{C}$ in the alternative economy would need to change to make a household with expenditure e indifferent between living there for one period or living one period in the steady state economy.[29] Formally, the necessary change in consumption is defined as

$$\Delta\mathcal{C}^{CF}(e) = \left(\frac{U(e) + \frac{1}{1-\phi}}{u\left(\mathcal{C}^{CF}(e)\right) + \frac{1}{1-\phi}} \right)^{\frac{1}{1-\phi}} - 1,$$

where $U(e)$ is the steady state one-period utility of spending e, $u(\cdot)$ the CRRA utility function, and $\mathcal{C}^{CF}(e)$ the consumption level associated with spending e in one of the counterfactual economies.

Figure 1.8 reports the results, averaging the consumption change within each expenditure quintile. It shows that the bottom quintile of expenditures is almost equally well off when paying average prices but not exerting shopping effort (shutting off the direct effect of shopping). The cost of higher prices is offset by reducing the disutility of effort to zero. The top quintile would forego 0.85%

[29]Results for infinite horizon welfare measures are reported in Appendix A.3.3.

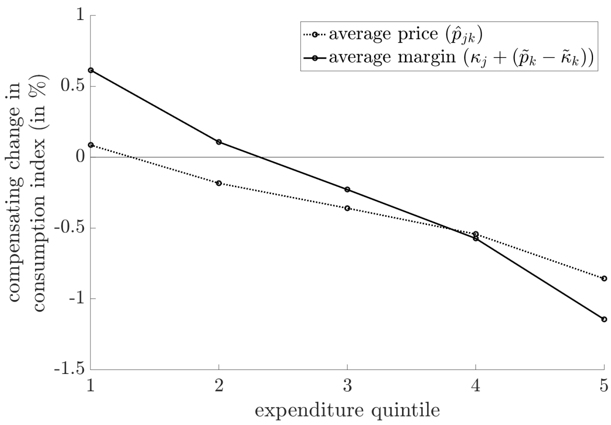


Figure 1.8: Welfare Effects of Shopping Effort

Note: Change in total consumption index ($\mathcal{C}$) under alternative prices and zero shopping effort to make a household indifferent between living one period in an alternative economy and one period in the steady state economy. Counterfactuals fix the expenditure distribution and let households choose a basket assuming they have to pay (i) the average price within each grocery variety (shutting down the direct effect of shopping) or (ii) marginal cost plus the average margin across varieties (shutting down the direct and equilibrium effect).

of consumption to face the average price paid for a variety without exerting shopping effort. If households instead were to pay the average margin (shutting off direct and equilibrium effects), the bottom quintile would need to receive an increase of 0.6% in their consumption to be indifferent to the steady state, while the top quintile would forego up to 1.15% of consumption. These findings imply that the direct and equilibrium effect of shopping jointly reduce inequality in (one-period) welfare by 1.75%. 0.9pp. of the effect are attributable to the equilibrium response of posted prices. Again, this is a lower bound as I restrict search frictions to groceries.

Externalities. As retailers' price posting targets the average buyer in the market, each agents' search effort imposes an externality on the prices faced by all other households. Non-homotheticities and the partial separation of demand into different varieties reduce this externality. The model economy allows for a quantification of this reduction in externalities and of the remaining externality due to the non-zero overlap in households' consumption baskets. To do so, I construct counterfactual price distributions based on alternative values of demand-weighted shopping effort $\bar{s}_j$ across varieties.[30]

[30] Sangani (2022) conducts a similar exercise in a model economy with a single good and, despite a different calibration approach, finds similar magnitudes for the difference between the pooling and full separation scenario. In addition, I provide a quantification of how much the externality is reduced due to partial separation in the goods market, which is not possible in a single-good economy.

For a *full pooling* counterfactual, I assume demand-weighted effort $\bar{s}_j$ is the same across all varieties and let retailers post prices based on the average composition of demand. To equalize $\bar{s}_j$ across varieties, I compute the average shopping effort weighting individuals' shopping policies with their share in total demand across all varieties such that $\bar{s}_j = \int\int \frac{\lambda(z,a)\sum_{j=1}^{J} c_j(z,a)}{\sum_{j=1}^{J} C_j} s(z,a)\, dz\, da$. I fix households' consumption baskets and shopping policies at their equilibrium values and compute the percentage change in their cost of consumption if they were drawing from the distributions retailers post when facing the counterfactual $\bar{s}_j$ for all varieties.[31] The *full pooling* line in Figure 1.9 plots the results over the range of the model implied expenditure distribution. If $\bar{s}_j$ would be identical across all varieties, households at the bottom of the expenditure distribution would spend around 2% more on their consumption bundle while households at the top could save around 1% while buying the same basket, in line with the equilibrium effect of shopping.

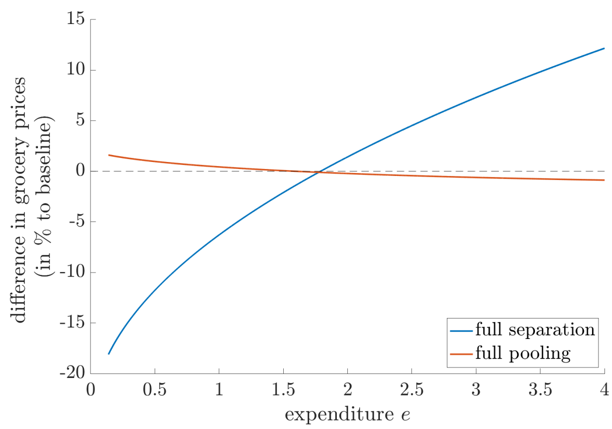


Figure 1.9: Shopping Externality

Note: Change in households' prices paid under alternative posted price distributions. Baseline is the calibrated steady-state. Counterfactuals fix the expenditure distribution, consumption baskets and shopping effort. *Full pooling* assumes price distributions for each variety determined as the best response to average search effort $\bar{s}_j = \int\int \frac{\lambda(z,a)\sum_{j=1}^{J} c_j(z,a)}{\sum_{j=1}^{J} C_j} s(z,a) dz da$. *Full separation* assumes individually targeted price distributions for each variety determined as the best response to individual search effort $\bar{s}_j = s(z,a)$.

Next, I allow households to draw from price distributions targeted to their individual shopping effort. In this counterfactual a household in state (z,a) draws for each variety from a price distribution that retailers' would post if all other households were exerting the same shopping effort, i.e. $\bar{s}_j = s(z,a)$. In this

[31]The exercise can be interpreted as the change in a Laspeyres price index.

counterfactual, I obtain *full separation*. Figure 1.9 shows that the remaining externality is sizeable. Fixing shopping effort and consumption baskets, households at the bottom of the expenditure distribution could save an additional 18% if they were to face targeted price distributions, i.e. if retailers could perfectly discriminate between household types. The high-spending households on the other hand would pay up to 12% more in a world with perfect discrimination.[32]

The large size of the remaining externality can be accounted for by the generally higher spending levels at the top of the expenditure distribution. While high-spending households consume relatively less of the goods that are important in low-spending households' baskets, they still account for a sizeable share of expenditures across all varieties due to their higher level of expenditures.[33]

Overall, the findings show that differences in demand composition across goods and the ensuing equilibrium effects of heterogeneity in shopping effort on posted prices have substantial implications for how we should interpret inequality in expenditures in terms of consumption and welfare. The previous literature, such as Aguiar and Hurst (2007), Arslan et al. (2021), and Pytka (2022), has focussed on how households can reduce the price they pay for a given variety. The findings outlined above suggest that shopping provides additional insurance of similar magnitued through the effect of low-income households' collective search effort on the prices posted for the products they purchase.

1.6 Implications for Average Prices and Markups

The relationship between demand-weighted shopping effort and posted prices makes the average price and markup in the economy a function of the distribution of expenditure across households. The distribution of households can change, e.g. when the economy is hit by aggregate shocks or when policies change the economic environment. In this section, I consider first how the response of posted prices to shopping effort has contributed to price dynamics during the Great Recession and how shifts in demand composition can affect the cyclicality

[32]Retailers do not leave money on the table due to a lack of price discrimination. Holding households' policy functions constant, the difference in total sales is less than 0.01% across scenarios. Lower profits on low-search households are almost perfectly offset by higher profits on high-search households.

[33]Appendix A.2.4 shows that this is in line with the data.

of average retail prices and markups in general. In addition, I provide evidence on how the response of posted prices to demand composition can affect the cost of redistributive taxation for high-earning households.

1.6.1 Shopping and Prices over the Business Cycle

A growing empirical literature studies the cyclical properties of retail prices and markups in response to aggregate demand shocks, but so far remains inconclusive. E.g. Anderson et al. (2020) find acyclical prices and markups in response to local unemployment shocks while Stroebel and Vavra (2019) find strongly procyclical responses to changes in local house prices. I revisit these findings in the model economy by focussing on the Great Recession period around 2008.

The Great Recession saw both substantial earnings losses due to an increase in unemployment and losses in wealth in response to the decline in house prices. I construct a similar shock and hit the model economy with an unexpected one-time loss in households' net worth and persistent earnings losses differentiated by households' labor productivity state z. I choose an equal loss in wealth of 15% for all households, to match the decline in households' net worth between the last quarter of 2007 and first quarter of 2009 as reported in the US Financial Accounts (Table Z.1).[34] For losses in labor earnings along the income distribution, I build on the findings of Heathcote et al. (2020a). I take their estimates for earnings changes at different points of the income distribution in 2008-2010 for the first three periods after the shock hits and let earnings return to their steady state level by $t = 6$. A mapping of the findings in Heathcote et al. (2020a) into the labor productivity states of the model is provided in Table A.6 in Appendix A.3.5. It shows that losses are heavily concentrated at the bottom of the earnings distribution. Overall, the two components of the shock are comparable in magnitude. The decline in wealth amounts to roughly 32% of aggregate annual income in the model economy and the cumulated earnings loss to 24%.

[34]This choice is in line with the decline in wealth growth for the top two quintiles of the wealth distribution (those households holding significant wealth) reported in Krueger et al. (2016).

To measure the cyclical properties of retail prices, I focus on a Laspeyres index of average prices across grocery varieties, given by

$$P_t^l = \frac{\sum_{j=1}^J C_j^{SS} \mu_{jt}^l}{\sum_{j=1}^J C_j^{SS}}.$$

The Laspeyres index abstracts from changes in households' baskets when aggregating prices and is therefore ideally suited to isolate prices changes. I will consider separately changes in the Laspeyres index for two definitions of average prices per variety: average posted prices P_t^F and average prices paid P_t^G. Throughout the exercises conducted in this section I keep all parameters at their steady-state values. Any response of posted and paid prices is therefore by construction driven by changes in households' (demand-weighted) shopping effort and its effect on posted and paid prices.

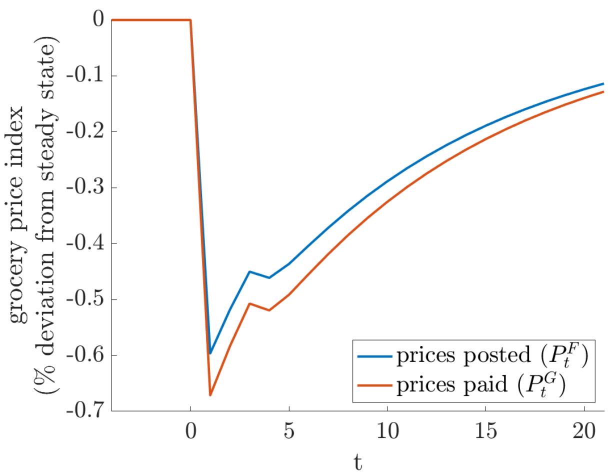


Figure 1.10: Prices Posted and Prices Paid during the Great Recession

Note: Model implied response of an aggregate Laspeyres index $P_t^l = \frac{\sum_{j=1}^J C_j^{SS} \mu_{jt}^l}{\sum_{j=1}^J C_j^{SS}}$ of prices posted (P_t^F) and prices paid (P_t^G) to the Great Recession shock (15% loss in wealth and earnings losses from Heathcote et al. (2020a)).

Figure 1.10 plots the response of the aggregate Laspeyres index of grocery prices, separately for prices posted P_t^F and prices paid P_t^G. It shows that changes in households' shopping behavior reduced prices paid by about 0.7% during the Great Recession. This effect is predominantly driven by a reduction in posted prices as retailers' respond to households' choices in equilibrium. Posted prices declined by 0.6 percentage points while changes in paid prices relative to posted prices account for only 0.1 percentage points of the overall decline in prices paid.

Hence, taking into account the equilibrium effect on posted prices in response to changes in demand-weighted shopping effort is quantitatively important for fluctuations in retail prices during the Great Recession.[35]

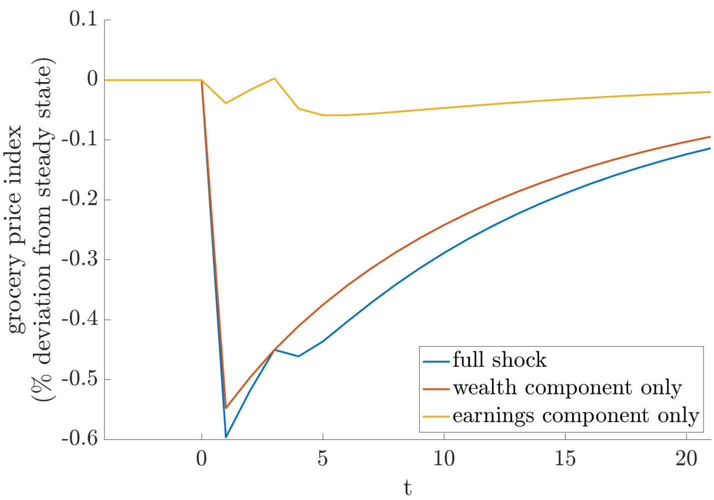


Figure 1.11: Response of Posted Prices to Losses in Earnings and Wealth

Note: Model implied response of an aggregate Laspeyres index $P_t^F = \frac{\sum_{j=1}^{J} C_j^{SS} \mu_{jt}^F}{\sum_{j=1}^{J} C_j^{SS}}$ of prices posted to the Great Recession shock (15% loss in wealth and earnings losses from Heathcote et al. (2020a)). Full shock decomposed into response to loss in earnings and loss in wealth.

Next, I study whether the change in the index of posted prices is driven by the decline in wealth or losses in labor earnings. To isolate their respective effects on posted prices, I hit the economy only with a loss in wealth or only with a loss in labor earnings. Figure 1.11 plots the response of the posted price index P_t^F for each of the two components separately and for the combined response. It shows that the response of posted prices is almost entirely driven by the reduction in wealth, while retailers barely react to the decline in earnings. With all parameters, including marginal cost κ_j, fixed at steady-state levels, these price responses are driven entirely by changes in posted markups. Hence, the model yields procyclical responses of retail prices and markups to the decline in wealth, but acyclical responses to the change in labor earnings during the Great Recession. This finding shows that the model can reconcile the conflicting

[35]Despite the absence of any adjustment cost, the dynamics of the model are not inconsistent with significant price stickiness. The price distribution for each variety on impact of the shock overlaps to about 96% with its steady-state counterpart. Therefore, the model can be consistent with up to 96% of retailers not adjusting their prices, as they are indifferent between any price on the support of the posted distribution. This is in line with the findings of Burdett and Menzio (2018).

empirical evidence presented in Stroebel and Vavra (2019) and Anderson et al. (2020).[36]

To understand the mechanism behind this result, I decompose the price response into the forces shaping it in equilibrium. In the model, posted prices (markups) respond only to changes in demand-weighted shopping effort

$$\bar{s}_{jt} = \int\int \frac{\lambda_t(z,a)c_{jt}(z,a)}{C_{jt}}\ s_t(z,a)\ dz\ da.$$

Demand weighted effort either changes because of adjustments in individual search effort, or because changes in demand composition alter how retailers take the effort of different households into account. The effect of individual search behavior is captured by the response of $\bar{s}_{jt}$ to changes in shopping policies $s_t(z,a)$. Shifts in demand composition arise due to changes in households' consumption policies $c_{jt}(z,a)$ or the distribution of households across the state space $\lambda_t(z,a)$. I consider the contribution of each of the three separately, fixing the others at steady state level.

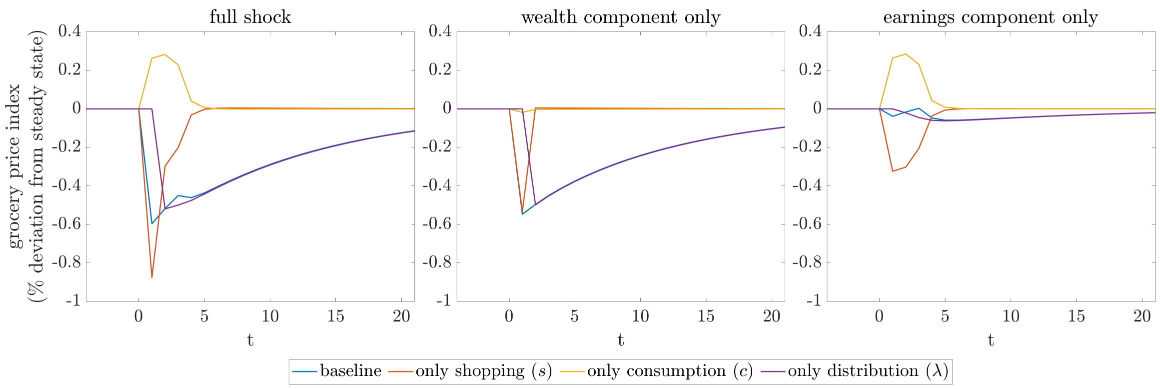


Figure 1.12: Decomposition of Posted Price Responses

Note: Response of an aggregate Laspeyres index of posted prices $P_t^F = \frac{\sum_{j=1}^{J} C_j^{SS}\mu_{jt}^F}{\sum_{j=1}^{J} C_j^{SS}}$ to the Great Recession shock (15% loss in wealth and earnings losses from Heathcote et al. (2020a)). Panels report responses to the full shock, only the wealth, and only the earnings component. Complete response as a baseline, for each panel decomposed into the response to changes in consumption policies, shopping policies, and the distribution of households, holding the respective others constant at steady state levels.

[36]The model also resonates with the literature quantitatively. Figure A.11 in Appendix A.3.5 shows that the decline in aggregate grocery prices (markups) paid in response to a 1% decline in households' wealth is about 0.04%. This elasticity is at the lower end but of similar magnitude as the range of estimates for the elasticity of retail prices to house prices reported in Stroebel and Vavra (2019), who find values of 0.02-0.2. Assuming that most household wealth is held in real estate and that households' have a levered position in housing due to mortgages implies that a 1% wealth shock is a conservative choice to capture the consequences of a 1% decline in house values.

Figure 1.12 plots the decomposition of price responses separately for the full shock and only the wealth and earnings component respectively. Changes in households' shopping policies alone reduce prices in response to both the wealth and earnings component, as each affected household increases her search effort to insure against an income loss. What accounts for the differences in cyclicality are differential responses of demand composition, driven by changes in households' consumption policies and the distribution of agents across the state space. This is due to the incidence of the shocks, earnings losses being concentrated among low income households and high-income households facing a larger absolute decline in wealth.

The stronger low-income households are affected the more they have to reduce consumption and the lower becomes their share in overall demand. Retailers now face relatively more high-income buyers and respond to this shift in demand composition by attaching more weight to their (lower) shopping effort. In response, they increase prices. For the earnings component, this demand composition effect is strong enough to offset the direct increase in shopping effort. As high-income households are disproportionately affected by the wealth component, the effect of changes in households' consumption policies goes in the opposite direction and reallocates relative demand towards low-income (high-effort) households. In addition, high-income households significantly reduce their savings which increases their future shopping effort as they have become relatively poorer. This effect is captured by changes in the distribution of agents across the state space and drives the response of prices to the wealth component from the second period onwards. In combination of both components the change in households' consumption policies dampens the response of retail prices to the increase in individual shopping effort by about one third, emphasizing the quantitative importance of shifts in demand composition during the Great Recession.[37]

The implications for inequality are in sharp contrast to the results of the previous section. While heterogeneity in demand composition across varieties

[37] Appendix A.3.4 provides additional evidence on the effect of demand composition on the cyclicality of retail prices by simulating the same decline in aggregate income but distributing it differentially along the earnings distribution. It shows that if losses are sufficiently concentrated among the bottom of the distribution, retail prices can increase in response to a decline in aggregate earnings.

provides additional insurance by reducing the prices charged on goods bought by low-income households, shifts in demand composition over the business cycle amplify inequalities. High-earning households are partially compensated through a decline in posted prices when they are hit by an aggregate income loss. Low-earning households might see prices rise if they lose income as retailers adjust posted prices to their declining share in aggregate demand.

The findings presented in this section extend the work of Kaplan and Menzio (2016) who introduce demand composition effects in a model with two types of agents: Employed and unemployed. In order to sustain multiple equilibria and self fulfilling unemployment fluctuations, their framework requires a decline in aggregate income to be associated with an increasing role for (high-search-effort) unemployed households and a resulting decline in prices and profit margins for firms. The results derived in this paper from a framework with rich household heterogeneity suggest that shifts in demand composition can dampen or amplify the responses of retail prices to households' shopping effort, depending on the incidence of aggregate income shocks.

1.6.2 Demand Composition Across Policy Regimes

When posted prices become a function of the distribution of income and wealth in the economy, policies (re-)shaping these distributions affect households' cost of consumption. This section shows how posted prices respond to redistributive taxation. For that purpose, I introduce a flat tax on earnings and redistribute the proceeds equally as a lump sum payment to all households. Introducing the policy alters the budget constraint to

$$e + a' \leq (1+r)a + (1-\tau)zw + T,$$

where the government-budget-clearing transfer satisfies $T = \int\int \lambda(z,a)zw\, dzda$.[38]

[38] As the process for z is calibrated to households' earnings post taxes and transfers, the introduction of τ should be interpreted as additional redistribution relative to the current US system.

I solve for the steady state of the model for given τ and compute the ensuing changes in households' earnings post taxes and transfers

$$\Delta\text{earn(z,a)} = \frac{(1-\tau)zw + T - zw}{zw} = \frac{T - \tau zw}{zw}$$

as well as changes in an individual Laspeyres index for total consumption $P^{lasp}(z,a)$

$$\Delta P^{lasp}(z,a) = \frac{\tilde{e}^{\tau}(z,a)}{e^{0}(z,a)} - 1 = \frac{c_O^0 + \sum_{j=1}^{J} p_j^{\tau}(s^0(z,a))c_j^0(z,a)}{c_O^0 + \sum_{j=1}^{J} p_j^{0}(s^0(z,a))c_j^0(z,a)} - 1$$

and grocery consumption $p_G^{lasp}(z,a)$

$$\Delta p_G^{lasp}(z,a) = \frac{\tilde{e}_G^{\tau}(z,a)}{e_G^{0}(z,a)} - 1 = \frac{\sum_{j=1}^{J} p_j^{\tau}(s^0(z,a))c_j^0(z,a)}{\sum_{j=1}^{J} p_j^{0}(s^0(z,a))c_j^0(z,a)} - 1,$$

where $p_j^{\tau}(s)$ is the price paid for grocery variety j by a household exerting effort s in an economy with redistributive tax τ. The index should be interpreted as the counterfactual expenditure level $\tilde{e}^{\tau}(z,a)$ a household in state (z,a) needs to buy the same basket as in the original steady state with the same shopping effort. Again, I focus on changes in a Laspeyres index and keep all policy functions at the original steady state with $\tau = 0$ to isolate changes in posted prices. For a fall in prices, the Laspeyres index provides a lower bound on the welfare impact as households can gain further by adjusting their choices.

The change in households' real income without reoptimizing policy functions is approximated by Δearn(z,a)$-\Delta P^{lasp}(z,a)$. I aggregate changes in earnings and prices by expenditure quintile using the distribution of households in the original steady state. Table 1.8 presents results for a 5% earnings tax ($\tau = 0.05$). Overall, prices decline in response to the policy change as with more redistribution a larger share of demand is accounted for by relatively low-income (high-shopping-effort) households. This drives up demand-weighted search effort and hence drives down the prices of grocery goods. The effect is even stronger for varieties with higher quality q_j, yielding larger declines in the price index of high-spending households.

While the transfer dominates changes in real income at the bottom of the expenditure distribution, price changes are relatively more important at the top of the distribution and can compensate net contributors for a significant share of their earnings loss. Due to the assumption of a perfectly competitive outside good market, the reported changes in the aggregate price index P provides a lower bound while the results for p_G provide an upper bound if all consumption was subject to the same frictions. Table 1.8 shows that households at the top of the expenditure distribution are compensated for 5-14% of the loss in their post-tax earnings due to the response of posted price distributions.

Table 1.8: Earnings and Price Changes under Redistributive Policies ($\tau = 0.05$)

		quintile of expenditures				
		Q1	Q2	Q3	Q4	Q5
income	Δearn	16.66%	5.21%	1.91%	-0.15%	-2.06%
prices	ΔP	-0.08%	-0.09%	-0.09%	-0.09%	-0.10%
	Δp_G	-0.23%	-0.25%	-0.26%	-0.27%	-0.28%
share ΔP	$\frac{-\Delta P}{\lvert\Delta \text{earn}\rvert}$	0.5%	1.6%	4.7%	62.3%	4.8%
share Δp_G	$\frac{-\Delta p_G}{\lvert\Delta \text{earn}\rvert}$	1.4%	4.7%	13.4%	178%	13.7%

Note: Average change in post-tax earnings (Δearn), grocery (Δp_G), and aggregate Laspeyres price index (ΔP) within each expenditure quintile in response to a 5% earnings tax and budget neutral transfer.

Table A.8 in Appendix A.3.5 reports price and income changes for alternative values of τ. While both the income and price effects increase in the degree of redistribution (with higher values of τ), their relative contributions to the overall change in real income remains similar across all policy regimes considered.

The findings show that the effect of redistribution on demand composition reduces estimates of the loss from taxation at the top. Net contributors to redistribution schemes can benefit from lower price levels as retailers' place more weight on the purchase behavior of low-income households, reducing the real burden of redistributive policies.

1.7 Conclusion

This paper develops an equilibrium theory of expenditure inequality and price dispersion, featuring search for prices, heterogeneous households' with non-homothetic preferences, and endogenous price distributions for multiple varieties. I provide analytical results on retailers' best response to households' shopping effort and show that average posted prices decline in the share of demand stemming from high-search-effort households. Theoretical predictions on the skewness of posted price distributions are in line with empirical evidence from the Nielsen Consumer Panel. The calibrated model replicates salient features of expenditure inequality and price dispersion. It shows that the response of posted prices across varieties doubles the contribution of shopping effort to the difference between inequality in expenditure and consumption. After a shock similar to the Great Recession, posted prices respond to losses in wealth but not to losses in earnings. By showing the importance of accounting for the incidence of aggregate shocks, the model reconciles conflicting evidence on the cyclicality of retail markups. Finally, I show that endogenous price changes in response to redistributive policies reduce the loss from redistribution at the top by up to 14%. All of these results highlight the importance of accounting for equilibrium effects of heterogeneity in households' shopping effort and demand composition when thinking about retail prices.

The focus of this paper is on households' shopping effort and heterogeneity in price elasticities across retailers for a given product. Recent evidence by Auer et al. (2022) suggests additional heterogeneity in price elasticities across products, generating an additional role of demand composition for posted prices. In the interest of tractability and introducing rich household heterogeneity, retailers' price posting problem has been deliberately kept simple. An alternative approach would be to take the results of this paper as motivation and introduce stylized heterogeneity in price elasticities and demand composition into a more evolved price setting problem. Possible extensions include e.g. multiproduct price posting as in Kaplan et al. (2019) or state dependent price adjustments as in Golosov and Lucas (2007), Midrigan (2011), and Burdett and Menzio (2018). I leave all these extensions for future work.

Chapter 2

Joint Search over the Life Cycle

Joint with Annika Bacher and Philipp Grübener

Abstract This paper studies how the added worker effect - intra-household insurance through increased spousal labor market participation - varies over the life cycle. We show in U.S. data that the added worker effect is much stronger for young than for old households. A stochastic life cycle model of two-member households with job search in a frictional labor market is capable of replicating this finding. The model suggests that a lower added worker effect for the old is driven primarily by better insurance through asset holdings. Human capital differences between employed young and old contribute to the difference but are quantitatively less important, while differences in job arrival rates play a limited role.[1]

[1]This paper uses data from the Survey of Income and Program Participation sponsored by the US Census Bureau, IPUMS CPS (Flood et al., 2020) and the Panel Study of Income Dynamics (Panel Study of Income Dynamics, 2021), the collection of which was partly supported by the National Institutes of Health under grant number R01 HD069609 and R01 AG040213, and the National Science Foundation under award numbers SES 1157698 and 1623684.

2.1 Introduction

Household earnings dynamics vary strongly over the life cycle. Recent literature documents that key moments of the earnings growth distribution exhibit significant age-dependency (De Nardi et al., 2020; Guvenen et al., 2021b). Earnings variability is highest for young individuals as they change jobs frequently before settling into a stable job. However, the earnings growth distribution is more left-skewed for older individuals: Most of the time older individuals are employed in stable employment relationships at relatively high wages. If they lose this job, however, this fall off the job ladder implies very large earnings losses. In this paper we take a complementary perspective: Instead of investigating how risks change over the life cycle, we study how insurance against individual earnings risk varies over the life cycle. Specifically, we focus on an insurance margin against individual earnings and unemployment risk available to couples, the added worker effect (AWE), where a previously non-participating spouse enters the labor force upon job loss of the primary earner to stabilize joint earnings.

While the added worker effect has in general been widely documented, our focus on how it varies over the life cycle is novel to the literature.[2] Age differentials in the AWE are important for a variety of reasons: Observed heterogeneity along this margin improves our understanding of how well households at different ages are insured against income losses. Therefore, disparities in the availability of this self-insurance margin can alter the optimal provision of public insurance over the life cycle. Moreover, in light of demographic change any difference in the labor market behavior of old versus young households can alter aggregate labor market dynamics in the future.

We begin by providing empirical evidence on the added worker effect over the life cycle: Using data for the United States from the Current Population Survey (CPS), we show that the likelihood of a non participating spouse entering the labor force increases substantially when the primary earner loses her job compared to when she remains employed. We find, however, a strong age-dependency in this effect. In particular, the added worker effect is largest for young households and continuously declines over the life cycle. For the age

[2] See the related literature below for a detailed discussion.

group just before retirement, the added worker effect is almost non-existent. For young households, job loss of the primary earner is associated with a significant increase in the likelihood of an out of the labor force spouse entering the labor force both directly to employment and to unemployment. This finding is robust across education levels, the presence of children in the household, different reasons for being out of the labor force, different reasons for an employment to unemployment transition of the primary earner, and holds also when looking at only one cohort.

Still, there remain several candidate explanations for the observed change in the AWE over the life cycle. It might be that older households have accumulated sufficient asset holdings that allow them to smooth consumption during a potentially temporary job loss of the primary earner. In this case, older households do *not need* the added worker effect as an (additional) insurance margin. An out of the labor force spouse could in principle join the labor force, find employment, and stabilize joint earnings, but chooses not to do it. Alternatively, it could be that older spouses have been out of the labor force for a long time such that their labor market qualifications have become less valuable than those of younger individuals. In this case, spousal labor supply is *unavailable* as an insurance margin if the spouse can provide little marketable skills. In order to distinguish between the *need for* and the *availability of* the spousal insurance margin, we build a quantitative model of joint labor supply over the life cycle in a frictional labor market.

In the model, a household consists of two members, each of whom can be either employed, unemployed (and actively searching for a job), or out of the labor force. The labor market is frictional, an individual can only take up employment if she has a job offer. While both out of the labor force and unemployed individuals can receive job offers, unemployed members increase the chance of finding a job through costly search. Employed individuals face the risk of (exogenous) separation and wage changes due to match quality shocks. Human capital is accumulated while employed but depreciates during non-employment. A couple can jointly save in a risk-free bond. Job arrival rates are endogenous and determined by the solution to the vacancy posting problem of single-worker firms.

These model ingredients allow us to differentiate between the different candidate explanations for the age dependency in the added worker effect. Household savings are a key alternative insurance mechanism against individual unemployment risk. With a realistic life cycle savings profile the model can speak to whether differences in asset holdings between young and old are sufficient to explain the difference in the observed AWE. On the other hand, human capital accumulation and endogenous arrival rates allow for the possibility that older households might have fewer opportunities to provide insurance against individual risk, as human capital depreciates over long spells out of the labor force. Furthermore, firms might be less willing to hire older individuals as there is only little time remaining to recover hiring costs before their entry into retirement.

We calibrate the model to match key features of the U.S. labor market and of inequality over the life cycle. For the labor market, we focus on matching average transition rates across labor market states as well as the joint distribution of couples across labor market states. For inequality, we match life cycle income profiles and asset holdings over the life cycle. Without targeting them, the model reproduces reasonably well life cycle profiles of labor market transitions as well as very closely the age-dependency in the added worker effect. The model captures very well that the effect is largest for the young and smallest for the age group just before retirement.

With the calibrated model at hand, we perform counterfactuals to evaluate which mechanisms are important in explaining the age-dependency in the added worker effect. Our results suggest a significant influence of larger asset holdings of older households, which can serve as a cushion against temporary job loss. Higher human capital levels of old employed spouses relative to their younger counterparts – accumulated during a longer working life – make spousal labor supply less valuable as an insurance margin but are quantitatively less important. Differences in job arrival rates for young and old out of the labor force spouses play a limited role, as they turn out to be relatively low for both age groups.

In future work, we will evaluate the consequences of these mechanisms for the provision of optimal life cycle unemployment insurance. For such an analysis it is key to match the risk exposure of households over their life cycle as well as the private insurance mechanisms, which could be crowded out through public transfer payments. As our model covers a wide range of insurance mechanisms available to households at different stages of their life cycle, the framework naturally lends itself to this question. Michelacci and Ruffo (2015) study optimal life cycle unemployment insurance using a single earner life cycle search model.[3] They argue that unemployment insurance should be more generous for the young than for the old, as the insurance value is very high for individuals with little assets and the moral hazard problem is limited, as young individuals need to accumulate labor market experience. Studying this question in a search model of couples is relevant because unemployment insurance could crowd out the added worker effect, which is an important insurance margin among young households.

Related Literature. The added worker effect is widely studied in the empirical literature, going back to the seminal contribution of Lundberg (1985). The early literature following this paper does not find much evidence supporting the presence of the added worker effect in the data (Maloney, 1987, 1991). More recent literature, however, documents a positive added worker effect as a relevant insurance mechanism against the primary earner's job loss (Bredtmann et al., 2018; Guner et al., 2020; Halla et al., 2020; Stephens, 2002), using data for a variety of countries. Mankart and Oikonomou (2016b) and Mankart et al. (2021) show that the added worker effect has become more important in the U.S. over the last decades. The literature argues that the size of the added worker effect crucially depends on the institutional environment and the state of the business cycle. For example, Cullen and Gruber (2000) show that generous unemployment insurance crowds out a spousal labor supply response. Expanding upon previous work, we argue that there is a sizeable age-dependency in the added worker effect.

While the added worker effect has been studied extensively in the empirical literature, the vast majority of the large macro-labor literature focuses on the job search problem of a single earner household. Guler et al. (2012) is among the first papers to study the joint search problem of a couple by extending the

[3] Optimal age-dependent policies are also commonly studied in public finance. See for example Erosa and Gervais (2002), Weinzierl (2011), and Heathcote et al. (2020b).

classic single-agent search problems of McCall (1970), Mortensen (1970), and Burdett (1978). A number of recent papers introduces asset accumulation into the joint search framework, expanding on the single agent search problem with asset accumulation as in Lentz (2009), Krusell et al. (2010), and Krusell et al. (2017). The focus of these papers is mostly on business cycle dynamics. Mankart and Oikonomou (2016a) build a search model with two member households to explain the cyclical properties of employment and labor force participation. Wang (2019) builds a model showing that joint household search is crucial for accounting for the countercyclicality of womens' unemployment rate. Ellieroth (2019) argues that there is precautionary labor supply by spouses whose partners face an increased job loss risk in recessions. Garcia-Perez and Rendon (2020) focus on the role of household wealth for the added worker effect. Birinci (2019), Choi and Valladares-Esteban (2020), and Fernández-Blanco (2020) investigate the implications of joint search for optimal unemployment insurance. Bardóczy (2020) focuses on the role of spousal labor supply as an automatic stabilizer for aggregate consumption. Relative to these papers, we focus on the life cycle dimension of the joint search problem to analyze whether the age-dependency in the added worker effect is explained by changing opportunities or changing insurance margins.

Life cycle search problems have been studied in the literature, but mostly in single earner frameworks. Chéron et al. (2011, 2013) extend the random search framework of Mortensen and Pissarides (1994) to a life cycle setting. Menzio et al. (2016) build a directed search life cycle model in the tradition of Moen (1997) and Menzio and Shi (2011). Griffy (2021) extends their model by incorporating risk averse workers and borrowing constraints. More closely related to our paper, Haan and Prowse (2017) propose a structural life cycle model of labor supply, consumption, and savings of married couples. They focus on the optimal mix of unemployment insurance and social assistance but do not discuss any age-dependency in the added worker effect. Finally, the current paper is related to a number of studies analyzing life cycle labor supply decisions of couples in incomplete market frameworks (Blundell et al., 2016; Ortigueira and Siassi, 2013; Wu and Krueger, 2021).

Roadmap. The paper proceeds as follows. Section 2.2 contains the empirical evidence. In Section 2.3 we introduce the model setup. Section 2.4 contains the calibration and section 2.5 the results. Section 2.6 concludes.

2.2 Evidence

The following section first explains the data and the sample selection criteria. Afterwards, we provide empirical evidence of the AWE in our sample and show that its magnitude is decreasing in age.

2.2.1 The Sample

To compute joint labor market transitions, we work with data from the Current Population Survey (CPS), provided by the Integrated Public Use Microdata Series (IPUMS) (Flood et al., 2020).[4] The CPS is a monthly rotating panel which is representative for the U.S. population. Households enter the survey for four consecutive months, drop out for eight months, and are re-interviewed for another four months. In our setting, the unit of observation is a couple. Our final sample spans from 1994 until 2020 (pre-Covid) and is restricted to couples who are both between 25 and 65 years old. We focus on couples with one spouse working and the other spouse out of the labor force. We include both legally married as well as cohabiting couples, irrespectively of their sex. In contrast, we drop couples who report that one spouse lives permanently outside of the household or is institutionalized. Moreover, we only keep couples for whom we observe the labor market status of both spouses in every month that they are interviewed. Throughout the analysis, we weigh each observation by the provided survey weights.

2.2.2 Uncovering the AWE from Joint Labor Market Transitions

We follow Guner et al. (2020) in our method to calculate the added worker effect from the data. First, we classify all individuals either as *employed* (E), *unemployed* (U) or *non-participating* (N) as outlined in the CPS. Hence, there exist

[4]To analyze the AWE by asset holdings, we complement our analysis with data from the Survey of Income and Program Participation (SIPP). See Section 2.2.4 for details.

nine possible combinations of labor market states for each couple. A common issue when considering multiple non-employment states is misclassification between unemployment and non-participation, resulting in implausibly high transition rates across these two. We therefore adjust labor market flows as in Elsby et al. (2015) and re-classify individuals who report to be unemployed (non-participating) in one month but to be out of the labor force (unemployed) in both the following and in the previous month as non-participating (unemployed).

In a next step, we pool all observations and construct a 3×3 matrix of joint labor market transition probabilities, conditional on the couple having one member previously employed and one out of the labor force. Table 2.1 and Table 2.3 display our main results. In each table, the columns refer to the monthly labor market transition of the household's primary earner, that is either employment-to-employment (EE), employment-to-unemployment (EU), or employment-to-non-participating (EN). In contrast, each row indicates the probability of the spousal labor market transition, conditional on the respective transition of the primary earner. Given that for this exercise we only include couples with one member employed and the other one non-participating, spouses can either transition from non-participating to employment (NE), from non-participating to unemployment (NU) or remain out of the labor force (NN). We define the added worker effect as the change in the conditional probability of the spouse transitioning from non-participating to employment (NE) or from non-participating to unemployment (NU) if the primary earner becomes unemployed (EU) in contrast to when the primary earner remains employed (EE). Referring to Table 2.1, we compute the added worker effect as the difference between the second and first column, adding up the first and the second row.

Overall Effect

Table 2.1 shows the overall strength of the added worker effect in our sample. The likelihood that a spouse enters the labor force increases by 5.9 percentage points, if the primary earner becomes unemployed compared to when the primary earner remains employed, confirming the existence of the added worker effect in

Table 2.1: Joint Labor Market Transitions (Full Sample)

	Primary earner transition		
	EE	EU	EN
Cond. prob. of spousal NE transition	6.03%	8.01%	16.79%
Cond. prob. of spousal NU transition	1.63%	5.55%	1.33%
Cond. prob. of spousal NN transition	92.34%	86.44%	81.88%

Notes: This table shows the probability of a spousal transition from out of the labor force conditional on primary earner transitions for the entire population.

our sample.[5] This result is in line with Guner et al. (2020), who find an overall AWE of 6.89 percentage points with CPS data spanning from 1976 to 2018 for couples between 25 and 54 years.

Zooming in on the precise margin of adjustment, we find that the conditional probability of the spouse transitioning directly into employment increases by 1.98 percentage points, whereas the conditional probability of the spouse transitioning into unemployment increases by 3.92 points. Thus, around two thirds of the overall AWE arise from individuals transitioning into unemployment, highlighting the importance of explicitly distinguishing between unemployed and non-participating individuals. Some couples may wish to leverage spousal labor supply as an insurance margin against job loss but labor market frictions (or the lack of appropriate job offers) prevent them from doing so.

Table 2.2 splits primary earners by the reason for why they became unemployed. We distinguish between laid-off workers (who face a high chance of being recalled), job losers, workers whose temporary contracts ended, and voluntarily job leavers. Table 2.2 shows that the strength of the AWE is similar across household members who voluntarily quit (column *Job Leavers*, especially with spouse NE) and those who are exogenously separated (*Job Losers*). In contrast, we find a slightly smaller effect for households in which the head's job loss can be seen as expected (*Temp. Job ended*) or temporary in nature (*Layoff*).

[5] In this paper we focus on the transitions of out of the labor force spouses conditional on the labor market transitions of primary earners. In the appendix, Tables B.1 and B.2 we also report the conditional transition probabilities of unemployed and employed spouses, respectively. There is a slightly higher likelihood that unemployed spouses transition to employment or stay unemployed rather than leave the labor force if the primary earner loses the job compared to the primary earner staying employed. However, evidence for insurance through spousal labor supply is strongest when considering out of the labor force spouses, which we focus on.

Table 2.2: AWE by reasons of Unemployment for Household Head

	EE	EU (by reasons for U)			
		Layoff	Job Loser	Temp. Job ended	Job Leaver
NE	6.03%	6.13%	8.81%	7.56%	10.47%
NU	1.63%	3.51%	6.66%	6.59%	7.68%
NN	92.34%	90.35%	84.53%	85.85%	81.86 %

Notes: This table shows the added worker effect (as defined in the main text) by reason for the EU transition of the primary earner.

Appendix B.1 reports additional results for couples that start as both employed or with one employed and one unemployed member. We find that unemployed spouses are slightly more likely to enter employment or keep looking for jobs rather than dropping out of the labor force if the primary earner moves from employment to unemployment. In addition, we observe couples making joint transitions: The likelihood of a spouse dropping out of the labor force is drastically increased when the primary earner does the same.

The Added Worker Effect by Age

Table 2.3: Joint Labor Market Transitions by Age

	Primary earner transition		
	EE	EU	EN
Age Spouse 25-35:			
Cond. prob. of spousal NE transition	6.66%	9.30%	26.93%
Cond. prob. of spousal NU transition	2.00%	6.89%	2.02%
Cond. prob. of spousal NN transition	91.34%	83.81%	71.05%
Age Spouse 36-45:			
Cond. prob. of spousal NE transition	6.73%	9.32%	26.69%
Cond. prob. of spousal NU transition	1.86%	6.37%	2.00%
Cond. prob. of spousal NN transition	91.41%	84.31%	71.30%
Age Spouse 46-55:			
Cond. prob. of spousal NE transition	6.13%	7.96%	16.62%
Cond. prob. of spousal NU transition	1.62%	4.79%	1.72%
Cond. prob. of spousal NN transition	92.25%	87.25%	81.66%
Age Spouse 56-65:			
Cond. prob. of spousal NE transition	4.29%	3.73%	8.69%
Cond. prob. of spousal NU transition	0.90%	2.75%	0.56%
Cond. prob. of spousal NN transition	94.81%	93.52%	90.76%

Notes: This table shows the probability of a spousal transition from out of the labor force conditional on primary earner transitions by age group.

Next, we split our sample into four age brackets and construct joint labor market transitions for each group in the same manner as above. Table 2.3 displays the results. We find a strong age-dependency in the strength of the AWE: For the youngest group (25 to 35 years), the likelihood that the spouse enters the labor force upon the job loss of the primary earner increases by 7.53 percentage points, for the young middle aged (36 to 45 years) it increases by 7.10 points, for the older middle aged (46 to 55 years) by 5.00 points, and eventually only slightly increases by 1.29 points for the oldest group (56 to 65 years). Thus, spousal labor supply adjustments of the youngest age group are more than five times larger than for the oldest age group.

For the young, we find behavioral responses both from non-participating directly into employment (2.64 percentage points) as well as into unemployment (4.89 percentage points). Thus, the relative share of young individuals transitioning directly into employment is slightly larger than for the entire sample. For the oldest age group, we only find small behavioral responses into unemployment (1.85 percentage points) and no response directly into employment (-0.56 points), suggesting that the AWE is not only a weaker margin of insurance for older workers through its decreased magnitude but also through its lower relative share of spouses actually finding employment.

2.2.3 Dynamic Response

So far, we have focused on the contemporaneous spousal labor supply response, that is, the probability that a spouse enters the labor force in the *same month* as the head transitions into unemployment. This most likely understates the overall strength of the added worker effect since spousal labor supply responses may occur in prior months (anticipation effects) or with some delay. In fact, Ellieroth (2019) documents spousal insurance not only in response to actual job loss of the primary earner but also in anticipation of such event, a phenomenon that she names “precautionary labor supply”. To analyze the strength of both anticipation and lagged responses, we run the following linear regression specification:

$$\Delta LFS_{it}^{sp} = \alpha_j + \beta_j \Delta ES_{it+j}^{h} + \gamma_j X_{it} + \epsilon_{jit}, \tag{2.1}$$

where ΔLFS_{it}^{sp} is a dummy that takes the value 1 if the non-participating spouse of couple i transitions either into employment or into unemployment between month $t-1$ and t, and 0 if she or he remains out of the labor force. Similarly, ΔES_{it}^{h} is defined as a dummy taking the value 1 if the primary earner transitions from employment into unemployment whereas it is 0 if the head stays in employment. X_{it} further controls for month fixed-effects, year fixed-effects, state fixed-effects, sex, race, education, children as well as the quarterly unemployment rate in the couple's state of residence.

Our coefficient of interest is β_j, indicating the likelihood that the spouse enters the labor force in month t if the household head transitions into unemployment in month $t+j$ versus when he or she remains employed (i.e. the strength of the AWE in month $t+j$). We conduct the analysis for $j = \{-2, -1, 0, 1, 2\}$. In the CPS, we observe the same couple for at most four consecutive months and hence a maximum of three consecutive labor market transitions, preventing us from considering more distant leads and lags. Figure 2.1 reports the results for the entire sample, whereas Figure 2.2 splits the observations by age.

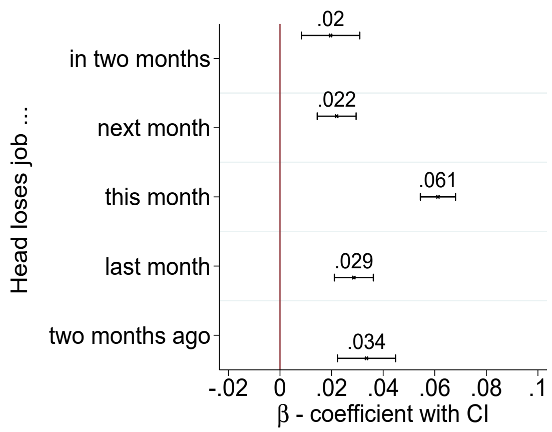
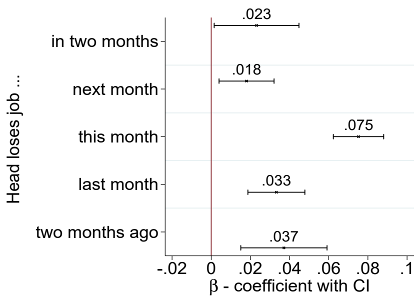


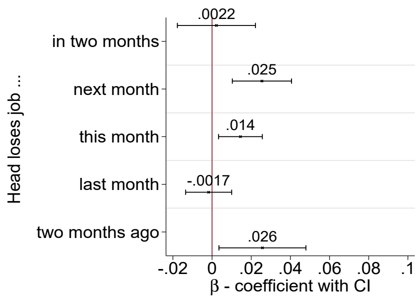
Figure 2.1: Δ Pr(Spouse enters LF) this month

Notes: Figure 2.1 shows the change in probability that a non-participating spouse enters the labor force (either as unemployed or as employed) if the household head loses/lost the job in two months, next month, this month, last month or two months ago, respectively, relative to the baseline in which the household head remains employed. The sample includes couples in which one spouse is working and one spouse is out of the labor force between age 25 and 65 from the Current Population Survey (CPS), waves 1994 until 2020. The regression producing the coefficients is Equation 2.1.

In line with section 2.2.2, Figure 2.1 confirms the overall strength of the AWE of around 6.1 percentage points in the contemporaneous month. Moreover, this effect is statistically significantly different from zero. In addition to the



(a) Age 25 to 35



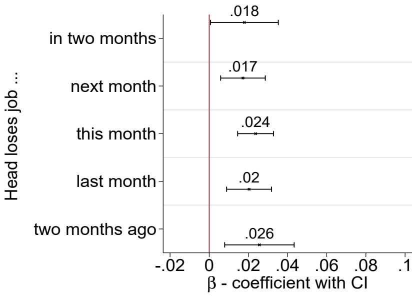
(b) Age 56 to 65

Figure 2.2: Δ Pr(Spouse enters LF) this month

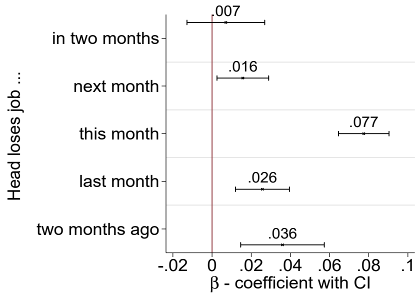
Notes: Figure 2.2 shows the change in probability that a non-participating spouse enters the labor force (either as unemployed or as employed) if the household head loses/lost the job in two months, next month, this month, last month or two months ago, respectively, relative to the baseline in which the household head remains employed. The sample includes couples in which one spouse is working and one spouse is out of the labor force between age 25 and 35 (Figure 2.2a) and between age 56 and 65 (Figure 2.2b) from the Current Population Survey (CPS), waves 1994 until 2020. Age refers to the non-participating spouse. The regression producing the coefficients is Equation 2.1.

contemporaneous effect, we find strong support of both anticipation and lagged effects, albeit of lower magnitude. Our results indicate that spousal labor supply responses in the months preceding and in the months after the primary earner's job loss are around half as strong as the direct response. When splitting the sample by age (Figure 2.2), we find that the contemporaneous effect is statistically significant for all age groups, however it is around five times stronger for the young than for the old. Moreover, young households display both lagged responses as well as anticipation effects, whereas we cannot confirm any clear pattern of those among households between 56 and 65 years. We relegate the results for the two middle age groups to Figure B.1 in the appendix.

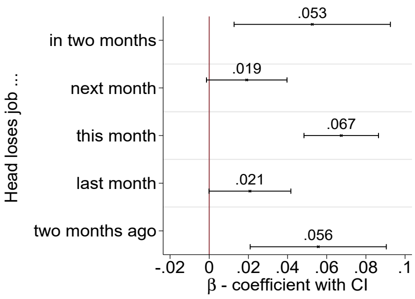
Lastly, in Figure 2.3, we again split the sample by reasons for unemployment of the primary earner (as in Table 2.2). As before, the probability that a non-participating spouse enters the labor force increases most if the EU transition of the primary earner is due to a quit or job loss, and less so in case of a layoff when there is a chance of being recalled. For spouses of household heads who voluntarily leave their job the effect two months ahead and two months lagged are smaller, while the effect in the month before and after the primary earner transition is larger. This finding can be taken as indication that these labor market transitions are coordinated choices within a short time span.



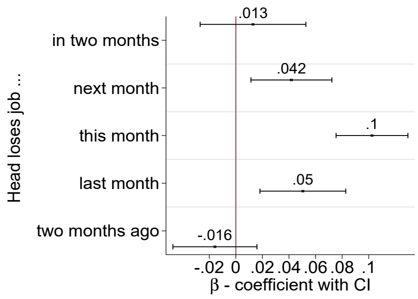
(a) Job Loser – On Layoff



(b) Job Loser



(c) Temporary Job Ended



(d) Job Leaver

Figure 2.3: Δ Pr(Spouse enters LF) this month

Notes: Figure 2.3 shows the change in probability that a non-participating spouse enters the labor force (either as unemployed or as employed) if the household head loses/lost the job in two months, next month, this month, last month or two months ago, respectively, relative to the baseline in which the household head remains employed; split by reasons for unemployment of the household head. Specifically, Figure 2.3a shows the results if the household head is on layoff, Figure 2.3b if the household head lost his job, Figure 2.3c if a temporary job ended and Figure 2.3d if the head voluntarily quit his or her job. The sample includes couples in which one spouse is working and one spouse is out of the labor force between age 25 and 65 from the Current Population Survey (CPS), waves 1994 until 2020. The regression producing the coefficients is Equation 2.1.

2.2.4 Supplementary Analysis with Asset Information

The Data. A deficiency of the CPS is that we do not observe asset holdings, which are a key insurance margin available to households. Therefore, we complement our analysis with data from the Survey of Income and Program Participation (SIPP) which collects asset information as well as the labor market state of both spouses. In the SIPP, households are interviewed every four months and report their monthly labor market states retrospectively, instead of being interviewed at a monthly frequency.[6] As a result, labor market transitions within each four-months interview wave tend to be underreported, whereas those across interview waves are overreported ("seam bias"). Facing this trade-off between datasets, and given the importance of correctly capturing monthly labor market flows for our research question, we perform our main analysis on CPS data,

[6]In panel 2016, interviews took place once a year.

and use the SIPP to report complementing evidence on the AWE by net liquid wealth.[7]

To do so, we work with waves 1994-2016 from the SIPP and apply the same sample restrictions as in the CPS. Our definition of net liquid wealth follows Chetty (2008) and is defined as total wealth minus home equity, vehicle equity, and unsecured debt. We use this measure because we are interested in wealth holdings that can be liquidated within a relatively short time frame, and hence provide insurance against temporary unemployment shocks.

The AWE by Net Liquid Wealth. Table 2.4 documents the AWE separately for the top and bottom 50% of net liquid wealth distribution within a given month. Among low wealth households, the job loss of the primary earner increases the likelihood of a spousal labor force entry by 6.56%-points. This response is reduced to 5.56%-points among high wealth households. Hence, in line with better insurance through asset holdings, richer households use spousal labor supply as an additional insurance margin relatively less.

When further splitting the sample by age (Table 2.5), we find a stronger AWE among the poorer half of the sample for old households. For young households, in contrast, the AWE is more pronounced among relatively high wealth couples. We hypothesize that this age-heterogeneity is driven by differences in the incentive to accumulate human capital. Young households with high asset holdings are most likely well educated and have a strong incentive to keep at least one member employed in order to not forgo the potential to accumulate human capital. Old households, in contrast, view spousal labor supply as a pure insurance mechanism which they execute less if they have other forms of insurance (such as assets) available to them.

Lastly, to ensure that potential seam bias is not confounding our results, we aggregate the data up to interview frequency. Within each aggregated time interval, we then assign individuals the labor market state that they report to be in most often. Table B.9 in Appendix B.1.8 confirms that the qualitative patterns

[7] In Appendix B.1.8, we show that the strength of the AWE is similar in magnitudes across both datasets, even though the baseline transitions tend to be underreported in the SIPP.

Table 2.4: Joint Labor Market Transitions by Net Liquid Wealth

	Primary earner transition		
	EE	EU	EN
Bottom 50% *of Net Liquid Wealth:*			
Cond. prob. of spousal NE transition	2.28%	5.16%	7.18%
Cond. prob. of spousal NU transition	1.37%	5.05%	2.65%
Cond. prob. of spousal NN transition	96.35%	89.79%	90.16%
Top 50% *of Net Liquid Wealth:*			
Cond. prob. of spousal NE transition	2.14%	5.36%	4.63%
Cond. prob. of spousal NU transition	0.84%	3.18%	1.02%
Cond. prob. of spousal NN transition	97.02%	91.46%	94.34%

Notes: This table shows the probability of a spousal transition from out of the labor force conditional on primary earner transitions by asset holdings.

of the AWE by age and by net liquid wealth on the aggregated sample are similar to those on monthly frequency.

2.2.5 Robustness

In this section, we explore further channels that could result in the observed age-dependency in the added worker effect without relating to life cycle heterogeneity in the insurance value of the AWE itself nor to other insurance margins that differ by age.[8] All corresponding tables are listed in Appendix B.1 and constructed with CPS data.

Education. If educational attainment differs by age and at the same time affects spousal labor supply responses, the stronger AWE for younger couples may simply arise from differences in education levels between old and young couples. Indeed, Table B.3 confirms that the AWE is larger for spouses with a college degree. However, when splitting the sample by age and education (Panel III to VI in Table B.3), the decreasing magnitude of the AWE over the life cycle holds both among spouses with a college degree and among those without a college degree.

[8] Some of these variables are also included as controls in the regressions. We still address the economically most important ones explicitly in this section.

Table 2.5: Joint Labor Market Transitions by Net Liquid Wealth & Age

	Primary earner transition		
	EE	EU	EN
Bottom 50% of Net Liquid Wealth (Young):			
Cond. prob. of spousal NE transition	2.97%	5.29%	10.84%
Cond. prob. of spousal NU transition	1.90%	5.57%	5.58%
Cond. prob. of spousal NN transition	95.13%	89.14%	83.58%
Top 50% of Net Liquid Wealth (Young):			
Cond. prob. of spousal NE transition	2.46%	5.99%	8.84%
Cond. prob. of spousal NU transition	0.95%	3.91%	0.00%
Cond. prob. of spousal NN transition	96.59%	90.09%	91.16%
Bottom 50% of Net Liquid Wealth (Old):			
Cond. prob. of spousal NE transition	1.34%	3.36%	4.58%
Cond. prob. of spousal NU transition	0.69%	4.03%	0.80%
Cond. prob. of spousal NN transition	97.97%	92.61%	94.62%
Top 50% of Net Liquid Wealth (Old):			
Cond. prob. of spousal NE transition	1.64%	2.29%	3.36%
Cond. prob. of spousal NU transition	0.49%	2.12%	1.19%
Cond. prob. of spousal NN transition	97.87%	95.59%	95.45%

Notes: This table shows the probability of a spousal transition from out of the labor force conditional on primary earner transitions by asset holdings and age group. “Young" refers to spouses below 40, and “Old" refers to spouses above 50.

Cohort Effects. If preferences for labor supply or within household insurance differ by cohorts (through, for example, changing gender norms), any age-dependency in the added worker may be driven by these underlying preference shifts. We address this concern in two ways. First, we split our sample by gender and age. If we can replicate the age-dependency in the AWE for couples in which the non-participating spouse is a man, possible cohort effects due to changing gender norms are less concerning. Table B.4 (Panels I and II) shows that although the overall probability of the spouse joining the labor force is higher when the non-participating household member is a man, we do not find significant changes in the strength of the AWE. Even when focusing only on male non-participating spouses, young households still show a stronger AWE than older ones.

Arguably, couples for which a man is non-participating could be a particular selection whose preferences differ from those of the remaining population. To address this concern, we extract one cohort and repeat the empirical exercise on this restricted sample. We focus on couples in which the non-participating spouse was born between 1960 and 1970. We choose this timespan to ensure sufficiently many observations both for the young and for the old age brackets. Table B.4 (Panel III and IV) confirms the decreasing magnitude of the AWE over the life cycle for this particular cohort, i.e. for the same cohort when young and when old.

Children. Young couples are more likely to have children living in their household, which arguably affects labor supply behavior. Couples with children might have stronger incentives to enter the labor force in response to the job loss of the primary earner because they have larger consumption commitments. On the other hand, if household members specialize in childcare and paid work, the willingness of spouses to switch tasks might be low. To address this issue, Table B.5 reports the AWE for couples below age 40 (to avoid picking up age-effects) with and without children as well as for couples below age 40 with and without children under age five (who require the most childcare). While out of the labor force spouses in couples without children have a higher baseline probability of entering the labor force, we do not find any (significant) differences in the overall strength of the AWE by age between couples with and without children.

Reasons for Non-Participation. If the non-participating spouse is retired, transitioning back into the labor force has a much smaller insurance value because of pension payments. Similarly, if the non-participating spouse dropped out because of bad health, she or he might simply not be able to start working again. Arguably, both retirement and health related non-participation are more prevalent among the old. Therefore, Table B.6 repeats the empirical analysis excluding retired spouses (Panels I and II), disabled or ill spouses (Panels III and IV), as well as excluding both retired and disabled/ill spouses (Panels V and VI). We do not find any significant impact on the strength of the AWE among the old, increasing our confidence that the observed age-heterogeneity is not driven by age-dependent reasons for non-participation.

Business Cycle. Next, we investigate whether our results differ by the state of the economy. If the primary earner loses a job in a recession, it might be harder to find a job again, so that insurance through spousal labor supply could be more important during downturns. On the other hand, it could also be harder for an out of the labor force spouse to find a job and provide this insurance. In Table B.7 we split the sample by NBER recessions and expansions. We do not find find large differences in the AWE across young and old for different states of the business cycle.

2.3 Model

The empirical evidence presented so far suggests that there is a significant age-dependency in the added worker effect: Spousal labour supply is a more important insurance margin for young than for old couples. We now build a life cycle search model with two-member households in order to better understand why the added worker effect is more prevalent among the young.

2.3.1 Environment

The economy is populated by two-member households. We assume that both members have the same age. Households live for T periods, after which they die deterministically. Households retire jointly after a working life of T_W periods, so that retirement lasts $T - T_W$ periods.

During working life an individual can be in one of four labor market states. An individual can be employed (E), in which case the agent receives a wage payment. If the individual does not have a job, there are three other labor market states: First, an agent may be unemployed and receive benefits (U). Second, the agent can be unemployed without receiving benefits (S). In both these states, the agent exerts costly search effort in order to increase the probability of finding a job. Third, an agent may choose to not exert this costly search effort. In that case, the agent is considered to be out of the labor force (N). Individuals who are not actively searching can never receive unemployment benefits. Given these

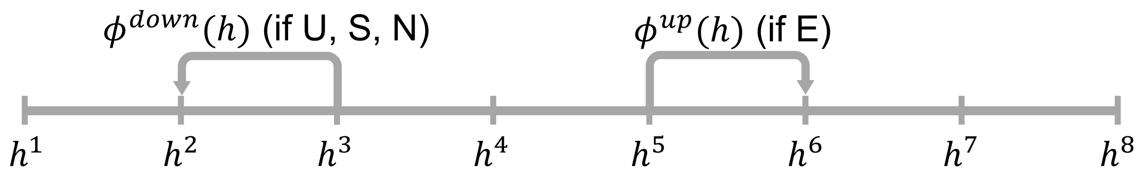


Figure 2.4: Human Capital Transitions

Notes: Figure 2.4 illustrates human capital transitions in the model.

four individual labor market states, there are 16 joint labor market states for a two-member household: $jk \in \mathcal{J} = \{E,U,S,N\} \times \{E,U,S,N\}$.

Each household member is endowed with a level of human capital, which evolves stochastically depending on the agent's employment status and current human capital level. If an individual member is employed, the human capital will go up by one unit with probability $\phi^{up}(h)$. For non-employed agents, human capital drops by one unit with probability $\phi^{down}(h)$. This process is illustrated in Figure 2.4.

While employed, an individual is additionally characterized by match quality z, which evolves according to a first-order Markov process. The match quality and the human capital level jointly determine the wage an individual receives. Non-employed individuals do not have a match quality, however they draw one upon finding a new job.

Individual labor market transitions are illustrated in Figure 2.5. An employed agent can receive an exogenous separation shock with probability $\delta(h)$, which depends on the level of human capital. If such a separation shock occurs, the agent transitions to unemployment and receives unemployment benefits. Note that in case of a separation shock an agent can choose to immediately leave the labor force instead of becoming unemployed and receiving benefits. This can be beneficial because no costly search effort is exerted while out of the labor force. If there is no separation shock, the individual can choose between staying employed and quitting. If she chooses to quit, she can either become unemployed without receiving benefits or leave the labor force entirely.

An unemployed agent who receives benefits can transition to all other labor market states. First, she receives a job offer with probability $\lambda^U(x_i)$ and transitions to employment if she chooses to accept the offer. The arrival rates with which non-employed agents receive job offers are endogenously determined as the solution to an optimal vacancy posting problem of firms (see below) and for household member i depend on state $x_i = \{h_i, h_{-i}, z_{-i}, a', jk\}$. An agent can also choose to reject the offer and might do so if the initial match quality draw is low. In that case, it may be preferable to wait for a new offer with a potentially better match quality draw. Second, an unemployed worker who receives benefits can stochastically lose benefit eligibility with probability ϕ^{US}, capturing that unemployment benefits run out after a certain time period. Third, she can choose to stop searching and leave the labor force. Similarly, an unemployed worker without benefits receives job offers with probability $\lambda^S(x_i)$ and can quit the labor force.

Finally, out of the labor force agents receive job offers with probability $\lambda^N(x_i)$, even though they do not exert active search effort. This assumption is necessary to capture the empirical observation that individuals directly transition from out of the labor force into employment. Moreover, non-participating agents can rejoin the labor force as unemployed without benefits.

While each household member has an individual labor market state, human capital level, and match quality shock when employed, households jointly have access to a risk-free bond. They can save in this bond at the exogenous interest rate r. Borrowing is not allowed.

2.3.2 Household Search Problem

Timing in the model is as follows: In each period, households first receive their labor income (wages or unemployment benefits) as well as their asset income from investing in the risk-free bond. Given their budget constraint, households then make a consumption-savings choice. Afterwards, separation shocks, job offers and potential losses of benefit eligibility are realized for both household members in parallel. Next, match quality shocks and human capital transitions are revealed. Finally, households choose their joint future labor market state from the feasible subset of $\mathcal{J}$, which is determined by their previous labor market state and job offers, separations, and benefit eligibility losses.

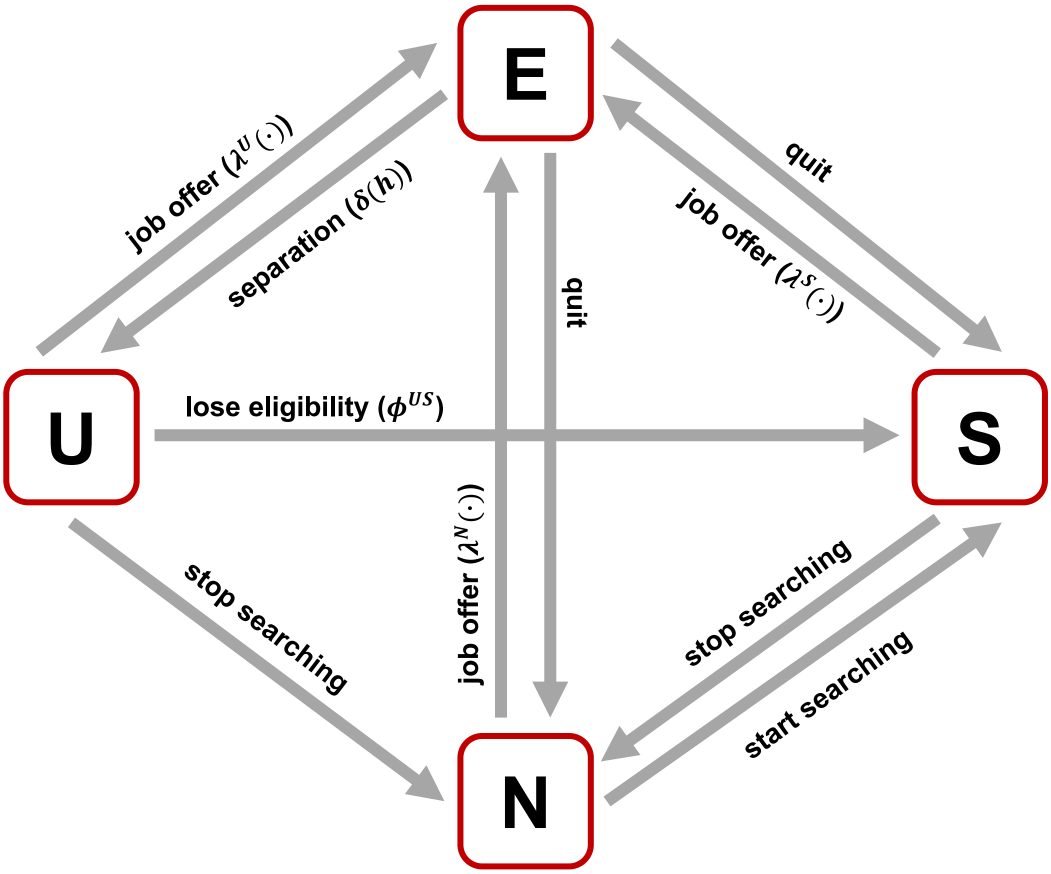


Figure 2.5: Labor Market Transitions in the Model

Notes: Figure 2.5 illustrates possible labor market transitions in the model. $x_i = \{h_i, h_{-i}, z_{-i}, a', jk\}$ is the relevant state for the arrival rate of household member i.

Table 2.6 summarizes all possible combinations of job opportunities and unemployment benefit eligibility of the two household members along with the associated choice sets over joint labor market states. The superscripts to $\mathcal{J}$ indicate whether the household members have the opportunity to be employed. An employment opportunity arises either because an agent was employed in the previous period and did not receive a separation shock or because an agent received a job offer while non-employed. If both members have the opportunity to be employed, the superscript is EE. In contrast, X indicates that a member cannot be employed. Hence, EX and XE are the cases in which only one member has a job opportunity, whereas XX indicates that neither household member can be employed in the following period.

Table 2.6: Labor Supply Choice Sets

Benefit Eligibility	Job (Offer) Both	Member 1	Member 2	None
Both	$\mathcal{J}_{UU}^{EE} = \{E,U,N\} \times \{E,U,N\}$	$\mathcal{J}_{UU}^{EX} = \{E,U,N\} \times \{U,N\}$	$\mathcal{J}_{UU}^{XE} = \{U,N\} \times \{E,U,N\}$	$\mathcal{J}_{UU}^{XX} = \{U,N\} \times \{U,N\}$
Member 1	$\mathcal{J}_{UX}^{EE} = \{E,U,N\} \times \{E,S,N\}$	$\mathcal{J}_{UX}^{EX} = \{E,U,N\} \times \{S,N\}$	$\mathcal{J}_{UX}^{XE} = \{U,N\} \times \{E,S,N\}$	$\mathcal{J}_{UX}^{XX} = \{U,N\} \times \{S,N\}$
Member 2	$\mathcal{J}_{XU}^{EE} = \{E,S,N\} \times \{E,U,N\}$	$\mathcal{J}_{XU}^{EX} = \{E,S,N\} \times \{U,N\}$	$\mathcal{J}_{XU}^{XE} = \{S,N\} \times \{E,U,N\}$	$\mathcal{J}_{XU}^{XX} = \{S,N\} \times \{U,N\}$
None	$\mathcal{J}_{XX}^{EE} = \{E,S,N\} \times \{E,S,N\}$	$\mathcal{J}_{XX}^{EX} = \{E,S,N\} \times \{S,N\}$	$\mathcal{J}_{XX}^{XE} = \{S,N\} \times \{E,S,N\}$	$\mathcal{J}_{XX}^{XX} = \{S,N\} \times \{S,N\}$

Notes: This table shows the labor supply choice sets of households.

The logic for the subscripts is similar. However, they refer to unemployment benefit eligibility of the individual household member. Again, U indicates eligibility, while X refers to non-eligibility.

We are now in the position to formally state the household search problem. The value function of a household of age t in joint labor market state jk is

$$V_t^{jk}(z,h,a) = \max_{a'} u(c^{jk}(z,h,a,a')) + \psi_t^{jk} + \beta \Theta_{t+1}^{jk}(z,h,a'), \tag{2.2}$$

where the additional state variables are the match quality shocks of both household members ($z = (z_1, z_2)$), their human capital levels ($h = (h_1, h_2)$), and joint asset holdings a. Households value consumption c according to the utility function $u(c)$. Consumption is pooled within the household. Additionally, instantaneous utility is affected by ψ which is allowed to depend on the labor market state and age. It captures disutility from search and the utility of staying at home. Households discount their continuation value Θ, which is described in detail below, with discount factor β.

Households choose assets for the next period subject to their budget constraint

$$c^{jk}(z,h,a,a') = \underbrace{\mathbb{I}_{j=E} w(z_1,h_1) + \mathbb{I}_{k=E} w(z_2,h_2)}_{\text{labor income}} + \underbrace{\mathbb{I}_{j=U}\bar{b} + \mathbb{I}_{k=U}\bar{b}}_{\text{unemployment benefits}} - \underbrace{(a' - (1+r)a)}_{\text{net savings}}. \tag{2.3}$$

Depending on their employment status households receive wage and benefit income. In addition to this, a household can use its assets and interest income to finance consumption and new purchases of the risk-free bond.

To write the continuation utility for one labor market state explicitly, we consider a household with two employed members today. Since both members are employed, the relevant state variables are two match quality shocks and two human capital levels. In addition, the continuation utility depends on the asset choice.

We express the continuation value in two steps. First, we take expectations over separation shocks and and the resulting choice sets for future labor market states:

$$\begin{aligned}
&\Theta^{EE}_{t+1}(z_1,z_2,h_1,h_2,a') = \\
&\quad (1-\delta(h_1))(1-\delta(h_2))\,\widetilde{V}_{t+1}(z_1,z_2,h_1,h_2,a',\mathcal{J}^{EE}_{XX}) \\
&\quad +\delta(h_1)(1-\delta(h_2))\,\widetilde{V}_{t+1}(z_1,z_2,h_1,h_2,a',\mathcal{J}^{XE}_{UX}) \\
&\quad +(1-\delta(h_1))\delta(h_2)\,\widetilde{V}_{t+1}(z_1,z_2,h_1,h_2,a',\mathcal{J}^{EX}_{XU}) \\
&\quad +\delta(h_1)\delta(h_2)\,\widetilde{V}_{t+1}(z_1,z_2,h_1,h_2,a',\mathcal{J}^{XX}_{UU}).
\end{aligned} \tag{2.4}$$

If neither member is exogenously separated (first line), both household members have the opportunity to work, but neither of them is eligible for benefits if he or she chooses to voluntarily quit. Hence, the feasible set of labor market states is denoted by $\mathcal{J}^{EE}_{XX}$. Lines 2 and 3 deal with the cases in which one member is exogenously separated whereas the last line considers the case in which both members receive the separation shock. In these instances, the exogenously separated member is eligible for benefits but cannot be employed in the next period.

In a second step, we consider transitions for match quality z and human capital h as well as the household's discrete choice over feasible future labor market states:

$$
\begin{aligned}
&\widetilde{V}_{t+1}(z_1,z_2,h_1,h_2,a',\mathcal{J}_{QR}^{OP}) = \\
&\qquad \phi^{up}(h_1)\phi^{up}(h_2)\,\mathbb{E}_{z_1'|z_1}\mathbb{E}_{z_2'|z_2}\,\mathbb{E}_\epsilon \max_{\widehat{jk}\in\mathcal{J}_{QR}^{OP}} \left\{V_{t+1}^{\widehat{jk}}(z_1',z_2',h_1+1,h_2+1,a') + \sigma\epsilon^{\widehat{jk}}\right\} \\
&\quad +\phi^{up}(h_1)(1-\phi^{up}(h_2))\,\mathbb{E}_{z_1'|z_1}\mathbb{E}_{z_2'|z_2}\,\mathbb{E}_\epsilon \max_{\widehat{jk}\in\mathcal{J}_{QR}^{OP}} \left\{V_{t+1}^{\widehat{jk}}(z_1',z_2',h_1+1,h_2,a') + \sigma\epsilon^{\widehat{jk}}\right\} \\
&\quad +(1-\phi^{up}(h_1))\phi^{up}(h_2)\,\mathbb{E}_{z_1'|z_1}\mathbb{E}_{z_2'|z_2}\,\mathbb{E}_\epsilon \max_{\widehat{jk}\in\mathcal{J}_{QR}^{OP}} \left\{V_{t+1}^{\widehat{jk}}(z_1',z_2',h_1,h_2+1,a') + \sigma\epsilon^{\widehat{jk}}\right\} \\
&\quad +(1-\phi^{up}(h_1))(1-\phi^{up}(h_2))\,\mathbb{E}_{z_1'|z_1}\mathbb{E}_{z_2'|z_2}\,\mathbb{E}_\epsilon \max_{\widehat{jk}\in\mathcal{J}_{QR}^{OP}} \left\{V_{t+1}^{\widehat{jk}}(z_1',z_2',h_1,h_2,a') + \sigma\epsilon^{\widehat{jk}}\right\}
\end{aligned}
\tag{2.5}
$$

For employed individuals human capital can either remain constant or increase. Each line of equation 2.5 corresponds to one of the resulting four combinations of possible human capital transitions. Moreover, in each case, expectations are also taken with respect to match quality shocks.

The possible choices of future labor market states can be read off Table 2.6. $\epsilon \in \mathbb{R}^{|\mathcal{J}_{QR}^{OP}|}$ is a vector of iid, Type-I extreme value (Gumbel) shocks with mean zero. We introduce these taste shocks for computational purposes, as they smooth out kinks and discontinuities in the policy functions that arise from the discrete choices over labor market states. We choose the variance of these taste shocks to be small enough such that they do not affect the solution to the problem in an economically meaningful way.

While we outline here the continuation value for a household with two members currently employed, the problem for all other current joint labor market states evolves in a very similar manner: In equation 2.4, instead of separation shocks expectations are formed over job offer arrivals and potential losses of benefit eligibility for non-employed members. Equation 2.5 remains mostly unaffected except for initial draws of z out of non-employment, which stem from an initial distribution and are independent of past realizations of z.

2.3.3 Vacancy Posting and Endogenous Arrival Rates

To determine the job arrival rates of households endogenously we consider the optimal vacancy posting problem of single-job firms. We assume free entry of firms and a cost κ of posting a vacancy. A vacancy lasts for one period and if not filled can be renewed by paying κ again.

A match with quality z between a firm and a worker with human capital h produces per period output $y(z,h)$, of which the worker receives a constant share χ as a wage $w(z,h) = \chi y(z,h)$, yielding firms' per period profit of such match as $\pi(z,h) = (1-\chi)y(z,h)$.

The expected future value to a firm of a match with a worker i from a household with current state $x_i = (t, z_i, z_{-i}, h_i, h_{-i}, a, jk)$ and asset choice for next period a' , given that the household can choose the joint future labor market state from set $\mathcal{J}_{QR}^{OP}$, is defined as

$$\begin{aligned} &EJ_{t+1}^{jk}(z_i, z_{-i}, h_i, h_{-i}, a', \mathcal{J}_{QR}^{OP}) = \\ &\qquad \mathbb{E}_{h_i'|h_i}\mathbb{E}_{h_{-i}'|h_{-i}}\mathbb{E}_{z_i'|z_i}\mathbb{E}_{z_{-i}'|z_{-i}}\mathbb{E}_{\hat{j}\hat{k}\in\mathcal{J}_{QR}^{OP}}\mathbb{I}_{\hat{j}=E|x'}J_{t+1}^{\hat{j}\hat{k}}(z_i', z_i', h_{-i}', h_{-i}', a'), \end{aligned} \tag{2.6}$$

where $\mathbb{E}_{\hat{j}\hat{k}\in\mathcal{J}_{QR}^{OP}}\mathbb{I}_{\hat{j}=E|x'}$ is firms' expectation of the household's joint labor market choice and an indicator of whether for each joint state member i stays with the firm, i.e. firms ' expectation over endogenous acceptances and quits. The contemporaneous value to the firm is then given by

$$\begin{aligned} &J_t^{jk}(z_i, z_{-i}, h_i, h_{-i}, a) = \\ &\qquad \pi(z_i, h_i) + \frac{1}{1+r}(1-\delta(h_i))\mathbb{E}_{P,R}EJ_{t+1}^{jk}(z_i, z_{-i}, h_i, h_{-i}, a', \mathcal{J}_{XR}^{EP}), \end{aligned} \tag{2.7}$$

where $\mathbb{E}_{P,R}$ is a firm's expectation over job loss, job finding, and eligibility transitions of the spouse and $a' = a(t, z_1, z_2, h_1, h_2, a, jk)$ is the household's asset choice.

We discuss the determination of endogenous arrival rates using the example of a household with both members unemployed but not eligible for benefits, i.e. a household with initial labor market state SS. Define member i's arrival rate as

$$\lambda_t(h_i, h_{-i}, a, jk) = \lambda_S p(\theta_t(h_i, h_{-i}, a, jk)) \tag{2.8}$$

with arrival rate $p(\theta) = m(1, \theta)$ and corresponding vacancy filling rate $q(\theta) = m(\frac{1}{\theta}, 1)$, where $m(U,V)$ is the standard Cobb-Douglas matching function, with market tightness θ denoting the ratio of vacancies over searchers in any given submarket. Hence $p(\theta) = \theta^{1-\alpha}$, $q(\theta) = \theta^{-\alpha}$, and $p(\theta) = \theta q(\theta)$. λ_S is an exogenous shifter that only depends on the previous labor market state and reflects the consequences of differences in search effort between unemployed (U or S) and out of the labor force (N). This distinction is necessary because – conditional on the remaining states of the household – firms will not differentiate whether they hire a worker out of unemployment or from out of the labor force.

Free entry imposes that the expected value of a vacancy (probability of filling times the value if filled) has to equal the cost of posting κ. This condition determines relevant market tightness $\theta_t(h_i, h_{-i}, a, jk)$. The free entry condition needs to satisfy

$$\kappa = q(\theta_t(h_i, h_{-i}, a, jk))\mathbb{E}_P E J_{t+1}^{jk}(z_i, z_{-i}, h_i, h_{-i}, a', \mathcal{J}_{XX}^{EP}). \tag{2.9}$$

Here $\mathbb{E}_P$ captures expectations over the spouse's job finding and is an equation in the spouse's $\theta_t(h_{-i}, h_i, a, jk)$ as the spouse is also currently not employed. Hence, in all cases with currently two non-employed household members we have to solve a system of two non-linear equations in two unknowns.

With slight abuse of notation the two equations solving for two θs can be written as

$$\kappa = q(\theta_i)[\lambda(\theta_{-i})\underbrace{EJ_{t+1}^{SS}(h_i, h_{-i}, a', \mathcal{J}_{XX}^{EE})}_{EJ_i^{EE}} + (1-\lambda(\theta_{-i}))\underbrace{EJ_{t+1}^{SS}(h_i, h_{-i}, a', \mathcal{J}_{XX}^{EX})}_{EJ_i^{EX}}], \tag{2.10}$$

$$\kappa = q(\theta_{-i})[\lambda(\theta_i)\underbrace{EJ_{t+1}^{SS}(h_{-i}, h_i, a', \mathcal{J}_{XX}^{EE})}_{EJ_{-i}^{EE}} + (1-\lambda(\theta_i))\underbrace{EJ_{t+1}^{SS}(h_{-i}, h_i, a', \mathcal{J}_{XX}^{EX})}_{EJ_{-i}^{EX}}]. \tag{2.11}$$

This yields

$$\theta_{-i} = \left[\frac{\kappa}{\lambda(\theta_i)EJ_{-i}^{EE} + (1-\lambda(\theta_i))EJ_{-i}^{EX}}\right]^{-\frac{1}{\alpha}} \tag{2.12}$$

and hence

$$\kappa = q(\theta_i)\left[\lambda_S\left[\frac{\kappa}{\lambda(\theta_i)EJ_{-i}^{EE}+(1-\lambda(\theta_i))EJ_{-i}^{EX}}\right]^{\frac{\alpha-1}{\alpha}}EJ_i^{EE}\right.$$
$$\left.+\left(1-\lambda_S\left[\frac{\kappa}{\lambda(\theta_i)EJ_{-i}^{EE}+(1-\lambda(\theta_i))EJ_{-i}^{EX}}\right]^{\frac{\alpha-1}{\alpha}}\right)EJ_i^{EX}\right], \tag{2.13}$$

which is a non linear equation in one unknown and can be solved numerically.

The endogenous arrival rates can be derived in a similar fashion for other cases of original labor market states. The exogenous component of λ needs to be adjusted to reflect whether an agent is unemployed or out of the labor force. Solving for endogenous arrival rates gets substantially easier if one spouse has been previously employed since in this case we only have one θ and hence we only need to solve one equation with one unknown.

Given this setup, job finding probabilities of an individual depend on all the state variables, including assets, age, and own human capital, but also the spouse's human capital, employment status, and potentially match quality. With regard to age, our setup is hence able to capture that it may be harder for older workers to find a new job. In the model, firms are less willing to hire older workers because they have to retire at a certain age, leaving less time to recover the vacancy posting cost. In our calibration, this effect is strong close to retirement but relatively weak at young ages because in these cases it is quite likely that the match is dissolved before retirement in any case.

It is also intuitive that arrival rates depend on an individual's human capital. It is potentially less appealing that we also condition on the spouse's state variables. It is necessary, however, because it influences the probabilities of an individual accepting a certain job and quitting later on. Having different submarkets and free entry in each active submarket simplifies computation drastically, as we do not need to know the distribution of individuals across states to solve for arrival rates.

This setup for determining age-dependent arrival rates in the labor market generally implies arrival rates decreasing in age, decreasing in assets because richer individuals are more likely to quit, increasing in human capital because the value of the match is higher and individuals are less likely to quit, increasing in match quality for the same reasons, and decreasing in a spouse's employment, human capital, and match quality because having a spouse earning high wages increases the quit probability and lowers the value of a match to the firm.

2.3.4 Numerical Implementation

In our setup, agents do not face risk during retirement. This assumption renders the household problem during retirement very simple. We solve the retirement problem using the endogenous grid method of Carroll (2006) to obtain a terminal condition for the household problem during working life.

The household problem during working life is high-dimensional because of the many combinations of labor market states and the fact that we have to keep track of match quality shocks and human capital for both members. Furthermore, given our focus on labor market transitions, the model has a monthly frequency. For computational efficiency, we therefore solve the household problem following Iskhakov et al. (2017), who extend the endogenous grid-point method of Carroll (2006) to problems with discrete and continuous choices. Thus, their approach is well suited for our problem with a discrete choice over labor market states and a continuous asset choice.

The algorithm proceeds as follows: Within each period, given future value functions of both the household and firm, we begin by determining households' choices over future labor market states for each potential choice set. With this, we are able to solve firms' vacancy posting problem and determine endogenous arrival rates. Endogenous arrival rates given, we can solve households' consumption-savings problem as described above. In a final step, we update households' and firms' value functions making use of households' policy functions and again the endogenous arrival rates.

2.4 Calibration

We solve the model at a monthly frequency. This assumption is in line with the frequency at which we observe labor market transitions in the data and necessary because the U.S. labor market exhibits high rates of turnover. We assume that the period of working life is 40 years, corresponding to 480 months. The retirement period is another 120 months, i.e. 10 years.

2.4.1 Functional Form Assumptions

Households value consumption with a standard CRRA utility function

$$u(c) = \frac{c^{1-\gamma} - 1}{1-\gamma}, \tag{2.14}$$

where γ is the coefficient of relative risk aversion. The second part of instantaneous utility that has to be parameterized is the parameter ψ_t^{jk} which differs across joint labor market states, reflecting disutility of work and search. Furthermore, we allow it to vary by age.[9]

Output is assumed to be the product of human capital and the match quality shock:

$$y(h,z) = hz. \tag{2.15}$$

Human capital is defined on an equidistant grid. The probabilities of moving to a higher (lower) human capital level when employed (non-employed) are given by the following processes:

$$\phi^{up}(i) = \bar{\phi}^{up} i^{\underline{\phi}^{up}} \tag{2.16}$$

$$\phi^{down}(i) = \bar{\phi}^{down} i^{\underline{\phi}^{down}}, \tag{2.17}$$

where i indicates the grid point rather than the level of human capital. This process is able to capture falling or rising probabilities of moving up or down the human capital ladder. The match quality shock while employed is assumed to follow an autoregressive process of order 1 in logs. We discretize the process using the method of Tauchen (1986).

Finally, we have to make an assumption on the arrival rates of job offers and separation rates in the labor market. We restrict $\lambda_S, \lambda_U, \lambda_N$ to be constant across

[9] In the current calibration, the disutility of search parameter is mostly constant across age. In fact, we make an exception only for one labor market state, as discussed below.

age.[10] We allow the separation rate to vary with human capital according to a similar process as the probabilities of moving up or down the human capital ladder:

$$\delta(i) = \bar{\delta} i^{\underline{\delta}}. \tag{2.18}$$

2.4.2 Parameters and Moments

To compare the model to the data, we simulate the full life cycle of 40,000 households and compute model-implied moments of this simulation. We initialize the distribution of households across labor market states such that it is consistent with the data. We assume that all agents start with one of the lowest asset levels. For employed individuals, we draw the match quality shock from the stationary distribution of the match quality process. For human capital, even though this is mostly supposed to capture work experience in our model, we assume some heterogeneity in the initial distribution to obtain sufficient dispersion in incomes. Human capital levels are, however, concentrated on the lower rungs of the human capital ladder.

While in the model all parameters jointly determine all moments, we now discuss which parameters are most closely related to which moments. Table 2.7 summarizes the parameter values. We start by setting a number of parameters without solving the model. We exogenously fix the coefficient of relative risk aversion to two, a standard value in the literature. We set the monthly net interest rate to 0.17%, corresponding to an annual interest rate of roughly 2%. We assume a probability of losing unemployment benefits of $\phi^{US} = 1/6$, consistent with an average duration of benefit receipt of six months. Finally, we set the elasticity of the matching function α to 0.5, as in Petrongolo and Pissarides (2001), and the share of match output going to the worker χ to 0.7.

We target key moments of the U.S. labor market that are related to a large number of parameters. First, we target individual transition rates between labor market states. These are closely related to the parameters $\lambda_N, \lambda_S, \lambda_U$, the exogenous upper bounds on arrival rates depending on labor market states. We impose the restriction $\lambda_S = \lambda_U$, as these two states only differ in whether an

[10] Even though the exogenous component of arrival rates is constant in age, the solution to firms' vacancy posting problem endogenously yields arrival rates falling in age t conditional on households' remaining states.

individual receives unemployment benefits or not. Individual transition rates are closely related to the vacancy posting cost κ. The EU rate in particular pins down parameters of the job loss process. The model captures well the magnitude of the transitions between employment and unemployment. In contrast, it undershoots the magnitude of transitions between non-employment and employment/unemployment, as we will discuss in more detail in the next section when looking at the added worker effect in the model.

Another important set of targeted labor market moments is the distribution of households over joint labor market states for four ten-year age groups. Because the arrival rates are endogenously determined from the firm problem we treat the preference parameters ψ that govern the disutility of work and search as free parameters to match joint labor market states by age. We keep all these parameters constant by age, except for $\psi^{EN} = \psi^{NE}$, which we assume to be decreasing with age. Specifically, we assume $\psi^{EN} = \psi^{NE}$ to start at a level of 1.30 at age 25 and to decay logistically to a level of 90 with a half-life of 100 months. Imposing this age-dependency is necessary in order to avoid that too many young households have both members employed. Economically, we justify a higher utility of having one member at home for young households because this is the age group who are most likely to have young children. As we do not model children explicitly, introducing age-dependency in $\psi^{EN} = \psi^{NE}$ is a parsimonious way of capturing this motive and helps us to match a high enough share of young households with one member employed and one member out of the labor force.

In addition to these labor market moments, we target life cycle profiles of income and assets. The pension level p and the discount factor β are mostly determined by the shape of the life cycle asset profile. Specifically, we target mean asset holdings for four age groups. An important question is which assets to consider in the data when constructing the moments to be matched. For insurance reasons, the relevant concept is liquid assets. In particular, because a model period is one month, it would be desirable to consider only assets that can be liquidated at a monthly frequency. However, given the life cycle dimension of our setup, retirement is an important driver of savings. Imposing too strict requirements on asset liquidity would exclude much of households' retirement savings. Therefore, considering the trade-off between asset liquidity and retirement savings, we choose to target financial assets including retirement accounts net of debt. In addition, we include vehicle equity because it can be accessed very quickly. In contrast, we exclude houses and mortgages because tapping into home equity is difficult for unemployed and might take longer, so it is not as useful for insurance purposes on a monthly frequency. Business equity is excluded for the same reason. We construct asset-related data moments from the Panel Study of Income Dynamics (PSID).

The parameters of the human capital process are chosen to match the income profile over the life cycle. In the data, these moments are constructed from the PSID. The probability of moving up the human capital ladder is decreasing in the human capital level which is a way of achieving a concave income profile: When young, an agent moves up the human capital ladder quickly such that the wage increase is steeper. After a few steps on the human capital ladder, the likelihood of a further increase in human capital decreases quite significantly such that the income profile becomes flatter. The probability of losing human capital, by contrast, is constant across human capital levels. Human capital decay of non-employed allows us to capture the empirical observation that newly employed individuals have lower wages than long-time employed and that job losses lead to persistent wage drops (Davis and von Wachter, 2011; Jarosch, 2015; Kospentaris, 2021).

The parameters of the match quality shock process are chosen to match the variance in income levels by age group. Additionally, we have to pin down the distribution from which newly employed draw their match quality, which we set to the stationary distribution of the discretized Markov chain.

The only remaining parameters to be set are the level of the unemployment benefit and the variance of the taste shock. We assume the unemployment benefit to be constant and set its level to be roughly 50% of median income. For the taste shock, we set $\sigma_{\varepsilon} = 0.1$. Using 0.05 instead does not meaningfully impact or results.

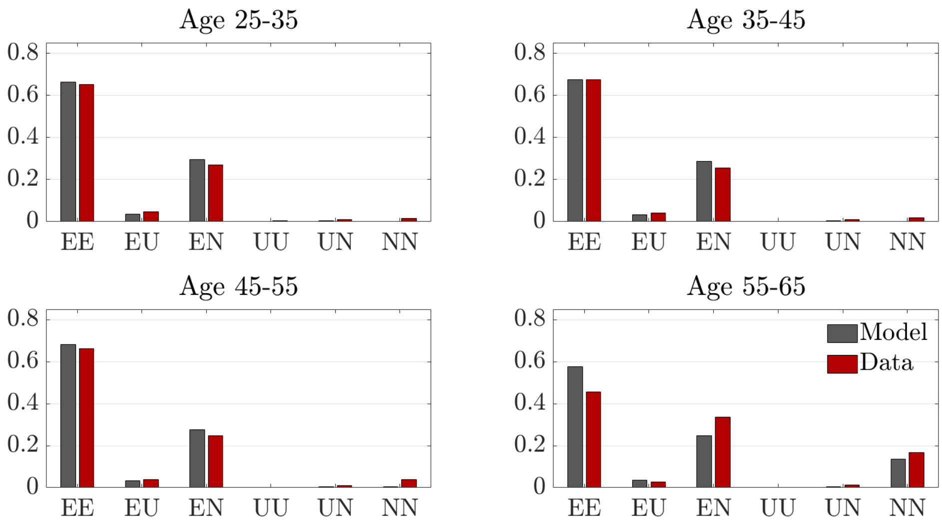


Figure 2.6: Joint Labor Market States of Couples (Model vs. Data)

Notes: Figure 2.6 shows the joint labor market states of couples in the model and in the data. For the model, U includes both unemployed receiving benefits and searchers who do not receive benefits. The data is from the CPS.

2.4.3 Fit of Targeted Moments

In this section, we present the model fit for key targeted moments. First, Figure 2.6 shows the share of households in joint labor market states by age group in the model and in the data. To compare the model to the data, we pool all agents who are unemployed with and without benefits into one group, labeled U. In all age groups, the most common joint labor market state is that both members are employed. This share is, however, strongly decreasing in age, with around 65% of households being in that group among the two young groups and just 45% in the oldest age group. By contrast, the share of households where at least one member is out of the labor force is increasing over the life cycle. Among the youngest there are very few households with both members out of the labor force. Among the oldest, almost 20% of all couples are jointly non-participating. In addition, the share of households with one member employed and one member out of the labor force is slightly increasing in age. Overall, the model matches very well the distribution of households over joint labor market states. It also captures that the share of two earner households is decreasing in age and that the share of households with at least one member out of the labor force is increasing in age, though it somewhat understates the magnitude of these changes over the life cycle.

Moreover, the model is able to replicate average asset holdings over the life cycle, as shown in Table 2.8. Averaging over all age groups, we match the average asset level of the population well. However, the model slightly underpredicts the mean asset holdings of the medium age groups. However, it captures that average asset holdings are strongly increasing in age.

Finally, we consider the model fit for mean income levels and the dispersion in income across age groups. Table 2.9 shows the comparison between data and model. Again, when averaging over all age groups, the model is close to the income level in the data but as of now undershoots the dispersion. Moreover, the model is able to replicate the increase in mean income for the age groups 25-35, 35-45, and 45-55. It fails, however, in generating a fall in income for the oldest group. This mismatch for the oldest age group arises from a strong selection effect in the model with respect to who stays in the labor force. Many agents with relatively low human capital and/or match quality prefer to drop out of the labor force, which drives up the average income among the employed. In contrast, the model replicates that income dispersion within age group is higher among the old than among the young.

Table 2.7: Parameter Values

Parameter	Interpretation	Value
Demographics		
T	Length of life in months	600
T_W	Length of working life in months	480
Preferences		
β	Discount factor	0.9955
γ	Risk aversion	2.0000
$\psi^{EE}, \psi^{EU}, \psi^{UE}, \psi^{ES}, \psi^{SE}$	Disutility of work/search	0.0000
$\psi^{UU}, \psi^{SS}, \psi^{SU}, \psi^{US}$	Disutility of work/search	0.5000
$\psi^{UN}, \psi^{NU}, \psi^{SN}, \psi^{NS}$	Disutility of work/search	1.2000
ψ^{NN}	Disutility of work/search	2.6000
ψ^{EN}, ψ^{NE}	Disutility of work/search	$1.3 + \frac{0.9-1.3}{1+e^{-0.05(t-100)}}$
Financial Assets		
r	Interest rate	0.0017
Labor Market		
$\bar{\delta}$	Level parameter separation rate	0.0200
$\underline{\delta}$	Curvature parameter separation rate	-0.5000
λ_U, λ_S	Probability of job offer for unemployed	0.4500
λ_N	Probability of job offer out of labor force	0.3000
Human Capital		
$\underline{h}$	Lower bound h	0.2000
$\bar{h}$	Upper bound h	0.8000
$\bar{\phi}^{up}$	Level parameter prob. h rise	0.0500
$\underline{\phi}^{up}$	Curvature parameter prob. h rise	-1.2000
$\bar{\phi}^{down}$	Level parameter prob. h fall	0.3316
$\underline{\phi}^{down}$	Curvature parameter prob. h fall	0.0000
Match Quality Shocks		
ρ_z	Persistence	0.9000
σ_z	Standard deviation	0.1000
Firms		
χ	Labor share of output	0.7000
κ	Cost of vacancy posting	8.0000
α	Matching elasticity	0.5000
Government		
b	Unemployment benefit	0.2500
ϕ^{US}	Probability of losing benefits	0.1667
p	Pension	0.2000
Gumbel shock		
σ_ε	Standard deviation of taste shock	0.1000

Notes: Table 2.7 summarizes the parameter values.

Table 2.8: Asset Levels

	Model	Data
All	10.4	11.8
Age 25-35	2.8	3.0
Age 36-45	4.9	7.0
Age 46-55	10.6	14.6
Age 55-65	23.3	24.1

Notes: Table 2.8 compares mean asset holdings by age group in the model and in the data. The data is from the PSID. In the data, assets include financial assets net of debt and vehicle equity. 1 unit corresponds to $10,000.

Table 2.9: Income Levels and Dispersion

	Level		Standard deviation	
	Model	Data	Model	Data
All	0.3596	0.3424	0.1363	0.2374
Age 25-35	0.3296	0.3020	0.1172	0.2009
Age 36-45	0.3538	0.3572	0.1341	0.2456
Age 46-55	0.3752	0.3629	0.1429	0.2486
Age 56-65	0.3826	0.3400	0.1511	0.2466

Notes: Table 2.9 compares mean and standard deviation of labor income by age group in the model and in the data. The data is from the PSID. 1 unit corresponds to $10,000.

2.5 Results

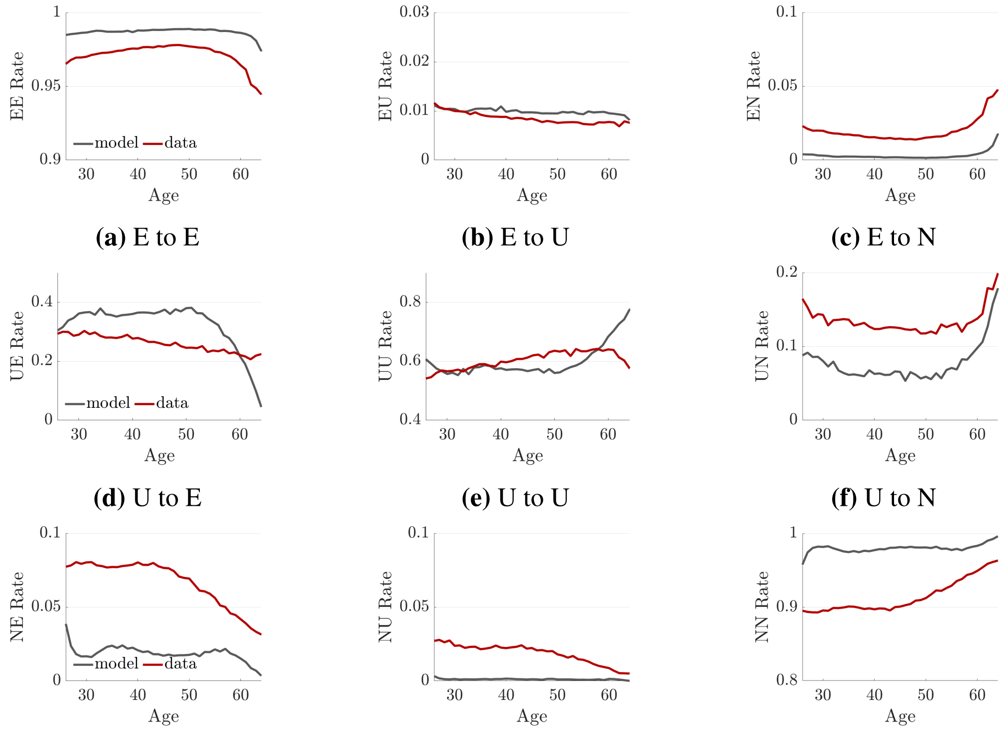
In this section we first present the model implications for untargeted moments. Second, we show that our model can replicate the decreasing magnitude of the added worker effect over the life cycle. Third, we use the model to construct counterfactuals and analyze which channels are responsible for the age-dependency in the added worker effect.

2.5.1 Untargeted Moments

We begin this section by presenting untargeted life cycle profiles of individual labor market transitions in Figure 2.7. Again, in the model U comprises both the group of unemployed who receive benefits and those who exert costly search effort without receiving benefits.

First, consider transitions from employment over the life cycle (Figure 2.7a to 2.7c). The model captures that the likelihood of remaining in employment falls quite rapidly towards the end of working life, though the monthly transition probability out of employment never falls below 95%. The counterpart to this in model and data is a corresponding increase in the likelihood of moving from employment to out of the labor force. As agents get closer to the retirement age, it is not worthwhile for them to stay employed when they receive a bad match quality shock or have low human capital. By contrast, young agents continue to work even in these cases. Several model mechanisms account for this. First, young agents have a longer time horizon until retirement, so that they need labor income to cover consumption needs during working life. In contrast, old agents hold much higher levels of assets which they can use to finance consumption. Second, human capital is only accumulated while employed. Thus, higher human capital is more valuable for the young as they can benefit from it for a longer time period. The model performs very well in matching the slightly decreasing path of E to U transitions over the life cycle.

Next, consider the transitions out of unemployment (Figure 2.7d to 2.7f). The model replicates that across the entire life cycle the most likely transition is to remain unemployed. It also matches well that the probability of transitioning to employment declines with age, whereas the probability of giving up on searching



(a) E to E **(b)** E to U **(c)** E to N

(d) U to E **(e)** U to U **(f)** U to N

(g) N to E **(h)** N to U **(i)** N to N

Figure 2.7: Labor Market Transitions over the Life Cycle

Notes: Figure 2.7 shows individual labor market transitions in the data and in the model. For the model, U includes both unemployed receiving benefits and searchers who do not receive benefits. The data is from the CPS.

and leaving the labor force increases with age. Finally, the model generates a fall in transitions from out of the labor force into employment (Figure 2.7g) but understates the likelihood to transition into unemployment (Figure 2.7h) over the life cycle, while it matches well the high persistence of non-participation (Figure 2.7i).

Again, it is apparent from these figures that the model generates too few transitions between out of the labor force and employment/unemployment. This is most likely due to the fact that we leave many important life events such as child birth, marital transitions, and health shocks unmodeled. We will show next, however, that the model captures well the impact of one key life event, job loss of the primary earner, on the labor force participation of out of the labor force spouses, the added worker effect.

Table 2.10: Joint Labor Market Transitions by Age (Model vs. Data)

	Primary earner transition	
	EE	EU/ES
Young (25-35):		
Cond. prob. of spousal NE transition	2.26%	3.12%
	6.66%	9.30%
Cond. prob. of spousal NS transition	0.40%	5.28%
	2.00%	6.89%
Cond. prob. of spousal NN transition	97.34%	91.60%
	91.34%	83.81%
Old (55-65):		
Cond. prob. of spousal NE transition	1.95%	2.24%
	4.29%	3.73%
Cond. prob. of spousal NS transition	0.11%	1.16%
	0.90%	2.75%
Cond. prob. of spousal NN transition	97.95%	96.60%
	94.81%	93.52%

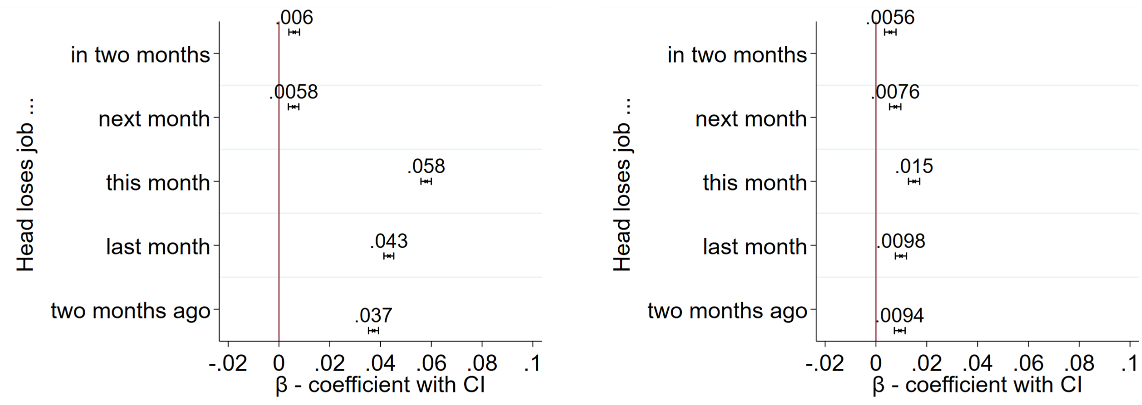
Notes: This table compares joint labor market transitions by age in the model and in the data.

2.5.2 The Added Worker Effect over the Life Cycle in the Model

We now evaluate whether the model can replicate our main empirical finding: the age dependency in the added worker effect. To compare model to data, we replicate Table 2.3 from Section 2.2 with simulated model data in Table 2.10. For ease of comparison, we also report empirical transition probabilities.

For the young, the model is capable of producing a strong increase in the probability of moving from out of the labor force directly into employment and into unemployment upon job loss of the primary earner. The model generally underestimates the probability of spousal transitions directly into employment independently of the primary earner's transition. However, it captures very well the difference in probabilities depending on the primary earner transition, which is the added worker effect.

In the model, as in the data, there is a much smaller added worker effect for the old. The model reproduces that there is no substantially increased likelihood of transitioning from out of the labor force directly into employment when the primary earner loses a job for the old. Furthermore, the increased probability of searching for a job by exerting costly effort is much lower than for the young, in line with the data.



(a) Young: 25-35 years

(b) Old: 55-65 years

Figure 2.8: Dynamic Response: AWE by Age in the Model

Notes: Figure 2.8 shows the change in the probability that a non-participating spouse enters the labor force (either as unemployed or as employed) this month if household head loses/lost the job in two months, next month, this month, last month, two months ago, respectively, relative to the baseline in which the household head remains employed. Figure 2.8a shows the model results for young households; Figure 2.8b shows the model results for old households. The regression producing the coefficients is Equation (2.1).

Hence, the model performs well in generating the instantaneous added worker effect over the life cycle. To analyze anticipation effects and lagged responses, Figure 2.8 replicates Equation (2.1) on model simulated output, separately by age. In line with the data, the model produces larger contemporaneous and lagged effects for the young than for the old. The lead effects are, however, of similar size across both age groups.

The model mechanisms that produce lagged responses are threefold. First, after becoming unemployed the primary earner may lose human capital which decreases potential human capital differences across spouses. Consequently, it may be optimal that both spouses search or to re-optimize on the actively searching household member. Second, unemployment benefits can expire, making employment a more desirable state. Third, households without any employed member may run down their assets to finance consumption, which increases the need to search for a new job to re-accumulate assets for precautionary reasons and for retirement.

While the model produces some anticipation effect in the two months prior to a primary earner's job loss, these lead effects are smaller than in the data. Job loss is predictable because the exogenous separation probability depends on human capital. Spouses of low human capital employed individuals may enter the labor force because a future separation is relatively likely, whereas spouses of high human capital individuals choose not to do so because the chance of an exogenous separation is low. By the law of large numbers, these separations do in fact realize at higher rates for low human capital primary earners, producing the effect that spouses are more likely to enter the labor force in anticipation of a job loss. In addition, persistence in match quality might induce non-participating spouses to enter the labor force upon a decline in match quality for the employed spouse, preparing a potential future quit if match quality remains low.

2.5.3 Counterfactuals

Finally, we use the model to construct counterfactuals and analyze which channels are important in driving the age-dependency in the added worker effect. For that purpose, we start with the added worker effect of the young and then change individual model elements towards the counterparts of old households. Table 2.11 reports the results for three such counterfactuals together with the baseline results for young households.

Table 2.11: Joint Labor Market Transitions Counterfactuals

	Primary earner transition	
	EE	EU/ES
Young (25-35):		
Cond. prob. of spousal NE transition	2.26%	3.12%
Cond. prob. of spousal NS transition	0.40%	5.28%
Cond. prob. of spousal NN transition	97.34%	91.60%
Counterfactual meeting probabilities		
Cond. prob. of spousal NE transition	2.14%	2.93%
Cond. prob. of spousal NS transition	0.41%	5.36%
Cond. prob. of spousal NN transition	97.46%	91.71%
Counterfactual human capital		
Cond. prob. of spousal NE transition	1.70%	3.02%
Cond. prob. of spousal NS transition	0.24%	3.09%
Cond. prob. of spousal NN transition	98.06%	93.89%
Counterfactual assets		
Cond. prob. of spousal NE transition	1.04%	1.73%
Cond. prob. of spousal NS transition	0.31%	3.31%
Cond. prob. of spousal NN transition	98.66%	94.96%

Notes: This table shows the counterfactual joint labor market transition probabilities.

The first counterfactual adjusts job arrival rates for young households. More specifically, we first compute the average job arrival rate for old and for young households in the model, restricting the sample to households with one member employed and one member out of the labor force. Afterwards, we adjust the individual arrival rates of each young household in our simulation by the difference between these previously computed means. This approach moves the average arrival rate of young households to that of their old counterparts, but preserves the relative distribution of arrival rates among the young. The second block of Table 2.11 shows that adjusting arrival rates has a limited impact on the added worker effect. This result arises because the average arrival rates for young and old are very similar: As most non-participating spouses are unlikely to accept a job offer, firms are only offering low arrival rates in order to satisfy their free entry condition. Nevertheless, the average arrival rate is slightly lower for older households resulting in fewer employment transitions both in the EE and in the EU case.[11]

In the second counterfactual, we adjust the human capital level of young households. Similar to above, we compute the difference in mean human capital levels across age groups separately for employed and non-participating spouses and adjust the human capital level of each young household by the difference. In our simulation, the employed spouse among older households has a higher human capital due to on average longer cumulative employment spells. In contrast, human capital levels for non-participating spouses are very similar across age groups. This is partially driven by selection (low human capital individuals are more likely to be non-participating when they have an employed spouse) and partially by fast depreciation of human capital during non-employment in order to match empirical wage losses from non-employment spells. Thus, the results of the second counterfactual can be attributed to a higher human capital level of the employed spouse during old age.

The third block of Table 2.11 shows that the increase in human capital of the employed spouse reduces transition probabilities into participation for both the

[11]This result may be partially due to the timing assumptions in the model. At the moment firms post vacancies in all the submarkets before separation shocks occur. Hence, out of the labor force spouses do not consider that their partner loses the job, translating into low acceptance probabilities and in turn low vacancy posting rates. In future work, we will investigate the robustness of the finding to different timing assumptions.

EE and the EU case, but also dampens the added worker effect. When the human capital of a separated spouse is higher, this spouse is more likely to find a new job (arrival rates are increasing in human capital) and the difference in human capital levels across spouses is potentially larger, making a switch in the prime earner position less likely.

In a third counterfactual, we adjust the asset levels of young households. This time, we adjust each young household asset level by the relative difference in average asset holding among the old and the young. Since old households have on average substantially higher asset levels we make young households richer. The fourth block of Table 2.11 shows that this reduces the incentive for a non-participating spouse to transition into participation. Hence, the added worker effect becomes smaller. Young households with asset holdings of the old are relatively rich for their age, reducing the incentive to work also in the baseline EE case, and are well insured against any labor market shock such that they do not have to rely on the added worker effect as a margin of insurance.

Taking all three counterfactuals together, we find that the substantially lower added worker effect among the old predominantly arises through higher wealth levels. Hence, older households exhibit a weaker AWE because they have better access to self-insurance through savings and are therefore less in need of other insurance margins, as opposed to a lack of opportunity to make use of the AWE.

2.6 Conclusion

In this paper, we provide evidence that the added worker effect is an important insurance margin against job loss of the primary earner for two-member households, but that the prevalence of this insurance channel strongly differs over the life cycle. When the primary earner transitions from employment to unemployment, an out of the labor force spouse is much more likely to enter the labor force in order to offset the income loss compared to when the primary earner remains employed. In particular, this spousal labor supply response is very strong for young households and becomes continuously weaker as households age.

To analyze the mechanisms that drive this age-dependency, we build a stochastic life cycle model of two-member households with a frictional labor market. We calibrate the model economy to match salient features of the US labor market. The model endogenously generates the added worker effect and its decreasing magnitude over the life cycle. Model counterfactuals reveal that the added worker effect is weaker for old than for young households mainly because older households are better insured through larger asset holdings, so that their need for spousal insurance is lower. In addition, human capital of employed spouses is higher for the old, making the spousal labor supply less valuable, though this channel is quantitatively smaller. Differences in arrival rates across age groups contribute little to the difference in the added worker effect due to a general reluctance of firms to offer jobs to non-participating workers.

Chapter 3

Distributive Effects of Banking Sector Losses

Joint with Caterina Mendicino and Marcel Peruffo

Abstract Using data from the Consumer Expenditure Survey, we document that in response to declines in bank equity returns the consumption of low-income households decreases by roughly twice as much as that of the average household. To understand this result, we develop a heterogeneous-agent model featuring rich income and portfolio heterogeneity and a banking sector subject to financial frictions. The model matches the empirically observed inequality in consumption responses following a shock to banks' asset returns. Households at the bottom of the income distribution suffer from losses in labor earnings and from an increase in the cost of borrowing. In contrast, high-income consumers can take advantage of temporarily low asset prices and high future returns and increase their savings to sustain higher consumption in the medium term. In fact, a fraction of households benefit from distress in the banking sector. A debt-financed asset purchase program can improve welfare, especially for low-income individuals, by dampening the increase in credit spreads and stabilizing investment.[1]

[1]This paper makes use of the Consumer Expenditure Survey Public Use Micro Data provided by the US Bureau of Labor Statistics.

3.1 Introduction

Which households are most exposed to severe disruptions to the financial sector? Do shocks to banks increase inequality? The severe economic distress in the wake of the 2007–9 financial crisis renewed interest in the consequences of large disruptions to banks and sparked a debate about the unequal impact of recessions. Inequality is now a critical concern for policy makers: in its 2020 strategy review, the Federal Reserve emphasizes the importance of considering the distributive consequences of economic fluctuations.[2] A comprehensive analysis of the real economic consequences of financial sector distress must therefore contemplate its heterogeneous effects across households.

Disruptions in the banking sector cause a reduction in financial intermediation, fluctuations in interest rate spreads and asset prices, and ultimately a general decline in economic activity (Gertler and Kiyotaki, 2010). Households are exposed to these factors in heterogeneous ways, depending on the composition of their income between labor earnings and financial returns, whether they are savers or borrowers, and how exposed they are to interest rate and asset price changes. A clear assessment of these heterogeneous effects is critical for understanding which households are impacted the most by banking sector disruptions, and consequently who ultimately benefits from government support to distressed financial institutions.

While the implications of severe impairments to banks' intermediation for aggregate economic outcomes are widely studied, the literature is silent about the distributive effects of banking sector losses on household consumption and welfare. Our paper fills this gap. First, we document a novel empirical fact about banking sector conditions and household consumption: Banking sector distress is associated with a stronger consumption response at the bottom of the income distribution, relative to the aggregate. Second, we build a model economy featuring rich household heterogeneity and an explicit banking sector. The model replicates the empirically observed consumption responses to banking sector losses along the income distribution. In addition, it allows us to uncover

[2] In the press conference following the release of the Fed's *Statement on Longer-Run Goals and Monetary Policy Strategy*, Jerome Powell referred to the benefits that a strong economy brings to *low- and moderate-income communities* (Powell, 2020).

the mechanisms behind those movements, to consider the welfare implications of bank losses, and to evaluate the role of policy interventions.

Our empirical analysis combines consumption data from the Consumer Expenditure Survey with the bank equity index provided by Baron et al. (2021). We estimate local projections of consumption by quintile of total after-tax income in response to changes in bank equity returns, controlling for the return to nonfinancial equities. Thus, our results capture the response to banking sector conditions over and above the impact of overall economic conditions.

We find that the decline in consumption for households in the lowest income quintile is almost twice as strong as the aggregate, while responses are roughly homogeneous over the upper half of the income distribution. On average, a one–standard deviation drop in returns is associated with a cumulative decline in consumption of 4.9 percent over the following twelve quarters. Focusing on transmission mechanisms, we find that declines in bank returns are associated with falls in investment, labor earnings, and asset prices, as well as an increase in consumer credit spreads.

To understand these findings, we construct a two-asset heterogeneous-agent model featuring a banking sector subject to financial frictions. Households face uninsurable income risk and a portfolio decision between assets with different degrees of liquidity: deposits are liquid and can be adjusted in every period, while capital holdings are subject to liquidity frictions. Banks collect deposits and lend funds both to nonfinancial firms and to households. They are subject to a leverage constraint restricting their future liabilities to a fraction of their assets. These features allow us to capture the interactions between banks and households and to explore the effects of banking sector losses on consumption and welfare along the income distribution.

We calibrate the model to US data and use it to study the effects of an unanticipated, exogenous shock to banks' asset returns. The shock causes a decline in banks' net worth of 20 percent on impact, corresponding to the fifth percentile of equity returns in the data – i.e., an episode of severe distress in the banking sector. It severely impairs banks' intermediation capacity, resulting in a fire sale of their assets to reduce the size of their balance sheet and satisfy their leverage constraint. In equilibrium, lending spreads increase and asset prices decline, generating further losses and triggering a financial accelerator (Bernanke et al., 1999). Ultimately, the reduced investment activity causes a decline in output and a recession. The responses are in line with our empirical results on potential transmission mechanisms.

Importantly, the model-implied consumption responses across income quintiles also align with our empirical findings, both qualitatively and quantitatively: While consumption of all income groups declines on impact and gradually recovers from the shock, households in the lowest income quintile experience the largest change. They see their consumption decrease by a cumulative 14 percent over twelve quarters, roughly twice as much as the average fall. Over the upper half of the income distribution, consumption responses are homogeneous, again consistent with our empirical findings.

We decompose the consumption responses into the contributions of different transmission mechanisms. Low-income households are especially exposed to fluctuations in the cost of borrowing and in labor earnings. They are often borrowers, are poorly insured against income shocks, and are highly dependent on labor income to finance their consumption. In contrast, for high-income households movements in financial income, particularly in the returns to holding capital, are the most important drivers of the observed responses. A substantial portion of the initial decline in their consumption is due to an increase in savings following temporarily low asset prices and high future returns on holding deposits and capital.

In addition, we study how banking sector losses affect consumers' welfare. On average, households would be willing to permanently give up 0.4 percent of their consumption to avoid the consequences of the shock. While those in the lowest income quintile would forgo 1 percent of their consumption to avoid the shock, those in the top quintile would only give up 0.1 percent. In fact, we find that 11 percent of the population stands to gain from the shock. These are typically high-income, wealthy households, with a high proportion of their income stemming from financial sources.[3] Despite their exposure to the

[3]A small fraction of the wealthiest households in our economy hold claims to banks' dividend payments and suffer substantially from their direct exposure to the banking sector.

initial sharp decline in asset prices, they are able to make up for their losses by adjusting their savings behavior. Overall, they take advantage of movements in financial variables, enabling them to sustain higher future consumption. This is why the heterogeneity in welfare changes is even more pronounced than that in the response of consumption.

Finally, we study the distributive consequences of a policy intervention aimed at alleviating the impact of banking sector losses. We consider an asset purchase program along the lines of the Troubled Asset Relief Program (TARP) instituted by the US government in the aftermath of the Great Recession. Similarly to Gertler and Karadi (2011), in response to losses in the banking sector the government intervenes and temporarily acts as a financial intermediary, issuing bonds to households and financing investments. Such an intervention dampens the increase in the lending spread as well as the decline in investment activity and asset prices caused by initial bank losses. Our baseline policy, calibrated to the size of TARP, is able to reduce the welfare impact of the original shock by 23 percent, with gains concentrated in the bottom quintile of the income distribution.

3.1.1 Related Literature

Our paper relates to the empirical literature studying micro-level consumption dynamics in response to macroeconomic fluctuations. Meyer and Sullivan (2013) examine the evolution of US consumption inequality during the Great Recession. Using a factor model, De Giorgi and Gambetti (2017) find consumption inequality to be procyclical. Coibion et al. (2017) and Cloyne et al. (2020) study consumption responses to monetary policy shocks. In contrast to this literature, our contribution is to examine the inequality in consumption in response to changes in banking sector conditions. In this regard, our paper is similar to Baron et al. (2021), which studies the dynamics of macroeconomic aggregates in response to banking sector distress and from which we draw our measure of conditions in the banking sector.

We also contribute to a series of contemporaneous works combining heterogeneous households and a banking sector: Arslan et al. (2020) study a house price boom and bust in a small open economy framework; Ferrante and Gornemann (2021) analyze the heterogeneous pass-through of exchange rate shocks; Fernández-Villaverde et al. (2019) show how interacting financial frictions and household heterogeneity can generate endogenous aggregate volatility; Lee et al. (2021) study how countercyclical borrowing wedges amplify business cycles. We share with them the joint consideration of financial intermediaries and household heterogeneity, but our focus lies on understanding the distributive effects of losses originating in the banking sector. Our model differs from those of the above studies in that households can hold capital both directly and indirectly (through bank deposits). This allows them to rebalance their portfolio in response to asset price movements, an important mechanism driving our results.

More generally, we build on the seminal work of Kiyotaki and Moore (1997) and Bernanke et al. (1999), as well as subsequent studies on the consequences of financial shocks (e.g., Christiano et al., 2014; Eggertsson and Krugman, 2012; Jermann and Quadrini, 2012; Justiniano et al., 2019) and frictions in the financial intermediation sector (e.g., Brunnermeier and Sannikov, 2014; Gertler and Karadi, 2011; Gertler and Kiyotaki, 2010; He and Krishnamurthy, 2019; Iacoviello, 2015; Mendicino et al., 2020) for the aggregate economy. While this line of research has not focused on the role of household heterogeneity, a parallel strand of the literature studies the implications of aggregate shocks for heterogeneous households (e.g., Glover et al., 2020; Kaplan et al., 2018; Krueger et al., 2016; Krusell and Smith, 1998). This literature either abstracts from a banking sector and financial frictions entirely or considers exogenous movements in borrowing limits or credit spreads (Antunes et al., 2020; Guerrieri and Lorenzoni, 2017). Considering both an explicit banking sector and household heterogeneity, we generate endogenous movements in credit spreads and asset prices and provide novel results on the distributional consequences of financial recessions.[4]

The remainder of the paper is structured as follows: Section 3.2 describes our empirical analysis; Section 3.3 presents the model; Section 3.4 discusses the

[4]Methodologically, we also build on heterogeneous-agent models with endogenous portfolio choices (e.g., Bayer et al., 2019; Kaplan and Violante, 2014). Our work expands on their framework in that we explicitly model a financial intermediation sector that transforms illiquid capital holdings into liquid deposits.

model's quantitative implementation; Section 3.5 presents the dynamics of the economy in response to a shock to banks' asset returns; and Section 3.6 explores the consequences of credit policy interventions.

3.2 Bank Losses and Consumption Inequality

We begin with an empirical assessment of how banking sector conditions affect consumption along the income distribution. Using household-level data from the Consumer Expenditure Survey and a measure of bank equity returns from Baron et al. (2021), we document a novel fact: households at the bottom of the income distribution exhibit a stronger consumption response to changes in bank returns.

3.2.1 Data

Household-Level Data. We use household survey data from the US Consumer Expenditure Survey (henceforth CEX). The survey is available monthly since 1980 and is based on a rotating sample of about 1,500–2,500 households selected to be representative of the US population. The CEX gathers information on household expenditures through interview and diary surveys. We focus on the former, which cover a broad set of consumption categories, while the latter only cover small but frequent purchases. Each household is interviewed once per quarter and for no more than five consecutive quarters. In each interview, separate information is collected for the previous three months. Our sample consists of the waves from 1980 to 2010. In cleaning and aggregating the micro data into expenditure categories at the household level, we follow closely Coibion et al. (2017). We define household consumption as the sum of nondurable and durable expenses and services and use the OECD equivalence scale to adjust for household composition.

In addition to data on consumption, the CEX also provides information on household income, from both labor and nonlabor sources. We define total after-tax income as the sum of labor earnings, financial and business income, and transfers less taxes, where taxes are imputed using TAXSIM. We use this information to group households into income quintiles and aggregate the expenditure data into five per capita series at the quintile level, taking monthly averages across households.[5] Finally, we transform the series to quarterly frequency by summing up expenditures for each quintile across months, and we deflate the expenditures with the All Urban CPI.

Previous research (Aguiar and Bils, 2015) has shown a mismatch of the CEX with consumption reported in national accounts. We follow Cloyne et al. (2020) in addressing this concern: First, to ensure consistency between the survey and national accounts we compute the ratio between the national statistics series and the corresponding aggregate consumption from the CEX and rescale the expenditure data for each of the five groups as well as the aggregate series with the (same) factor. With this transformation, the source of variation in aggregate consumption in our data is the national accounts, whereas the relative variation in consumption across income quintiles originates from the micro data. Second, all our empirical specifications feature income-quintile-specific time trends, which are aimed at capturing slow-moving changes in reporting within income brackets. This is again in line with the approach taken in Cloyne et al. (2020).

Bank Equity Returns. To measure conditions in the banking sector we use the index of bank equity returns provided by Baron et al. (2021). They show that bank equity declines capture early signs of banking crises in real time and predict large and persistent contractions in output and in bank credit to the private sector. Compared to other financial variables, such as credit spreads, bank equity returns are a convenient measure of banking distress since they are more sensitive to early losses.[6] This is because bank equity has the lowest payoff priority among bank stakeholders. Baron et al. (2021) also show that bank equity returns have predictive content for future macroeconomic dynamics even excluding episodes with narrative evidence of panics or widespread bank failures. In addition, the use of a continuous measure to identify periods of bank

[5] In all aggregation steps, we apply the sample weights provided by the CEX throughout.

[6] Baron et al. (2021) document that bank equity has a better signal-to-noise ratio than other financial and macroeconomic variables, in terms of identifying banking crises in real time (identified by narrative accounts). In particular, large bank equity declines tend to precede credit spread spikes across one hundred banking crises. In addition, conditional on a particular historical crisis episode, the magnitude of the peak-to-trough bank equity decline is correlated with the economic severity of the ensuing crisis.

Table 3.1: Summary Return Indices

Series	Mean	Std	Min	P25	Median	P75	Max	AC
r^B	0.0174	0.1229	-0.4666	-0.0465	0.0288	0.0943	0.2946	0.0168
r^{NF}	0.0197	0.0976	-0.2988	-0.0231	0.0347	0.0786	0.2069	0.0371

Notes: r^B: return of bank index (capital gains and dividends), r^N: return of nonfinancial corporations index (capital gains and dividends). AC: autocorrelation of series. Data series are taken from Baron et al. (2021) for the United States from 1980 to 2010.

distress instead of a narrative approach (Laeven and Valencia, 2013; Reinhart and Rogoff, 2009) allows us to focus the analysis on a single country.[7]

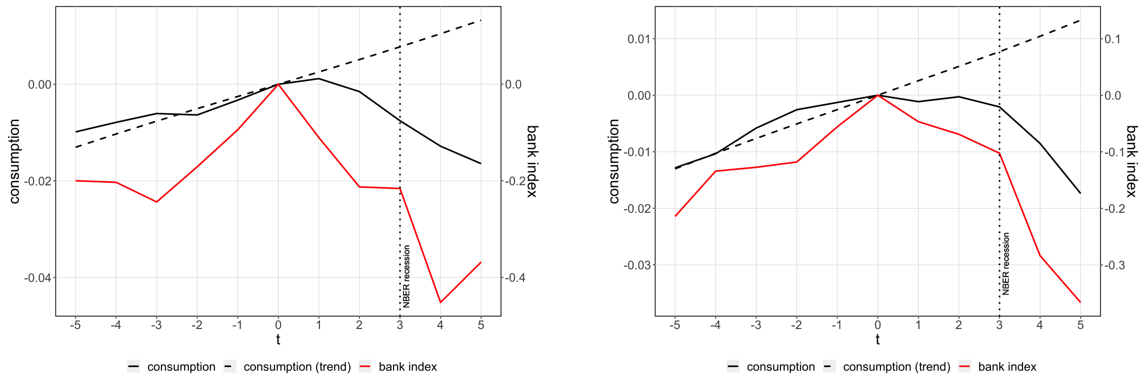
Table 3.1 shows summary statistics of returns to the US bank equity index (r^B) at quarterly frequency, as well as its counterpart for nonfinancial corporations (r^{NF}).[8] Both series feature a similar, slightly positive mean, but the banking series features more volatility, materialized in a higher standard deviation and more extreme realizations – both in the left and right tails of the return distribution. In addition, both series display very low autocorrelation, attesting to a lack of predictability based on past realizations as one would expect for financial market return series. This gives us confidence to treat sudden changes in bank equity returns as reflecting new information about the banking sector.

To provide some intuition for our data measures, Figure 3.1 shows the evolution of the US bank equity return index (red line) and log real aggregate consumption (black solid line) around two dates of bank equity crashes over our sample period.[9] Both consumption and the bank equity return index are normalized to zero in the year of the first decline in bank equity returns (t=0), and for reference we also plot the average dynamics (trend) of consumption over the entire sample. For both episodes, bank equity starts to decline well ahead of the official start of the recession date, as identified by the NBER. In the quarters before the banking sector distress, the evolution of aggregate consumption tracks

[7] Large bank equity declines line up closely with the narrative approach. However, Baron et al. (2021) show that relying on bank equity returns allows one to uncover a number of episodes of banking distress that do not appear in previous data sets. The bank equity index for the United States, which we use for our analysis, corresponds to the S&P 500 for banks and is adjusted for dividend payouts.

[8] We use the index of returns on NFC stocks as a control in our regressions, as we explain below. The latter is also obtained from Baron et al. (2021) and consists of the S&P 500 Industrials adjusted for dividends.

[9] Baron et al. (2021) define a bank equity crash as a decline in the bank equity index of more than 30 percent. Since 1980, there have been two of those in the United States – in 1990 and in 2007. The former corresponds to the Rhode Island banking crisis (Pulkkinen and Rosengren, 1993) and the latter to the global financial crisis.



(a) 1990 **(b) 2007**

Figure 3.1: Bank Equity Return Index
Notes: Dynamics of real aggregate consumption (black solid line) and bank equity return index (red solid line) around bank equity crashes in the US. Bank equity declines are defined to begin in quarter t=0. The dotted vertical line denotes the NBER recession start date. For comparison, the average consumption trend over the full sample period is presented by the dashed black line.

the average (trend) closely. After the decline in bank equity returns, however, consumption starts to fall slowly, opening a gap to trend growth even before the start of the NBER-dated recessions. With this descriptive evidence in mind, we now proceed to a formal investigation of the dynamic relation between equity returns and consumption.

3.2.2 Estimation Strategy

To examine the predictive power of bank equity returns for household consumption at different points of the income distribution, we follow Baron et al. (2021) and estimate the ensuing local projections specification in the spirit of Jordà (2005):

$$c_{i,t+h} = \alpha_i^h + \gamma_i^h(t+h) + \sum_{j=0}^{J} \beta_i^{h,j} r_{t-j}^B + \sum_{s=0}^{S} \delta_i^{h,s} r_{t-s}^{NF} + \sum_{k=0}^{K} \lambda_i^{h,k} c_{i,t-k} + \epsilon_{i,t}^h. \quad (3.1)$$

Here $c_{i,t+h}$ is the log of real household consumption by income quintile $i \in \{1,2,3,4,5\}$, $h \in \{0,1,2,...,H\}$ denotes horizons ahead of t, r_t^B and r^{NF} are returns to the bank and nonfinancial corporation indices respectively, and J, S, and K are the number of lags included for each series.[10] Our baseline specification includes one lag on each variable; i.e., $J = S = K = 1$. Coefficients

[10]Our baseline results are based on a smoothed version of $c_{i,t}$ using a four-quarter moving average as in Cloyne et al. (2020). This adjustment is meant to absorb noise inherent to the survey data. In the baseline specification we use a centered moving average. Results are also robust to the use of a forward- or backward-looking moving average and (qualitatively) to other means of seasonal adjustment such as X-13-ARIMA-SEATS.

α and γ represent a constant and a time trend, which are specific to the income quintile. The baseline specification is estimated for total household consumption.

The key parameters of interest are $\{\beta_i^{h,0}\}_{i,h}$, which characterize the sequence of local projection impulse responses of consumption to bank equity returns at time t. In line with the specification of Baron et al. (2021), we control for nonfinancial returns r_t^{NF} to adjust for the influence of contemporaneous (and lagged) general economic conditions (Stock and Watson, 2003). Hence, coefficients $\{\beta_i^{h,0}\}_{i,h}$ capture the response of household consumption over and above the response to overall conditions in the non-financial sector.

3.2.3 Results

Figure 3.2 displays the impulse-response functions for a one–standard deviation *decline* in bank returns by income quintile, as well as aggregate consumption in the bottom-right panel. The bands correspond to one–standard deviation and 95 percent confidence intervals respectively. Responses for every quintile, as well as the aggregate, are statistically significant for at least one quarter at the 95 percent level. We find that a one–standard deviation decline in quarterly bank stock returns (0.123) is associated with a cumulative fall of 5.6 log points in aggregate consumption over a three-year horizon, an economically sizable response.[11]

The main takeaway from Figure 3.2 and from our empirical analysis is that the consumption response for households at the bottom of the income distribution to changes in bank equity returns is stronger than that of other households. We are the first to document this empirical relationship. The peak response of consumption of households in the first quintile is twice as strong as that of the highest income group. Similarly, the cumulative three-year-horizon response is roughly twice as high for low-income households compared to their high-income counterparts. Figure 3.3 compares the cumulative responses over time. After twelve quarters, the bottom income quintile exhibits a cumulative response of 9 log points, while the responses for the other quintiles stay between 5.2 and 4.5 log points.

[11] Recall that the source of variation for the aggregate series comes from the national accounts and not from the CEX.

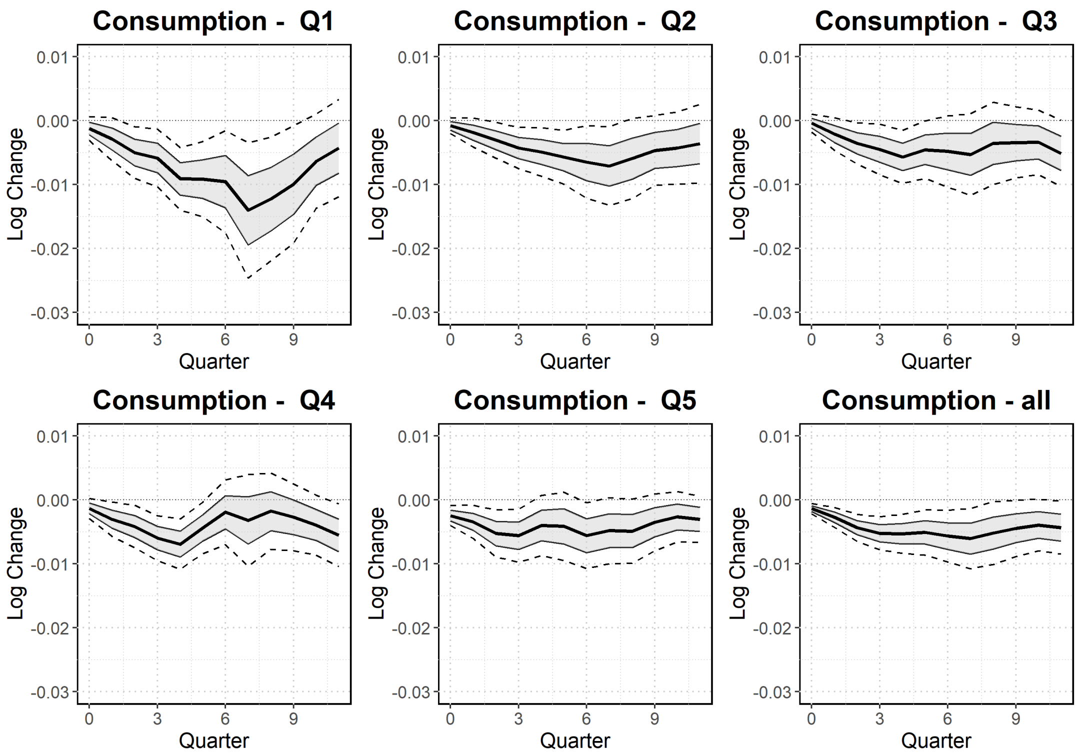


Figure 3.2: Effects of Bank Equity Returns on Household Consumption
Notes: Impulse responses of household consumption by income quintile and aggregate using data starting for 1980–2010 to a negative one–standard deviation change in r^B. The shaded areas indicate one– standard deviation confidence intervals; dashed lines represent 95 percent confidence bands. Robust, Newey-West standard errors.

Robustness Checks. We estimate a range of alternative specifications to test for robustness of our main result. These include using a monthly series, varying lag structures based on the Akaike criterion, analysis by consumption categories (durables, nondurables), splitting the sample according to housing tenure, and restricting the bank returns to below-median returns to test for nonlinearities. We provide detailed results in Appendix C.1.1. Our main finding is robust across all alternative specifications considered: consumption is more responsive to bank equity returns at the bottom of the income distribution.

Mechanisms. Figure 3.4 provides some evidence on potential transmission mechanisms following movements in bank equity returns. We repeat the same local projection as in (3.1) for the following dependent variables: total compensation of employees, the credit card rate spread, real investment, and the Dow Jones Industrials index as a proxy for asset prices. Details of the specifications and the data series are provided in Appendix C.1.2. Negative bank returns are associated with a decline in the total wage bill, investment, and the Dow Jones

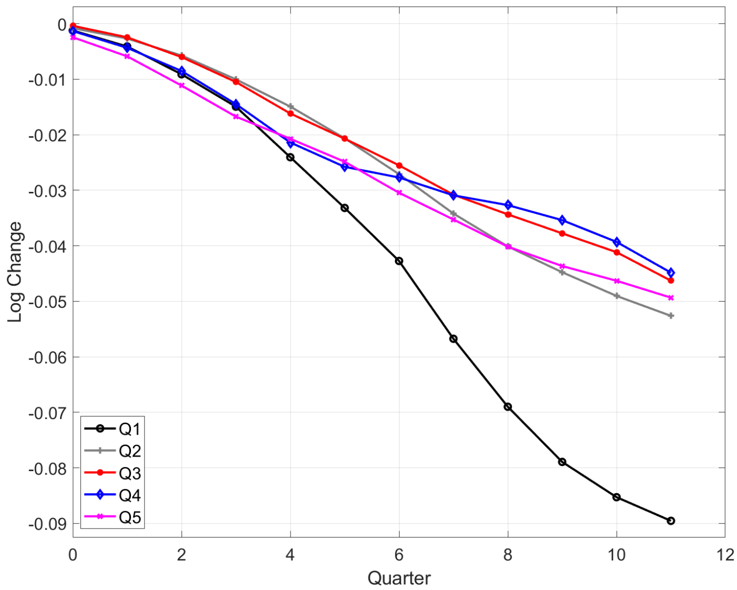


Figure 3.3: Cumulative Responses of Consumption by Quintile
Notes: Cumulative impulse responses of household consumption by income quintile, using data for 1980–2010, for a one–standard deviation decline in r^B.

Industrials index. Credit card spreads, on the other hand, rise following negative bank returns, reflecting the deterioration in credit conditions (Baron et al., 2021).

In sum, our empirical analysis provides new evidence on the dynamic relation between bank equity returns and consumption across the income distribution. We find that low-income households are more responsive to banking sector conditions, particularly in the lowest income quintile. We also provide suggestive evidence on the mechanisms operating behind these responses, and we find that banking sector distress is associated with declines in aggregate labor income, investment, and the stock market, as well as an increase in the consumer credit spread. Building on these findings, we now move on to analyzing the distributive effects of banking sector distress through our model.

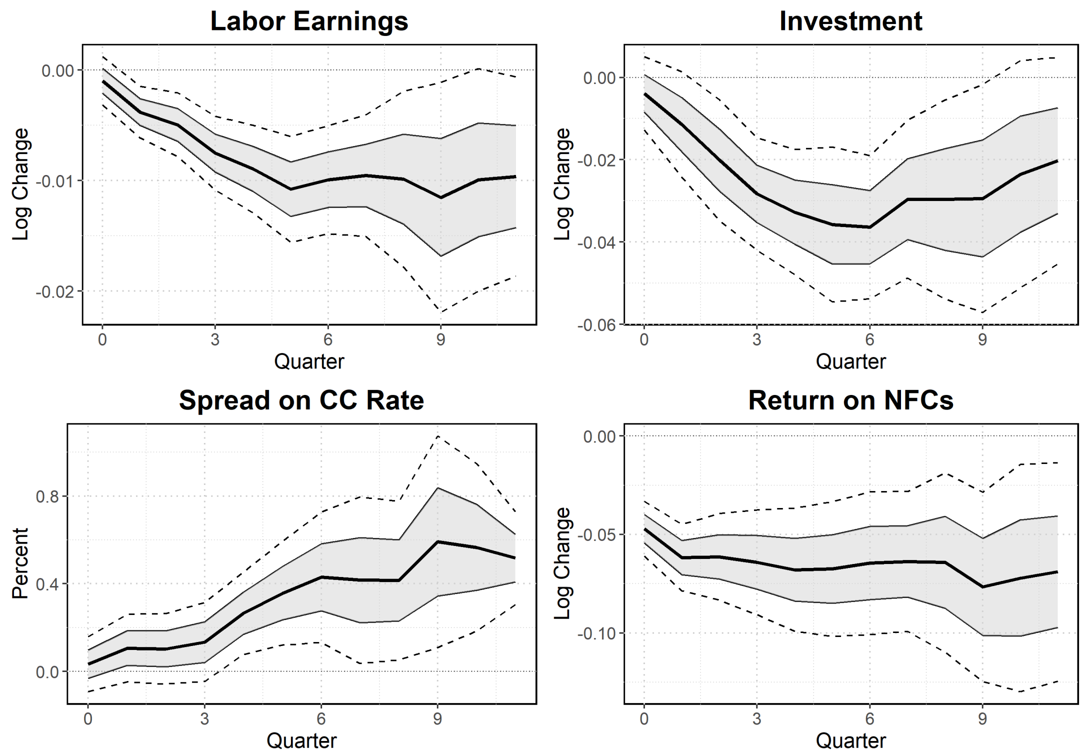


Figure 3.4: Effects of Bank Equity Returns on Selected Variables

Notes: Impulse responses of total employment compensation, investment, the spread on credit card rates, and the Dow Jones Industrials index for a one–standard deviation decline in r^B. Details of the data series are provided in Appendix C.1.2. The shaded areas indicate one–standard deviation confidence intervals; dashed lines represent 95 percent confidence bands. Robust, Newey-West standard errors.

3.3 Model

To analyze the distributive effects of banking sector losses in more detail, we build a model economy featuring both household heterogeneity and an explicit banking sector. The model enables us to go beyond the empirical exercise: we consider how the observed heterogeneity in consumption responses translates into changes in welfare, study the relative contribution of different transmission mechanisms, and evaluate policy responses to banking sector losses.

The model economy features five types of agents: competitive production firms produce intermediate consumption goods, which are differentiated into final goods by monopolistically competitive retailers; competitive capital producers transform consumption goods into capital goods; a continuum of ex ante identical households facing idiosyncratic income risk can save or borrow through a liquid asset intermediated by banks and can also invest directly in illiquid capital; finally, banks collect deposits from and lend to households, invest directly in capital, and are subject to a leverage constraint. We outline the problems solved by each type of agent in detail below.

3.3.1 Production

Intermediate Goods Producers. A continuum of identical production firms combine capital input K and labor input N to produce intermediate goods using production technology

$$Y_t = A_t K_t^{\alpha} N_t^{1-\alpha}, \tag{3.2}$$

where A_t represents total factor productivity.

Production firms sell the intermediate consumption good at price p_t^I to retailers. Assuming competitive input and output markets, profit maximization of production firms yields factor prices as

$$w_t = p_t^I (1-\alpha) A_t K_t^{\alpha} N_t^{-\alpha} \tag{3.3}$$

$$r_t^K = p_t^I \alpha A_t K_t^{\alpha-1} N_t^{1-\alpha}. \tag{3.4}$$

Retailers. Monopolistically competitive retailers differentiate the intermediate consumption good into varieties of final goods. Final goods are combined into households' consumption baskets with a standard CES aggregator such that $c_t = \left[\int_j c_{jt}^{R\,\frac{1}{\mu}} dj\right]^{\mu}$, where $\mu > 1$. The demand for each variety is given as

$$c_{j,t}^R = \left(\frac{p_{j,t}}{P_t}\right)^{\frac{\mu}{1-\mu}} c_t. \tag{3.5}$$

Normalizing the price of a unit of the consumption bundle c_t to $P_t = 1$ and imposing a symmetric equilibrium, the profit maximization problem of retailers yields the price for the intermediate good as

$$p_t^I = \frac{1}{\mu}. \tag{3.6}$$

Retailers' profits are distributed to households as dividends given by

$$div_t^Y = \frac{\mu - 1}{\mu} Y_t. \tag{3.7}$$

Capital Producers. A continuum of identical, competitive capital producers transform the final consumption good into the next period's capital, which they sell to households and banks at price q. They live for one period and are subject to adjustment costs relative to the stock of capital in steady state K^{SS} and choose investment I to maximize profits

$$\max_{I_t}\left\{(q_t-1)I_t-\frac{\phi_K}{2}\left(\frac{I_t}{K^{SS}}-\delta\right)^2 K^{SS}\right\}. \tag{3.8}$$

The resulting first-order optimality condition yields the price of capital as

$$q_t=1+\phi_K\left(\frac{I_t}{K^{SS}}-\delta\right). \tag{3.9}$$

This pricing equation highlights how adjustment costs to the aggregate capital stock are important to generate fluctuations in the price of capital. Capital producers' optimality implies a steady-state value of $q=1$, while $q>1$ whenever investment is above its steady-state level ($I>\delta K^{SS}$) and $q<1$ whenever investment is below its long-run level ($I<\delta K^{SS}$). The profits from capital production given by equation (3.8) are distributed to households as dividends div_t^I.

3.3.2 Banking Sector

Banks are run by managers, which are assumed to be of zero mass and whose discount factor is β_B. Banks fund their investments through short-run deposits D, along with their own net worth E. They hold two types of assets: claims to nonfinancial capital K^B, and consumer loans L. Managers maximize the following objective function:

$$V_t^B(E_t)=\max_{\substack{K_{t+1}^B\geq 0,\, L_{t+1}\geq 0\\ D_{t+1}\geq 0,\, div_t^B\geq 0}} \log(div_t^B)+\beta_B\mathbb{E}_tV_{t+1}^B(E_{t+1}), \tag{3.10}$$

subject to

$$E_t=\underbrace{(1+r_t^L)L_t}_{\text{repayment from borrowing HHs}}+\underbrace{((1-\delta)q_t+\xi_t^B r_t^K)K_t^B}_{\text{repayments from NFCs}}-\underbrace{(1+r_t^D)D_t}_{\text{repaying depositors}} \tag{3.11}$$

$$div_t^B+L_{t+1}+q_tK_{t+1}^B=D_{t+1}+E_t \tag{3.12}$$

$$\underbrace{(1+r^D_{t+1})D_{t+1}}_{\text{future liabilities}} \leq \chi \underbrace{\mathbb{E}_t\Big((1+r^L_{t+1})L_{t+1} + ((1-\delta)q_{t+1} + \xi^B_{t+1} r^K_{t+1})K^B_{t+1}\Big)}_{\text{expected future assets}}. \tag{3.13}$$

Equation (3.11) is the law of motion for banks' beginning-of-period equity E_t. The shock ξ^B_t is a disturbance to the productive capacity of banks' capital holdings, similar to the capital quality shock in Gertler and Kiyotaki (2010) but restricted to the capital intermediated by banks. We take this as a reduced-form way to generate losses in the banking sector, and we assume $\xi^B_{SS} = 1$. In the context of the model, this shock can be interpreted as an (unexpected) realization of lower returns on bank equity, triggering a recession.[12]

Equation (3.12) represents banks' flow of funds identity, with assets (and dividends) on the left-hand side and liabilities on the right-hand side. Finally, equation (3.13) imposes a leverage constraint, restricting future bank liabilities to a fraction of the expected value of future assets.[13]

Bankers' optimal behavior implies a no-arbitrage condition between lending to households and holding capital given by

$$\mathbb{E}_t\left(\frac{(1-\delta)q_{t+1} + \xi^B_{t+1} r^K_{t+1}}{q_t}\right) = 1 + r^L_{t+1}. \tag{3.14}$$

Note here that the bank forms expectations about the future return to capital, while the return on lending to households (as well as the interest paid on deposits) is predetermined. In addition, the leverage constraint creates a wedge between deposit and lending rates:

$$r^L_{t+1} - r^D_{t+1} = \frac{1}{\chi\gamma_{t+1} + \frac{\mathbb{E}_t div^B_{t+1}}{\beta^B div^B_t}} - \frac{1}{\gamma_{t+1} + \frac{\mathbb{E}_t div^B_{t+1}}{\beta^B div^B_t}} > 0. \tag{3.15}$$

This wedge is positive as long as the leverage constraint is binding, and thus the associated multiplier γ is positive.

[12] In the appendix to their paper, Baron et al. (2021) provide a brief description of the banking crises identified in their data set, with references to detailed accounts. Common causes are exposure to (ex post) troubled sectors, either domestically or internationally. Our shock thus can be interpreted as exposure to a particular sector whose assets turned out to produce returns below expected.

[13] Our setup for the banking sector follows closely Iacoviello (2015).

Since bank managers are assumed to be of zero mass, the payments they receive require zero resources and will not affect the resource constraint of the economy. Dividends from banking activities are distributed in full to households.

3.3.3 Households

The demand side of the economy is modeled similarly to Bayer et al. (2019). Households are ex ante identical but ex post heterogeneous due to idiosyncratic shocks to their labor productivity z. They can save (deposit) or borrow in a liquid asset a and invest directly in capital k. Investment in capital is subject to a stochastic illiquidity: in any given period, the utility cost of adjusting θ_t is determined by an i.i.d. draw from a logistic distribution with mean μ_θ and variance σ^2_θ. Households in productivity state $z = z^*$, which we refer to as *capitalists*, receive additional income in the form of dividends $div_t = \frac{div_t^Y + div_t^I + div_t^B}{\sum_{(a,k)} \lambda_t(a,k,z^*)}$.[14] Here, λ denotes the distribution of households across the idiosyncratic state space at the beginning of each period and hence the term $\sum_{(a,k)} \lambda_t(a,k,z_*)$ summarizes the mass of capitalist households. Throughout the paper, we refer to noncapitalist households as *workers*.

At the beginning of a period, households are aware of their current portfolio position and learn about the realization of their idiosyncratic productivity state z, as well as their current cost of adjusting the illiquid portfolio. They first decide on whether to adjust their capital holdings in this period (extensive margin), and in a second stage they decide jointly on borrowing/saving in the liquid asset a, investing in capital k (intensive margin, if they chose to adjust), and consuming.

A non-adjusting household does not incur the utility cost θ but must keep capital holdings constant at $k_{t+1} = k_t$. It solves the dynamic optimization problem given by

$$V_t^n(a_t,k_t,z_t) = \max_{c_t \geq 0, a_{t+1} \geq \underline{a}} \left\{ u(c_t) + \beta \mathbb{E}_t V_{t+1}(a_{t+1},k_t,z_{t+1}) \right\} \tag{3.16}$$

$$\text{s.t. } c_t + a_{t+1} \leq (1 + r_t^{HH}(a_t))a_t + (r_t^K - \delta q_t)k_t + w_t z_t + \mathbb{I}_{z_t = z^*} div_t,$$

[14] As in Bayer et al. (2019), households can transition into and out of the capitalist state. We detail this process in Section 3.4, when we describe the model's quantitative implementation.

with $\underline{a}$ as the (exogenous) borrowing limit. The return on the liquid asset $r_t^{HH}(a_t)$ depends on whether the household holds deposits ($a_t \geq 0$) or is a borrower ($a_t < 0$):

$$1 + r_t^{HH}(a_t) = \begin{cases} 1 + r_t^D & \text{if } a_t \geq 0 \\ 1 + r_t^L + \tau & \text{if } a_t < 0 \end{cases} \tag{3.17}$$

Here $\tau > 0$ is a proportional transaction cost of issuing a loan, which is treated as a deadweight loss. The return to capital less the replacement cost of depreciation is credited to households' liquid account; i.e., the liquidity friction only applies to households' stock of capital. Value functions are indexed by t as they depend on prices, which might fluctuate over time.

If households instead chose to incur the utility costs of adjusting, they can select any positive value of k_{t+1}. With all notation as above, their problem is given by

$$V_t^a(a_t, k_t, z_t) = \max_{c_t \geq 0, a_{t+1} \geq \underline{a}, k_{t+1} \geq 0} \left\{ u(c_t) + \beta \mathbb{E}_t V_{t+1}(a_{t+1}, k_{t+1}, z_{t+1}) \right\} \tag{3.18}$$

$$\text{s.t. } c_t + a_{t+1} + q_t k_{t+1} \leq (1 + r_t^{HH}(a_t)) a_t + ((1-\delta) q_t + r_t^K) k_t + w_t z_t + \mathbb{I}_{z_t = z^*} div_t.$$

The value function of a household after the revelation of its current labor productivity z_t and portfolio adjustment cost draw θ_t is given by

$$V_t(a_t, k_t, z_t, \theta_t) = \max\{V_t^a(a_t, k_t, z_t) - \theta_t, V_t^n(a_t, k_t, z_t)\}. \tag{3.19}$$

Here the max operator summarizes households' decision of whether or not to adjust their portfolios. Before the current draw for adjustment costs is revealed, the probability of adjusting conditional on state (a, k, z) is hence given by

$$F_\theta \left(V_t^a(a_t, k_t, z_t) - V_t^n(a_t, k_t, z_t) \right),$$

where F_θ is the CDF of the logistic distribution.

The framework with capital holdings subject to illiquidity frictions at the household level provides an explicit microfoundation for why households are willing to hold capital indirectly through banks. The adjustment friction paired with idiosyncratic income risk makes the liquidity provided by holding deposits valuable to households. Contrary to models featuring banks and representative households, there is no need to abstract from households' ability to invest in capital directly in order to allow for a wedge between deposit rates and the return on capital. In fact, households in our model economy – as in the data – hold both deposits and capital simultaneously.

3.3.4 Market Clearing

Market clearing requires that the quantities chosen by bankers align with households' choices of the liquid asset such that

$$L_{t+1} = \sum_{(a_t,k_t,z_t)} \mathbb{I}_{a_{t+1}(a_t,k_t,z_t)<0} \left(-a_{t+1}(a_t,k_t,z_t)\right) \lambda_t(a_t,k_t,z_t) \tag{3.20}$$

$$D_{t+1} = \sum_{(a_t,k_t,z_t)} \mathbb{I}_{a_{t+1}(a_t,k_t,z_t)\geq 0}\, a_{t+1}(a_t,k_t,z_t)\lambda_t(a_t,k_t,z_t). \tag{3.21}$$

In addition, aggregate capital holdings of households are given by

$$K_{t+1}^{HH} = \sum_{(a_t,k_t,z_t)} k_{t+1}(a_t,k_t,z_t)\lambda_t(a_t,k_t,z_t). \tag{3.22}$$

Total efficiency units of capital demanded have to equal total capital supplied such that

$$K_t = \xi_t^B K_t^B + K_t^{HH}, \tag{3.23}$$

where capital supplied by bankers is adjusted for the capital productivity shock ξ_t^B. Additionally, the law of motion for total capital in the economy has to be consistent with the investment choices of capital-producing firms,

$$K_{t+1}^{HH} + K_{t+1}^B = I_t + (1-\delta)(K_t^{HH} + K_t^B). \tag{3.24}$$

Market clearing in the goods market requires

$$C_t + I_t + \Xi_t = Y_t, \tag{3.25}$$

where Ξ_t consists of a series of deadweight losses from the cost of capital adjustment and loan issuance given by

$$\Xi_t = \frac{\phi_K}{2}\left(\frac{I_t}{K_{ss}} - \delta\right)^2 K_{ss} + \tau L_t. \tag{3.26}$$

Finally, as households inelastically supply z_t effective labor units, labor market clearing is given by

$$N_t = \sum_{(a_t,k_t,z_t)} z_t \lambda_t(a_t,k_t,z_t). \tag{3.27}$$

We define an equilibrium in the economy formally in Appendix C.2.

3.4 Quantitative Implementation

In this section, we outline our quantitative implementation of the model. We start by describing the solution method, and we then discuss the calibration strategy and quantitative fit of the model.

3.4.1 Solution Method

The main exercise in this paper simulates a one-time unexpected ("MIT") shock, followed by a transition back to steady state. Thus our equilibrium consists of a perfect-foresight transition path for all aggregate variables, households' policies, and the distribution of households across the state space. The solution method requires first solving for a steady-state equilibrium and then computing the transitional dynamics following the shock.

Finding the stationary equilibrium entails (i) solving the households' problem and (ii) satisfying equilibrium conditions under the assumption of stationarity. We solve the households' problem by implementing a version of the algorithm described in Hintermaier and Koeniger (2010). This methodology involves combining the endogenous grid method of Carroll (2006) with a no-arbitrage condition between the marginal values of holding deposits and capital.[15] The latter determines households' portfolio choice. We use the implied policy

[15]This method requires concavity of the value function, which is not generally guaranteed in a model with an extensive margin of portfolio adjustment, especially for low values of σ_θ. We test our solutions for concavity and find it to be preserved both in the steady state and along all transition paths for our calibration.

functions to compute aggregates. To compute the distribution across households we proceed as in Young (2010) and use linear interpolation whenever the policy values do not coincide with grid points – which happens almost surely, with the exception of boundaries and the kink in the return of liquid assets at $a = 0$.

Beyond the market clearing conditions (equations (3.20)–(3.25)), computing the steady state involves satisfying both the banker's leverage constraint and consistency in the implied dividends. We iterate on r^D and on div^B using a quasi-Newton method, extracting the remaining equilibrium prices from firms' and bankers' optimality conditions, until a fixed point is achieved.

As our setup for the banking sector features the standard financial accelerator, we solve for transitional dynamics of the economy exactly to account for non-linearities in response to aggregate shocks.[16] We begin by selecting a horizon T, after which we assume the economy has returned to its steady state. We set $T = 1000$. We then guess a path of endogenous variables, compute the deviations from the equilibrium conditions at each $t = \{1, 2, ..., 1000\}$, and iterate on the endogenous variables until all equilibrium conditions are satisfied. We obtain an update for the path of endogenous variables again through a quasi-Newton method, where we compute the required Jacobian of equilibrium conditions – including non-analytical aggregates from heterogeneous households – following the methodology of Auclert et al. (2021).

3.4.2 Calibration

We assume a model period corresponds to one quarter. For the calibration we proceed in two steps: First, we set a range of parameters to values commonly used in the literature. We assume CRRA utility such that $u(c) = \frac{c^{1-\sigma}}{1-\sigma}$, and we set $\sigma = 2$. Furthermore, we set the capital share to $\alpha = 0.33$ and the capital adjustment cost to $\phi_K = 40$, in line with the elasticity of investment with respect to the price of capital reported in Gertler and Karadi (2011). Similarly to Kaplan et al. (2018), we set households' borrowing limit $\underline{a}$ to average quarterly earnings, which we normalize to 1 by scaling households' labor productivity process.

[16]For instance, a shock with 50 percent of the magnitude of our baseline shock produces a 60 percent lower initial decline in bank equity.

Earnings Process. The process for idiosyncratic income is split into two components: The first is a process for *workers'* idiosyncratic labor productivity z. This process is crucial in determining households' incentives to hold each type of asset. Households subject to high earnings risk tend to hold a relatively larger portion of liquid assets in their portfolio to insure against the risk of negative income realizations. To capture this important channel and match the rich earnings dynamics present in the data as precisely as possible, we assume that labor productivity follows an AR(1) process with innovations consisting of a mixture of normal distributions, given by

$$\log(z_t) = \rho \log(z_{t-1}) + \varepsilon_t,$$

with

$$\varepsilon_t \sim \begin{cases} \mathcal{N}(\mu_1, \sigma_1^2) \text{ with probability } p \\ \mathcal{N}(\mu_2, \sigma_2^2) \text{ with probability } 1-p. \end{cases}$$

The earnings process introduces six parameters, $\{\rho, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2, p\}$. We calibrate these via simulated method of moments, targeting moments of the earnings distribution. Specifically, we target (i) the cross-sectional variance of log annual earnings, (ii) the standard deviation, (iii) the skewness and (iv) kurtosis of log annual earnings changes, and the (v) ratio of the 90th to the 10th percentile of log changes. Furthermore, we normalize $\mu_2 = -\frac{p}{1-p}\mu_1$.

Our baseline calibration does not feature a system of tax and transfers, and thus we target after-tax, household-level earnings. We obtain the values for our five targets from De Nardi et al. (2020). The moments are computed from the PSID waves for 1962 to 1992, restricting attention to households whose head is aged between twenty-five and sixty.[17] Household-level earnings are adjusted by year fixed effects, as well as family size.[18]

The model-implied moments are obtained by simulating the evolution of quarterly earnings for a panel of workers and aggregating them to annual frequency.

[17]The PSID provides annual data only up until 1992 and was adjusted to lower frequency afterward.

[18]See De Nardi et al. (2020), Section 2, for full details. We thank Gonzalo Paz-Pardo for kindly making the specific target values available to us.

Table 3.2: Calibration—Earnings Process

Target	Model	Data
Cross-Sectional Variance	0.57	0.57
Standard Deviation of Changes	0.33	0.33
Skewness of Changes	-0.99	-0.98
Kurtosis of Changes	10.5	10.3
P90-P10 of Changes	0.65	0.64

Notes: Data moments are computed with annual log earnings using the PSID waves from 1962 to 1992, restricted to households whose head is of age twenty-five to sixty. Associated parameter values are $\rho = 0.963$, $\sigma_1 = 0.50$, $\sigma_2 = 0.01$, $p = 0.156$, $\mu_1 = -0.105$, and $\mu_2 = 0.019$.

We are able to match all five targets precisely with implied parameter values $\rho = 0.963$, $\sigma_1 = 0.50$, $\sigma_2 = 0.01$, $p = 0.156$, $\mu_1 = -0.105$, and $\mu_2 = 0.019$. We discretize the workers' labor productivity on a grid with eleven earnings states, using the algorithm introduced in Farmer and Toda (2017). Table 3.2 summarizes the results of the earnings process calibration.

For the second component of idiosyncratic income, we assume the existence of a *capitalist* state at the top of the discretized labor productivity process. Households under this category are the claimants to all dividends in the economy.[19] In every period, there is a probability ν^i that a worker in the highest-productivity state will become a capitalist, which we assume to account for 1 percent of the population. With probability $\nu^o = 0.0625$ they transition back into the highest-productivity worker state, corresponding to the probability of falling out of the top 1 percent of the income distribution found in Guvenen et al. (2021a). The discretized Markov process for idiosyncratic labor productivity together with parameter ν^o and the assumption that capitalists correspond to 1 percent of households implies $\nu^i = 0.025$. Finally, we set labor productivity in the capitalist state to the median labor productivity in the economy.

Internally Calibrated Parameters. In a second step, the remaining parameters (δ, β, τ, μ, β_B, χ, $\mu_\theta, \sigma_\theta$) are calibrated internally. We target an annual $\frac{K}{Y}$ ratio of 3 based on data from Penn World Tables. The steady-state interest rate on deposits r^D is calibrated to an annualized three-month Treasury bill rate of

[19]Castaneda et al. (2003) were the first to introduce a top earner state to account for US income and wealth inequality. Distributing dividends at the top of the income distribution is in line with Bayer et al. (2019), whose calibration strategy we follow.

2 percent, and the wedge between deposits and lending rates is calibrated to $r^L - r^D = 2$ percent annually, in line with the results of Philippon (2015) on the returns to intermediation. We target an (annual) $\frac{L}{Y}$ ratio of 3 percent, as in Kaplan et al. (2018), as well as a $\frac{D}{Y}$ ratio of 0.4 and $\frac{K^B}{Y}$ ratio of 0.6 to match data on deposit-taking institutions' balance sheets from the Federal Reserve Board's data table H.8 for 2004. In addition, we target a Gini coefficient for net wealth of 0.8 from the 2004 wave of the Survey of Consumer Finances (SCF).[20]

Even though the internal calibration procedure identifies all parameters jointly, each one is more closely related to some of the targets. The depreciation rate is immediately pinned down from the intermediate producer's capital demand in combination with bankers' arbitrage conditions, given our targets for capital-to-output ratio and r^L. The household discount factor β regulates the overall desire to save and thus is identified by the deposit-to-output ratio, given a target for r^D. The parameter μ regulates the relative share of profits in the economy. A higher μ increases the dividend income of capitalist households and by consequence their equilibrium wealth as well as the degree of wealth inequality in the economy. The patience of bank managers affects their required equilibrium return on equity. The latter is determined by the lending spread, thus identifying β_B. The parameter χ is selected to ensure that the banker's leverage constraint (3.13) holds with equality, given our targets for deposits, consumer loans, banker's capital, and interest rates. The parameter τ affects the cost of consumer credit and thus is identified by total borrowing in the economy. The parameter μ_θ regulates the cost of adjusting capital holdings, which ultimately determines total demand for capital by households, thus strongly affecting $\frac{K^{HH}}{K} = 1 - \frac{K^B}{K}$. We are left with the parameter σ_θ, which regulates the dispersion in households' probability of adjusting their capital holdings. Since empirical evidence on this moment is scarce, we set $\sigma_\theta = 4$ but ensure that our results are not driven by this choice by repeating our main counterfactual with different values of σ_θ.[21] We find reasonable variations on this parameter to be inconsequential for our results.

[20] The Gini coefficient for net worth is computed based on households with positive net worth both in the data and in the model.

[21] Bayer et al. (2019) use a value of $\sigma_\theta = 22,500$, achieved by targeting the second quintile of portfolio liquidity. In practice, we find that σ_θ has little influence over that moment in the model, which motivates our decision to set it exogenously.

Table 3.3: Summary of Calibration Procedure

Target	Model	Data	Closest Parameter	Source
$\frac{K}{Y}$ Ratio	3	3	$\delta = 0.016$	Penn World Tables
Deposit-to-Output $\frac{D}{Y}$	0.40	0.40	$\chi = 0.6318$	Fed H.8 2004
(Liquid) Debt-to-Output $\frac{L}{Y}$	3%	3%	$\tau = 1.23\%$	Fed H.8 2004
Bank Investment-to-Output $\frac{K^B}{Y}$	0.60	0.60	$\mu_\theta = 5.453$	Fed H.8 2004
Annual r^D	2%	2%	$\beta = 0.9676$	Annualized 3M Tbill rate
Annual Spread $(r^L - r^D)$	2%	2%	$\beta_B = 0.9816$	Philippon (2015)
Net-Worth Gini	0.80	0.80	$\mu = 1.122$	SCF 2004
Risk Aversion			$\sigma = 2$	see text
Capital Share			$\alpha = 0.33$	see text
K Adjustment Cost			$\phi_K = 40$	Gertler and Karadi (2011)
Borrowing Limit			$\underline{a} = -1$	Kaplan et al. (2018)
P(Entering Star Earner)			$\nu^i = 0.025$	1% of households are capitalists
P(Quitting Star Earner)			$\nu^o = 0.0625$	Guvenen et al. (2021a), Bayer et al. (2019)
Dispersion of Adjustment Cost			$\sigma_\theta = 4$	see text

Notes: The first block of parameters is calibrated internally by matching the reported data targets. The second block of parameters is set externally. See text for explanations.

The data moments and their model counterparts, as well as the complete set of parameter values, are reported in Table 3.3.

Model Validation. Table 3.4 compares untargeted distributional statistics in the model with their data counterparts. All wealth data are from the 2004 wave of the Survey of Consumer Finances, while income data are obtained from the Congressional Budget Office. We define liquid wealth as the sum of checking, savings, and money market accounts net of credit card debt. We then compute illiquid assets residually by subtracting liquid assets from net worth.[22] Income is defined as total after-tax household income, including labor earnings and business and financial income. The first two sets of columns refer respectively to the quintile shares of the distribution of liquid assets and total net worth, and the last two columns report the distribution of income. Recall that the only moments of the wealth distribution that we target in the calibration are the Gini coefficient of net worth as well as the aggregate amount of debt, deposits, and capital held by households, while for income we only target moments of the distribution of labor earnings (growth).

The calibration does a very good job in matching not only the distribution of overall net worth, but also the quintile shares of the distribution of liquid asset holdings. In addition, it matches almost exactly the bottom-quintile share of

[22] Consistent with our definition of deposits, we do not include bonds and stocks as liquid assets. In computing the moments in the data we only keep households whose head is aged between twenty-five and sixty-five.

Table 3.4: Moments of the Wealth Distribution—Model vs. Data

	Liquid		**Net Worth**		**Total Income**	
	Model	Data	Model	Data	Model	Data
	(1)	(2)	(3)	(4)	(5)	(6)
Q1	-7.9	-7.7	-0.8	-0.2	4.0	7.0
Q2	0.1	0.0	1.6	1.1	8.9	10.5
Q3	4.0	1.4	5.3	4.2	13.8	14.9
Q4	11.9	7.9	11.9	11.5	20.1	20.8
Q5	91.9	98.5	82.1	83.3	53.11	47.7

Notes: Data for columns 1–4 come from the 2004 wave of the Survey of Consumer Finances. Data for columns 5–6 come from the Congressional Budget Office, (The Distribution of Household Income, publication no. 56575). Quintile shares are for 2004.

liquid assets, as well as the share of households with negative liquid holdings – 25.5 percent in the model versus 25.2 percent in the data (not reported in Table 3.4). Matching these two moments is important in capturing households' exposure to changes in lending rates. The model can also match the substantial degree of concentration in liquid assets (columns 1–2) and each of the five shares of the distribution of net worth (columns 3–4). Finally, columns 5–6 show that it also does well in capturing the distribution of total after-tax household income (including both labor earnings and financial returns).

To evaluate the joint distribution of liquid and illiquid assets, Figure 3.5 plots the average portfolio composition for distinct quintiles of the distribution of net worth. We are able to capture the general pattern of portfolio composition in the data, especially for the bottom quintile. Low-net-worth individuals hold a lower share of their savings in the form of illiquid assets. Yet we understate the average share of illiquid assets. This is because our calibration target for aggregate deposits – the liquid asset in our economy – is obtained from banks' balance sheets, instead of households'.[23]

[23] Our choice is conservative for the analysis we conduct, as restricting the supply of liquid assets further would mean that households in general would be less able to insure against shocks, which would increase the (welfare) consequences of bank losses, especially at the bottom of the income distribution.

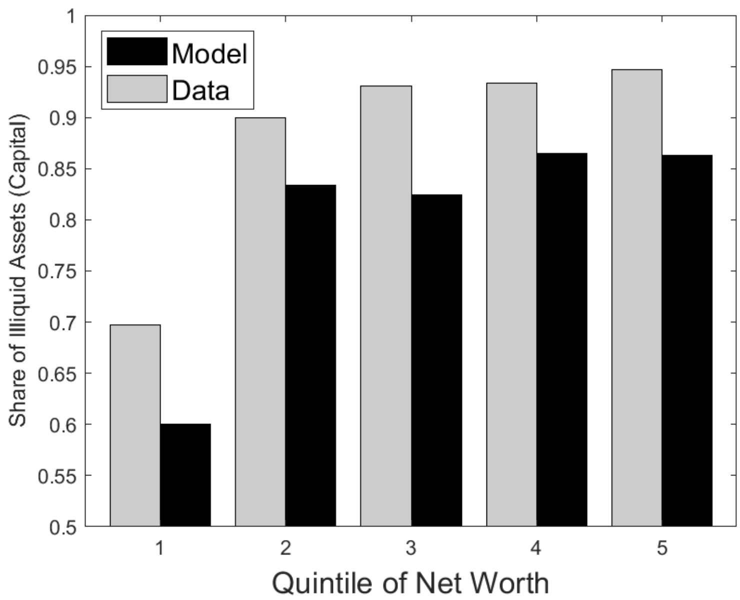


Figure 3.5: Portfolio Composition by Quintile of Net Worth
Notes: Data come from the Survey of Consumer Finances in 2004 and authors' own calculations. Both model and data samples are restricted to households with strictly positive net worth. Furthermore, the model sample is restricted to households with non-negative liquid assets, and (net) liquid assets in the data correspond to the sum of checking, savings, and money market accounts net of credit card debt. Illiquid assets are obtained by subtracting liquid assets from total net worth.

3.5 Quantitative Results

To study the distributive consequences of losses in the banking sector, we simulate the response of the economy to a one-time, unexpected ("MIT") shock to bankers' capital productivity ξ_t^B, reverting back to its steady state value of 1 at rate ρ_ξ. Specifically, we assume

$$\xi_t^B = \begin{cases} \epsilon \text{ if } t = 1 \\ (1-\rho_B) + \rho_\xi \xi_{t-1}^B \text{ if } t > 1. \end{cases}$$

We calibrate ϵ and ρ_ξ to jointly generate an initial decline in bank equity corresponding to the fifth percentile of empirical bank equity returns and the twelve-quarter cumulative consumption response to a shock of that magnitude.[24] This corresponds to a roughly 20 percent decline in initial bank equity and a cumulative decline of 8.6 percent in aggregate consumption. The implied parameter values are $\epsilon = 0.5$ and $\rho_\xi = 0.72$.

[24] We rescale the impulse response reported in the bottom-right panel of Figure 3.2 by a factor of 1.74, as the fifth percentile of bank returns (r^B) corresponds to 1.74 standard deviations.

Note that only the banking sector is directly exposed to this shock. Its impact on households works entirely through the general-equilibrium responses of market prices, interest rates, and dividends. Our analysis thus isolates the distributive effects of banking sector losses. In complex advanced economies, households might be directly exposed to the same sources of disturbances as the banking sector, with reinforcing or mitigating effects in addition to those highlighted below. We abstract from this direct exposure to focus on the bank loss channel.

3.5.1 Aggregate Responses

We begin by reporting the dynamics of macroeconomic aggregates. Figure 3.6 reports responses of the components of banks' balance sheet. On impact, the shock causes a surprise loss to bankers' beginning-of-period net worth. In response, banks have to reduce the size of their balance sheet and increase the cost of borrowing r^L sharply while reducing the interest paid on deposits r^D, causing a decline both in deposits D and in banks' claims on productive capital K^B (movements in prices are shown in Figure 3.8). Despite an increase in the cost of borrowing, household loans L increase, driven by households' desire to smooth consumption over time.

Figure 3.7 reports the dynamics of aggregates in the real economy. As banks are forced to reduce their balance sheet, investment falls in response to the shock and in consequence so does the aggregate capital stock in the economy. The decline in the demand for investment leads to a sharp drop in the price of capital, as seen in Figure 3.8. Investment falls by less than the capital held by the banking sector, as households' aggregate capital holdings increase in response to capital's lower price and ensuing high returns going forward. Since the value of banks' assets depends on the price of capital, a decline in q further constrains banks' intermediation capacity, amplifying the decline in investment and the increase in spreads.[25] Finally, aggregate output declines, both because the shock leads to a fall in the effective units of capital available for production and because of the reduction in investment activity, and aggregate consumption falls, albeit by less than output and investment.

[25] This is the standard financial accelerator mechanism (Gertler and Kiyotaki, 2010).

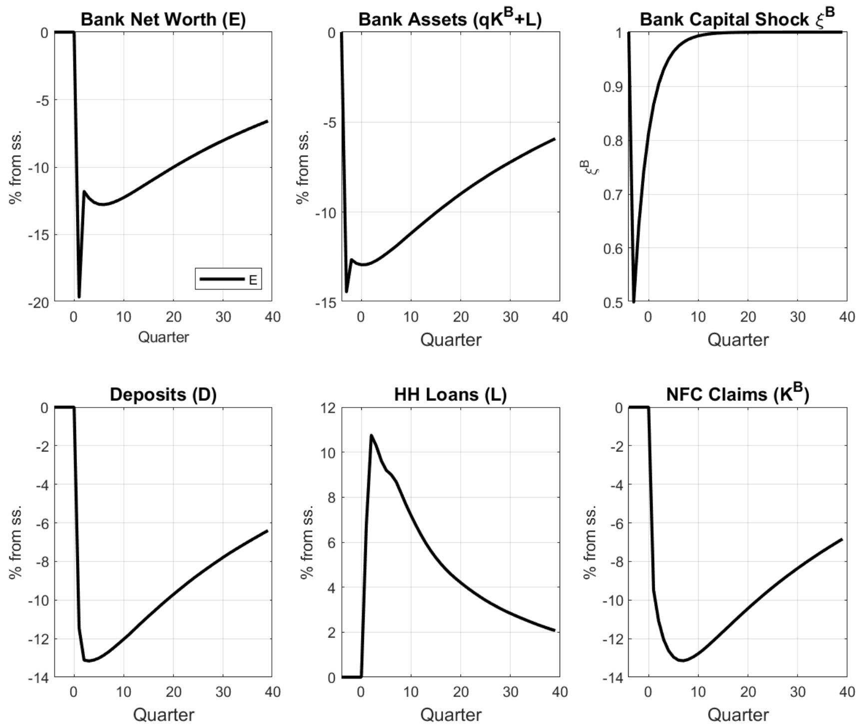


Figure 3.6: Dynamics of Banks' Balance Sheet
Note: Responses of components of banks' balance sheet. The shock is plotted in the top-right panel.

Figure 3.8 shows the effects of the shock on interest rates, prices, and dividends. As mentioned, on impact the interest charged on borrowing (r^L) increases and the return on deposits r^D falls as the leverage constraint tightens and banks reduce their balance sheet. r^D increases shortly afterward as banks have to compete with a now-higher return on holding capital R^K, defined as $R_t^k \equiv \frac{r_t^K+(1-\delta)q_t}{q_{t-1}} - 1$, in order to collect deposits. These higher returns on capital are partly driven by an increase in the marginal product of capital r^K, as capital effectively becomes scarcer, and partly driven by capital gains from an increasing price of capital q as it recovers from its sharp drop. Dividends experience a steep decline, driven by both the decline in output and the reduction in proceeds from banks' intermediation. Finally, wages decrease as the marginal productivity of labor falls with the capital stock.

3.5.2 The Distributive Implications of Banking Sector Losses

In Section 3.2, we showed that in the data households at different income levels react heterogeneously to bank losses. To relate our model to these results, Figure

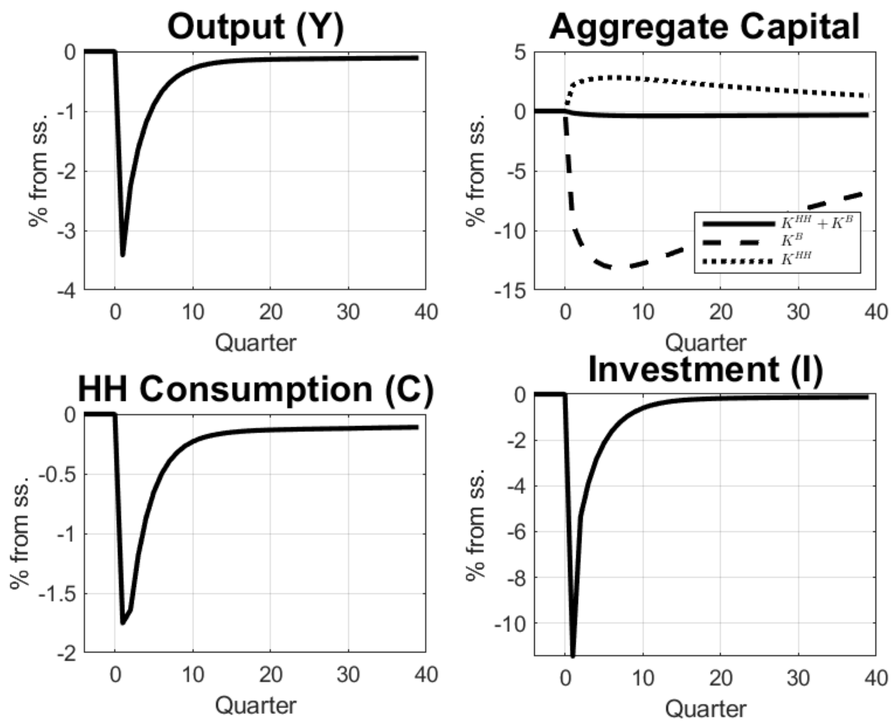


Figure 3.7: Dynamics of Macroeconomic Aggregates
Note: Responses of macroeconomic aggregates to the shock on banks' asset returns. All variables reported in percentage deviation from their respective steady-state levels.

3.9 reports the model-implied consumption responses by quintile of total (labor and financial) income, as well as aggregate consumption in the bottom right.[26]

The heterogeneous responses along the income distribution align well with our empirical results: First, while consumption of all income groups declines on impact and gradually recovers from the shock, households in the lowest income quintile experience the largest decline. In addition, over the upper half of the income distribution, consumption responses resemble each other when measured against steady-state consumption levels, similarly to our findings in Section 3.2. Finally, our model can also account for the quantitative magnitude of differences in consumption responses in the data. Figure 3.10 compares the model-implied

[26]To compute the impulse responses by income quintile, we follow households belonging to each group over time and compare their realized path of consumption to the counterfactual scenario in which the shock never materializes. For each state triplet we compute the expected value of consumption over time in the steady state and in the case of the shock. We then take the relative difference between these two series and aggregate within each group using the steady-state distribution over idiosyncratic states. This is equivalent to following a large panel of households over time.

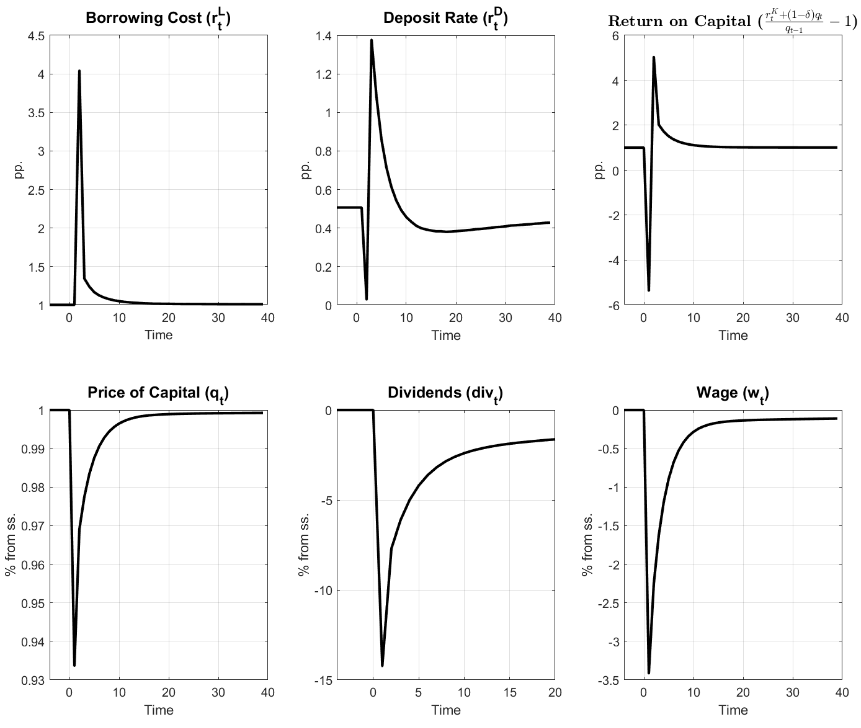


Figure 3.8: Dynamics of Equilibrium Prices
Note: Model-implied general equilibrium responses of prices. The top three panels are measured in percentage points. The three bottom panels consist of percent deviations from their respective steady-state values. The return on capital is defined as $R_t^k \equiv \frac{r_t^K+(1-\delta)q_t}{q_{t-1}} - 1$

cumulative impulse responses with their empirical counterparts.[27] The overall magnitude of the cumulative decline aligns well in each of the six panels.

Having matched the empirical patterns of consumption responses along the income distribution, we now investigate how consumption responses translate into changes in households' welfare as well as which transmission channels explain the heterogeneity displayed in Figures 3.2, 3.9, and 3.10.

Measuring Welfare Changes. To measure the welfare implications of banking sector losses, we compute households' expected value functions immediately after the shock is realized and compare them with the respective values in steady state. To express welfare changes as consumption equivalence units, we follow Bayer et al. (2019) and normalize the difference by the expected value of the discounted consumption stream for each household state triplet.[28] This allows

[27] In this figure, we rescale the impulse responses shown in Figure 3.2 to match the size of the shock in the model.
[28] Due to the utility cost of portfolio adjustment, households' value functions differ from the expected discounted stream of utility from consumption.

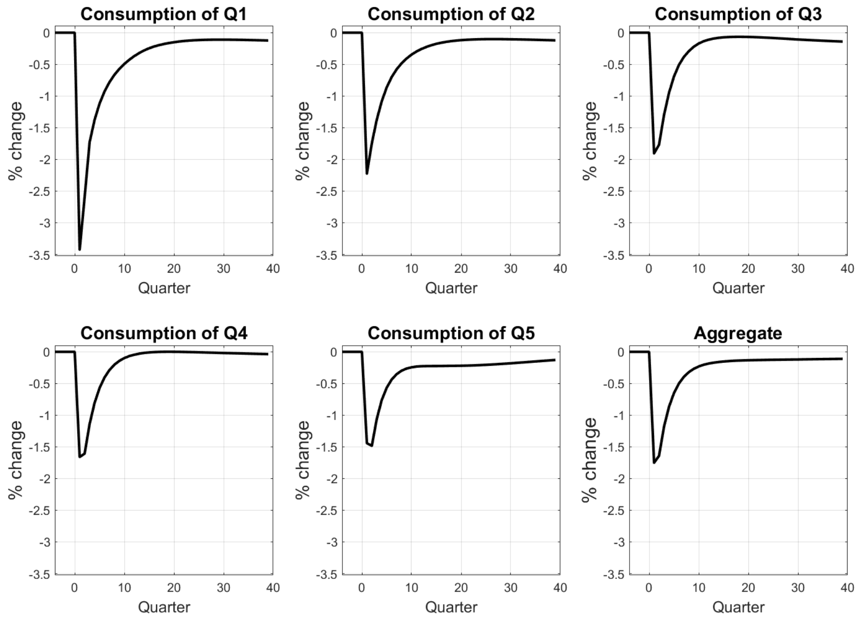


Figure 3.9: Consumption Responses by Income Quintile
Note: Impulse responses are displayed relative to the counterfactual evolution of consumption for each group in the absence of the shock. Income quintiles sorted based on total income in the steady state, including earnings, interest received, and dividends.

us to interpret changes in welfare as the fraction of consumption a household would be willing to forgo permanently to avoid the consequences of the shock and have the economy remain in steady state.

In percentage terms, the consumption-equivalent (CE) measure is calculated as follows:

$$CE(a,k,z) = 100 \times \left[\left(\frac{V_1(a,k,z) - V^{ss}(a,k,z)}{\mathbb{E}U(a,k,z)} + 1 \right)^{\frac{1}{1-\sigma}} - 1 \right]. \tag{3.28}$$

Here,

$$\mathbb{E}U(a,k,z) = \mathbb{E} \sum_{t=0}^{\infty} \beta^t u(c_t^{ss}(a,k,z)).$$

In the expressions above, V_1 and V^{ss} refer respectively to households' associated value functions after the shock hits and in steady state respectively. In addition, $\mathbb{E}U(a,k,z)$ is the expected discounted utility from consumption in the steady state.

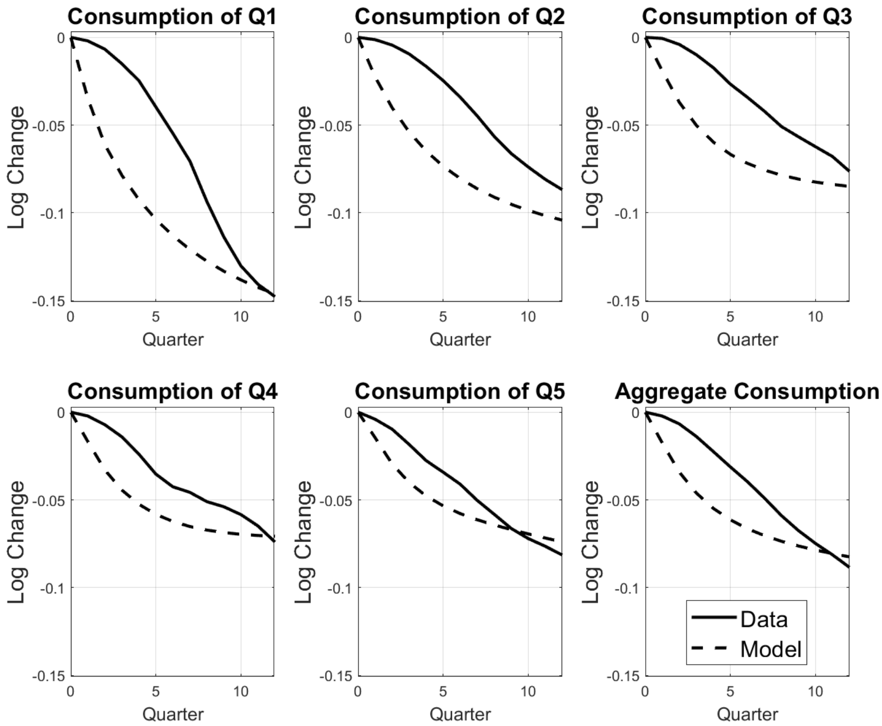


Figure 3.10: Consumption Responses by Income Quintile—Model vs. Data
Note: Model- and data-implied consumption responses. The former are obtained by rescaling the responses in Figure 3.2 by a factor of 1.72 to match the shock size in the model. The model- and data-implied responses are represented as log deviations from steady state.

Distribution of Welfare Changes. Figure 3.11 represents the distribution of welfare changes as computed by equation (3.28). The figure has two main takeaways: First, there is considerable heterogeneity in welfare changes. Second, even though the distribution is centered around a negative value – the average CE change is -0.39 percent – 11 percent of households exhibit a positive change in welfare and are actually better off in the presence of the bank shock.

Table 3.5 compares households who are worse off following the shock with those who benefit from it. Relative to the former group, individuals who experience a positive welfare change are more productive, wealthier, more dependent on income from financial sources, and have a more liquid portfolio.

Table C.1 in the appendix shows the breakdown of household characteristics for quintiles of the distribution of welfare changes. Overall, the conclusions are the same as those from Table 3.5: losses are decreasing in wealth, earnings, and portfolio liquidity and increasing in households' reliance on labor income.

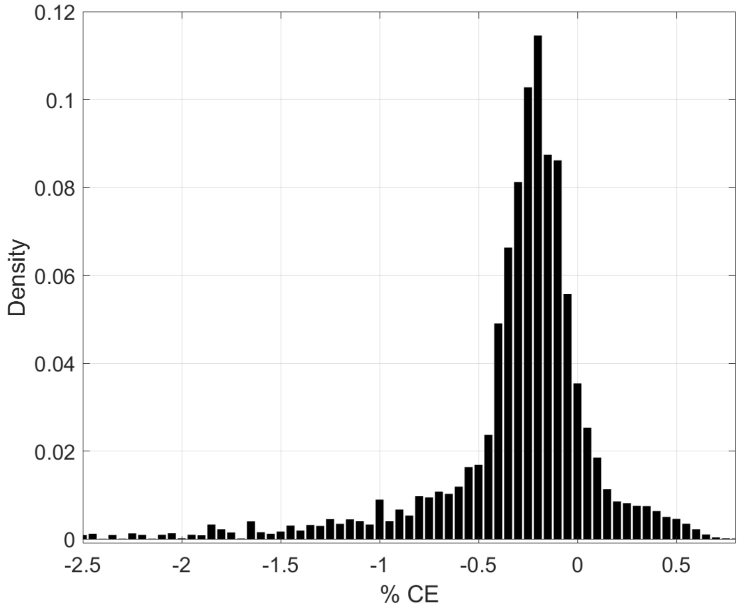


Figure 3.11: Distribution of Welfare Changes
Note: Distribution of welfare changes, measured in consumption equivalent units, as in equation 3.28.

Table 3.5: Characteristics of Gainers and Losers from Bank Losses

	Negative CE	Positive CE
Average Liquid Assets	0.41	5.7
Average Capital Holdings	0.4	3.9
Average Earnings	0.98	1.14
Average (Total) Income	0.94	1.44
Average Portfolio Liquidity	0.98	1.12
Dependence on Labor Income	95%	62%

Note: "Dependence on labor income" refers to the average share of earnings in households' total income. With the exception of the last row, numbers are displayed as a multiple of economy-wide averages.

Before we investigate the mechanisms behind these results, we turn our attention to heterogeneity in welfare changes along the income distribution and how they compare to the observed consumption responses.

Welfare Changes along the Income Distribution. Figure 3.12 shows that the changes in welfare caused by the shock are more unevenly distributed than those of consumption. For welfare (black bars), there is a clear monotonic pattern with households at the bottom of the income distribution suffering the largest welfare losses. While agents in the first quintile (Q1) would be willing to permanently forfeit 1 percent of their consumption to avoid the consequences of the shock, households at the top would give up only 0.08 percent. On the other hand, the

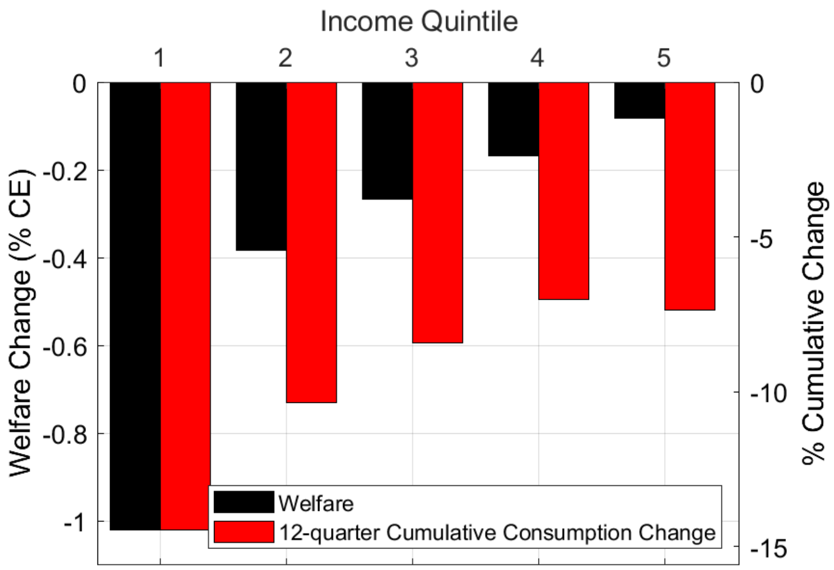


Figure 3.12: Welfare and Consumption Changes by Income Quintile
Note: Welfare changes, whose scale is on the left y-axis, are computed according to equation (3.28) and aggregated within each income quintile. The twelve-quarter cumulative consumption changes are measured on the right y-axis.

inequality in initial consumption responses is not nearly as pronounced: while the total decline for Q1 is 14.7 percent, for the fifth quintile (Q5) it is 7.4 percent.

Transmission Mechanisms. What mechanisms explain the patterns in Figure 3.12? Why do the rich suffer much less than what their initial consumption response suggests? How can a considerable fraction of households gain from a negative shock to the economy? To examine these questions, following Kaplan et al. (2018), we decompose the general-equilibrium responses of consumption and welfare into their partial-equilibrium changes due to movements in different prices, interest rates, and dividends. We compute counterfactuals in which we change *only* (i) labor earnings (w_t), (ii) the cost of borrowing (r_t^L), or (iii) financial income (r_t^D, R_t^K, and div_t jointly) to their realized general-equilibrium path and keep all other prices, rates, and dividends at their steady-state level.

Figure 3.13 decomposes the welfare changes by income quintile into these three components. The figure reveals substantial heterogeneity in transmission channels affecting different households. First, low-income households are exposed to changes in borrowing rates, which account for more than half of their welfare losses. These households use short-term debt to insure against temporary

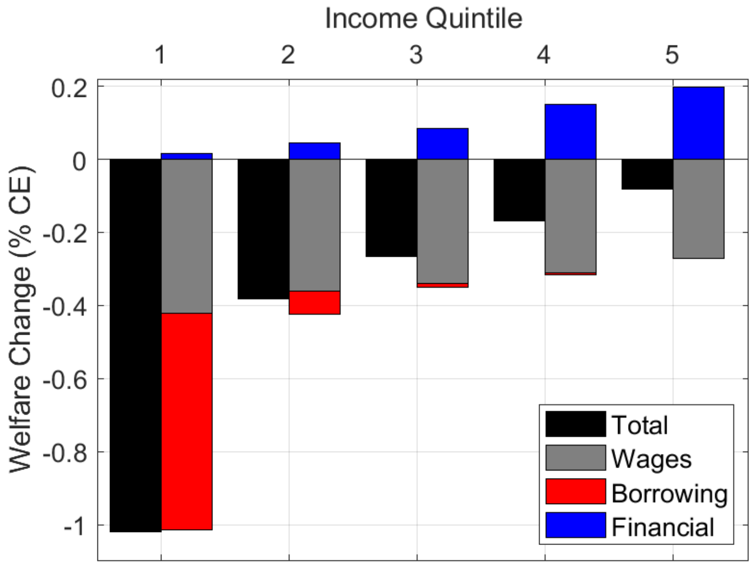


Figure 3.13: Decomposition of Welfare Changes by Income Quintile
Note: Decomposition of welfare changes due to wages $\{w_t\}_{t=0}^{T}$, the lending rate $\{r_t^L\}_{t=0}^{T}$, and financial variables (jointly $\{r_t^D, r^K, q_t, div_t\}_{t=0}^{T}$). The black bar represents the general-equilibrium welfare changes, replicating Figure 3.12. Each of the gray and colored bars is obtained by simulating the economy in response to the general-equilibrium path of one variable (or all four, in the case of financial variables).

income losses, which becomes more expensive in response to banking sector distress. Second, although all quintiles are substantially affected by changes in wages, those at the bottom are once again more exposed to wage variation. This is due both to their inability to insure against income shocks and to the fact that wages account for a larger proportion of household income for them. Financial variables, on the other hand, display a positive contribution for all the quintiles, with welfare gains increasing in household income.[29]

Figure 3.14 shows the consumption counterpart to the decomposition described above. In line with the decomposition for welfare, consumption at the bottom is mostly affected by the cost of borrowing and by labor income, while these channels have a limited impact on consumption at the top.

Financial income, on the other hand, plays a lesser role for the consumption responses of low-income households but increases in importance the further we move up the income distribution. Note that in response to movements in

[29] Capitalists are included throughout, and their income places them in the fifth quintile. Figure C.2 in the appendix presents capitalists, which represent 1 percent of the population, as a separate category. For them, the contribution of financial variables is negative due to the losses in dividends.

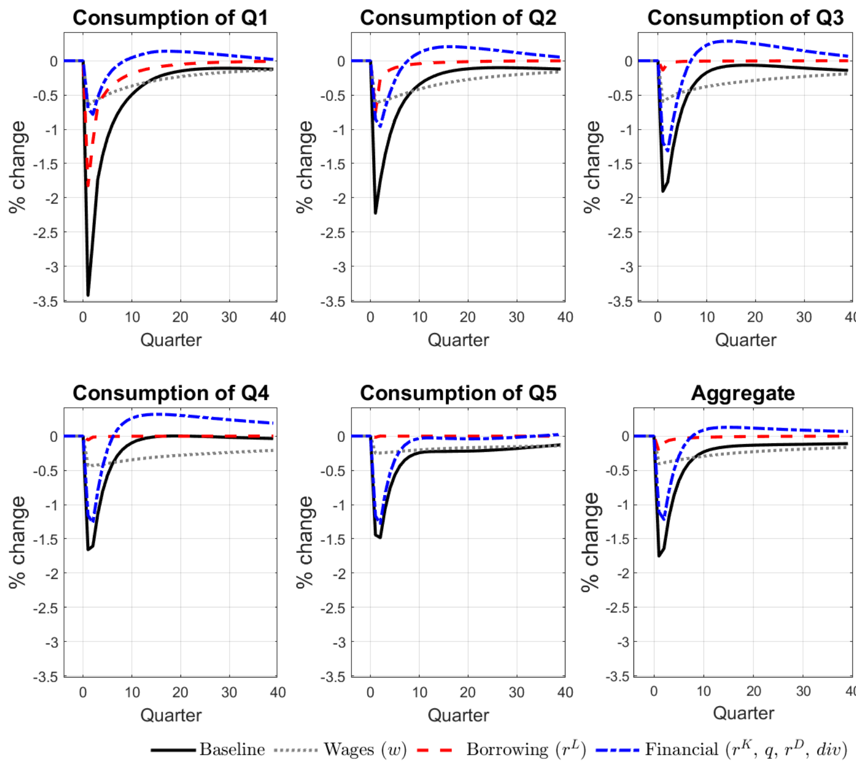


Figure 3.14: Consumption Decomposition by Income Quintile

Note: Impulse responses are displayed relative to the counterfactual evolution of consumption for each group in the absence of any price variation. Income quintiles are sorted based on total income in steady state, including earnings, interest received, and dividends. Consumption responses are decomposed into partial-equilibrium effects of wages $\{w_t\}_{t=0}^T$, the lending rate $\{r_t^L\}_{t=0}^T$, and financial variables (jointly $\{r_t^D, r^K, q_t, div_t\}_{t=0}^T$)

financial variables, households initially reduce their consumption. In the future, however, consumption overshoots for all quintiles except Q5. As we shall see, this overshooting is behind the positive changes in welfare induced by movements in financial variables.

The Role of Financial Variables. Figure 3.15 breaks down the financial component of welfare changes into those due to deposit rates r_t^D, the return on holding capital R_t^K, and dividends. The welfare impact of changes in deposit rates is positive for the first two quintiles and negative for the remaining ones. This is due to the initial decline and later overshooting in deposit rates. Households in the two lowest quintiles are largely insulated from the consequences of the initial decrease because they hold little savings and many of them are borrowers. They benefit from future increases in r^D because this gives them the opportunity to save at a higher return in the future. In contrast, high-income individuals suffer

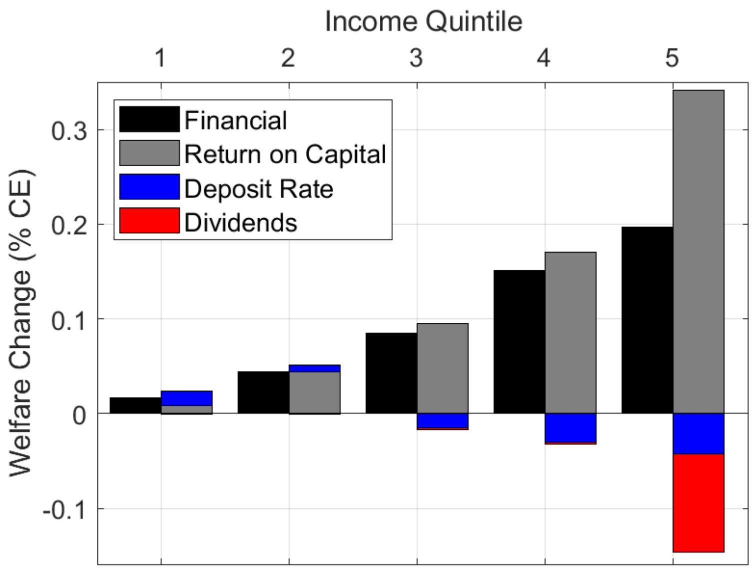


Figure 3.15: Decomposition of Welfare Changes—Financial Variables
Note: Decomposition of welfare changes due to financial variables (jointly $\{r_t^D, R_t^K, div_t\}_{t=0}^T$, in the black bar) and each of its separate components (gray and colored bars). Each of the gray and colored bars is obtained by simulating the economy in response to the partial-equilibrium path of one variable (or all four, in the case of the black bar).

from movements in r^D. This is because even though their portfolios consist mostly of capital, such households do hold a considerable amount of deposits, exposing them to the initial decline in rates.

Movements in the return on capital, on the other hand, benefit households across the board, and particularly those at the top of the income distribution. High-income households in fact take advantage of movements in the price of capital q_t as well as in the increased return on capital r_t^K and invest to finance higher consumption moving forward. This is clearly seen in Figure 3.16, where we contrast the general-equilibrium consumption responses with a counterfactual scenario in which the return on savings is kept fixed. In other words, for this counterfactual we fix both the return on holding deposits, r_t^D, and the return on holding capital, given by $R_t^k \equiv \frac{r_t^K + (1-\delta)q_t}{q_{t-1}} - 1$, at their steady-state values.

Across the entire income distribution, the immediate impact of the shock on consumption is reduced for the case of fixed returns on saving, relative to the general-equilibrium responses. This is because part of the initial decline in consumption is driven by households' increased desire to save when future returns are high. This mechanism becomes more important as we move up the

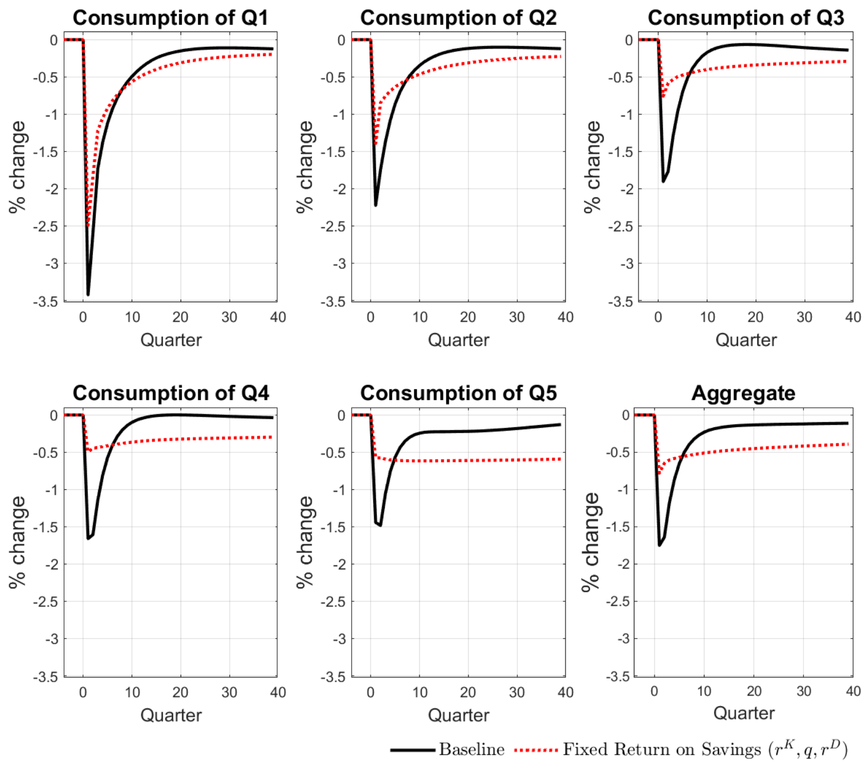


Figure 3.16: Consumption Decomposition—The Role of Savings Returns
Note: Model-implied consumption responses in general equilibrium (solid line) and partial equilibrium (dotted line). Income quintiles are sorted based on total income in steady state, including earnings, interest received, and dividends. Impulse responses are displayed relative to the counterfactual evolution of consumption for each group in the absence of any price variation. The dotted line shows the partial-equilibrium response to changes *only in* wages, the lending rate, and dividends ($\{w_t, r_t^L, div_t\}_{t=0}^T$).

income distribution, as low-income households often want to dis-save or borrow, as illustrated by the difference between the dotted and solid lines on impact ($t = 1$), which is largest for Q5 and smallest for Q1.

While changes in the return on savings have a very limited effect on low-income households' future consumption, they have important consequences for those at the top. This can be seen from the difference between the two lines for high-income households: absent changes in returns to savings, their consumption would be substantially lower in the medium term. In other words, high-income individuals take advantage of the movements in financial variables and save more on impact to sustain a relatively higher future consumption.

Finally, returning to Figure 3.15, we see that the decline in dividends imposes a direct income loss to some households at the top of the distribution – the capitalists. Movements in dividends explain why the consumption response to

Table 3.6: Welfare Changes—Heterogeneity

Quintile	Q1	Q2	Q3	Q4	Q5
by Income	-1.021	-0.383	-0.266	-0.167	-0.081
by Net Worth	-1.169	-0.363	-0.239	-0.150	0.012

Notes: Changes in welfare measured in consumption equivalent units, as in equation 3.28.

financial variables by households in Q5 (Figure 3.14) does not overshoot as it does for the other quintiles. The reduction in dividends persists for some time (see Figure 3.8) as long as both output and banks' net worth are suppressed, which contributes to lower consumption in Q5 in response to financial variables. Note that, on average, high-income agents still benefit from movements in financial markets, even though some of them are hurt by a drop in dividends.

Heterogeneity along the Distribution of Net Worth. Table 3.6 compares changes in welfare across quintiles of income and net worth. Net worth is defined as the sum of capital, liquid assets, and the net present value of the stream of dividends.[30] Welfare falls even more for the bottom quintile of the distribution, if sorted by net worth instead of income. This is because these households are mostly borrowers and therefore exposed to variations in r^L. Heterogeneity across the other quintiles of the net-worth distribution closely resembles that of the income distribution. Remarkably, those in the top quintile of the distribution of net worth on average benefit from the shock. This once again highlights the role of financial income in helping these households cushion – and in fact take advantage of – disruptions to the banking sector.

Taken together, the results in this section show that disruptions to banks have substantial redistributive consequences. Along with those who hold a direct claim to bank dividends, the ultimate losers from bank losses are low-income households, who are highly exposed to changes in wages and in the lending rate. Rich households, on the other hand, take advantage of movements in returns to savings. Even though these individuals experience a significant decline in consumption on impact, this is compensated by relatively higher future consumption. Thus, the welfare impact of the shock on high-income

[30] Figure C.1 displays the responses of consumption by quintile of net worth.

individuals is small, with some of them even standing to gain from disruptions to the banking sector.

3.6 Policy Response

In this section we examine which households benefit from policy interventions in response to banking sector losses. We consider an asset purchase program along the lines of the US government's Troubled Asset Relief Program (TARP).[31]

Government. To study policy interventions, we need to introduce a government into the model. We assume that the government can (i) impose a system of taxes and transfers on households and (ii) engage in financial intermediation by issuing debt in form of one-period *liquid* bonds to fund loans to NFCs. The government promises to pay the deposit rate r^D_{t+1} on the bond and earns the market return on holding capital for its loans. Let B_{t+1} be the total value of government-intermediated assets – i.e., the total amount of short-run debt issued to households. At the end of period t, the government then holds claims to K^g_{t+1} units of capital:

$$K^g_{t+1} = \frac{B_{t+1}}{q_t}.$$

As in Gertler and Karadi (2011), the government is not subject to a leverage constraint. Further, we assume that the productivity of government-intermediated capital equals $\xi^G \in [0, 1]$. This assumption captures the fact that the government might face higher costs of raising funds or have difficulties in identifying productive projects. We experiment with different values of ξ^G.

Our objective is to derive positive implications concerning a particular credit policy intervention, with a focus on its redistributive effects. For that reason, we consider an exogenously set policy where the government immediately reacts to the shock by issuing $B_2 = \bar{B}$ in the first period of the transition ($t = 1$) and

[31]TARP was introduced in the Emergency Economic Stabilization Act of 2008 to support the US financial sector in the global financial crisis through purchases or guarantees of distressed assets by the Department of Treasury. Until 2011, about \$410 billion was disbursed. The Congressional Budget Office estimates that the government has earned a net profit on its support to financial institutions through TARP during the crisis (CBO, 2021).

deterministically repays the debt according to

$$B_{t+1} = \rho_b B_t, \; \rho_b \in (0,1).$$

The government is subject to a budget constraint given by

$$\left\{\left[\xi^G r_t^k + (1-\delta)q_t\right] K_t^g - (1+r_t^d)B_t\right\} = \mathcal{T}_t.$$

The left-hand-side term consists of the excess return on intermediation (in braces) and (ii) transfers $\mathcal{T}_t$. So long as bankers' leverage constraint binds – which is the case throughout our simulations – the government actually makes positive revenues from intermediation, which are transferred back to other economic agents; hence the term $\mathcal{T}_t$. The way in which these resources are rebated matters for the redistributive consequences of the proposed credit policy. For this reason, we consider three distinct possibilities: (i) a lump-sum transfer to all households, (ii) a transfer to all households that is proportional to their total income, and (iii) a transfer to banks.[32] In Appendix C.3.3 we describe how households' budgets, model aggregation, and equilibrium conditions change when we include the government.

We consider a policy in which in response to the shock the government's intervention is of similar magnitude to TARP – roughly $400 billion, or 10 percent of quarterly GDP. The parameter ρ_b is set to the same value as the autoregressive coefficient of the shock; i.e., the government policy is phased out as the banking sector distress fades. In our baseline specification, proceeds from intermediation are rebated lump sum and $\xi^G = 1$.[33]

By absorbing a portion of the demand for liquid assets in the economy, the credit intervention makes it easier for banks to reduce their leverage (see Figure C.15 in the appendix). As a consequence, the equilibrium increase in the spread is lower than it would be absent the policy. This mechanism is responsible for the lower decline in consumption at the bottom of the income distribution, as seen in Figure 3.17. Furthermore, the increased deposit rate is responsible for a steeper decline in initial consumption for households at the top of the

[32]The proportional transfer (ii) is meant to capture a reduction in overall tax rates made feasible through profits from intermediation.

[33]In Appendix C.3.3 we report the results from alternative schemes.

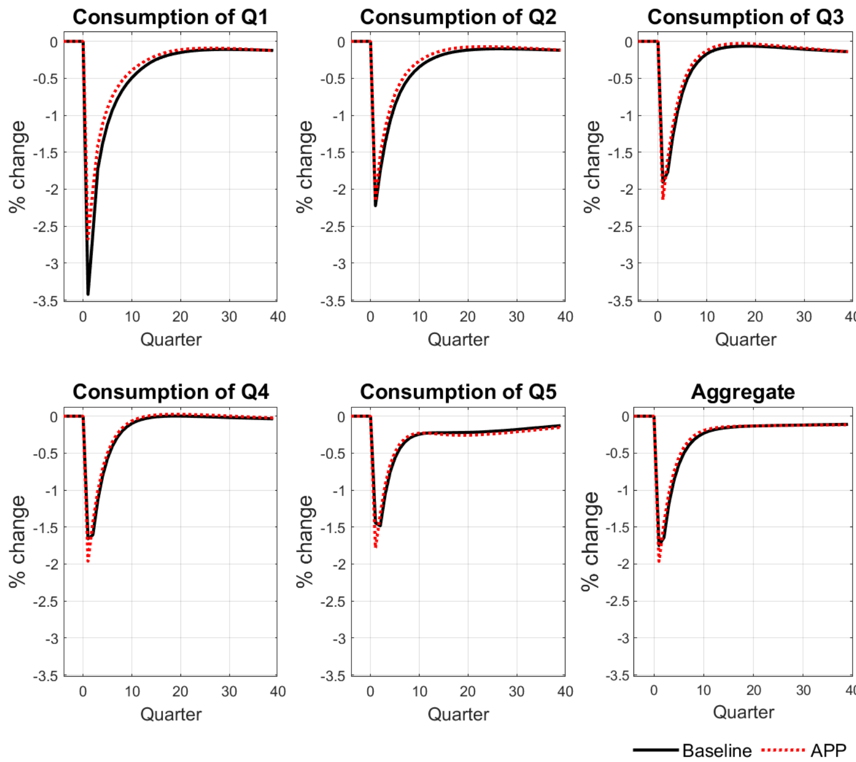


Figure 3.17: Consumption Responses—Credit Policy
Note: Model-implied consumption responses to the baseline shock, with the policy (red dotted line) and in its absence (black solid line). Income quintiles are sorted based on total income in the steady state, including earnings, interest received, and dividends. Impulse responses are displayed relative to the counterfactual evolution of consumption for each group in the absence of the shock.

income distribution as well as in the aggregate. The resulting heterogeneity in consumption responses is smaller.

The policy increases the on-impact decline in consumption, which is compensated by higher investment driven by government's capital holdings. This ensures higher future output. As a consequence, the credit policy reduces the overall welfare losses from banking sector distress by roughly one-fourth (from -0.39 percent to -0.30 percent). The reduction in welfare losses is remarkable, given that the government is unable to counter the decline in banks' capital productivity directly.[34] But it can prevent the consequences of a sudden and severe contraction on bank intermediation and dampen the associated price fluctuations.

[34] The reason for the relatively large reduction in the welfare decline is not that government's capital is more productive than banks'. The second row of Table C.4 shows that if the government-financed capital were as productive as the banks', the credit policy would still mitigate a fifth of the welfare consequences of the shock.

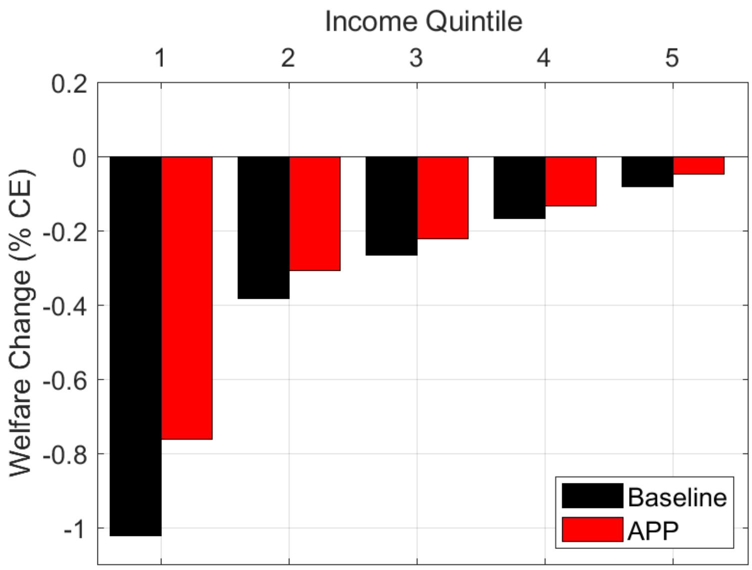


Figure 3.18: Welfare Changes—Credit Policy
Note: Welfare changes due to shock as in Section 3.5, with the policy (gray bars) and in its absence (black bars), computed according to equation (3.28) and aggregated within which income quintile.

As for the distribution of welfare gains from the policy intervention, the impact of the shock is strongly mitigated especially for those at the bottom of the income distribution. The welfare impact of the shock when the credit policy is in place is compared to our baseline results in Figure 3.18.

Last, the capitalists – claimants to banks' dividends – are worse off after a policy intervention. While this may seem counterintuitive, it is because government-intermediated assets crowd out deposits, which causes a reduction in spreads and a slower recovery of banks. This particular result can, however, be overturned under other rebate schemes (see Table C.4 in the appendix).

3.7 Conclusion

In this paper, we examine the distributive effects of banking sector losses. We document a novel empirical relationship between consumption along the income distribution and conditions in the banking sector: distress in the latter is associated with a stronger consumption response at the bottom of the income distribution. To understand these results, we build a two-asset heterogeneous-agent model featuring banks subject to a leverage constraint. The model is successful in replicating the pattern of heterogeneity observed in the data following a disruption in the banking sector.

We find that the relevant transmission channels vary substantially with income: low-income households suffer from an increase in borrowing cost and a decline in labor earnings; high-income households increase their savings in response to temporarily low asset prices and high future returns to sustain higher consumption in the medium term. This is why we find 11 percent of households to be better off after the shock. These are high-income, wealthy individuals, with a high share of income from financial sources.

Finally, we study the effects of a credit policy intervention aimed at alleviating the impacts of banking sector losses, along the lines of the Troubled Asset Relief Program. The policy reduces the negative welfare effect of bank losses by roughly one-fourth, with gains concentrated among low-income households.

While in this paper we take a positive approach in analyzing government interventions, understanding how the design of optimal policy should account for its redistributive effects is a promising avenue for future research.

Chapter 4

Who Cares about Inflation?

Endogenous Expectation Formation of Heterogeneous Households

Abstract This paper builds a joint theory of endogenous inflation expectations and consumption-savings choices of heterogeneous households. We introduce imperfect information about future inflation rates in a consumption-savings model and allow households to exert costly effort to reduce uncertainty about future price changes. High wealth households are more exposed to future inflation due to its effect on real interest rates and hence choose to be better informed. The joint distribution of wealth and inflation expectations generated by the model is consistent with key features of the data. The implied consumption response to news about inflation is hump shaped in wealth: Wealthier households pay closer attention and update their expectations more in response to any signal received, but change their consumption less after any given update in expectations due to the income effect of future inflation. We show this mechanism to reduce the on-impact aggregate consumption response to forward guidance policies by up to 55% compared to an attentive counterfactual.[1]

[1] In this paper use is made of data of the DNB Household Survey administered by CentERdata (Tilburg University, The Netherlands) and of data provided by the University of Michigan, Survey Research Center, Surveys of Consumers.

4.1 Introduction

In recent years, central banks have begun to rely more and more on forward guidance – influencing households' behavior through signals about future macroeconomic outcomes such as inflation – as a policy instrument to stimulate current demand. Who adjusts their expectations in response to news about future inflation remains as much an open issue as how heterogeneous households respond to changes in their expectations. Understanding both of these points is essential to evaluate the effectiveness of forward guidance policies.

When central banks want to stimulate current demand by signaling higher inflation for future periods, the response of consumption is determined in two steps: First, households need to update their expectations based on the signal they receive. Second, they respond to their updated expectations by adjusting their consumption behavior. This paper considers both of these steps jointly by introducing endogenous expectation formation in an otherwise standard consumption-savings model with heterogeneous households. We show that wealth is an important determinant of both households' incentives to pay close attention to signals about future inflation rates and their consumption response to any given change in their inflation expectations. Based on this finding we argue that allowing for an endogenous wealth effect on the formation of households' inflation expectations substantially reduces the responsiveness of aggregate consumption to signals about future inflation rates, because those likely to adjust their expectations are less responsive in their consumption. Explicitly accounting for heterogeneity in wealth and its impact on the formation of inflation expectations hence suppresses the effectiveness of forward guidance policies at its origin.

For a discussion of the incentives to pay attention to future inflation it is important to note that heterogeneous households are not exposed to inflation uniformly. Through its effect on real interest rates households are more affected by inflation the more they borrow or save between periods. If expectation formation is in any way costly, we should therefore consider households with higher wealth

– and hence larger exposure – to be more willing to face these cost.[2] How heterogeneous households respond to expected inflation likewise depends on their wealth holdings. Any given change in (expected) inflation rates will have different consequences for households with different asset levels. While an increase in expected inflation is good news for debtors since it reduces the real value of their future repayments, it is bad news for savers who would be the recipients of those payments. This heterogeneous income effect makes households' consumption response to any given change in inflation expectations a declining function of their wealth. Taken together, both effects imply a negative correlation between households' updating their expectations and their potential consumption response to changes in expectations, dampening the responsiveness of aggregate consumption to signals about future inflation rates.

This paper formalizes the intuitive arguments above in a theoretical framework, disciplining it with empirical observations on inflation expectations along the wealth distribution. Our approach is novel in considering the joint formation of inflation expectations and consumption-savings decisions of heterogeneous households.

We begin by developing a theory of endogenous inflation expectations. Households are assumed to understand the underlying inflation process and to be uncertain only about future inflation rates. To reduce this uncertainty, they can exert costly effort. The proposed framework is sufficiently tractable to integrate it into a heterogeneous agents model while capturing key features of the data. Analytical results show that this model of expectation formation implies the standard deviation of forecast errors as well as the mean absolute error across a group of households to be decreasing in the effort they exert. We use these results to discipline our model with cross-sectional statistics from the joint distribution of inflation expectations and wealth, making use of the Dutch National Bank's Household Survey. We find the standard deviation of forecast errors and the mean absolute error across households to be decreasing in absolute wealth in the data. Both richer as well as indebted households have more precise and

[2]In this paper, the focus lies on inflation as a risk to the real interest rate that affects all saving and borrowing uniformly. Wages are real and all households face the same effective inflation rate. This assumption is discussed in later sections.

less dispersed expectations compared to those around zero net wealth. Integrating the proposed model of expectation formation into an infinite-horizon consumption-savings problem allows us to study jointly the formation of and response to households' inflation expectations. The calibrated model matches the empirically observed pattern of forecast errors along the wealth distribution. Households with higher net savings or debt endogenously choose to be better informed about future inflation as they are more exposed to inflations' effect on real interest rates.

The theoretical framework allows us to back out the consumption response to any signal about future inflation rates – the marginal propensity to consume on signal (MPCS) – from households' policy functions. We show that the MPCS depends on two factors: How much a household updates its expectations in response to any signal received, and how it reacts to any given change in its expectations. Households' consumption response to any given change in expectations is decreasing in wealth due to an expected income effect. This income effect arises as higher expected inflation ceteris paribus reduces the expected future value of savings. In contrast, as richer households are endogenously paying closer attention to inflation rates, they update their expectations more in response to any signal received, making their consumption more responsive to news about inflation. Combined, these two forces yield a hump shaped pattern for MPCS' along the wealth distribution.[3]

To highlight the importance of our findings at the aggregate level, we conduct a forward guidance exercise within our framework. Capturing the on-impact effect of forward guidance, we simulate the aggregate consumption response to a one percentage point increase in all signals received by households about next period's inflation. We show that under endogenous expectation formation, forward guidance misses out on up to 55% of the effect it could have if all households choose to be as informed as the most attentive. This result is driven by low wealth households, who are potentially most responsive to any change in their inflation expectations, but fail to update their expectations in response to the signal as they do not pay close attention to news about future price changes. The

[3] The group of indebted households who would be both likely to update their expectations and strongly respond in their consumption is small and therefore quantitatively less important.

(richer) households paying attention to the signal and updating their expectations in response perceive higher inflation as a loss in their real income, yielding a relatively lower consumption response.

Most existing models of inflation expectations, summarized in Coibion and Gorodnichenko (2012), study expectation formation in isolation from other household choices and often abstract as well from underlying heterogeneity among agents, Madeira and Zafar (2015) being one of few exceptions. We extend this work to include wealth as a direct determinant of expectation formation when precise expectations are costly. In this regard we are closest in spirit to the literature on rational inattention founded by Sims (2003) and surveyed in Mackowiak et al. (2018). In a heterogeneous agent framework, Carroll et al. (2020) and Auclert et al. (2020) introduce sticky expectation formation but abstract from endogeneity of expectations with respect to households' idiosyncratic state. We share the analysis of heterogeneous incentives to form precise expectations with Broer et al. (2018). They discuss the endogenous choices of households to use precise laws of motion for aggregate capital in an economy à la Krusell and Smith (1998) and find substantial heterogeneity in the utility loss from not using full information. In their model, choices are based on simulated lifetime utilities and forecasting capital impacts forecasts about both returns and wages, accounting for the difference to our findings. Recent work has found households' consumption responses to income shocks to be an important determinant of Macroeconomic outcomes, leading to a large and growing literature on heterogeneous households' marginal propensity to consume out of transitory income (MPC), sampled e.g. in Kaplan and Violante (2021). In contrast, we focus on households marginal propensity to consume in response to signals about future inflation (MPCS) and show that heterogeneity along this margin is important to consider for policy analysis. Studying forward guidance in a framework with heterogeneous agents, McKay et al. (2016) show how occasionally binding borrowing constraints can reduce the responsiveness of aggregate consumption to interest rate changes in the distant future, alleviating the so called *Forward Guidance Puzzle*. Compared to their framework with full information, we show how an endogenous correlation between expectation updating and consumption responses can dampen the effects of forward guidance also in the short run.

Similar to previous theoretical approaches, most empirical work on inflation expectations has as well abstracted from wealth as a potential determinant of expectation formation. Closest to our analysis is Ben-David et al. (2018) who study the relation between uncertainty about macroeconomic variables such as inflation or house prices and socio-economic status of households. Their data does not include households' asset holdings but they find uncertainty about macroeconomic variables to decrease in income and employment – both highly correlated with wealth. Another strand of the empirical literature considers the impact of expectations on households' consumption savings choices. Among others, Armantier et al. (2015), Crump et al. (2015), Dräger and Nghiem (2018), or Vellekoop and Wiederholt (2019) evaluate the consistency of households' choices with their expectations. Coibion et al. (2019) study how expectations and consumption respond to exogenous news about inflation. They find a negative consumption response to higher inflation, driven by high wealth households in line with our model. Also in line with our theory, Lieb and Schuffels (2019) find the likelihood of positive durable consumption expenditure in response to higher inflation expectations to be decreasing in wealth. We contribute to this literature by highlighting the importance of considering jointly heterogeneity in households' incentives to form precise expectations and their potential response to such expectations.

The remainder of the paper is organized as follows: Section 4.2 introduces a framework for endogenous inflation expectations. Section 4.3 presents empirical findings on the joint distribution of wealth and inflation expectations. Section 4.4 incorporates the endogenous expectation framework in a consumption-savings model with heterogeneous households. Section 4.5 analyses households' consumption responses to news about inflation and discusses aggregate consequences of endogenous expectation formation. Section 4.6 concludes.

4.2 Modeling Endogenous Inflation Expectations

To allow for two-way interactions between households' consumption-savings choices and their inflation expectations, we require a model with endogenous expectation formation. This section provides an endogenous expectation framework, which will later be incorporated into a consumption-savings problem. In the interest of a clear exposition and computational tractability, we keep the expectation formation process as simple as possible while at the same time rich enough to account for key features of the data. Households are assumed to understand the underlying process and perfectly observe current inflation but to be uncertain about the shock component to future inflation rates. This uncertainty can be reduced endogenously by households exerting costly effort. The setup yields heterogeneous effort choices if the gains of reducing uncertainty about future inflation rates are distributed unevenly across households. In this section, focus lies on how effort transmits into individual expectations and cross-sectional moments of expectation errors. Section 4.4 discusses households' effort choice.

Assume inflation follows a first-order autoregressive process

$$\pi_{t+1} = (1-\rho)\mu + \rho\pi_t + e_{t+1} \qquad e_{t+1} \sim \mathcal{N}(0, \sigma_e^2). \tag{4.1}$$

π_t is inflation in period t, μ is the long run mean of inflation and ρ its persistence across periods. e_t is a shock to inflation, which is i.i.d across time.

Households know that inflation follows (4.1) and agree about (the true) μ, ρ and σ_e^2. In contrast to most of the literature on expectation formation, households perfectly observe current and all past inflation rates.[4] In period t, π_τ is known for all $\tau \leq t$. This assumption keeps the state space of the household problem small. We believe this is justified, given that information about current and past inflation rates is easily accessible online.[5] Furthermore, when embedding the expectation formation process in a consumption-savings model, it will be important for households to know current prices in order to pin down their budget set in real terms. Therefore, households are assumed to be uncertain only about future inflation rates.

Households form expectations with respect to the shock to future inflation, e_{t+1}. In period t, household i can exert some effort n_t^i to influence the noise in a signal

[4]See e.g. Vellekoop and Wiederholt (2019).

[5]One can reinterpret our assumption as the first marginal bit of effort providing full information about present and past inflation rates. With the assumptions imposed below on the cost of effort, the first marginal bit of information is costless and hence always obtained.

$\hat{e}^i_{t+1}$ he receives about next period's shock. The signal he receives follows

$$\hat{e}^i_{t+1} = e_{t+1} + s^i_{t+1} \qquad s^i_{t+1} \sim \mathcal{N}(0, \sigma^2_s(n^i_t)), \tag{4.2}$$

where the noise component s^i_{t+1} can be influenced by households' effort choice. We assume its standard deviation to be a decreasing but convex function of effort ($\sigma'_s(n) < 0,\ \sigma''_s(n) > 0$) and s^i_{t+1} to be pure noise, i.e.

$$e_t \perp\!\!\!\perp s^i_t \ \forall i,t \qquad s^i_t \perp\!\!\!\perp s^j_t \ \forall i,j,t \qquad s^i_t \perp\!\!\!\perp s^i_{t+s} \ \forall i,t,s. \tag{4.3}$$

Households have identical priors about the shock corresponding to the true unconditional distribution $e_{t+1} \sim \mathcal{N}(0, \sigma^2_e)$.[6] Based on the signal received, the household updates his prior belief according to Bayes Rule. Let $\omega^i_{t+1}(n^i_t) = \frac{\sigma^2_e}{\sigma^2_e + \sigma^2_s(n^i_t)}$ be the weight he attaches to the signal, yielding his posterior belief about the shock as

$$e_{t+1}|_{\hat{e}^i_{t+1}, n^i_t} \sim \mathcal{N}\left(\omega^i_{t+1}(n^i_t)\hat{e}^i_{t+1}, \omega^i_{t+1}(n^i_t)\sigma^2_s(n^i_t)\right). \tag{4.4}$$

Household i's expected value for inflation is determined by his expectation about the future shock and is given as

$$\mathbb{E}_t[\pi_{t+1}|\hat{e}^i_{t+1}, n^i_t] = (1-\rho)\mu + \rho\pi_t + \omega^i_{t+1}(n^i_t)\hat{e}^i_{t+1} \tag{4.5}$$

implying an ex-post forecast error of

$$\begin{aligned} err^i_{t+1} &= \mathbb{E}_t[\pi_{t+1}|\hat{e}^i_{t+1}, n^i_t] - \pi_{t+1} \\ &= \omega^i_{t+1}(n^i_t)s^i_{t+1} - (1 - \omega^i_{t+1}(n^i_t))e_{t+1}. \end{aligned} \tag{4.6}$$

The first term in the error captures households' over-reaction to noise while the second term captures under-reaction to news contained in the signal, as is standard in models with Bayesian updating.

[6] Heterogeneity in prior variance, especially if correlated with households wealth, would complicate the analysis substantially. Assuming a common (unbiased) prior about the mean of the shock is without loss of generality. Relaxing this assumption would yield similar results as introducing heterogeneous beliefs about μ, see appendix D.2.1.

The standard deviation of households' posterior belief about future inflation can be referred to as their *subjective uncertainty* (SU) and is given by

$$SU_{t+1}^{i} = \sqrt{\omega_{t+1}^{i}(n_t^i)\sigma_s^2(n_t^i)} = \sqrt{\frac{\sigma_e^2\sigma_s^2(n_t^i)}{\sigma_e^2 + \sigma_s^2(n_t^i)}}. \tag{4.7}$$

Under the assumptions that $\sigma_s'(n) < 0,\ \sigma_s''(n) > 0$, one can show that SU is a decreasing and convex function of households' effort n.

We can also derive theoretical moments for a group g of households with equal choices for $n_t^i = \bar{n}_t^g$. An equal choice for n implies identical weights $\omega_{t+1}^i(n_t^i) = \omega_{t+1}^g(\bar{n}_t^g)$. The model implies that the forecast errors within a group g will be normally distributed with a variance, across households and time, given by

$$\begin{aligned} \text{Var}^g(err_{t+1}^i) &= (\omega_{t+1}^g(\bar{n}_t^g))^2\sigma_s^2(\bar{n}_t^g) + (1-\omega_{t+1}^g(\bar{n}_t^g))^2\sigma_e^2 \\ &= \frac{\sigma_e^2\sigma_s^2(\bar{n}_t^g)}{\sigma_e^2 + \sigma_s^2(\bar{n}_t^g)} = \left(\overline{SU}_{t+1}^g\right)^2, \end{aligned} \tag{4.8}$$

where we make use of the assumption that noise is uncorrelated across households. Hence the within group standard deviation of forecast errors across households exerting effort $\bar{n}_t^g$ can be interpreted as the standard deviation of the posterior belief about future inflation of households in that group $\overline{SU}_{t+1}^g$. With our model, disagreement among households becomes a measure of how noisy a signal about future inflation rates these households chose to receive.

As an additional measure of forecast precision, we can derive the mean absolute error of households by using the fact that s and e are normally distributed and uncorrelated with each other and over time. This implies a normal distribution for the expectation error among a group of households with mean zero and variance given in (4.8). By the properties of folded normal distributions, the average absolute error is given as

$$\mathbb{E}^g[|err^i|] = \sqrt{\text{Var}^g(err_{t+1}^i)\frac{2}{\pi}} = \overline{SU}_{t+1}^g\sqrt{\frac{2}{\pi}}. \tag{4.9}$$

Therefore, our model predicts a strong co-movement of the standard deviation and mean absolute error for a group of households exerting similar effort, driven by how noisy a signal they chose to receive about future inflation.

An important implication of our theoretical results is that they rationalize the use of cross-sectional moments to learn something about households' expectation formation. They allow us to discipline a model of joint consumption-savings choices and expectation formation with the standard deviation of expectation errors at different points of the wealth distribution. Before we turn to incorporating endogenous expectations into a consumption-savings framework we therefore report in the next section on the joint distribution of expectation errors and wealth in the data.

In order to reduce the state space of the problem and incorporate it into a heterogeneous agent framework, we have kept the expectation formation process as simple as possible, but sufficiently rich to account for key features of the data. In Appendix D.2.1 we show that the results are robust to additional sources of heterogeneity in expectations, such as fundamental disagreement about the long run mean of inflation.

4.3 Expectations Along the Wealth Distribution

This section presents empirical observations on the joint distribution of households' wealth levels and inflation expectations, which we will use to discipline our model. As suggested by the results of the previous section, we study the cross-section of households at different points of the wealth distribution and focus on two statistics: The standard deviation of forecast errors and the average absolute forecast error. After outlining the data used and methodology applied we present our baseline findings before concluding with some additional robustness tests.

4.3.1 Data and Methodology

To gain insight into the joint distribution of inflation expectations and wealth we use data from the Dutch National Bank's Household Survey (DHS). This dataset is unique in providing comprehensive data on both households' wealth and their inflation expectations. We combine observations from the survey waves 2010-2018. The choice of period reflects changes made to the questionnaire on inflation expectations in the 2008 wave and excludes the financial crisis episode. We use data at the individual level, as presented in the DHS, but restrict our sample to household heads to avoid within household correlations. We take heads' answers to be representative for their household.

In the survey, households are simultaneously asked about their current wealth in a variety of asset classes and their expectations of one year ahead inflation. We compute households' net financial wealth as the sum of all assets less liabilities reported in the DHS, excluding houses and related mortgages, business equity and vehicles. Whenever referring to "wealth" in the remainder of this paper, we apply this definition. In the baseline results, we focus on financial wealth as we believe it to capture best the resources out of which the household decides to consume or save in response to changes in inflation rates.[7] Using the described wealth measure, we construct decile groups based on households' position in the wealth distribution in the year of observation. We pool observations across waves that are in the same wealth decile for their wave. Table D.1 in the appendix reports summary statistics for these groups.[8] It also shows that results are robust to pooling all observation across years and defining wealth deciles based on the full sample.

Participants in the DHS are asked to report a point forecast for the inflation rate over the following 12 months, choosing from the set of whole numbers between 1 and 10. Ex-post errors are computed by subtracting the realized inflation rate over the next 12 months from this forecast. As the exact month of the observation is unknown (the survey generally takes place between April and October each year), we subtract June-to-June inflation as an approximation to

[7] Previous real estate or durable goods purchases are unlikely to be re-considered in response to small fluctuations in expected inflation rates.

[8] The table shows different numbers of missing observations for inflation expectations across wealth deciles, with the highest number of missings in the second decile. To test robustness with respect to differential numbers of missings, we have constructed bounds in the spirit of Lee (2009). All main findings are robust to the number of missing values across deciles. Results are omitted for brevity but are available upon request.

the forecasted rate. As an example, for an observation of the 2016 wave inflation is the change in the Dutch CPI between June 2016 and June 2017.[9]

4.3.2 Empirical Observations

Our focus is on observations at the wealth decile group level. For each wealth decile group, we report the within-group standard deviation of forecast errors and the mean absolute error.[10] Figure 4.1 presents our baseline empirical results. Both the mean absolute error and the within-group standard deviation increase between the first and second decile and decline as wealth increases further until reaching a stable level in the upper half of the wealth distribution. At its lowest level, both variables are about 0.6 pp. lower than at their peak in the second decile. Both the initial increase and the subsequent decline are statistically significant at the 95% level.

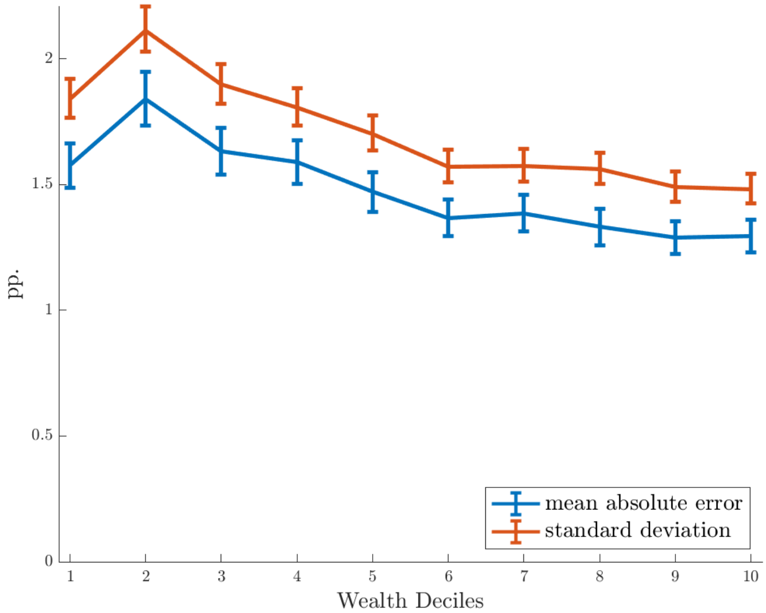


Figure 4.1: Expectation Errors by Wealth Decile Groups

The figure plots the within-decile group standard deviation of errors and the mean absolute forecast error. Bars provide confidence bands at the 95% level. Expectation errors are ex-ante point forecasts minus ex-post realizations. Data from DNB Household Survey waves 2010-2018.

[9] None of the results is altered qualitatively if June-to-June inflation is replaced by inflation in the current year (annual inflation in 2016 in our example) or the following year (annual inflation in 2017).

[10] Baseline results are unweighted. Using household weights has no significant impact on our findings.

For the interpretation of these findings it is important to note that the second wealth decile group is centered around zero net financial wealth, i.e. net debtors are concentrated in the first decile group. Through the lens of our model of expectation formation, the results suggest that wealthier as well as indebted households choose to exert higher effort in order to form precise expectations about future inflation rates.

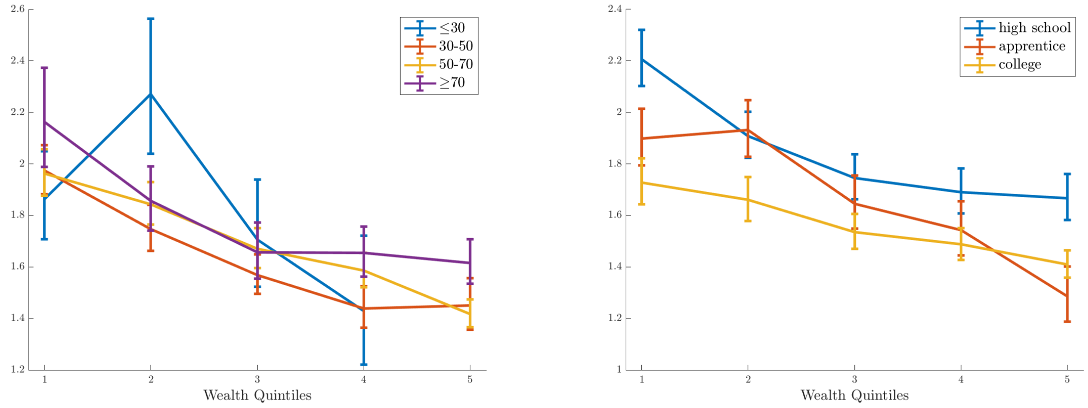
To validate our approach to modeling households' expectation formation, Figure D.6 in the appendix plots the histograms of errors by decile. Our theory would suggest that these errors should be normally distributed within decile groups. Despite limitations such as the discreteness and truncation of expectation data due to the sample question in the DHS, the fitted kernel densities align well with their respective normal counterparts.

The role of age and education

It is well established that other demographic characteristics are highly correlated with positions in the wealth distribution.[11] The two most important for our analysis are age and education. An argument can be made that more experienced (as older) people could be better at forming expectations. Similar argument applies for more educated individuals. As education and age correlate positively with wealth this could be driving the finding in Figure 4.1. As we include neither education nor age in our model, we test for robustness and repeat our analysis controlling for age and education respectively.

Testing for robustness towards age and education, we look at the data on quintile group level to allow for a sufficient number of observations within each age/wealth and education/wealth cell. At the quintile level, debtors are pooled with households around zero wealth. Figure 4.2 reports the within wealth quintile group standard deviation of errors by age groups and education groups. The general downward trend of disagreement in wealth persists after controlling for either age or education. Age appears to have little explanatory power beyond the impact of wealth, providing an argument against experience as a driving force for expectation formation. College education, however, appears to somewhat

[11] See e.g. Cooper and Zhu (2016).



(a) by age

(b) by education

Figure 4.2: Standard Deviation of Expectation Errors by Wealth Quintiles – Controls

The figure plots the within-quintile group standard deviation of errors by age (a) and education groups (b). Bars provide confidence bands at the 95% level. Expectation errors are ex-ante point forecasts minus ex-post realizations. Data from DNB Household Survey waves 2010-2018. Combination of youngest age and highest wealth quintile omitted due to lack of observations.

decrease disagreement compared to less educated groups. Similar findings hold for the mean absolute expectation error.[12]

Measure of wealth

To test robustness with respect to the considered measure of wealth, we repeat the analysis dividing households into decile groups based on two alternative measures: A first including both housing and associated mortgages as well as a second considering only positive financial assets. Including housing wealth leaves the results qualitatively unchanged, as Figure 4.3 shows. The peak of both mean absolute error and standard deviation of errors remains in the second decile group (again around zero net wealth). Both decline to either side and the overall decline between peak and trough in both variables is of similar magnitude as before. Different from previous results there is a hump shaped pattern between the 4th and 10th decile especially in the standard deviation of errors. This is perfectly in line with the correlation of financial wealth and housing wealth: Median financial wealth increases up to the 4th decile of wealth including housing, declines again for deciles 5 to 7 before increasing substantially for deciles 8 to 10. We take this as further support for financial wealth as the relevant measure to consider.

[12]These results are presented in Figure D.4 in the appendix.

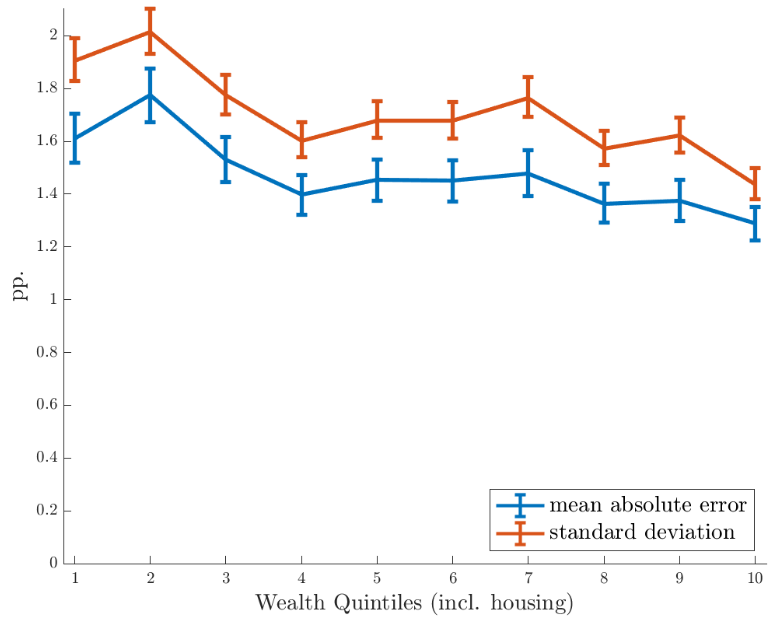


Figure 4.3: Expectation Errors by Wealth Decile Groups – Housing

The figure plots the within-group standard deviation of errors and mean absolute errors by net financial wealth decile groups, including housing and mortgages in the wealth measure. Bars provide confidence bands at the 95% level. Expectation errors are ex-ante point forecasts minus ex-post realizations. Data from DNB Household Survey waves 2010-2018.

Excluding debt from the wealth measure, we find that both mean absolute errors and the standard deviation of errors are declining in asset holdings as Figure 4.4 shows. This is as expected given the limited amount of financial debt (and hence limited netting of financial asset positions) in the DHS dataset.[13] Again, the overall decline in both measures along the wealth distribution is of similar magnitude as in the baseline results, and both flatten over the the 6th-10th decile of total financial assets.

Individual Level Analysis

While measuring the standard deviation of errors requires us to pool households into groups, differences in the absolute forecast error can also be tested at the household level. To do so, we regress households' absolute forecast error on indicators for their wealth decile and controls. The specification is given in

[13]Median debt is zero for all but the first decile of financial wealth and averages liabilities are below EUR 1,000 for deciles 2 to 10.

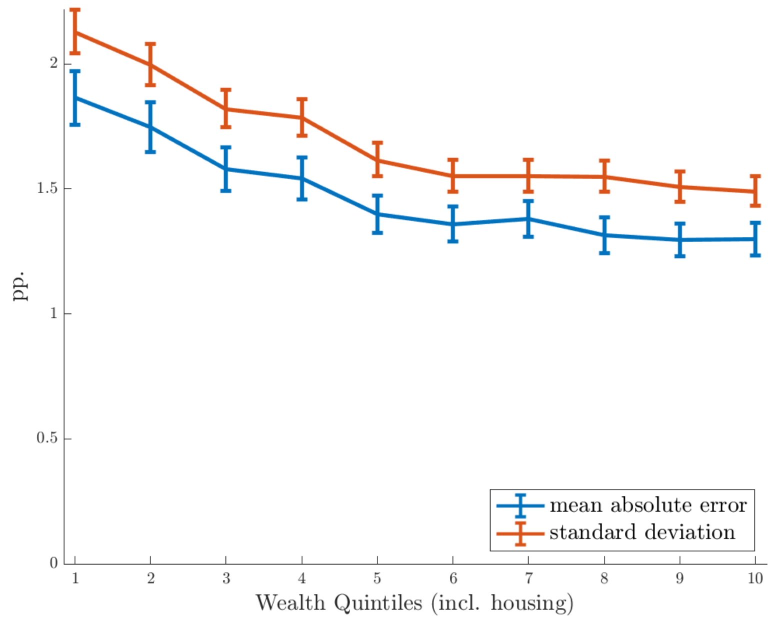


Figure 4.4: Expectation Errors by Wealth Decile Groups – No Debt

The figure plots the within-group standard deviation of errors and mean absolute errors by total financial assets. Bars provide confidence bands at the 95% level. Expectation errors are ex-ante point forecasts minus ex-post realizations. Data from DNB Household Survey waves 2010-2018.

equation (4.10).

$$abs(err^i_{t+1}) = \alpha + \sum_{d=2}^{10} \beta^d \mathbb{1}_{dec_{i,t}=d} + \gamma X_{i,t} + \epsilon_{i,t} \tag{4.10}$$

The controls $X_{i,t}$ include indicators for households age and education group as well as home ownership status.[14] Coefficients β^d must be interpreted as the difference in absolute forecast errors of households in decile d relative to households in the first wealth decile. Figure 4.5 plots the coefficients β^d along with the corresponding findings from Figure 4.1 for comparision. The similarity of both lines in Figure 4.5 suggests that controlling for age, education and homeownership does not alter the findings of absolute forecast errors along the wealth distribution. Full results are reported in Table D.2 in the appendix and show an insignificant effect of age and education (except for college degrees) but significantly lower errors for home owners.

[14]The estimation of β^d relies on variation in wealth across households. The median household is in the sample for 3 years and the analysis is carried out at annual frequency, making it difficult to obtain sufficient within household variation in wealth to be able to control for household fixed effects.

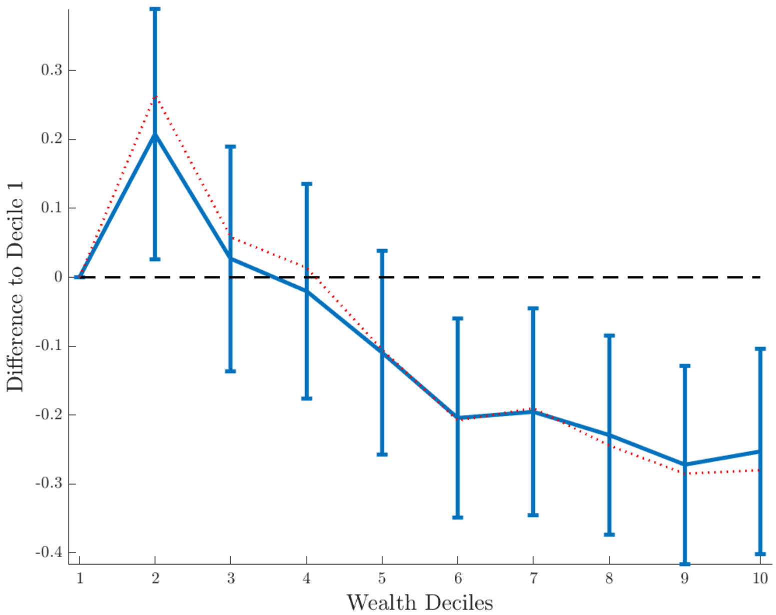


Figure 4.5: Mean Absolute Error - Individual Level Regression

The figure plots in blue the etimated coefficients β^d from running (4.10) at the household level. Bars provide confidence bands at the 95% level, based on standard errors clustered at the household level. For comparison, the red dotted line reports the corresponding results from Figure 4.1, i.e. the difference in mean absolute errors vs. the first decile group.

Additional data sources

For further robustness, we repeat our analysis with US data from the Michigan Survey of Consumers (MSC). Compared to the DHS data, the MSC contains substantially less information on household wealth. Reported stock market investment has to be used as an approximation of net financial wealth and hence there are no households with negative wealth, making it impossible to test for a hump shaped pattern. Detailed findings are provided in Appendix D.1.1. The patterns reported for the DHS are strongly supported by findings from the Michigan Survey of Consumers, i.e. both the standard deviation of errors and the mean absolute forcast error are strictly decreasing in households stock market investment.

Alternative Mechanisms

While the mechanism under study in this paper relies on wealth levels influencing households' expectation formation, alternative mechanisms might be proposed to explain the reported patterns. The literature often imposes causality to run from the level of expectations to wealth levels, higher inflation expectations implying lower savings, abstracting from any reverse effect (see e.g. Crump et al., 2015; Vellekoop and Wiederholt, 2019). There are two important differences, in timing and in the moments considered, compared to the present paper. First, while this literature speaks to how the level of expectations on inflation between yesterday and today impact today's wealth levels, we focus on how today's wealth impacts dispersion in expectations between today and tomorrow. Second, while previous work has focussed on the mean inflation forecast at the household level, we focus on the dispersion in forecast errors across households.

A recent literature has found heterogeneity in realized inflation rates due to heterogeneity in households' consumption baskets, higher income households experiencing lower inflation (see e.g. Argente and Lee, 2021; Jaravel, 2019; Kaplan and Schulhofer-Wohl, 2017). While systematic differences in realized inflation rates could account for the observed dispersion in expectations across groups, it is unlikely to account for the pattern of dispersion in expectations within wealth deciles along the distribution as consumption baskets are strongly correlated with income/wealth.

Another possible driver of the observed patterns could be that more financially literate households are at the same time wealthier and better able to form expectations about inflation. Direct measures of financial literacy are available in the DHS only for a special module in 2005, i.e. not in our sample (Deuflhard et al., 2019). However, patterns are robust by education and age group, both likely correlated with financial literacy.

In our analysis we focus on net financial wealth and abstract from portfolio composition. This is justified by looking at the share of wealth held through checking, savings or deposit accounts, savings certificates and deposit books in the DHS. These assets arguably have predetermined, nominal interest rates and therefore all carry the same one-for-one exposure to inflation. In our sample the share of these assets over total assets is around 80% for most parts of the wealth distribution except for the wealthiest households, where it drops to about 50% for the top decile. Therefore, changes in portfolio composition are unlikely to account for the decline in dispersion of forecast errors between the second and sixth decile of the wealth distribution. In theory, an active portfolio choice would also make incentives to learn about inflation an increasing function of beginning of period wealth. As long as we assume portfolio composition to be adjustable at annual frequency, what matters is not the initial exposure to different assets but overall wealth, as the total size of the portfolio determines the benefits of forming precise expectations and possibly adjusting its composition going forward.[15]

Taking stock

Our empirical findings show that both the standard deviation as well as the mean absolute error across households co-move with households' wealth in a meaningful way: Richer and indebted households exhibit lower dispersion and mean absolute errors in their forecasts of inflation compared to their counterparts around zero net wealth. The findings are robust to covariation with age or education as well as our definition of wealth and can be replicated in US data. These results are in line with our theory of expectation formation if richer and indebted households choose to exert higher effort to learn about future inflation. We use these findings to discipline a consumption-savings model with endogenous expectation formation of heterogeneous households in the following section.

4.4 Savings Choice and Endogenous Expectations

Building on the results of the previous sections, we are now in a position to incorporate the expectation formation presented in Section 4.2 into an infinite-horizon consumption-savings model. The model explicitly considers the effect of wealth on households' expectation formation and at the same time allows us to trace out their responses to changes in expectations. The dynamic setting generates a joint distribution of expectations and wealth, which can be validated against the empirical findings in Section 4.3.

[15] See Peress (2004) for theoretical results along these lines and the related discussion in Section 4.4.3.

4.4.1 Household Problem with Endogenous Expectations

At the beginning of each period, a household knows the assets carried over from the previous period a and learns about his real income y as well as the current inflation rate π, which are both stochastic over time. Together, these variables determine the available resources for consumption and saving. Based on this information, the household decides on his effort n. After deciding on n, he receives a signal about the shock to inflation between the current and the next period and updates his belief about future inflation. He will base his choice over consumption today and savings on the updated belief. Households' income is assumed to follow a Markov process with transition matrix Π_y. We assume income y to be real income.[16] Savings and borrowing are subject to a nominal interest rate. We abstract from interest rate risk and assume the nominal interest between any two periods to be constant at r^n. We do so to discuss the effect of inflation risk in isolation. Qualitatively, the findings presented below rely on this assumption only to the extent that nominal interest rates do not move one-for-one with inflation. As long as nominal rates co-move disproportionately, changes in inflation will induce fluctuations in the real interest rate. A constant nominal interest rate together with real income define inflation in our model effectively as a risk only to the real interest rate.

With all other notation as introduced above, households' information choice problem is given as

$$\tilde{V}(a, y, \pi) = \max_{n \in [0, \bar{n}]} \mathbb{E}_{\hat{e}'}[V(a, y, \pi, n, \hat{e}')|n], \tag{4.11}$$

where we restrict the choice of effort to be positive and impose an upper limit $\bar{n}$ on how much the households can learn about future inflation to rule out perfect foresight.

[16]This choice is motivated by the fact that labour income, the largest component of non-financial income, for the Netherlands over our sample period is to a large extend protected from inflation through collective bargaining agreements. According to OECD data, collective bargaining coverage in the Netherlands was well above 80% for the period under study.

The subsequent consumption-savings choice, conditional on chosen effort n and received signal $\hat{e}'$, can be described as the solution to

$$\begin{aligned} V(a,y,\pi,n,\hat{e}') = \max_{c,a'} & \left(c^{1-\gamma} + \beta \left(\mathop{\mathbb{E}}_{\pi',y'} [\tilde{V}(a',y',\pi')^{1-\alpha} | \hat{e}', n, \pi, y] \right)^{\frac{1-\gamma}{1-\alpha}} \right)^{\frac{1}{1-\gamma}} \\ s.t. \quad & c + a' = \frac{1+r^n}{1+\pi} a + y - \mathcal{F}(n) \\ & a' \geq \bar{a}, \; c \geq 0, \end{aligned} \tag{4.12}$$

where the budget constraint is written in real terms, a is today's nominal asset level divided by yesterday's prices and $\bar{a}$ is the borrowing limit.[17] Expectations over π' are based on households' updated belief taking into consideration π, $\hat{e}$ and the previous choice for n. The law of motion of inflation and the expectation formation based on the signal are as presented in Section 4.2. Preferences of the household are recursive as in Epstein and Zin (1989), allowing for independence of risk aversion and intertemporal substitution.

We model the cost of effort as a monetary cost, representing both the opportunity cost of spending time on forming expectations as well as the cost of acquiring information. For the cost of effort and the relationship between effort and noise in the signal, we assume functional forms

$$\sigma_s(n) = \frac{\chi}{1+n} \quad \text{and} \quad \mathcal{F}(n) = (\theta n)^{\phi}. \tag{4.13}$$

These choices yield convex cost of and convex gains from exerting effort.[18] Note that with these functional forms, χ is the variation in the noise if zero effort is exerted, i.e. the maximum variation possible, and that zero effort implies zero cost.

4.4.2 Calibration

The calibration of the model aims to replicate the patterns presented in Figure 4.1. Our calibration strategy is twofold: First, a range of parameters is set exogenously. These include preference parameters $\gamma = 1.5$ and $\alpha = 8$ which

[17]For details see appendix D.2.2.

[18]$\sigma_s'(n) < 0$, $\sigma_s''(n) > 0$ and $\mathcal{F}'(n) > 0$, $\mathcal{F}''(n) \geq 0$, iff $\phi \geq 1$.

we chose in line with previous work.[19] We furthermore assume the cost of information to be quadratic ($\phi = 2$). The inflation process is estimated from Dutch annual inflation rates for the period 1988-2018. This yields a long run mean of about 2% and an annual persistence of about 0.5, similar to the estimates of Vellekoop and Wiederholt (2019). The nominal interest rate r^n is set at 4% for a steady state real rate of 2%. Second, we calibrate β, $\bar{a}$, θ, χ and $\bar{n}$ as well as the process for y jointly for the model to fit the data on households' wealth and their expectation errors along the wealth distribution. Calibration targets include the position of the peak of households' errors in the second decile, the beginning of the flattened part of the standard deviation of errors in the sixth decile, the magnitude of the drop in error standard deviation of 0.57pp[20] as well as the share of wealth held by each decile of the wealth distribution. All parameters (and their interaction) influence a wide range of model statistics. Nevertheless, β and $\bar{a}$ are particularly important to determine the lower end of the wealth distribution while θ, χ and $\bar{n}$ reproduce the slope and level of errors along the wealth distribution. Bounding n generates a flat standard deviation of errors across high wealth groups. Under our calibration, the maximum possible effort $\bar{n}$ reduces the standard deviation of noise to 0.5pp, half the standard deviation of shocks to the inflation rate. Exerting effort $\bar{n}$ comes at a cost of less than 0.1% of average income, speaking to the fact that little is necessary to deter households from acquiring information about future inflation. As in Castaneda et al. (2003), the process for y is calibrated to generate the distribution of wealth. Similar to their results, one high earnings state with lower persistence is necessary to generate a long right tail of the wealth distribution. Table 4.1 summarizes our parameter choices.

Table 4.2 presents the fit of our model with respect to the wealth distribution. The model performs well along this dimension. It only struggles to match the strong concentration of wealth at the top as well as the total amount of debt. We argue, that the failure to match the concentration at the top has negligible relevance

[19]Papers applying Epstein-Zin preferences in a consumption-savings framework with idiosyncratic risk include Cooper and Zhu (2016), Ampudia et al. (2018), Campanale and Sartarelli (2018) and Kaplan and Violante (2014). They agree about the intertemporal elasticity of substitution. We chose the risk aversion from the lower end of the range of their estimates, a conservative choice closer to more standard CRRA preferences.

[20]We target the difference between the standard deviation of errors in the second decile group versus the average over deciles 6-10.

Table 4.1: Dynamic Model – Calibration

	Parameter	Value	Target
intertemp. substitution	γ	1.5	Literature
risk aversion	α	8.0	Literature
time preference	β	0.9779	fraction of debtors
borrowing limit	$\bar{a}$	-7.5	total debt
income states	y	[0.45 1 8]	wealth distribution
income transition	Π_y	$\begin{bmatrix} 0.975 & 0.025 & 0 \\ 0.057 & 0.931 & 0.012 \\ 0 & 0.15 & 0.85 \end{bmatrix}$	wealth distribution
nominal interest rate	r^n	0.04	2% SS real rate
persistence inflation	ρ	0.5	Dutch data, 1988-2018
long-run mean inflation	μ	0.02	Dutch data, 1988-2018
std. inflation shocks	σ_e	0.01	Dutch data, 1988-2018
curvature cost of effort	ϕ	2	quadratic cost of effort
scale cost of effort	θ	0.0015	range flat std. errors
maximum std. of noise	χ	0.1	peak std. errors
upper bound on effort	$\bar{n}$	17.5	low std. errors

for our results regarding expectations, since the model performs much better in matching the total fraction of wealth held by the flat part of the expectation distribution (wealth deciles 6-10 jointly). As expectation formation is similar within this range, not matching the correct distribution of wealth within the upper half of the wealth distribution will not have consequences for our results regarding expectation formation.[21] The failure to match the total amount of debt arises from the difficulty to match jointly total debt as well as the fraction of debtors, which we share with many similar models.[22] The model does well with respect to the fraction of indebted households, a feature important to match the position of the peak in the expectations distribution. It falls short in fully matching the amount of net liabilities of indebted agents. This imprecision is slightly biasing the standard deviation of expectation errors and mean absolute errors in the first wealth decile upwards, as we will see below.

[21]To match the top tail of the wealth distribution, the literature often introduces heterogeneity in time preferences (see e.g. (Krusell and Smith, 1998)). Introducing such a positive correlation between wealth levels and households' weight on future utility would only strengthen our results further as it would make high wealth households care even more about future inflation.

[22]To match both jointly we would need to introduce additional model features, such as e.g. a wedge between borrowing rates and the return on savings, from which we abstract here to keep the exposition as simple as possible.

Table 4.2: Wealth Distribution

Decile	1	2	3	4	5	6	7	8	9	10
Data	-6.14%	-0.01%	0.35%	1.01%	2.14%	3.82%	6.01%	10.07%	18.81%	63.94%
Model	-2.64%	-0.88%	0.65%	2.24%	4.01%	6.14%	9.00%	13.63%	22.45%	45.39%

Data refers to net financial wealth in the DNB Household Survey (waves 2010-2018). Compared to simulated, model implied wealth distribution.

4.4.3 Endogenous Expectations along the Wealth Distribution

Figure 4.6 presents the model implied equivalent to our baseline empirical findings in Figure 4.1. The model matches well qualitatively and quantitatively the differences in both the standard deviation of errors across households and their mean absolute forecast errors along the wealth distribution: A peak in the second wealth decile, a flattening over wealth deciles 6-10 and the quantitative magnitude of the decline between deciles 2 and 6. The model captures qualitatively the untargeted decline in both the standard deviation and mean absolute error for the first wealth decile vis-à-vis the second. As in the data, the first decile consists of households with negative net wealth. The shortfall in reproducing the quantitative magnitude of this decline is due to the left tail of the wealth distribution in the model being less spread out compared to the left tail of the net wealth distribution in the data. Where the model is off by the largest margin quantitatively is the level of both the standard deviation and the mean absolute error. In the data, both curves are about one percentage point higher than in the model. Note, however, that in order to isolate the effect of the proposed mechanism we abstract entirely from any exogenous dispersion in beliefs such as e.g. fundamental disagreement about the long run mean μ or heterogeneous biases in the signal. The level of error dispersion is in line with the fraction attributed to our mechanism in Figure D.1 after controlling for disagreement in long-run means. Exogenously imposing additional sources of dispersion would likely shift the reported measures up and towards the data equivalent.

We have shown in Section 4.2 that, with our model of expectation formation, the driving force behind changes in mean absolute errors and standard deviation of errors is the noise in signals households receive about future inflation rates and hence the effort they choose to reduce this noise. Our quantitative findings suggest that, indeed, wealthier and indebted households endogenously choose to

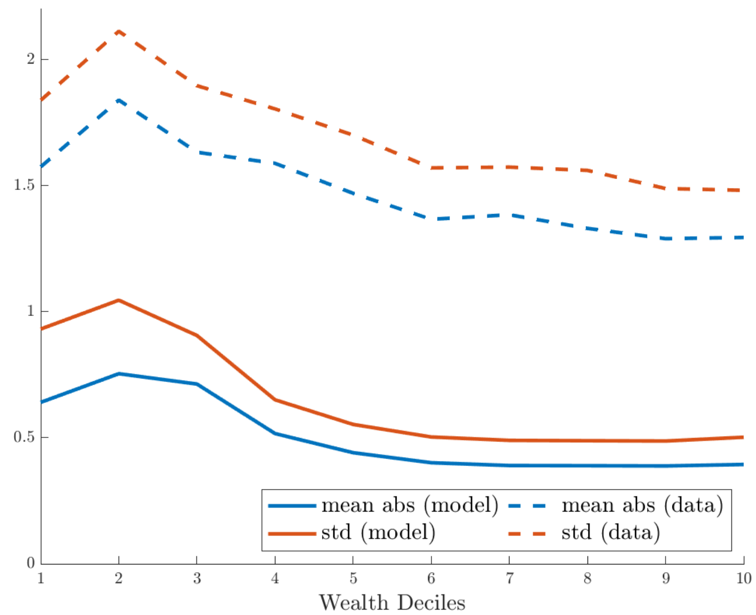


Figure 4.6: Expectation Errors by Wealth Decile Groups

Simulated, model implied statistics versus targeted data moments from Figure 4.1.

exert more such effort, enabling the model to replicate the empirical patterns. But why does the choice of effort vary with wealth? When inflation is a risk to the real interest rate, the more an agent wants to save or borrow between periods, the more he is exposed to fluctuations in the inflation rate. As future savings are positively correlated with current wealth, the richer (or the more indebted) a household is today, the more he will expose himself to inflation going forward. This exposure drives the incentives of households to exert effort and reduce the perceived uncertainty about future inflation.[23]

Discussion of Assumptions

Before we highlight potential consequences of endogenous expectations along the wealth distribution by studying households' consumption responses to signals about future inflation, we revisit four key assumptions underlying our results:

First is our choice of preferences. The qualitative finding of effort choices increasing in absolute wealth levels does not rely on the assumption of recursive preferences, it pertains also under more standard CRRA utility. A sufficient

[23]We provide a more detailed discussion of the exposure effect in a two period framework in Appendix D.3.

level of risk aversion is, however, important to quantitatively generate a steep decline in the standard deviation of errors as it leads to a stronger increase of the gains from effort with wealth. Epstein-Zin preferences allow for high risk aversion without marginalizing the intertemporal elasticity of substitution, which is important for our analysis of consumption responses to signals below.

Second is our empirical measure of wealth. What ultimately matters for the formation of households' inflation expectations in the model are beginning of period resources $\frac{1+r^n}{1+\pi}a + y$. These determine the *potential* exposure to inflation until the next period as they pin down the general range of future savings/borrowing. The *actual* exposure will then be given by the realized savings/borrowing choice within this range, but this happens only after the household has formed his expectations and is therefore endogenous to his effort choice. Motivated by the states relevant to households' expectation formation in the model, beginning of period wealth is the model-consistent empirical measure to consider.

Third, we have abstracted from modeling portfolio composition. In this regard, it is important to distinguish our analysis from work on the distributional consequences of surprises in inflation or monetary policy more general as e.g. in Doepke and Schneider (2006), Auclert (2019), and Tzamourani (2019). These papers focus on the *ex-post* distributional consequences of inflationary shocks, while we are concerned with the *ex-ante* anticipation of such shocks. In theory, households' exposure to future inflation is independent of the composition of beginning of period wealth as long as this composition is adjustable going forward. We argue that this is the case for financial wealth at annual frequency, the time horizon at which we have data and to which we calibrate the model. Households' balance sheets going forward are endogenous to their expectation formation. Including a portfolio choice into the model is likely to only strengthen results as the benefits from information in the presence of portfolio choice are increasing in wealth, shown e.g. in Peress (2004). His results suggest that when aggregate risk is distorting the relative returns of different assets, households with larger portfolios can gain more from acquiring information and rebalance their asset holdings optimally. Therefore, if inflation is distorting relative asset returns, again richer households would have higher incentives to form precise expectations.

Fourth, we have also abstracted from any exposure of non-asset (labour) income to inflation risk. Our results rely on this assumption to the extent that the exposure of labour income to inflation has to be sufficiently below the exposure of asset income. “Sufficiency" is determined by the levels of absolute risk aversion along the wealth distribution. What is important for our findings is that the residual absolute exposure to inflation, the exposure households face after controlling for all indexation of wages and asset returns to inflation, increases enough along the wealth distribution to outweigh the decrease in absolute risk aversion.[24] The high collective bargaining coverage along with the low portfolio share of potentially indexed assets or debt in our sample provide evidence for a sufficient difference in residual absolute exposure.

4.5 Expectations and Consumption Responses

To conclude the analysis, we turn to households' consumption responses to a signal about future inflation and how these depend on their wealth levels. Aggregating the individual responses yields the on-impact response of aggregate consumption to forward guidance policies, which we discuss in the final part of this section.

4.5.1 The Marginal Propensity to Consume upon Signal

The starting point to trace out aggregate effects of endogenous expectation formation is the relationship between wealth and households' marginal propensity to consume in response to a signal about future inflation rates. We will refer to this metric as *MPCS* and define it as the relative change in a household's consumption policy in response to a change in the signal he receives about tomorrows shock to inflation $\hat{\epsilon}$, when holding all other variables (a,y,π,n) constant. Defined in this way, the MPCS is the semi-elasticity of a household's consumption policy with respect to the signal he receives. This measure has two

[24]For an extended discussion of the interplay between absolute risk aversion and exposure in a two period example see Appendix D.3.

components,

$$\text{MPCS} = \underbrace{\frac{1}{c}\frac{\partial c}{\partial \mathbb{E}[\epsilon]}}_{\text{MPCE}} \times \left.\frac{\partial \mathbb{E}[\epsilon]}{\partial \hat{\epsilon}}\right|_{n}. \tag{4.14}$$

We will refer to the first term as the marginal propensity to consume in response to a change in expectations (MPCE). It is the percentage change in current consumption in response to a change in expectations about future inflation rates, i.e. the semi-elasticity of consumption with respect to expected inflation. The second term captures the change in expectations in response to a change in the signal, where $\mathbb{E}[\epsilon]$ stands for households' full subjective distribution over the future shock. Since we are concerned here only with a change in its value but not the fact that a household receives a signal and further under assumptions as above, the only moment of the subjective distribution affected by the realization of the signal is the posterior mean.[25] Applying Bayesian updating as before, in our framework the change in the subjective mean is given explicitly as

$$\left.\frac{\partial \bar{\epsilon}}{\partial \hat{\epsilon}}\right|_{n} = \omega(n) d\hat{\epsilon} = \frac{\sigma_e^2}{\sigma_e^2 + \sigma_s^2(n)} d\hat{\epsilon}. \tag{4.15}$$

It becomes clear immediately how the response of expectations to a signal depends on effort n: The more effort is exerted, i.e. the less noisy a signal is perceived to be, the more a household will respond to this signal by updating his expected mean of future inflation – a standard result of Bayesian updating.

We analyze households' MPCS' quantitatively and compute the change in current consumption for each household if he receives a signal of $\hat{\epsilon} = 0.01$ instead of $\hat{\epsilon} = 0$. We distinguish four different scenarios defined by how noisy they perceive the signal to be. An *endogenous* scenario follows our baseline model where noise is determined by the endogenous choice of effort and heterogeneous across households. To disentangle households' MPCEs from how their expectations respond to a signal, we compare this benchmark to three scenarios in which noise is equalized across households: An *inattentive* scenario, setting the perceived noise of all households equal to that of the endogenously least informed. An

[25]The standard deviation of the posterior distribution responds only to the fact that a signal is received and to the noise attached to such signal but is independent of the value the signal takes. Under our assumptions on how households' form their expectations, the change in mean and standard deviation are sufficient to characterize the response of the entire distribution.

attentive scenario, assigning to all agents the noise of the endogenously most informed households. A *flat* scenario, in which the noise of all agents is chosen in order to match the unconditional standard deviation of errors in our baseline economy. All three have in common that the second term in (4.14) is constant across households and forces them to update their expectations in response to the signal in the same way, isolating differences in their MPCE. In the endogenous scenario, we also let $\left.\frac{\partial \mathbb{E}[\epsilon]}{\partial \hat{\epsilon}}\right|_n$ vary according to the endogenous effort choices of households.

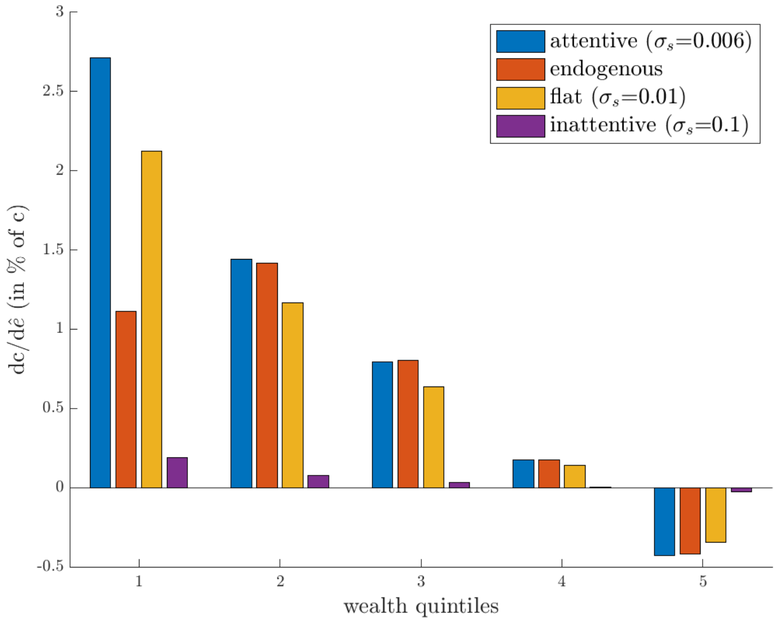


Figure 4.7: Marginal Propensity to Consume on Signal

Percentage change in consumption (aggregated by wealth quintile) on impact if $\pi = 2$ and $\hat{e}$ changes from 0 to 1pp. Endogenous: Noise as endogenously chosen. Attentive: All HHs σ_s=0.006. Inattentive: All HHs σ_s=0.1. Flat: All HHs σ_s=0.01.

Figure 4.7 plots the MPCS' aggregated by quintile of the wealth distribution.[26] We begin by looking at the three cases in which we keep $\left.\frac{\partial \mathbb{E}[\epsilon]}{\partial \hat{\epsilon}}\right|_n$ constant across households. The figure shows the MPCS for the inattentive, attentive and flat scenarios to be decreasing in wealth. Remember that from equation (4.14) $\text{MPCS} = \text{MPCE} \times \left.\frac{\partial \mathbb{E}[\epsilon]}{\partial \hat{\epsilon}}\right|_n$. Following this decomposition, it has to be the MPCE that is declining in wealth. This is due to the interaction of income and substitution effects in expectation of future inflation rates. For a household who previously would not have held any savings or debt between periods, a

[26]To aggregate, we use the stationary wealth distribution of the converged economy if inflation is constant at two percent.

change in expected inflation comes down to a change in the expected relative price of consumption today versus tomorrow, generating a substitution effect on current consumption. For a household who initially would have held either savings or debt, the substitution effect is accompanied by an income effect as a change in expected inflation implies a change in expected real financial income in the future. This income effect counteracts the substitution effect for saving households while it reinforces the substitution effect for borrowing households. For the case considered in Figure 4.7, a change in the signal from zero to one percentage point reveals to saving households that they will tomorrow be poorer than previously expected, hence diminishing their consumption response compared to households with little savings or debt. A good predictor for future savings in the model is the current asset level of a household, implying a MPCE declining in wealth.

The difference in the magnitude of MPCS' between the three cases with constant noise across households is driven by how much they update their expectations in response to the signal. The least informed ("inattentive") households choose a standard deviation of noise (σ_s) as high as 0.1 compared to a standard deviation of 0.01 of the actual shock (σ_e). Therefore, they attach little weight to any signal they receive ($\omega^{inatt} \approx 0.01$), do not update their beliefs in response and hence do not change their consumption behavior. This is why the MPCS for inattentive households remains low. For the flat scenario, σ_s decreases to 0.01 and hence $\omega^{flat} \approx 0.5$. For the attentive scenario we assume the standard deviation of the noise to be 0.006. Therefore, they attach more weight to any signal they receive ($\omega^{att} \approx 0.74$) and respond stronger in terms of consumption. The increase in the MPCS is not linear in ω across scenarios since a change in effort also affects household's uncertainty, i.e. the standard deviation of their inflation expectations, and hence their precautionary saving motive.

In the endogenous scenario, both terms in equation (4.14) interact. From our analysis so far we know the MPCE to be decreasing in wealth. From section 4.4.2 we know households effort choice and hence $\left.\frac{\partial \mathbb{E}[\epsilon]}{\partial \hat{\epsilon}}\right|_n$ to be increasing in wealth. The interaction between these two forces yields a hump shaped pattern of MPCS' along the wealth distribution. At the lowest wealth levels the increase in effort following an increase of resources outweighs the decline in MPCEs.

From the second quintile onwards the decline in MPCEs dominates as effort is almost constant over the upper half of the wealth distribution. The figure shows that at low levels of wealth endogenous effort leads to a substantially lower consumption response compared to the counterfactual attentive scenario. This gap is how the influence of wealth on expectation formation has an impact on macroeconomic aggregates.

4.5.2 A Forward Guidance Exercise

Campbell et al. (2012) famously coined the terms of *odyssean* and *delphic* forward guidance, the former referring to policy makers commitment to some future policy action and the latter standing in for an attempt to influence expectations about the future path of economic variables. Our model naturally lends itself to a discussion of the channel behind delphic forward guidance, as it provides an understanding into how heterogeneous households respond to signals about future inflation rates. More specifically, we can provide an approximation to how much endogenous expectation formation can decrease the effectiveness of such forward guidance policies. While a full general equilibrium analysis is beyond the scope of our setup, we will be able to capture the initial consumption response to a change in households' inflation expectations. Following Auclert and Rognlie (2020), any demand shock can be decomposed into a partial equilibrium consumption response on impact and a general equilibrium multiplier. Our results should be interpreted as capturing the initial partial equilibrium increase in aggregate demand which is then amplified through a general equilibrium multiplier.

To highlight households' response to delphic forward guidance we conduct a quantitative exercise: Assume the economy to be stationary at $\pi = 0.02$. In this economy we shift the signal of every household by 0.01, such that all signals are drawn from $\mathcal{N}(0.01, \sigma_s^2(n_t^i))$ instead of $\mathcal{N}(0, \sigma_s^2(n_t^i))$. For each household we compute the change in consumption compared to the original signal and obtain an aggregate response using the stationary distribution of households. We do so under two different assumptions about how noisy households perceive their signals to be: Our benchmark scenario, where households' choose their effort endogenously, as well as the counterfactual attentive scenario, where all households' are as informed as the most informed inside the model economy. The attentive scenario provides an upper bound on how effective forward guidance could be as it assumes all households to attach the highest possible weight to any signal received and hence a strong updating of expectations. The difference between the two scenarios provides us an estimate for the potential consumption response that forward guidance misses out on due to some households not paying attention to inflation.

Table 4.3: Forward Guidance Exercise

calibration	**attentive**	**endogenous**	**missing potential**
baseline	0.20	0.09	0.11 (55%)
adjusted	0.13	0.08	0.05 (42%)

The table reports aggregated MPCS' in pp as defined in (4.14) if signals are drawn from $\mathcal{N}(0.01, \sigma_s^2(n_t^i))$ instead of $\mathcal{N}(0, \sigma_s^2(n_t^i))$ when the economy is stationary at $\pi = 0.02$. The first row reports results for our baseline calibration, the second row for an alternative calibration with $\bar{n} = 10$.

The first row in table 4.3 presents results for our baseline calibration. It shows that due to endogenous expectation formation forward guidance loses approximately 55% of its consumption response on impact, a sizable drop in the partial equilibrium response necessary to trigger any general equilibrium effects. As outlined in the previous section and especially Figure 4.7, the missing potential lies with households around zero net wealth who exert little effort in forming precise expectations, perceive any signal about future inflation as noisy, and hence do not update their expectations despite having the largest potential consumption response if they would do so. Any higher order (general equilibrium) effects that rely on this initial trigger will also be attenuated. Reaching those households' to whom higher future inflation does not imply a decrease in future income from asset holdings could therefore substantially increase the effectiveness of delphic forward guidance policies. Central banks should take this into account when designing the communication of their policies.

It is important to set this result in relation to previous work on the role of frictional expectation formation for the effectiveness of forward guidance policies, such as e.g. Wiederholt (2015) or Angeletos and Lian (2018). While most of this literature has focussed on the overall effect of imperfect expectation formation compared to a full information counterfactual, the result highlighted in this section is driven by differentials in expectation formation across households and how they are correlated with the general responsiveness to the policy announcement. This is also why our counterfactual is not a full information economy but one where we eliminate differences in attention across households.

Our baseline calibration has attributed the entire decline in the standard deviation of errors along the wealth distribution to endogenous factors and, in this regard, provides an upper bound on the effect of differences in expectation formation across households on forward guidance.[27] To test the robustness of our estimate to this assumption, we adjust the calibration to match the decline in subjective uncertainty after controlling for dispersion in beliefs about the long-run mean of inflation μ as presented in Appendix D.2.1. This provides some lower bound as it assumes any decline in dispersed beliefs about μ to be entirely exogenous, restricting the endogenous gap of attention between high and low wealth households. Instead of a decline of 0.57 between peak and low of the standard deviation of errors, we now target a drop of only 0.34. This target is met by adjusting $\bar{n}$ to 10 and keeping all other parameters as they were before. The second row of Table 4.3 presents results for this alternative calibration. While in general the response of consumption is weaker than under the baseline calibration due to the reduced attentiveness (and hence reduced updating of expectations upon a signal) of the most informed households, forward guidance still loses about 42% of its initial effect on consumption when moving from maximum attention of all households' to endogenous expectation formation. This is due to the first marginal reduction in noise increasing ω more than the last and the non-linear effects of ω on consumption responses due to precautionary saving behavior.

4.6 Concluding Remarks

This paper provides a framework to discuss the joint formation of households' inflation expectations and savings choices. We argue that wealth levels are

[27]Along another dimension, the failure of the calibrated model to match the top of the wealth distribution dampens the consequences of the proposed mechanism. With higher inequality in wealth, as observed empirically, the effects of endogenous expectation formation would increase further. More wealth inequality implies larger dispersion in MPCEs and hence even more importance for who pays attention to future inflation rates.

important for both the formation of expectations and households' response to expected inflation. Looking at empirical observations from the DHS dataset, the standard deviation of forecast errors and mean absolute errors are declining in absolute wealth. We exploit changes in these cross-sectional statistics along the wealth distribution to discipline a consumption-savings problem with endogenous expectation formation, where households can exert effort to reduce uncertainty about future inflation rates. The model matches the empirical observations. The mechanism behind this finding works through the heterogeneous exposure to inflation that households at different points in the wealth distribution face. The model allows us to back out marginal propensities to consume in response to signals about future inflation. These MPCS' are hump shaped in wealth, driven by a negative correlation between households' consumption response to expected inflation and the change in their expectations in response to signals. At the aggregate level, small MPCS' of low wealth households (due to a lack of attention to inflation) can substantially reduce the effectiveness of forward guidance policies.

While an empirical analysis of MPCS' lies beyond the scope of this paper, others have conducted related work in the DHS dataset: Lieb and Schuffels (2019) find the likelihood of positive durable consumption expenditure in response to higher inflation expectations to be decreasing in wealth. Similarly, Coibion et al. (2019) report a stronger decline in durable consumption in response to (exogenously) higher inflation expectations for households with higher wealth levels. This can be seen as support for MPCEs declining in wealth due to the interaction of income and substitution effects. Unrelated to inflation but in line with our theory, Fuster et al. (2020) find more exposed participants to be willing to pay a higher cost for information about future house prices in an experiment. More work along these lines is necessary for a full empirical evaluation of our theory, especially with regard to the effect wealth has on how expectations respond to signals.

Our paper also leaves room for further theoretical work on the topic. One possible addition to the analysis presented here can be to include a portfolio choice into our model. As mentioned before, such an extension is unlikely to alter the findings presented in this paper. It might nevertheless yield interesting additional results on the implications of costly inflation expectations for wealth inequality, as suggested by the findings of Peress (2004) and Lei (2019). While we focus on uncertainty and endogenous expectations about the shock to inflation rates, the model can be extended to other sources of heterogeneity in expectations such as learning about the underlying model. Our extension to include fundamental disagreement provides a starting point for work in this direction. More importantly, a computationally demanding but interesting application of the mechanism described in this paper would be to introduce our model of expectation formation into a general equilibrium environment. Recent work by Carroll et al. (2020) and Auclert et al. (2020) has included imperfect expectations in general equilibrium models with heterogeneous households. These papers rely so far on exogenous updating of expectations. It would be important to understand the impact of our findings on MPCS' in their general equilibrium setting. We leave these extensions for future research.

References

Aguiar, M. and Bils, M. (2015). Has consumption inequality mirrored income inequality? *American Economic Review*, 105(9):2725–56.

Aguiar, M. and Hurst, E. (2005). Consumption versus expenditure. *Journal of Political Economy*, 113(5):919–948.

Aguiar, M. and Hurst, E. (2007). Life-cycle prices and production. *American Economic Review*, 97(5):1533–1559.

Aguiar, M., Hurst, E., and Karabarbounis, L. (2013). Time use during the great recession. *American Economic Review*, 103(5):1664–1696.

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

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.

Albrecht, J., Menzio, G., and Vroman, S. (2021). Vertical differentiation in frictional product markets.

Alessandria, G. (2009). Consumer search, price dispersion, and international relative price fluctuations. *International Economic Review*, 50(3):803–829.

Ampudia, M., Cooper, R., Le Blanc, J., and Zhu, G. (2018). MPC heterogeneity in Europe: Sources and policy implications. Working Paper 25082, National Bureau of Economic Research.

Anderson, E., Rebelo, S., and Wong, A. (2020). Markups across space and time.

Angeletos, G.-M. and Lian, C. (2018). Forward guidance without common knowledge. *American Economic Review*, 108(9):2477–2512.

Angelini, D. and Brès, M. (2022). Aging consumers, market structure, and growth.

Antunes, A., Cavalcanti, T., Mendicino, C., Peruffo, M., and Villamil, A. (2020). Tighter credit and consumer bankruptcy insurance. *CEPR Discussion Paper No. DP14330.*

Argente, D. and Lee, M. (2021). Cost of living inequality during the great recession. *Journal of the European Economic Association*, 19(2):913–952.

Armantier, O., Bruine de Bruin, W., Topa, G., van der Klaauw, W., and Zafar, B. (2015). Inflation expectations and behavior: Do survey respondents act on their beliefs? *International Economic Review*, 56(2):505–536.

Arslan, Y., Guler, B., and Kuruscu, B. (2020). Credit supply driven boom-bust cycles.

Arslan, Y., Guler, B., and Taskin, T. (2021). Price search, consumption inequality and expenditure inequality over the life-cycle. *International Economic Review*, 62(1):295–320.

Attanasio, O. P. and Pistaferri, L. (2016). Consumption inequality. *Journal of Economic Perspectives*, 30(2):3–28.

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

Auclert, A., Bardóczy, B., Rognlie, M., and Straub, L. (2021). Using the sequence-space jacobian to solve and estimate heterogeneous-agent models. *Econometrica*, 89(5):2375–2408.

Auclert, A. and Rognlie, M. (2020). Inequality and aggregate demand.

Auclert, A., Rognlie, M., and Straub, L. (2020). Micro jumps, macro humps: Monetary policy and business cycles in an estimated hank model. Working Paper 26647, National Bureau of Economic Research.

Auer, R., Burstein, A., Lein, S. M., and Vogel, J. (2022). Unequal expenditure switching: Evidence from switzerland.

Bai, Y., Rios-Rull, J.-V., and Storesletten, K. (2019). Demand shocks as technology shocks.

Bardóczy, B. (2020). Spousal insurance and the amplification of business cycles.

Baron, M., Verner, E., and Xiong, W. (2021). Banking Crises Without Panics. *Quarterly Journal of Economics*, 136:51–113.

Baxter, M. and King, R. (1991). Productive externalities and business cycles. Discussion Paper / Institute for Empirical Macroeconomics 53, Federal Reserve Bank of Minneapolis.

Bayer, C., Lütticke, R., Pham-Dao, L., and Tjaden, V. (2019). Precautionary savings, illiquid assets, and the aggregate consequences of shocks to household income risk. *Econometrica*, 87(1):255–290.

Ben-David, I., Fermand, E., Kuhnen, C. M., and Li, G. (2018). Expectations uncertainty and household economic behavior. Working Paper 25336, National Bureau of Economic Research.

Bernanke, B. S., Gertler, M., and Gilchrist, S. (1999). Chapter 21 the financial accelerator in a quantitative business cycle framework. volume 1 of *Handbook of Macroeconomics*, pages 1341–1393. Elsevier.

Bewley, T. (1977). The permanent income hypothesis: A theoretical formulation. *Journal of Economic Theory*, 16(2):252–292.

Bils, M. and Klenow, P. J. (2001). Quantifying quality growth. *American Economic Review*, 91(4):1006–1030.

Birinci, S. (2019). Spousal labor supply response to job displacement and implications for optimal transfers.

Bisgaard Larsen, R. and Weissert, C. J. (2020). Quality and consumption basket heterogeneity – transitory shocks and implications for consumption-savings behaviour.

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

Bredtmann, J., Otten, S., and Rulff, C. (2018). Husband's unemployment and wife's labor supply: the added worker effect across Europe. *ILR Review*, 71(5):1201–1231.

Broda, C., Leibtag, E., and Weinstein, D. E. (2009). The role of prices in measuring the poor's living standards. *Journal of Economic Perspectives*, 23(2):77–97.

Broda, C. and Parker, J. A. (2014). The economic stimulus payments of 2008 and the aggregate demand for consumption. *Journal of Monetary Economics*, 68:20–36.

Broer, T., Kohlhas, A., Mitman, K., and Schlafmann, K. (2018). Heterogenous information choice in general equilibrium.

Brunnermeier, M. K. and Sannikov, Y. (2014). A macroeconomic model with a financial sector. *American Economic Review*, 104:379–421.

Burdett, K. (1978). A theory of employee job search and quit rates. *The American Economic Review*, 68(1):212–220.

Burdett, K. and Judd, K. L. (1983). Equilibrium price dispersion. *Econometrica*, 51(4):955–969.

Burdett, K. and Menzio, G. (2018). The (q, s, s) pricing rule. *The Review of Economic Studies*, 85(2):892–928.

Burdett, K. and Mortensen, D. T. (1998). Wage differentials, employer size, and unemployment. *International Economic Review*, 39(2):257–273.

Butters, G. R. (1977). Equilibrium distributions of sales and advertising prices. *The Review of Economic Studies*, 44(3):465–491.

Campanale, C. and Sartarelli, M. (2018). Life-cycle wealth accumulation and consumption insurance.

Campbell, J. R., Evans, C. L., Fisher, J. D., and Justiniano, A. (2012). Macroeconomic effects of federal reserve forward guidance. *Brookings Papers on Economic Activity*, pages 1–80.

Carroll, C. D. (2006). The method of endogenous gridpoints for solving dynamic stochastic optimization problems. *Economics Letters*, 91(3):312–320.

Carroll, C. D., Crawley, E., Slacalek, J., Tokuoka, K., and White, M. N. (2020). Sticky expectations and consumption dynamics. *American Economic Journal: Macroeconomics*, 12(3):40–76.

Castaneda, A., Diaz-Gimenez, J., and Rios-Rull, J.-V. (2003). Accounting for the US earnings and wealth inequality. *Journal of Political Economy*, 111(4):818–857.

CBO (2021). Report on the troubled asset relief program - July 2021.

Chéron, A., Hairault, J.-O., and Langot, F. (2011). Age-dependent employment protection. *The Economic Journal*, 121(557):1477–1504.

Chéron, A., Hairault, J.-O., and Langot, F. (2013). Life-cycle equilibrium unemployment. *Journal of Labor Economics*, 31(4):843–882.

Chetty, R. (2008). Moral hazard versus liquidity and optimal unemployment insurance. *Journal of Political Economy*, 116(2):173–234.

Choi, S. and Valladares-Esteban, A. (2020). On households and unemployment insurance. *Quantitative Economics*, 11(1):437–469.

Christiano, L. J., Motto, R., and Rostagno, M. (2014). Risk shocks. *American Economic Review*, 104(1):27–65.

Cloyne, J., Ferreira, C., and Surico, P. (2020). Monetary policy when households have debt: new evidence on the transmission mechanism. *Review of Economic Studies*, 87(1):102–129.

Coibion, O., Georgarakos, D., Gorodnichenko, Y., and van Rooij, M. (2019). How does consumption respond to news about inflation? field evidence from a randomized control trial. Working Paper 26106, National Bureau of Economic Research.

Coibion, O. and Gorodnichenko, Y. (2012). What can survey forecasts tell us about information rigidities? *Journal of Political Economy*, 120(1):116–159.

Coibion, O., Gorodnichenko, Y., and Hong, G. H. (2015). The cyclicality of sales, regular and effective prices: Business cycle and policy implications. *American Economic Review*, 105(3):993–1029.

Coibion, O., Gorodnichenko, Y., and Koustas, D. (2021). Consumption inequality and the frequency of purchases. *American Economic Journal: Macroeconomics*, 13(4):449–482.

Coibion, O., Gorodnichenko, Y., Kueng, L., and Silvia, J. (2017). Innocent bystanders? monetary policy and inequality. *Journal of Monetary Economics*, 88:70–89.

Comin, D., Lashkari, D., and Mestieri, M. (2021). Structural change with long-run income and price effects. *Econometrica*, 89(1):311–374.

Cooper, R. and Ejarque, J. (2000). Financial intermediation and aggregate fluctuations: a quantitative analysis. *Macroeconomic Dynamics*, 4(4):423–447.

Cooper, R. and Zhu, G. (2016). Household finance over the life-cycle: What does education contribute? *Review of Economic Dynamics*, 20:63–89.

Crump, R. K., Eusepi, S., Tambalotti, A., and Topa, G. (2015). Subjective intertemporal substitution. *Federal Reserve Bank of New York Staff Report 734*.

Cullen, J. B. and Gruber, J. (2000). Does unemployment insurance crowd out spousal labor supply? *Journal of Labor Economics*, 18(3):546–572.

Davis, S. J. and von Wachter, T. (2011). Recessions and the costs of job loss. *Brookings Papers on Economic Activity*, pages 1–73.

De Giorgi, G. and Gambetti, L. (2017). Business cycle fluctuations and the distribution of consumption. *Review of Economic Dynamics*, 23:19–41.

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

DellaVigna, S. and Gentzkow, M. (2019). Uniform pricing in us retail chains. *The Quarterly Journal of Economics*, 134(4):2011–2084.

Deuflhard, F., Georgarakos, D., and Inderst, R. (2019). Financial literacy and savings account returns. *Journal of the European Economic Association*, 17(1):131–164.

Doepke, M. and Schneider, M. (2006). Inflation and the redistribution of nominal wealth. *Journal of Political Economy*, 114(6):1069–1097.

Dräger, L. and Nghiem, G. (2018). Are consumers' spending decisions in line with an euler equation?

Eggertsson, G. B. and Krugman, P. (2012). Debt, deleveraging, and the liquidity trap: A fisher-minsky-koo approach. *Quarterly Journal of Economics*, 127(3):1469–1513.

Ellieroth, K. (2019). Spousal insurance, precautionary labor supply, and the business cycle-a quantitative analysis.

Elsby, M. W. L., Hobijn, B., and Şahin, A. (2015). On the importance of the participation margin for labor market fluctuations. *Journal of Monetary Economics*, 72:64–82.

Epstein, L. G. and Zin, S. E. (1989). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica*, 57(4):937–969.

Erosa, A. and Gervais, M. (2002). Optimal taxation in life-cycle economies. *Journal of Economic Theory*, 105(2):338–369.

Faber, B. and Fally, T. (2022). Firm heterogeneity in consumption baskets: Evidence from home and store scanner data. *The Review of Economic Studies*, 89(3):1420–1459.

Farmer, L. E. and Toda, A. A. (2017). Discretizing nonlinear, non-gaussian markov processes with exact conditional moments. *Quantitative Economics*, 8(2):651–683.

Fernández-Blanco, J. (2020). Unemployment risks and intra-household insurance. *Working Paper*.

Fernández-Villaverde, J., Hurtado, S., and Nuno, G. (2019). Financial frictions and the wealth distribution. Technical report, National Bureau of Economic Research.

Ferrante, F. and Gornemann, N. (2021). Financial frictions and the re-distributive effects of exchange rate fluctuations.

Ferraro, D. and Valaitis, V. (2022). Wealth and hours.

Ferriere, A., Grübener, P., Navarro, G., and Vardishvili, O. (2022). On the optimal design of transfers and income-tax progressivity.

Flood, S., King, M., Rodgers, R., Ruggles, S., and Warren, J. R. (2020). Integrated public use microdata series, current population survey: Version 8.0 [dataset]. Minneapolis, MN: Ipums. *Dataset*.

Fuster, A., Perez-Truglia, R., Wiederholt, M., and Zafar, B. (2020). Expectations with Endogenous Information Acquisition: An Experimental Investigation. *The Review of Economics and Statistics*, pages 1–54.

Gaballo, G. and Paciello, L. (2021). Spending allocation under nominal uncertainty: A model of effective price rigidity.

Garcia-Perez, J. and Rendon, S. (2020). Family job search and wealth: the added worker effect revisited. *Quantitative Economics*, 11(4).

Gertler, M. and Karadi, P. (2011). A model of unconventional monetary policy. *Journal of monetary Economics*, 58(1):17–34.

Gertler, M. and Kiyotaki, N. (2010). Financial intermediation and credit policy in business cycle analysis. In *Handbook of monetary economics*, volume 3, pages 547–599. Elsevier.

Glover, A., Heathcote, J., Krueger, D., and Ríos-Rull, J.-V. (2020). Intergenerational redistribution in the great recession. *Journal of Political Economy*, 128(10):3730–3778.

Golosov, M. and Lucas, R. E. (2007). Menu costs and phillips curves. *Journal of Political Economy*, 115(2):171–199.

Griffith, R., Leibtag, E., Leicester, A., and Nevo, A. (2009). Consumer shopping behavior: How much do consumers save? *Journal of Economic Perspectives*, 23(2):99–120.

Griffy, B. S. (2021). Search and the sources of life-cycle inequality. *International Economic Review*.

Guerrieri, V. and Lorenzoni, G. (2017). Credit crises, precautionary savings, and the liquidity trap. *The Quarterly Journal of Economics*, 132(3):1427–1467.

Guler, B., Guvenen, F., and Violante, G. L. (2012). Joint-search theory: New opportunities and new frictions. *Journal of Monetary Economics*, 59(4):352–369.

Guner, N., Kulikova, Y., and Valladares-Esteban, A. (2020). Does the added worker effect matter? *Working Paper*.

Guvenen, F., Kaplan, G., and Song, J. (2021a). The glass ceiling and the paper floor: Changing gender composition of top earners since the 1980s. *NBER Macroeconomics Annual*, 35:309–373.

Guvenen, F., Karahan, F., Ozkan, S., and Song, J. (2021b). What do data on millions of us workers reveal about lifecycle earnings dynamics? *Econometrica*, 89(5):2303–2339.

Haan, P. and Prowse, V. L. (2017). Optimal social assistance and unemployment insurance in a life-cycle model of family labor supply and savings. *Working Paper*.

Hall, R. E. (2018). New evidence on the markup of prices over marginal costs and the role of mega-firms in the us economy.

Halla, M., Schmieder, J., and Weber, A. (2020). Job displacement, family dynamics, and spousal labor supply. *American Economic Journal: Applied Economics*, 12(4):253–87.

Handbury, J. (2021). Are poor cities cheap for everyone? non-homotheticity and the cost of living across us cities. *Econometrica*, 89(6):2679–2715.

He, Z. and Krishnamurthy, A. (2019). A macroeconomic framework for quantifying systemic risk. *American Economic Journal: Macroeconomics*, 11(4):1–37.

Heathcote, J., Perri, F., and Violante, G. L. (2020a). The rise of us earnings inequality: Does the cycle drive the trend? *Review of Economic Dynamics*, 37:S181–S204.

Heathcote, J., Storesletten, K., and Violante, G. L. (2020b). Optimal progressivity with age-dependent taxation. *Journal of Public Economics*, 189:104074.

Hintermaier, T. and Koeniger, W. (2010). The method of endogenous gridpoints with occasionally binding constraints among endogenous variables. *Journal of Economic Dynamics and Control*, 34(10):2074–2088.

Huo, Z. and Ríos-Rull, J.-V. (2015). Tightening financial frictions on households, recessions, and price reallocations. *Review of Economic Dynamics*, 18(1):118–139.

Iacoviello, M. (2015). Financial business cycles. *Review of Economic Dynamics*, 18(1):140–163.

Iskhakov, F., Jørgensen, T. H., Rust, J., and Schjerning, B. (2017). The endogenous grid method for discrete-continuous dynamic choice models with (or without) taste shocks. *Quantitative Economics*, 8(2):317–365.

Jaimovich, N., Rebelo, S., and Wong, A. (2019). Trading down and the business cycle. *Journal of Monetary Economics*, 102:96–121.

Jaravel, X. (2019). The unequal gains from product innovations: Evidence from the us retail sector. *The Quarterly Journal of Economics*, 134(2):715–783.

Jaravel, X. and Olivi, A. (2021). Prices, non-homotheticities, and optimal taxation.

Jarosch, G. (2015). Searching for job security and the consequences of job loss.

Jermann, U. and Quadrini, V. (2012). Macroeconomic effects of financial shocks. *American Economic Review*, 102(1):238–71.

Jordà, Ò. (2005). Estimation and inference of impulse responses by local projections. *American economic review*, 95(1):161–182.

Justiniano, A., Primiceri, G. E., and Tambalotti, A. (2019). Credit supply and the housing boom. *Journal of Political Economy*, 127(3):1317–1350.

Kaplan, G. and Menzio, G. (2015). The morphology of price dispersion. *International Economic Review*, 56(4):1165–1206.

Kaplan, G. and Menzio, G. (2016). Shopping externalities and self-fulfilling unemployment fluctuations. *Journal of Political Economy*, 124(3):771–825.

Kaplan, G., Menzio, G., Rudanko, L., and Trachter, N. (2019). Relative price dispersion: Evidence and theory. *American Economic Journal: Microeconomics*, 11(3):68–124.

Kaplan, G., Moll, B., and Violante, G. L. (2018). Monetary policy according to HANK. *The American Economic Review*, 108(3):697–743.

Kaplan, G. and Schulhofer-Wohl, S. (2017). Inflation at the household level. *Journal of Monetary Economics*, 91:19–38.

Kaplan, G. and Violante, G. (2021). The marginal propensity to consume in heterogeneous agents models. Technical report.

Kaplan, G. and Violante, G. L. (2014). A model of the consumption response to fiscal stimulus payments. *Econometrica*, 82(4):1199–1239.

Kiyotaki, N. and Moore, J. (1997). Credit cycles. *Journal of political economy*, 105(2):211–248.

Kospentaris, I. (2021). Unobserved heterogeneity and skill loss in a structural model of duration dependence. *Review of Economic Dynamics*, 39:280–303.

Krueger, D., Mitman, K., and Perri, F. (2016). Macroeconomics and household heterogeneity. In *Handbook of Macroeconomics*, volume 2, pages 843–921.

Krusell, P., Mukoyama, T., Rogerson, R., and Şahin, A. (2017). Gross worker flows over the business cycle. *The American Economic Review*, 107(11):3447–3476.

Krusell, P., Mukoyama, T., and Şahin, A. (2010). Labour-market matching with precautionary savings and aggregate fluctuations. *The Review of Economic Studies*, 77(4):1477–1507.

Krusell, P. and Smith, A. A. (1998). Income and wealth heterogeneity in the macroeconomy. *Journal of Political Economy*, 106(5):867–896.

Kryvtsov, O. and Vincent, N. (2021). The cyclicality of sales and aggregate price flexibility. *The Review of Economic Studies*, 88(1):334–377.

Laeven, L. and Valencia, F. (2013). Systemic banking crises database. *IMF Economic Review*, 61(2):225–270.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies*, 76(3):1071–1102.

Lee, S., Luetticke, R., and Ravn, M. O. (2021). Financial frictions: Macro vs micro volatility.

Lei, X. (2019). Information and inequality. *Journal of Economic Theory*, 184.

Lentz, R. (2009). Optimal unemployment insurance in an estimated job search model with savings. *Review of Economic Dynamics*, 12(1):37–57.

Lieb, L. and Schuffels, J. (2019). Inflation expectations and consumer spending: The role of household balance sheets.

Lundberg, S. (1985). The added worker effect. *Journal of Labor Economics*, 3(1, Part 1):11–37.

Mackowiak, B., Matejka, F., and Wiederholt, M. (2018). Survey: Rational inattention, a disciplined behavioral model.

Madeira, C. and Zafar, B. (2015). Heterogeneous inflation expectations and learning. *Journal of Money, Credit and Banking*, 47(5):867–896.

Maloney, T. (1987). Employment constraints and the labor supply of married women: A reexamination of the added worker effect. *Journal of Human Resources*, 22(1):51–61.

Maloney, T. (1991). Unobserved variables and the elusive added worker effect. *Economica*, 58(230):173–187.

Mankart, J. and Oikonomou, R. (2016a). Household search and the aggregate labour market. *The Review of Economic Studies*, 84(4):1735–1788.

Mankart, J. and Oikonomou, R. (2016b). The rise of the added worker effect. *Economics Letters*, 143:48–51.

Mankart, J., Oikonomou, R., and Pascucci, F. (2021). The rise of household insurance.

McCall, J. J. (1970). Economics of information and job search. *The Quarterly Journal of Economics*, 84(1):113–126.

McKay, A., Nakamura, E., and Steinsson, J. (2016). The power of forward guidance revisited. *The American Economic Review*, 106(10):3133–3158.

Mendicino, C., Nikolov, K., Supera, D., and Ramirez, J.-R. (2020). Twin defaults and bank capital requirements. *CEPR Discussion Paper DP14427*.

Mendicino, C., Nord, L., and Peruffo, M. (2022). Distributive effects of banking sector losses.

Menzio, G. (2021). Optimal product design: Implications for competition and growth under declining search frictions.

Menzio, G. and Shi, S. (2011). Efficient search on the job and the business cycle. *Journal of Political Economy*, 119(3):468–510.

Menzio, G., Telyukova, I. A., and Visschers, L. (2016). Directed search over the life cycle. *Review of Economic Dynamics*, 19:38–62.

Meyer, B. D. and Sullivan, J. X. (2013). Consumption and income inequality and the great recession. *American Economic Review*, 103(3):178–83.

Michelacci, C., Paciello, L., and Pozzi, A. (2022). The extensive margin of aggregate consumption demand. *The Review of Economic Studies*, 89(2):909–947.

Michelacci, C. and Ruffo, H. (2015). Optimal life cycle unemployment insurance. *The American Economic Review*, 105(2):816–59.

Midrigan, V. (2011). Menu costs, multiproduct firms, and aggregate fluctuations. *Econometrica*, 79(4):1139–1180.

Moen, E. R. (1997). Competitive search equilibrium. *Journal of Political Economy*, 105(2):385–411.

Mongey, S. and Waugh, M. (2022). Pricing inequality.

Mortensen, D. (2003). *Wage Dispersion: Why Are Similar Workers Paid Differently?* MIT press.

Mortensen, D. T. (1970). A theory of wage and employment dynamics. *Microeconomic foundations of employment and inflation theory*, pages 167–211.

Mortensen, D. T. and Pissarides, C. A. (1994). Job creation and job destruction in the theory of unemployment. *The Review of Economic Studies*, 61(3):397–415.

Nakamura, E. (2008). Pass-through in retail and wholesale. *American Economic Review*, 98(2):430–437.

Nevo, A. and Wong, A. (2019). The elasticity of substitution between time and market goods: Evidence from the great recession. *International Economic Review*, 60(1):25–51.

Ortigueira, S. and Siassi, N. (2013). How important is intra-household risk sharing for savings and labor supply? *Journal of Monetary Economics*, 60(6):650–666.

Panel Study of Income Dynamics (2021). Produced and distributed by the survey research center, institute for social research, University of Michigan, Ann Arbor, MI (2021). *Dataset*.

Peress, J. (2004). Wealth, Information Acquisition, and Portfolio Choice. *The Review of Financial Studies*, 17(3):879–914.

Petrongolo, B. and Pissarides, C. A. (2001). Looking into the black box: A survey of the matching function. *Journal of Economic Literature*, 39(2):390–431.

Petrosky-Nadeau, N. and Wasmer, E. (2015). Macroeconomic dynamics in a model of goods, labor, and credit market frictions. *Journal of Monetary Economics*, 72:97–113.

Philippon, T. (2015). Has the us finance industry become less efficient? on the theory and measurement of financial intermediation. *American Economic Review*, 105(4):1408–1438.

Pisano, L., Riva, L., and Stella, A. (2022). Price heterogeneity and consumption inequality.

Powell, J. H. (2020). New economic challenges and the fed's monetary policy review. At "Navigating the Decade Ahead: Implications for Monetary Policy," an economic policy symposium sponsored by the Federal Reserve Bank of Kansas City, Jackson Hole, Wyoming.

Pulkkinen, T. E. and Rosengren, E. S. (1993). Lessons from the Rhode Island banking crisis. *New England Economic Review*.

Pytka, K. (2022). Shopping frictions and household heterogeneity: Theory and empirics.

Reinhart, C. M. and Rogoff, K. S. (2009). *This time is different: Eight centuries of financial folly*. Princeton University Press.

Sangani, K. (2022). Markups across the income distribution: Measurement and implications.

Sara-Zaror, F. (2022). Expected inflation and welfare: The role of consumer search.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665–690.

Stephens, M. (2002). Worker displacement and the added worker effect. *Journal of Labor Economics*, 20(3):504–537.

Stock, J. H. and Watson, M. W. (2003). Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature*, 41(3):788–829.

Stroebel, J. and Vavra, J. (2019). House prices, local demand, and retail prices. *Journal of Political Economy*, 127(3):1391–1436.

Tauchen, G. (1986). Finite state markov-chain approximations to univariate and vector autoregressions. *Economics Letters*, 20(2):177–181.

Tzamourani, P. (2019). The interest rate exposure of euro area households.

Varian, H. R. (1980). A model of sales. *The American Economic Review*, 70(4):651–659.

Vellekoop, N. and Wiederholt, M. (2019). Inflation expectations and choices of households.

Wang, H. (2019). Intra-household risk sharing and job search over the business cycle. *Review of Economic Dynamics*, 34:165–182.

Weinzierl, M. (2011). The surprising power of age-dependent taxes. *The Review of Economic Studies*, 78(4):1490–1518.

Wiederholt, M. (2015). Empirical properties of inflation expectations and the zero lower bound.

Wu, C. and Krueger, D. (2021). Consumption insurance against wage risk: Family labor supply and optimal progressive income taxation. *American Economic Journal: Macroeconomics*, 13(1):79–113.

Young, E. R. (2010). Solving the incomplete markets model with aggregate uncertainty using the krusell–smith algorithm and non-stochastic simulations. *Journal of Economic Dynamics and Control*, 34(1):36 – 41.

Appendix A

Appendix to Chapter 1

A.1 Extensions to the Retailer Problem

A.1.1 Entry and Fixed Cost of Operating

Consider a version of the price posting problem outlined in Section 1.2.1 where retailers selling variety j are subject to a per-period fixed cost of operating K_j and the total mass of retailers M_j is determined endogenously by free entry. With all other notation as before, the total profits of a retailer posting price p in the adjusted model are given by

$$\tilde{\pi}_j(p) = \frac{C_j}{M_j}\left[(1-\bar{s}_j) + \bar{s}_j 2(1-F_j(p))\right](p-\kappa_j) - K_j = \frac{\pi_j(p)}{M_j} - K_j,$$

where $\pi_j(p)$ is retailers' profits of posting p in the version of the model without fixed cost of operating and a fixed mass one of retailers. To sustain an equilibrium distribution of posted prices retailers have to be again indifferent between all prices on the support of the posted distribution. Take two posted prices p_1 and p_2, indifference requires

$$\begin{aligned}
&\tilde{\pi}_j(p_1) = \tilde{\pi}_j(p_2) \\
\Rightarrow &\frac{\pi_j(p_1)}{M_j} - K_j = \frac{\pi_j(p_2)}{M_j} - K_j \\
\Rightarrow &\pi_j(p_1) = \pi_j(p_2).
\end{aligned}$$

The indifference condition between prices is independent of M_j and K_j, i.e. independent of entry and fixed cost of operating, and identical to the condition in the original model. This implies the distribution of posted prices $F_j(p)$ is identical to the model without entry and fixed cost. To solve the model with fixed cost and entry, one can therefore first recover the posted price distribution as well as the constant profits at any price on the support of $F_j(p)$, denoted $\bar{\pi}_j$, in the original model and solve for M_j given this solution. Free entry requires zero total profits of operating, i.e. $\tilde{\pi}_j(p) = 0$. The equilibrium mass of retailers is therefore given by $M_j = \frac{\bar{\pi}_j}{K_j}$.

A.1.2 Heterogeneous Marginal Cost

Take the setup from Section 1.2.1 but consider a continuous distribution of retailers over marginal cost, with CDF $\Gamma_j(\kappa)$ and support $[\underline{\kappa}_j, \bar{\kappa}_j]$ and assume $\bar{\kappa}_j = \bar{p}_j$. I.e. consider a distribution of active retailers for which the support has to end at the maximum willingness to pay. This assumption imposes no restriction on the solution as no retailer with marginal cost above $\bar{p}_j$ could ever make a sale with positive profits. Profits of a retailer with marginal cost κ of posting price p for variety j are given by

$$\pi_j(p,\kappa) = (p-\kappa)\left((1-\bar{s}_j) + 2\bar{s}_j(1-F_j(p))\right)C_j.$$

Define $p(\kappa)$ as the set of prices maximizing $\pi_j(p,\kappa)$ for given $F_j(p)$, i.e. the indifference set of posted prices for a retailer with marginal cost κ.

Solving for the Distribution of Posted Prices

To solve for the equilibrium distribution of posted prices I follow closely the steps of Burdett and Mortensen (1998) or Mortensen (2003) for a similar model of wage posting.

1. **Properties of the Distribution**

   By similar argument as in Burdett and Judd (1983), the posted distribution $F_j(p)$ has no mass points, has a connected support, and the upper bound of the support of $F_j(p)$ is $\bar{p}_j$. Intuitively, all three can be shown by providing a profitable deviation in price posting if an posted price distribution is violating one of the three conditions.

2. **Prices Posted Are Weakly Increasing in Marginal Cost**

   For any $\kappa'' > \kappa'$, $p' \in p(\kappa')$, and $p'' \in p(\kappa'')$ it has to hold that $\pi_j(p', \kappa') > \pi_j(p'', \kappa'')$ and $p'' \geq p'$, i.e. profits are strictly decreasing and prices are weakly increasing in marginal costs. To do so, note the following

$$\begin{aligned}
\pi_j'(p', \kappa') &= (p' - \kappa')\left((1 - \bar{s}_j) + 2\bar{s}_j(1 - F_j(p'))\right)C_j && (\star 1)\\
&\geq (p'' - \kappa')\left((1 - \bar{s}_j) + 2\bar{s}_j(1 - F_j(p''))\right)C_j && (\star 2)\\
&> (p'' - \kappa'')\left((1 - \bar{s}_j) + 2\bar{s}_j(1 - F_j(p''))\right)C_j = \pi_j(p'', \kappa'') && (\star 3)\\
&\geq (p' - \kappa'')\left((1 - \bar{s}_j) + 2\bar{s}_j(1 - F_j(p'))\right)C_j, && (\star 4)
\end{aligned}$$

   where the steps from $(\star 1)$ to $(\star 2)$ and $(\star 3)$ to $(\star 4)$ follow from the optimality of $p' \in p(\kappa')$ and $p'' \in p(\kappa'')$ respectively and the step from $(\star 2)$ to $(\star 3)$ from $\kappa'' > \kappa'$. From above, it is immediately clear that $\pi_j(p', \kappa') > \pi_j(p'', \kappa'')$, i.e. profits are strictly decreasing in κ. To see that $p'' \geq p'$ note that $(\star 1) - (\star 4) \geq (\star 2) - (\star 3) > 0$ and hence

$$(\kappa'' - \kappa')\left((1 - \bar{s}_j) + 2\bar{s}_j(1 - F_j(p'))\right) \geq (\kappa'' - \kappa')\left((1 - \bar{s}_j) + 2\bar{s}_j(1 - F_j(p''))\right),$$

   which yields $F_j(p'') \geq F_j(p')$ and therefore, as any cumulative distribution cannot be decreasing, $p'' \geq p'$. So any price optimal at κ' cannot be higher than any price optimal at κ''. Hence, $p(\kappa')$ and $p(\kappa'')$ can intersect in at most one boundary point. With a continuous distribution of marginal cost, the latter also implies that $p(\kappa)$ has to be single valued.

3. **The Price Distribution is a Shifted Distribution of Marginal Cost**

   By the single value property of $p(\kappa)$

$$F_j(p) = F_j(p(\kappa)) = \Gamma_j(\kappa)$$

   and hence

$$F_j'(p(\kappa)) = f_j(p(\kappa)) = \frac{\Gamma_j'(\kappa)}{p'(\kappa)}$$

4. **The Price Function $p(\kappa)$ Solves Retailers Profit Maximization**

   Analogue to before, the profits of a retailer with marginal cost κ posting price p are given by

$$\pi_j(p,\kappa) = (p-\kappa)\left((1-\bar{s}_j)+2\bar{s}_j(1-F_j(p))\right)C_j$$

   and the profit maximizing price satisfies

$$\frac{\partial \pi}{\partial p} = \big((1-\bar{s}_j)+2\bar{s}_j(1-F_j(p))-(p-\kappa)2\bar{s}_jF'(p)\big)\,C_j = 0,$$

   which yields

$$1 = \frac{(p-\kappa)2\bar{s}_jF'(p)}{(1-\bar{s}_j)+2\bar{s}_j(1-F_j(p))}$$

   and by the result of 3.)

$$p'(\kappa) = \frac{(p(\kappa)-\kappa)2\bar{s}_j\Gamma_j'(\kappa)}{(1-\bar{s}_j)+2\bar{s}_j(1-\Gamma_j(\kappa))}.$$

   This differential equation together with the boundary condition $p(\bar{\kappa}) = \bar{\kappa} = \bar{p}$ pins down the unique solution to $p(\kappa)$ and hence to $F'(p)$. The boundary condition holds because the upper bound of any price distribution has to be at $\bar{p}$ (else there are profitable deviations) and a firm with marginal cost $\kappa = \bar{\kappa} = \bar{p}$ will only be willing to post this price.

5. **Obtaining a Solution**

   Define

$$T(k) = log\left((1-\bar{s}_j)+2\bar{s}_j(1-\Gamma_j(\kappa))\right)$$

   such that

$$T'(k) = \frac{-2\bar{s}_j\Gamma_j(\kappa)}{(1-\bar{s}_j)+2\bar{s}_j(1-\Gamma_j(\kappa))}$$

   We can the rewrite the first difference equation pinning down the pricing function as

$$p'(\kappa) = -(p(\kappa)-\kappa)T'(\kappa) \Rightarrow p'(\kappa)+T'(\kappa)p(\kappa) = \kappa T'(\kappa).$$

Therefore, any solution has to satisfy (multiply both sides by $e^{T(\kappa)}$ before integrating)

$$p(\kappa)e^{T(\kappa)} = \int_{\underline{\kappa}}^{\kappa} xT'(x)e^{T(x)}dx + A = \kappa e^{T(\kappa)} - \underline{\kappa}e^{T(\underline{\kappa})} - \int_{\underline{\kappa}}^{\kappa} e^{T(x)}dx + A,$$

where the second equality follows from integration by parts. Hence

$$p(\kappa) = \kappa + e^{-T(\kappa)}\left[A - \underline{\kappa}e^{T(\underline{\kappa})} - \int_{\underline{\kappa}}^{\kappa} e^{T(x)}dx\right].$$

Using the boundary condition $p(\bar{\kappa}) = \bar{\kappa}$ it follows that

$$A = \underline{\kappa}e^{T(\underline{\kappa})} + \int_{\underline{\kappa}}^{\bar{\kappa}} e^{T(x)}dx.$$

The solution to the pricing function and the distribution of posted prices is hence given as

$$p(\kappa) = \kappa + e^{-T(\kappa)}\int_{\kappa}^{\bar{\kappa}} e^{T(x)}dx = \kappa + \int_{\kappa}^{\bar{\kappa}} \frac{(1-\bar{s}_j) + 2\bar{s}_j(1-\Gamma_j(x))}{(1-\bar{s}_j) + 2\bar{s}_j(1-\Gamma_j(\kappa))}dx$$

with derivative

$$p'(\kappa) = 1 - 1 - (-2\bar{s}_j\Gamma_j'(\kappa))\int_{\kappa}^{\bar{\kappa}} \frac{(1-\bar{s}_j) + 2\bar{s}_j(1-\Gamma_j(x))}{[(1-\bar{s}_j) + 2\bar{s}_j(1-\Gamma_j(\kappa))]^2}dx > 0$$

and therefore

$$F'(p) = \frac{\Gamma_j'(\kappa)}{p'(\kappa)} = \frac{1}{2\bar{s}_j\int_{\kappa}^{\bar{\kappa}} \frac{(1-\bar{s}_j)+2\bar{s}_j(1-\Gamma_j(x))}{[(1-\bar{s}_j)+2\bar{s}_j(1-\Gamma_j(\kappa))]^2}dx} = \frac{[(1-\bar{s}_j) + 2\bar{s}_j(1-\Gamma_j(\kappa))]^2}{2\bar{s}_j(\bar{\kappa}-\kappa)(1+\bar{s}_j) - 4\bar{s}_j^2\int_{\kappa}^{\bar{\kappa}}\Gamma_j(x)dx}.$$

We cannot conclude anything on how the profit margins (markups) per sale are changing with marginal cost κ. To see this note that

$$p(\kappa) - \kappa = \int_{\kappa}^{\bar{\kappa}} \frac{(1-\bar{s}_j) + 2\bar{s}_j(1-\Gamma_j(x))}{(1-\bar{s}_j) + 2\bar{s}_j(1-\Gamma_j(\kappa))}dx$$

and hence

$$\frac{\partial(p(\kappa)-\kappa)}{\partial\kappa} = p'(\kappa) - 1 = -1 + 2\bar{s}_j\Gamma_j'(\kappa)\int_{\kappa}^{\bar{\kappa}} \frac{(1-\bar{s}_j)+2\bar{s}_j(1-\Gamma_j(x))}{[(1-\bar{s}_j)+2\bar{s}(1-\Gamma_j(\kappa))]^2}dx.$$

So whether markups are increasing or decreasing in marginal costs depends on the shape of the distribution $\Gamma_j(\kappa)$. Intuitively, retailers' optimization trades off higher margins (markups) against a decrease in demand, where the latter depends on the distribution of prices which in turn depends on the distribution of marginal costs.

Quantitative Results under Uniform Marginal Costs

While an analytical characterization of how the moments of the price distribution respond to changes in demand-weighted shopping effort $\bar{s}_j$ under heterogeneous marginal cost is beyond the scope of this paper, I show robustness of the analytical results for the baseline model by reporting numerical simulations. I assume a uniform distribution of marginal cost over $[\underline{\kappa}_j, \bar{\kappa}_j]$ and consider parameterizations with $\bar{p}_j \in \{1,2,3,4,5\}$, $\kappa_{min} \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7\}$ such that $\underline{\kappa}_j = \kappa_{min}\bar{p}_j$ and $\bar{\kappa}_j = \bar{p}_j$. I take $\bar{p}_j = 3$ and $\kappa_{min} = 0$ as the baseline and change one parameter at a time, simulating 1,000,000 price draws for each combination of parameters and computing the mean and skewness of the posted price distribution.

To highlight the properties of a solution to the model with heterogeneous κ, Figure A.1 plots the pricing function $p(k)$ and CDF $F_j(p)$ as well as the analytical and simulated PDF of a single calibrated version with $\bar{\kappa} = \bar{p} = 2$, $\underline{\kappa} = 1$, $\bar{s} = 0.75$.

Figure A.2 recovers the result of skewness being a strictly increasing function of average search effort $\bar{s}$. Other parameters do not have considerable influence on the skewness of the price distribution. For the mean of posted prices the main mechanism pertains: For any combination of parameters considered the average posted price is decreasing in shopping effort. This is because the pricing function gets more and more concentrated at the maximum willingness to pay when $\bar{s}$ goes to zero.

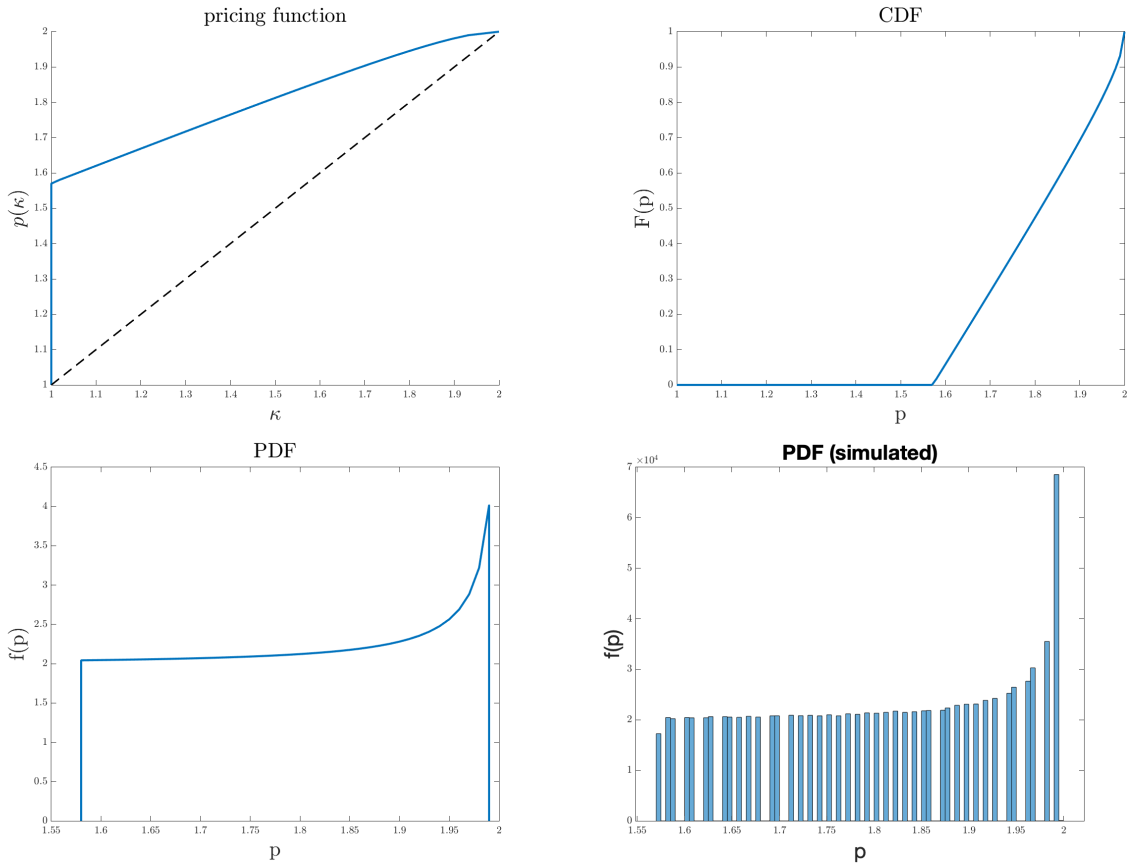


Figure A.1: Uniform Distribution - Example

Note: Model solution for a calibration with uniform distribution of marginal costs, $\bar{p} = 2$, $\underline{\kappa} = 1$, $\bar{s} = 0.75$.

Results for other types of distributions (exponential, logistic) as well as a version with a discrete set of marginal-cost types yield similar conclusions: While under some calibrations small regions of skewness decreasing in shopping effort are possible, these usually exist only for $\bar{s}_j \approx 1$ and are associated with counterfactually low levels of price dispersion. Exploiting the skewness of price distributions for an empirical test of the mechanism is therefore a reasonable approximation even in a world with potentially heterogeneous marginal cost.

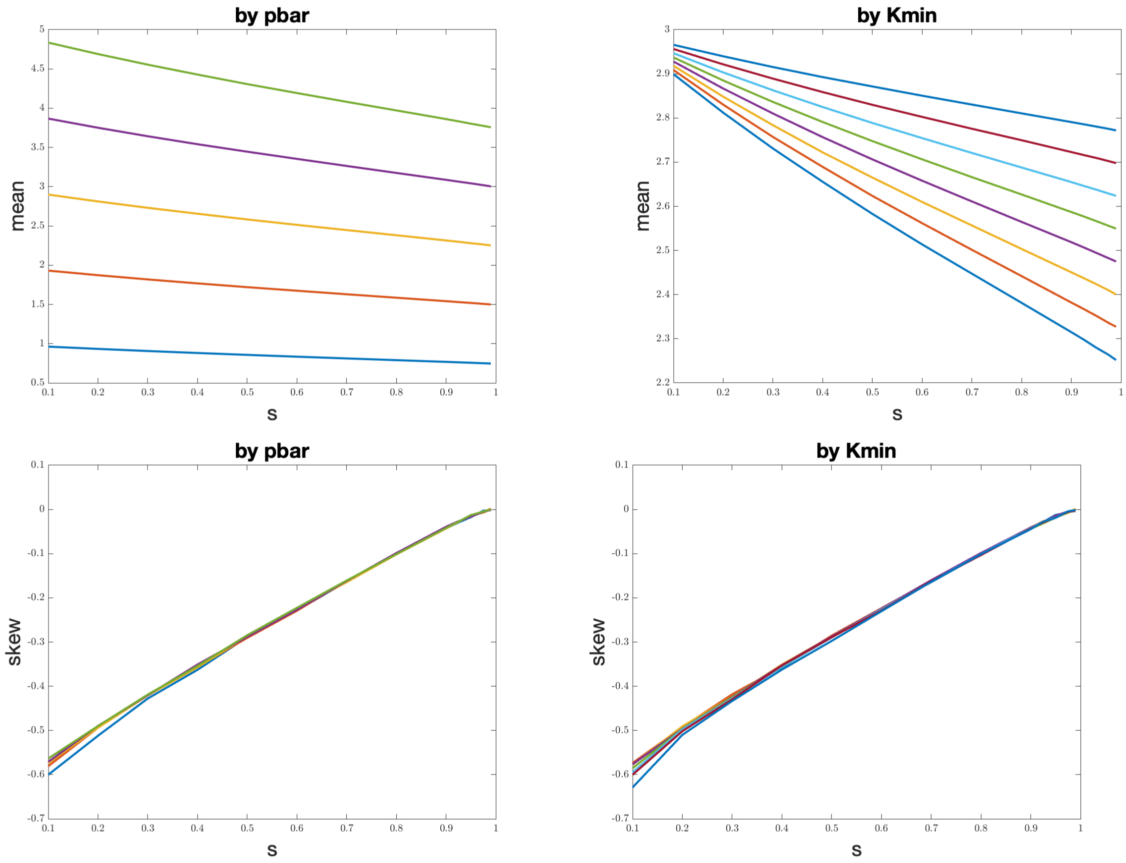


Figure A.2: Uniform Distribution - Simulations

Note: Moments of simulated price distributions over $\bar{s} \in [0, 1]$ for a calibration with uniform distribution and different values of $\bar{p}$ and $\underline{\kappa}$.

A.2 Empirical Appendix

A.2.1 The Nielsen Dataset

All empirical results presented in this paper are based on the Nielsen Consumer Panel, provided via the Kilts Center for Marketing at Chicago Booth. The dataset is a nationally representative, annual panel of around 60,000 US households who report on their grocery expenditures at daily as well as demographic information at annual frequency. Demographic variables include e.g. information on household composition, age, education, occupation, employment status, income, and location of residence. The dataset is constructed as a panel and the median household remains in the sample for about 3 consecutive waves. Nielsen applies several quality checks such as minimum reporting requirements to the sample before making data available. Households in the sample are provided with a device to record the prices and quantities of all purchases made in stores by scanning the barcodes of the items they bought (or record prices manually if the store is not participating in Nielsen's sample). The focus of the dataset is on grocery and drug stores, supermarkets and superstores, covering approximately 35% of spending excluding durable goods.[1]

Prices and quantities are reported at the barcode level. Nielsen organizes all barcodes into 10 departments (e.g. dry groceries or fresh foods), which are then divided into 125 product groups (e.g. snacks vs. pasta within dry groceries), and further split into about 1,100 product modules (e.g. potato chips vs. tortilla chips within snacks). Within product modules each variety is uniquely identified by its Universal Product Code (UPC), examples of a UPC are e.g. a box of Pringles Sour Cream and Onion or a bag of Lay's BBQ within the module potato chips. For each purchase of a barcode at a store at a given day, Nielsen records the quantity bought, the total price of the transaction, the value of all coupons used as well as the unique store identifier of the location where the purchase was made. Households' purchases can further be grouped into shopping trips, where a trip consists of all purchases of any barcode made by a household in a given store on a given day.

Data is provided in annual waves and I use the waves of 2007-2019. Data is also available for the period 2004-2006, but I focus on the later period due to a sample break between 2006 and 2007. All empirical results remain qualitatively unchanged if earlier waves are included. Across all households the dataset contains a total of about 7.5 million shopping trips and around 50 million purchases from a universe of 500,000 UPCs per wave.

No data on wealth is available in the Nielsen panel and income data is only available as categorical variable and reported as the tax base for the previous calendar year, i.e. refer to households taxable income two years prior to the sample. On the other hand, expenditures on the consumption categories covered in Nielsen are well measured. This is why for all baseline results on heterogeneity across households, I sort by their position in the expenditure distribution. Whenever I refer to expenditure, I adjust households' total annual expenditure measured in the Nielsen dataset by the square root of household size and (where applicable)

[1]For further details on the dataset and its application in Macroeconomic research see e.g. Argente and Lee (2021), Kaplan and Menzio (2015), Pisano et al. (2022), Broda and Parker (2014) or Michelacci et al. (2022).

sort them into quintiles/deciles based on their position in the expenditure distribution in the year of observation. Wherever dollar values are reported, these are adjusted to 2019 USD using the CPI for all Urban Consumers.

A.2.2 Local vs. National Average Prices

For the baseline analysis of households' shopping effort I measure the prices households pay relative to the national, annual average price across all households. The literature often defines price distributions and the relative price a household pays more narrowly, i.e. over Scantrack Market regions and by quarter (see e.g. Kaplan and Menzio, 2015). As also pointed out by Pytka (2022), this way of measuring households' shopping effort can be subject to a small sample bias. The bias can be alleviated by increasing the number of observations considered in computing average prices. In this appendix, I define the bias formally and show robustness of my main empirical findings to alternative definitions of average prices.

To measure shopping effort in the data, the literature generally compares relative prices paid and benchmarks household i's average price p_{ij} for barcode j against the average price paid $\bar{p}_j$ for the barcode across all households. This leads to a potential downward bias in the measured effect of shopping effort if household i accounts for a large share of transactions of barcode j. More formally, the price p_{ij} is defined as the quantity-weighted average over all transactions T_i of household i

$$p_{ij} = \frac{\sum_{\tau=1}^{T_i} p_{\tau ij} q_{\tau ij}}{\sum_{\tau=1}^{T_i} q_{\tau ij}},$$

where $p_{\tau ij}$ and $q_{\tau ij}$ are respectively the price paid and quantity purchased of barcode j by household i in transaction τ. The average price $\bar{p}_j$ is defined accordingly as

$$\bar{p}_j = \frac{\sum_i \sum_{\tau=1}^{T_i} p_{\tau ij} q_{\tau ij}}{\sum_i \sum_{\tau=1}^{T_i} q_{\tau ij}}.$$

One can rewrite the average price paid as

$$\bar{p}_j = \nu_{ij} p_{ij} + (1 - \nu_{ij}) p_{-ij},$$

where $\nu_{ij} = \frac{\sum_{\tau=1}^{T_i} q_{\tau ij}}{\sum_i \sum_{\tau=1}^{T_i} q_{\tau ij}}$ is household i's share in demand for variety j and

$$p_{-ij} = \frac{\sum_h \sum_{\tau=1}^{T_h} p_{\tau hj} q_{\tau hj} - \sum_{\tau=1}^{T_i} p_{\tau ij} q_{\tau ij}}{\sum_h \sum_{\tau=1}^{T_h} q_{\tau hj} - \sum_{\tau=1}^{T_i} q_{\tau ij}}$$

is the average price paid by all households except household i. The difference of household i's price relative to the average yields

$$\Delta p_{ij} = p_{ij} - \bar{p}_j = (1 - \nu_{ij})(p_{ij} - p_{-ij}).$$

While p_{-ij} is an unbiased measure of the true average price paid, the price difference will be biased towards zero by a factor $(1 - \nu_{ij})$, i.e. will be biased more the larger the demand share of household i for good j. A similar mechanism pertains for the household level price index described in Section 1.3.1.

A way to alleviate the bias is to increase the number of transactions considered to compute $\bar{p}_j$, thereby decreasing ν_{ij}. This can be done by either computing the average price for barcode j at the national, annual level or defining it at the local, quarterly level but only considering transactions for barcodes with a minimum number of transactions in the region and quarter. Alternatively, one could also drop a household's own transactions when computing the average price. However, as for many goods there are only few households consuming it in a narrow region this increases the noise in average prices and often effectively implies dropping the good if a household accounts for a significant share of local purchases of this barcode.

Table A.1 repeats the estimation in equation (1.6) for local average prices and barcodes with a minimum of 1, 25, 50 and 100 transactions respectively. The larger the minimum number of transactions, the closer the estimates (especially the coefficient on log-expenditures) get to the one obtained using national, annual average prices (column (1)). I take this as justification for focusing on the results based on national and annual average prices.

Table A.1: Shopping Effort across Households (by Number of Transactions)

	price index (national) ($N^{min} = 1$)	price index (local) ($N^{min} = 1$)	price index (local) ($N^{min} = 25$)	price index (local) ($N^{min} = 50$)	price index (local) ($N^{min} = 100$)
	(1)	(2)	(3)	(4)	(5)
log(expenditure)	0.706***	0.358***	0.456***	0.575***	0.678***
	(0.078)	(0.022)	(0.066)	(0.079)	(0.102)
income 30k-60k	0.080*	0.026	0.071	0.093	0.071
	(0.046)	(0.024)	(0.060)	(0.081)	(0.109)
income 60k-100k	0.178***	0.076**	0.208**	0.294***	0.251*
	(0.057)	(0.030)	(0.081)	(0.108)	(0.141)
income >100k	0.326***	0.113***	0.254**	0.354**	0.156
	(0.070)	(0.037)	(0.106)	(0.140)	(0.173)
1 non-employed household head	−0.236***	−0.104***	−0.163***	−0.214***	−0.287***
	(0.037)	(0.019)	(0.051)	(0.066)	(0.092)
2 non-employed household heads	−0.422***	−0.200***	−0.319***	−0.348***	−0.618***
	(0.068)	(0.036)	(0.087)	(0.111)	(0.156)
head's age 25-65	−0.013	0.021	0.095	0.093	0.211*
	(0.050)	(0.027)	(0.071)	(0.090)	(0.127)
sqrt(HH size)	0.399***	0.195***	0.222**	0.277**	0.403***
	(0.066)	(0.033)	(0.090)	(0.118)	(0.152)
FE year-state	X	X	X	X	X
FE household	X	X	X	X	X
Observations	801,398	801,398	800,320	797,812	780,476

Note: Regression of household-level price index on characteristics. Column (1) price index (prices paid vs. average price) defined based on national annual average price. Column (2) price index (prices paid vs. average price) defined based on local quarterly average price. Column (3) price index (prices paid vs. average price) defined based on local quarterly average price, restricted to products with at least $N = 25$ local quarterly observations. Column (4) price index (prices paid vs. average price) defined based on local quarterly average price, restricted to products with at least $N = 50$ local quarterly observations.. Column (5) price index (prices paid vs. average price) defined based on local quarterly average price, restricted to products with at least $N = 100$ local quarterly observations. Data obtained from Nielsen Consumer Panel waves 2007-2019. Observation weighted with Nielsen provided sample weights. Standard errors clustered at the household level. *p<0.1; **p<0.05; ***p<0.01.

A.2.3 Evidence on Demand Composition and Price Distributions

Table A.2: Demand Composition and the Skewness of Price Distributions (Robustness)

	baseline	Kelly	unweighted regression	unweighted skewness	only HH weights skewness
	(1)	(2)	(3)	(4)	(5)
expenditure quintile 2	-1.638^{***}	-0.135^{***}	-1.802^{***}	-1.450^{***}	-1.627^{***}
	(0.242)	(0.043)	(0.143)	(0.215)	(0.239)
expenditure quintile 3	-2.309^{***}	-0.226^{***}	-2.583^{***}	-2.100^{***}	-2.354^{***}
	(0.256)	(0.043)	(0.163)	(0.230)	(0.251)
expenditure quintile 4	-3.067^{***}	-0.282^{***}	-3.374^{***}	-2.793^{***}	-3.062^{***}
	(0.258)	(0.042)	(0.178)	(0.244)	(0.260)
expenditure quintile 5	-3.412^{***}	-0.382^{***}	-4.066^{***}	-3.151^{***}	-3.425^{***}
	(0.253)	(0.040)	(0.188)	(0.244)	(0.255)
FE module	X	X	X	X	X
FE quarter-SMC	X	X	X	X	X
Observations	3,026,551	2,832,442	3,026,551	3,026,551	3,026,551

Note: Regression of the skewness of price distributions on demand shares by expenditure quintile. Price distributions defined as all transactions of a barcode within a Scantrack Market Region and quarter. Demand shares defined as the share of national annual spending on a barcode by quintile. Column (1): Baseline result, observations weighted with distribution by household weights and quantities pruchased and across distributions by total expenditures included on given price distribution. Column (2): Baseline weights, Kelly's measure of skewness. Column (3): No weighting of price distributions in regressions. Column (4): No weighting of price observations within distributions. Column (4): Price observations within distributions weighted by household weights but not quantities. Data obtained from Nielsen Consumer Panel waves 2007-2019. Standard errors clustered at the barcode-year level. *p<0.1; **p<0.05; ***p<0.01.

Table A.3: Demand Composition and Price Distributions (Number of Transactions)

	Nmin = 25	Nmin = 50	Nmin = 100
	(1)	(2)	(3)
expenditure quintile 2	-1.638^{***}	-2.289^{***}	-3.162^{***}
	(0.242)	(0.528)	(1.174)
expenditure quintile 3	-2.309^{***}	-2.929^{***}	-4.268^{***}
	(0.256)	(0.570)	(1.212)
expenditure quintile 4	-3.067^{***}	-3.797^{***}	-5.186^{***}
	(0.258)	(0.569)	(1.225)
expenditure quintile 5	-3.412^{***}	-4.654^{***}	-6.436^{***}
	(0.253)	(0.556)	(1.219)
FE module	X	X	X
FE quarter-SMC	X	X	X
Observations	3,026,551	803,604	202,067

Note: Regression of the skewness of price distributions on demand shares by expenditure quintile. Price distributions defined as all transactions of a barcode within a Scantrack Market Region and quarter. Demand shares defined as the share of national annual spending on a barcode by quintile. Column (1): Only price distributions with at least $N = 25$ transactions. Column (2): Only price distributions with at least $N = 50$ transactions. Column (3): Only price distributions with at least $N = 100$ transactions. Data obtained from Nielsen Consumer Panel waves 2007-2019. Observations weighted by total expenditures included on given price distribution. Standard errors clustered at the barcode-year level. *p<0.1; **p<0.05; ***p<0.01.

A.2.4 Consumption Baskets and Separation in the Goods Market

Quantifying non-homotheticities in the data requires a measure for the similarity of consumption baskets. Define the consumption basket of any group g of households i via the share of their annual total expenditures allocated to each good ω_j^g. The expenditure share of good j for group g in a given year is given as

$$\omega_j^g = \frac{\sum_{i \in g} e_j^i}{\sum_{j \in J} \sum_{i \in g} e_j^i}.$$

The vector of expenditure shares for any given group can be seen as a distribution over a discrete set of alternatives – the universe of available products. The similarity of two such vectors, i.e. the consumption baskets of two groups of households g and h, can be measured by computing the histogram overlap $\Omega^{g,h}$

in expenditure shares, given as

$$\Omega^{g,h} = \sum_{j \in J} \min\left\{\omega_j^g, \omega_j^h\right\}.$$

Note that under homothetic preferences and the law of one price $\omega_j^g = \omega_j^h \ \forall j, g, h$ and hence $\Omega^{gh} = 1$, so any deviation of the overlap from one can be interpreted as a deviation from these assumptions. Conducting the analysis by groups of households accounts for variation in taste within groups and computing statistics at the annual frequency averages out seasonal fluctuations.

Figure A.3 reports the histogram overlap between the first and fifth quintile of the distribution of annual expenditures, defining a good at different levels of aggregation. If products are broadly defined, e.g. at the Nielsen department level, the overlap in consumption baskets is as high as 94% and even when considering product modules it is still as high as 86%. Only at the lowest level of aggregation where products are unique UPCs (the barcode level) the overlap decreases substantially to 63%. I.e. consumption baskets of high and low expenditure households exhibit a significant mismatch driven by variation in purchases of closely substitutable goods within Nielsen-defined product modules. For the empirical decomposition in Section 1.5.1 it is also important to note that conditioning on units of measurement within product modules does not alter the overlap substantially compared to considering the entire module, i.e. there are no notable non-homotheticities by unit of measurement. The overlap between any other two quintiles of the expenditure distribution exhibits similar patterns. Overlap at any level of aggregation decreases monotonically in the distance (difference in total expenditures) between two groups.

To complement the analysis based on Nielsen data for even broader categories of consumption goods, the final bar in Figure A.3 produces the overlap between the bottom and top quintile of the income distribution in the Consumer Expenditure Survey (CEX) defining goods at the 14 most aggregated categories.[2] The non-homotheticity in CEX categories is roughly at the level of Nielsen

[2] I use aggregated series for consumption by category and income quintile reported by the Bureau of Labor Statistics (BLS). The 14 expenditure categories considered include: *food at home, food away from home, alcoholic beverages, housing, apparel and services, transportation, healthcare, entertainment personal care products and services, reading, education, tobacco products and smoking supplies, miscellaneous expenditures, cash contributions, personal insurance and pensions*.

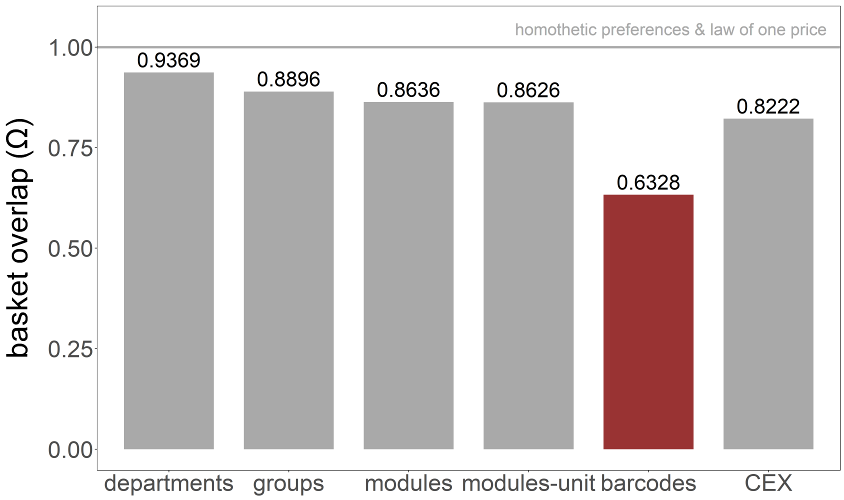


Figure A.3: Consumption Basket Overlap - Top vs. Bottom Expenditure Quintile
Note: Histogram overlap in the vector of expenditure shares for the bottom and top quintile of expenditures, by different definitions of a product. First five columns derived from Nielsen Consumer Panel, final column from Consumer Expenditure Survey (CEX). CEX column considers 14 spending categories.

defined product modules, while the barcode level overlap measured in Nielsen is approximately 25% lower.

Complementary evidence to the missing overlap in consumption baskets is a measure of how important the demand of other households is for the goods that any group of households buys. First, to determine how important demand from any group of households g is for a given good j, we define the *demand share* (DS) of group h for good j as

$$DS_j^h = \frac{\sum_{i \in h} e_j^i}{\sum_{g \in G} \sum_{i \in g} e_j^i}.$$

We can then weight the demand shares of group h with the basket of group g to compute the *cross market share* (CMS) of group h for the basket of group g, defined as

$$CMS^{gh} = \sum_{j \in J} \omega_{jt}^g DS_j^h.$$

This statistic can be interpreted as the average demand share of h in the basket of g and measures how important group h is for the demand of goods that group g buys.

Figure A.4 plots the cross market shares by quintile of the expenditure distribution at the barcode level. It shows that each group of households is substantially overrepresented in their own consumption baskets. E.g. the lowest expenditure quintile is twice as important for their own consumption basket as for the basket of the highest expenditure quintile.

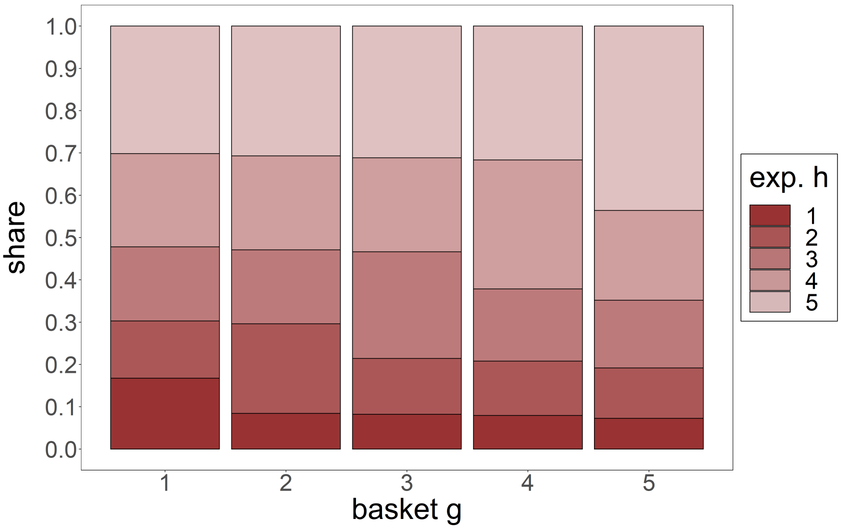


Figure A.4: Cross Market Shares

Note: Barcode-level cross market shares of expenditure quintile h for the basket of quintile g. Cross market shares are constructed weighting the share of demand for a product j coming from quintile h by the expenditure share ω_j^g of product j in the basket of quintile g. Data from Nielsen Consumer Panel.

A.2.5 Price Differences

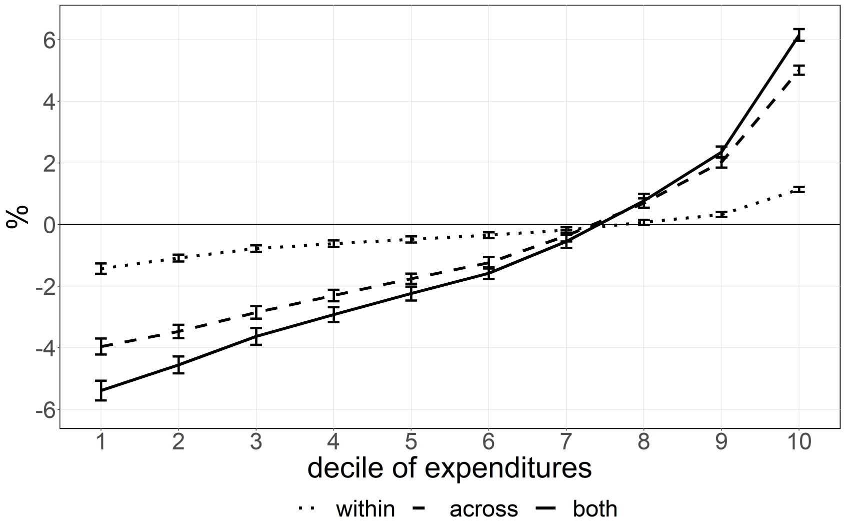


Figure A.5: Price Differences along the Expenditure Distribution – Deciles

Note: Price differences by expenditure quintile in the data, as a share of households' grocery spending. "Within" refers to differences between the price paid by a given household and the average price for a given product (direct effect of shopping), while "across" is the difference in average prices across different products. Data from the Nielsen Consumer Panel 2007-2019. The price within products is computed by barcode and the price across products is computed across barcodes within a product module (by unit of measurement). Confidence intervals bootstrapped with 100 draws at the household-year level.

Table A.4: Price Differences - Regressions

	within products	across products	trips per purchase
	(1)	(2)	(3)
log(expenditure)	0.961***	3.426***	−0.042***
	(0.069)	(0.147)	(0.001)
income	0.059	0.544***	−0.001*
30k-60k	(0.050)	(0.100)	(0.001)
income	0.186***	1.048***	−0.002**
60k-100k	(0.063)	(0.127)	(0.001)
income	0.363***	1.451***	−0.002**
>100k	(0.080)	(0.160)	(0.001)
1 non-employed	−0.284***	−0.665***	0.002***
household head	(0.041)	(0.083)	(0.000)
2 non-employed	−0.456***	−1.471***	0.004***
household heads	(0.072)	(0.160)	(0.001)
head's age	0.023	0.155	−0.001
25-65	(0.054)	(0.113)	(0.001)
sqrt(HH size)	0.544***	0.645***	−0.032***
	(0.073)	(0.148)	(0.001)
mean			0.15
FE year-state	X	X	X
FE household	X	X	X
Observations	801,398	801,398	801,398

Note: Regression of the contribution of differences between prices paid and the average price (1) within product or (2) across products to expenditure inequality on household characteristics. Contributions defined as a share of households' grocery expenditures. The price within products is computed by barcode and the price across products is computed across barcodes within a product module by unit of measurement. Column (3) number of annual shopping trips (stores visited) divided by number of annual purchases (transactions for a barcode-store-day pair). Data from the Nielsen Consumer Panel waves 2007-2019. Standard errors clustered at the household level. *p<0.1; **p<0.05; ***p<0.01.

A.3 Model Appendix

A.3.1 Income Process

For households' idiosyncratic labor productivity z, I assume an AR(1) process with innovations from a Gaussian mixture, formally defined as

$$\log(z') = \rho \log(z) + \varepsilon$$
$$\varepsilon \sim \begin{cases} \mathcal{N}(\mu_{\varepsilon,1}, \sigma^2_{\varepsilon,1}) \text{ with probability } \chi \\ \mathcal{N}(\mu_{\varepsilon,2}, \sigma^2_{\varepsilon,2}) \text{ with probability } 1-\chi \end{cases}$$

I discretize the process with 16 states for z following the method of Farmer and Toda (2017). The income process requires calibrating 6 parameters $(\rho, p, \mu_{\varepsilon,1}, \sigma^2_{\varepsilon,1}, \mu_{\varepsilon,2}, \sigma^2_{\varepsilon,2})$. I impose $\mu_{\varepsilon,2} = -\frac{\chi}{1-\chi}\mu_{\varepsilon,1}$ to obtain mean zero innovations and calibrate the remaining parameters to match five moments of annual, equivalence scale adjusted, post-tax household labor earnings: The cross-sectional variance of earnings, the standard deviation, skewness, kurtosis of annual earnings growth as well as the difference between the 90th and 10th percentile of annual earnings changes. Target values based on PSID data are obtained from De Nardi et al. (2020). For more information on how the target values are constructed see their Appendix A.3. All targets are reported in Table A.5 along with the model counterparts. The associated parameter values are $\rho = 0.91$, $\sigma_1 = 0.59$, $\sigma_2 = 0.23$, $\chi = 0.082$, and $\mu_1 = -0.57$.

Table A.5: Calibration Targets – Income Process

Targets (Annual)	**Model**	**Data**
Cross Sectional Variance (Levels)	0.61	0.57
Standard Deviation of Changes	0.33	0.33
Skewness of Changes	-0.99	-0.98
Kurtosis of Changes	10.6	10.3
P90-P10 of Changes	0.53	0.64

Note: Results of the calibration of an AR(1) income process with Gaussian-mixture shocks. The process is discretized with 16 states following Farmer and Toda (2017). Data moments for the PSID obtained from De Nardi et al. (2020).

A.3.2 Solution Method

The model is solved computationally. The solution to households' spending-savings problem is obtained by a version of the endogenous grid method (EGM) in the spirit of Carroll (2006). First, I solve for the consumption allocation and choice of shopping effort given expenditures, applying Broyden's method to the equations for households' consumption aggregator (1.10) and optimality condition for effort (1.11). The optimal choices for consumption allocation and shopping effort provide a solution for $U(e)$. I approximate the derivative $U'(e)$ numerically. With a numerical approximation of $U'(e)$ at hand, I can apply the standard EGM to solve for households' spending-savings decision.

To solve for a steady state, I iterate on a the vector of demand-weighted search effort for all varieties $\{\bar{s}_j\}_{j=1}^{J}$ jointly until a fixed point is reached. I make a guess for $\{\bar{s}_j\}_{j=1}^{J}$, solve for the implied price distributions and households' decisions, and finally aggregate households consumption and shopping policies to obtain an implied vector of demand-weighted shopping effort. I apply Broyden's method to update the guess for $\{\bar{s}_j\}_{j=1}^{J}$. Solving for aggregate dynamics requires finding a path for all demand-weighted shopping efforts $\left\{\{\bar{s}_{jt}\}_{j=1}^{J}\right\}_{t=1}^{T}$. I solve for these paths by applying the sequence-space Jacobian method of Auclert et al. (2021).

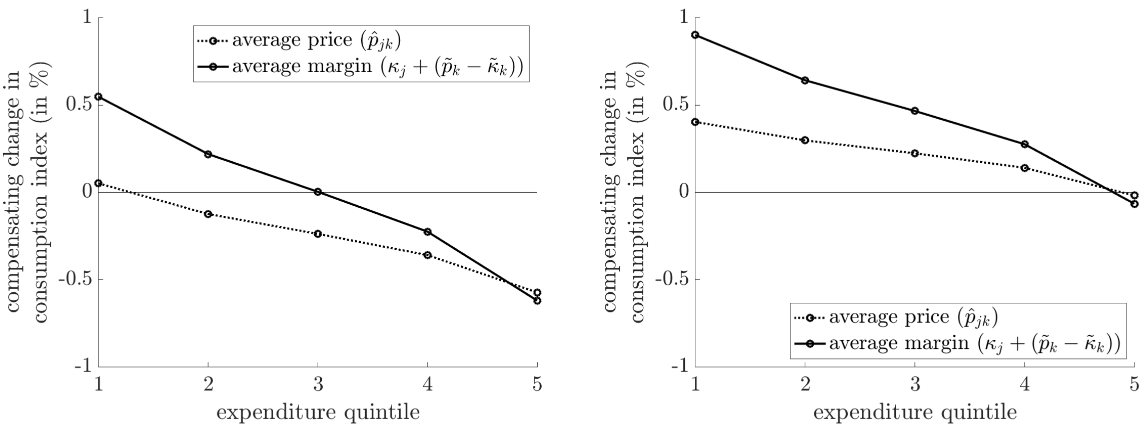
A.3.3 Welfare Implications

In addition to the one period results on welfare reported in Section 1.5, this appendix computes the long term effects of living in a counterfactual economy without search effort and alternative prices. As in the main text, I consider two counterfactuals, (i) all households paying the average price paid within each variety $\hat{p}_{jk}$ or (ii) paying the marginal cost plus average profit margin across varieties $\kappa_{jk} + (\tilde{p}_k - \tilde{\kappa}_k)$ for each variety. In both cases, there is no price dispersion for identical varieties, search effort is set to zero, and hence there is no disutility of effort.

I first compute welfare in terms of consumption-equivalent variation and ask by what percentage I would need to change consumption of a household with current idiosyncratic earnings z and assets a permanently to make her indifferent to the steady state in the counterfactual economy. Formally, this measure is defined as

$$\Delta\mathcal{C}^{CF}(z,a) = \left(\frac{V(z,a) + \frac{1}{(1-\phi)(1-\beta)}}{V^{CF}(z,a) + \frac{1}{(1-\phi)(1-\beta)}}\right)^{\frac{1}{1-\phi}} - 1,$$

where $V(z,a)$ is a households' value function in steady state and $V^{CF}(z,a)$ the value of living in the counterfactual economy.



(a) Full Effect

(b) Prices Only

Figure A.6: Permanent Welfare Effects of Shopping Effort

Note: Permanent change in total consumption index ($\mathcal{C}$) under alternative prices and zero shopping effort to make a household indifferent between living in the alternative economy and in the steady state economy. Counterfactuals allow households to optimally choose their savings and consumption baskets assuming they have to pay (i) the average price within each grocery variety or (ii) marginal cost plus the average margin across varieties. Panel (a) displays the full effect on welfare. Panel (b) uses a counterfactual measure of steady state welfare, abstracting from the disutility of effort, such that all differences are due to the alternative prices.

Figure A.6a reports the results for both counterfactuals. Its results are similar to the one-period findings presented in section 1.5.2. The consequences for the welfare of high spending households are attenuated, as taking into account the infinite horizon lets them internalize the probability of becoming a low spending household in the future. This forward looking effect reduces the impact of shopping on welfare inequality to about 1%.

Figure A.6b presents results using a counterfactual steady state value function without any disutility from shopping effort, computed as

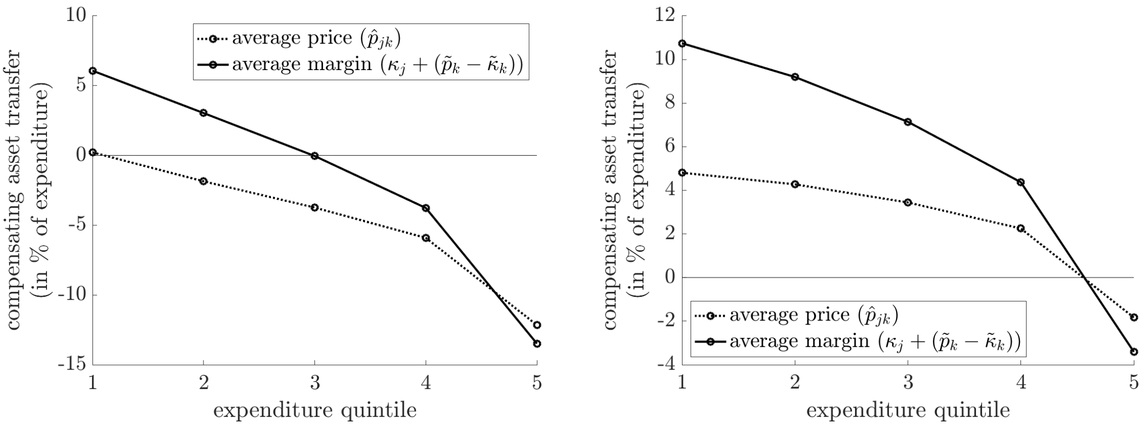
$$V^{noshop}(z,a) = u\left(\mathcal{C}^{SS}(z,a)\right) + \beta\mathbb{E}_{z'|z}V^{noshop}(z',a'),$$

where $\mathcal{C}^{SS}(z,a)$ is the steady state consumption level of a household in state (z,a). This counterfactual isolates the welfare effect of price differences as it eliminates any welfare gains from setting the disutility of shopping effort to zero under the alternative price regimes. Abstracting from the reduction in disutility makes the counterfactual economies more costly in welfare terms for all households, but more so for the top quintile (0.56% in consumption terms) as opposed to the bottom quintile (0.35%).

As the welfare results based on the non-homothetic aggregate consumption index $\mathcal{C}$ are not easily interpretable in terms of the quantities consumed, I compute an alternative measure by compensating households with a one-off change in their asset holdings $\Delta(z,a)$ for moving them between the counterfactual economies and the steady state, such that

$$V^{CF}(z,a+\Delta(z,a)) = V(z,a).$$

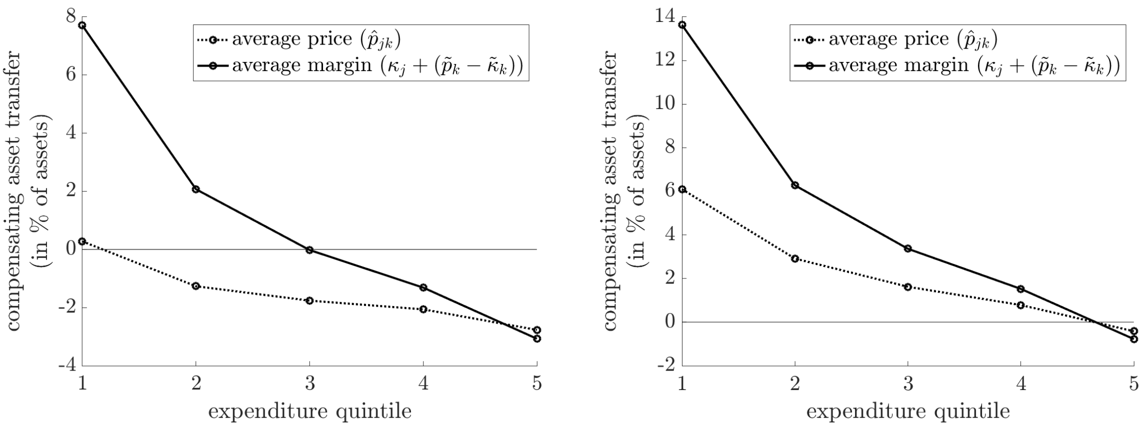
Figure A.7 reports the transfer as a share of households' expenditures and Figure A.8 as a share of households' assets. The conclusions from both figures with respect to the relative contributions of direct and equilibrium effects are similar to the computations based on compensating changes in households consumption index $\mathcal{C}$.



(a) Full Effect **(b)** Prices Only

Figure A.7: Asset Transfer as a Share of Expenditures

Note: Change in initial asset holdings as a fraction of total expenditure under alternative prices and zero shopping effort to make a household indifferent between living in the alternative economy and in the steady state economy. Counterfactuals allow households to optimally choose their savings and consumption baskets assuming they have to pay (i) the average price within each grocery variety or (ii) marginal cost plus the average margin across varieties. Panel (a) displays the full effect on welfare. Panel (b) uses a counterfactual measure of steady state welfare, abstracting from the disutility of effort, such that all differences are due to the alternative prices.



(a) Full Effect **(b)** Prices Only

Figure A.8: Asset Transfer as a Share of Assets

Note: Change in initial asset holdings as a fraction of assets under alternative prices and zero shopping effort to make a household indifferent between living in the alternative economy and in the steady state economy. Counterfactuals allow households to optimally choose their savings and consumption baskets assuming they have to pay (i) the average price within each grocery variety or (ii) marginal cost plus the average margin across varieties. Panel (a) displays the full effect on welfare. Panel (b) uses a counterfactual measure of steady state welfare, abstracting from the disutility of effort, such that all differences are due to the alternative prices.

A.3.4 Cyclicality of Retail Prices and Markups

To illustrate how the incidence of aggregate shocks can result in different responses of retail prices and markups even when the shock has identical size on aggregate, I simulate an unanticipated decline in aggregate earnings of 3% holding all parameters fixed at steady state values. Earnings revert back to their steady state level at a rate of 0.5 along a perfect foresight transition path. Consider three different scenarios: In a first scenario *all* households are affected, i.e. each household sees a 3% decline in her earnings. In a second scenario, aggregate income again falls by 3% but the losses are concentrated only on the *top* 25% households in the labor earnings distribution – each of them affected proportionately to their labor productivity. In a third scenario, the same aggregate loss affects only the *bottom* 25% of the labor earnings distribution.

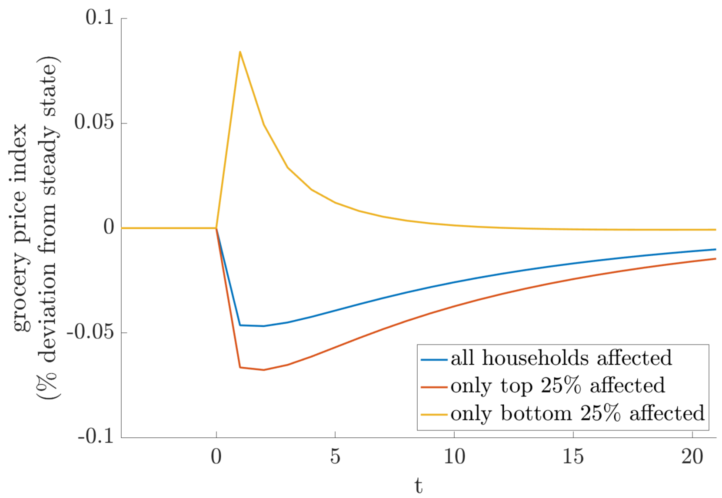


Figure A.9: Aggregate Prices under Varying Incidence

Note: Response of an aggregate Laspeyres index of posted prices $P_t^F = \frac{\sum_{j=1}^J C_j^{SS} \mu_{jt}^F}{\sum_{j=1}^J C_j^{SS}}$ to a 3% loss in aggregate labor earnings relative to the steady state, affecting (i) all households proportionately to their labor earnings, (ii) only the bottom quartile of labor earnings (proportionately to their earnings), or (iii) only the top quartile (proportionately to their earnings).

Figure A.9 plots the response of a Laspeyres index of average prices posted for all three scenarios. Focusing first on the case with all households affected, prices decline in response to a loss in aggregate earnings. If losses are concentrated at the top of the distribution, prices become more procyclical. In response to the same loss in aggregate earnings concentrated among the bottom of the earnings distribution, prices in the model economy become countercyclical. As all parameters including marginal cost κ_j are fixed at steady-state levels, these price responses are driven entirely by changes in markups.

Again, I decompose the response for each scenario into the contribution of changes in shopping policies, consumption policies, and the distribution of households. Figure A.10 plots the equilibrium response as a baseline together with the three counterfactuals for each of the scenarios considered. As in the Great Recession exercise presented in Section 1.6 an increase in the shopping effort of affected households reduces posted prices under all three scenarios. Shopping policies change by more when losses are concentrated among low-income households, as for them the same aggregate shock is comparably larger relative to their earnings. What accounts for the differences in cyclicality are differential responses of demand composition, driven by changes in households' consumption policies and the distribution of agents across the state space. In line with the findings for earnings and wealth losses during the Great Recession, the more low-income households are affected the more they have to reduce consumption and the lower becomes their share in overall demand. Retailers attach more weight to the lower effort of high-income buyers and increase prices in response. For the scenario affecting only the bottom of the income distribution, this demand composition effect is strong enough to outweigh the direct increase in shopping effort, yielding an overall countercyclical price response.

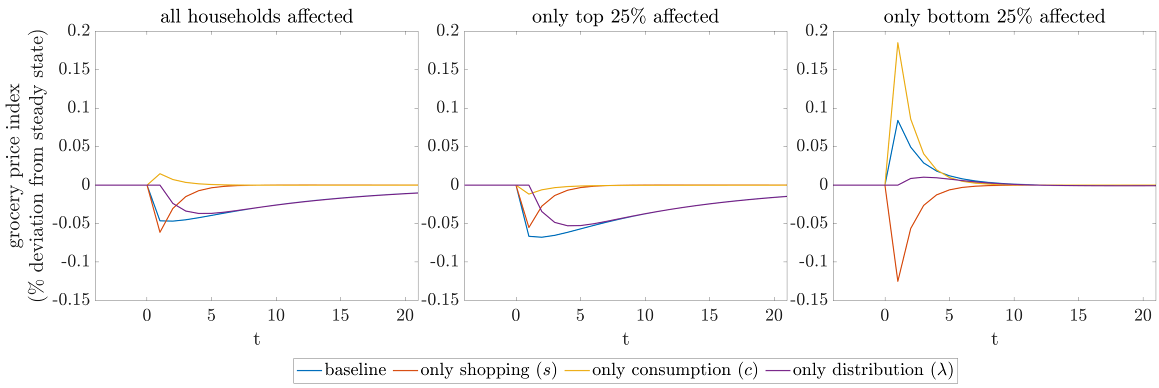


Figure A.10: Average Prices and Aggregate Income Losses

Note: Response of an aggregate Laspeyres index of posted prices $P_t^F = \frac{\sum_{j=1}^J C_j^{SS} \mu_{jt}^F}{\sum_{j=1}^J C_j^{SS}}$ to a 3% loss in aggregate labor earnings relative to the steady state, affecting (i) all households proportionately to their labor earnings, (ii) only the bottom quartile of labor earnings (proportionately to their earnings), or (iii) only the top quartile (proportionately to their earnings). Full response as baseline. Decomposed into response to changes in consumption policies (only c), shopping policies (only s), and the distribution of households (only dist), holding the respective others constant at steady state levels.

This additional exercise shows that even with an identical decline in aggregate earnings the incidence of the shock along the distribution of households matters for the cyclicality of retail prices and markups.

A.3.5 Additional Model Results

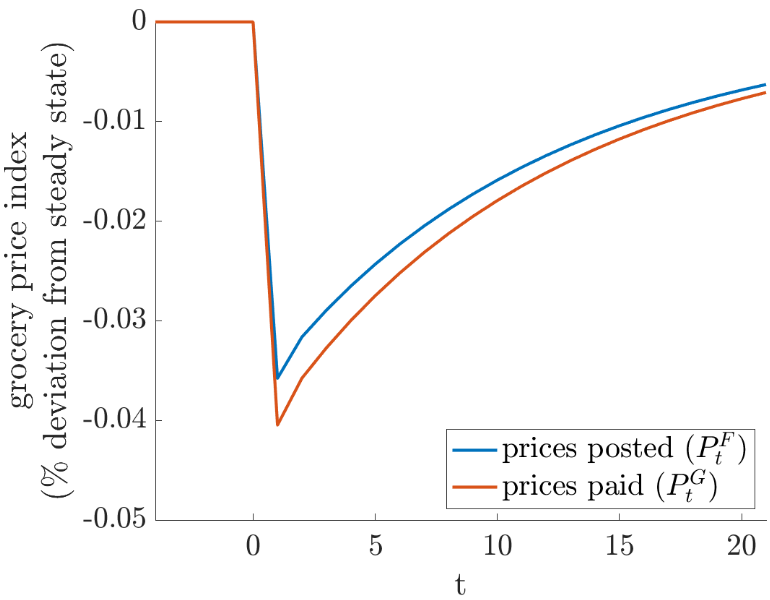


Figure A.11: Prices Posted and Prices Paid in Response to a 1% Loss in Wealth

Note: Model implied response of an aggregate Laspeyres index $P_t^l = \frac{\sum_{j=1}^J C_j^{SS} \mu_{jt}^l}{\sum_{j=1}^J C_j^{SS}}$ of prices posted (P_t^F) and prices paid (P_t^G) to a proportionate 1% decrease in beginning of period assets a for each household.

Table A.6: Earnings Losses for Great Recession Shock

		$z_1 - z_6$	z_7	z_8	z_9	z_{10}	z_{11}	$z_{12} - z_{16}$
cumulative share of households		24%	34%	47%	60%	74%	84%	100%
linked percentile in Heathcote et al. (2020a)		$P20$	$P30$	$P40$	$\frac{P50+P60}{2}$	$P70$	$P80$	$P90$
$\frac{z_t}{z^{SS}} - 1$	$t = 1$ (2008)	-17.3%	-7.4%	-3.0%	-4.3%	-3%	-3%	-0.9%
	$t = 2$ (2009)	-43.4%	-16.8%	-13.3%	-6.8%	-6.6%	-2.9%	-2.9%
	$t = 3$ (2010)	-55.6%	-23.7%	-15.1%	-8.5%	-6.4%	-4.6%	-2.5%

Note: Calibration of earnings losses by productivity state in the Great Recession. Data moments obtained from Heathcote et al. (2020a).

Table A.7: Cross Market Shares – Model vs. Data

		model					**data**				
		market share of exp. quintile					market share of exp. quintile				
		Q1	Q2	Q3	Q4	Q5	Q1	Q2	Q3	Q4	Q5
	Q1	0.09	0.15	0.19	0.24	0.33	0.17	0.14	0.18	0.22	0.30
by basket	Q2	0.08	0.14	0.19	0.24	0.35	0.08	0.21	0.18	0.22	0.31
of expenditure	Q3	0.07	0.13	0.18	0.24	0.37	0.08	0.13	0.25	0.22	0.31
quintile	Q4	0.07	0.13	0.18	0.24	0.39	0.08	0.13	0.17	0.30	0.32
	Q5	0.06	0.12	0.17	0.24	0.41	0.07	0.12	0.16	0.21	0.44

Note: Cross market shares are computed as the market share of a quintile for each variety averaged by the expedniture shares in the consumption basket of another quintile. Data obtained from Nielsen Consumer Panel waves 2007-2019, consumption baskets in the data defined at the barcode level.

Table A.8: Earnings and Price Changes after Redistributive Policies (alternative τs)

		$\tau = 0.05$	$\tau = 0.10$	$\tau = 0.15$	$\tau = 0.20$	$\tau = 0.25$
expenditure	Δearn	16.66%	33.31%	49.97%	66.63%	83.29%
quintile 1	ΔP	-0.08%	-0.16%	-0.23%	-0.30%	-0.37%
	Δp_G	-0.23%	-0.44%	-0.65%	-0.86%	-1.05%
	$\frac{-\Delta P}{\lvert\Delta \text{earn}\rvert}$	0.5%	0.5%	0.5%	0.5%	0.4%
	$\frac{-\Delta p_G}{\lvert\Delta \text{earn}\rvert}$	1.4%	1.3%	1.3%	1.3%	1.3%
expenditure	Δearn	5.22%	10.43%	15.65%	20.86%	26.08%
quintile 2	ΔP	-0.09%	-0.17%	-0.25%	-0.33%	-0.41%
	Δp_G	-0.25%	-0.49%	-0.72%	-0.95%	-1.17%
	$\frac{-\Delta P}{\lvert\Delta \text{earn}\rvert}$	1.6%	1.6%	1.6%	1.6%	1.6%
	$\frac{-\Delta p_G}{\lvert\Delta \text{earn}\rvert}$	4.7%	4.7%	4.6%	4.6%	4.5%
expenditure	Δearn	1.91%	3.83%	5.74%	7.65%	9.57%
quintile 3	ΔP	-0.09%	-0.18%	-0.27%	-0.35%	-0.43%
	Δp_G	-0.26%	-0.51%	-0.76%	-1.00%	-1.24%
	$\frac{-\Delta P}{\lvert\Delta \text{earn}\rvert}$	4.7%	4.7%	4.6%	4.6%	4.5%
	$\frac{-\Delta p_G}{\lvert\Delta \text{earn}\rvert}$	13.4%	13.3%	13.2%	13.1%	13.0%
expenditure	Δearn	-0.15%	-0.30%	-0.45%	-0.60%	-0.75%
quintile 4	ΔP	-0.09%	-0.19%	-0.28%	-0.37%	-0.46%
	Δp_G	-0.27%	-0.53%	-0.80%	-1.06%	-1.31%
	$\frac{-\Delta P}{\lvert\Delta \text{earn}\rvert}$	62.3%	62.0%	61.8%	61.4%	60.9$
	$\frac{-\Delta p_G}{\lvert\Delta \text{earn}\rvert}$	178.0%	177.3%	176.5%	175.4%	174.0%
expenditure	Δearn	-2.06%	-4.12%	-6.18%	-8.24%	-10.30%
quintile 5	ΔP	-0.10%	-0.20%	-0.30%	-0.39%	-0.49%
	Δp_G	-0.28%	-0.55%	-0.85%	-1.13%	-1.40%
	$\frac{-\Delta P}{\lvert\Delta \text{earn}\rvert}$	4.8%	4.8%	4.8%	4.8%	4.8%
	$\frac{-\Delta p_G}{\lvert\Delta \text{earn}\rvert}$	13.7%	13.7%	13.7%	13.7%	13.6%

Note: Average change in post-tax earnings (Δearn), grocery (Δp_G), and aggregate Laspeyres price index (ΔP) within each expenditure quintile in response to an earnings tax τ and budget neutral transfer.

Appendix B

Appendix to Chapter 2

B.1 Empirical Robustness Exercises

B.1.1 Employed and Unemployed Spouses

Table B.1: Joint Labor Market Transitions (Full Sample): Spouse Unemployed

	Primary earner transition		
	EE	EU	EN
Cond. prob. of spousal UE transition	25.29%	26.27%	34.11%
Cond. prob. of spousal UU transition	61.97%	63.33%	46.01%
Cond. prob. of spousal UN transition	12.74%	10.41%	19.87%

Notes: This table shows the probability of a spousal transition from unemployment conditional on primary earner transitions for the entire population.

Table B.2: Joint Labor Market Transitions (Full Sample): Spouse Employed

	Primary earner transition		
	EE	EU	EN
Cond. prob. of spousal EE transition	97.61%	91.49%	88.84%
Cond. prob. of spousal EU transition	0.77%	5.78%	1.25%
Cond. prob. of spousal EN transition	1.62%	2.72%	9.92%

Notes: This table shows the probability of a spousal transition from employment conditional on primary earner transitions for the entire population.

B.1.2 Education

Table B.3: Joint Labor Market Transitions by Spousal Education

	Primary earner transition		
	EE	EU	EN
I. Spouse College Degree (All):			
Cond. prob. of spousal NE transition	6.91%	11.40%	20.88%
Cond. prob. of spousal NU transition	1.59%	6.43%	1.04%
Cond. prob. of spousal NN transition	91.50%	82.18%	78.08%
II. Spouse No College Degree (All):			
Cond. prob. of spousal NE transition	5.55%	7.20%	15.08%
Cond. prob. of spousal NU transition	1.65%	5.34%	1.45%
Cond. prob. of spousal NN transition	92.81%	87.46%	83.47%
III. Spouse College Degree (Young):			
Cond. prob. of spousal NE transition	7.31%	13.25%	33.25%
Cond. prob. of spousal NU transition	1.70%	7.22%	1.29%
Cond. prob. of spousal NN transition	90.99%	79.53%	65.46%
IV. Spouse College Degree (Old):			
Cond. prob. of spousal NE transition	6.04%	7.72%	11.81%
Cond. prob. of spousal NU transition	1.35%	4.87%	0.86%
Cond. prob. of spousal NN transition	92.61%	87.41%	87.33%
V. Spouse No College Degree (Young):			
Cond. prob. of spousal NE transition	6.30%	8.34%	21.76%
Cond. prob. of spousal NU transition	2.01%	6.28%	2.21%
Cond. prob. of spousal NN transition	91.69%	85.37%	76.03%
VI. Spouse No College Degree (Old):			
Cond. prob. of spousal NE transition	4.19%	4.20%	9.41%
Cond. prob. of spousal NU transition	0.99%	2.83%	0.80%
Cond. prob. of spousal NN transition	94.82%	92.97%	89.79%

Notes: This table shows the probability of a spousal transition from out of the labor force conditional on primary earner transitions by education of the spouse.

B.1.3 Cohort Effects

Table B.4: Joint Labor Market Transitions

	Primary earner transition		
	EE	EU	EN
I. Spouse is a Man (Young) :			
Cond. prob. of spousal NE transition	13.54%	14.07%	44.10%
Cond. prob. of spousal NU transition	6.19%	11.69%	2.59%
Cond. prob. of spousal NN transition	80.27%	74.24%	53.31%
II. Spouse is a Man (Old):			
Cond. prob. of spousal NE transition	4.50%	4.59%	10.36%
Cond. prob. of spousal NU transition	1.13%	3.23%	0.63%
Cond. prob. of spousal NN transition	94.37%	92.18 %	89.01%
III. Spouse born between 1960-70 (Young):			
Cond. prob. of spousal NE transition	6.98%	8.62%	21.67%
Cond. prob. of spousal NU transition	1.89%	6.70%	2.42%
Cond. prob. of spousal NN transition	91.13%	84.68%	75.92%
IV. Spouse born between 1960-70 (Old)			
Cond. prob. of spousal NE transition	4.28%	2.94%	12.86%
Cond. prob. of spousal NU transition	1.11%	3.68%	1.04%
Cond. prob. of spousal NN transition	94.61%	93.38%	86.10%

Notes: This table shows the probability of a spousal transition from out of the labor force conditional on primary earner transitions by gender and cohort.

B.1.4 Children

Table B.5: Joint Labor Market Transitions (< Age 40)

	Primary earner transition		
	EE	EU	EN
I. Have Children:			
Cond. prob. of spousal NE transition	6.26%	8.71%	28.30%
Cond. prob. of spousal NU transition	1.75%	6.65%	2.31%
Cond. prob. of spousal NN transition	91.98%	84.64%	69.40%
II. No Children:			
Cond. prob. of spousal NE transition	9.68%	12.68%	23.69%
Cond. prob. of spousal NU transition	3.40%	8.54%	1.59%
Cond. prob. of spousal NN transition	86.91%	78.78%	74.72%
III. Have Children below 5:			
Cond. prob. of spousal NE transition	5.63%	8.55%	30.09%
Cond. prob. of spousal NU transition	1.47%	6.14%	1.96%
Cond. prob. of spousal NN transition	92.90%	85.31%	67.95%
IV. No Children below 5:			
Cond. prob. of spousal NE transition	8.08%	9.95%	24.82%
Cond. prob. of spousal NU transition	2.60%	7.80%	2.35%
Cond. prob. of spousal NN transition	89.32%	82.24%	72.82%

Notes: This table shows the probability of a spousal transition from out of the labor force conditional on primary earner transitions by presence of children in the household.

B.1.5 Reasons for Non-Participation

Table B.6: Joint Labor Market Transitions

	Primary earner transition		
	EE	EU	EN
I. Excluding Retirement (Young):			
Cond. prob. of spousal NE transition	6.66%	9.32%	27.13%
Cond. prob. of spousal NU transition	2.00%	6.91%	2.06%
Cond. prob. of spousal NN transition	91.33%	83.77%	70.81%
II. Excluding Retirement (Old):			
Cond. prob. of spousal NE transition	4.95%	4.15%	11.45%
Cond. prob. of spousal NU transition	1.18%	3.33%	1.00%
Cond. prob. of spousal NN transition	93.87%	92.52%	87.54%
III. Excluding Disabled/Ill (Young):			
Cond. prob. of spousal NE transition	6.55%	9.34%	27.02%
Cond. prob. of spousal NU transition	1.96%	6.94%	2.01%
Cond. prob. of spousal NN transition	91.49%	83.72%	70.97 %
IV. Excluding Disabled/Ill (Old):			
Cond. prob. of spousal NE transition	4.17%	3.42%	8.53%
Cond. prob. of spousal NU transition	0.88%	2.77%	0.50%
Cond. prob. of spousal NN transition	94.95%	93.81%	90.97%
V. Excluding Retired and Disabled/Ill (Young):			
Cond. prob. of spousal NE transition	6.55%	9.36%	27.23%
Cond. prob. of spousal NU transition	1.97%	6.96%	2.05%
Cond. prob. of spousal NN transition	91.48%	83.68%	70.72%
VI. Excluding Retired and Disabled/Ill (Old):			
Cond. prob. of spousal NE transition	4.74%	3.62%	11.20%
Cond. prob. of spousal NU transition	1.16%	3.40%	0.89%
Cond. prob. of spousal NN transition	94.11%	92.99%	87.91%

Notes: This table shows the probability of a spousal transition from out of the labor force conditional on primary earner transitions by reasons for non-participation.

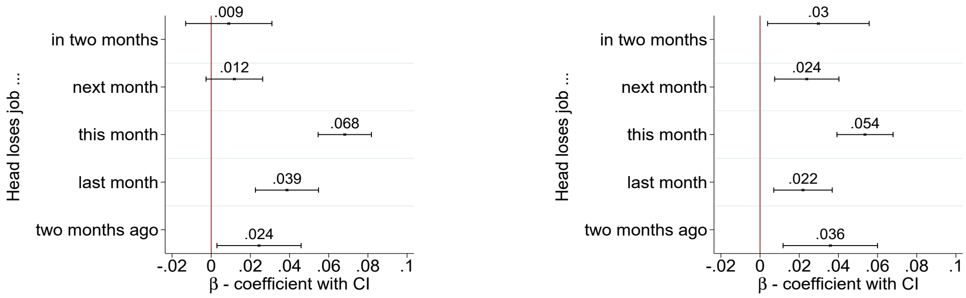
B.1.6 Business Cycle

Table B.7: Joint Labor Market Transitions

	Primary earner transition		
	EE	EU	EN
NBER Recession, Young			
Cond. prob. of spousal NE transition	6.48%	7.74%	22.38%
Cond. prob. of spousal NU transition	1.98%	8.73%	0.99%
Cond. prob. of spousal NN transition	91.55%	83.53%	76.63%
NBER Recession, Old			
Cond. prob. of spousal NE transition	4.14%	5.43%	7.71%
Cond. prob. of spousal NU transition	0.83%	2.76%	0.68%
Cond. prob. of spousal NN transition	95.03%	91.81%	91.61%
No NBER Recession, Young			
Cond. prob. of spousal NE transition	6.68%	9.53%	27.45%
Cond. prob. of spousal NU transition	2.00%	6.63%	2.14%
Cond. prob. of spousal NN transition	91.31%	83.85%	70.41%
No NBER Recession, Old			
Cond. prob. of spousal NE transition	4.30%	3.46%	8.80%
Cond. prob. of spousal NU transition	0.91%	2.75%	0.54%
Cond. prob. of spousal NN transition	94.79%	93.79%	90.66%

Notes: This table shows the probability of a spousal transition from out of the labor force conditional on primary earner transitions by state of the business cycle.

B.1.7 Dynamics Response for Other Age Groups



(a) Age 36 to 45 **(b)** Age 46 to 55

Figure B.1: Δ Pr(Spouse enters LF) this month

Notes: Figure B.1 shows the change in probability that a non-participating spouse enters the labor force (either as unemployed or as employed) if the household head loses/lost the job in two months, next month, this month, last month or two months ago, respectively, relative to the baseline in which the household head remains employed. The sample includes couples in which one spouse is working and one spouse is out of the labor force between age 36 and 45 (Figure B.1a) and between age 46 and 55 (Figure B.1b) from the Current Population Survey (CPS), waves 1994 until 2020. Age refers to the non-participating spouse. The regression producing the coefficients is Equation 2.1.

B.1.8 Additional Results on SIPP Data

Table B.8: Joint Labor Market Transitions – CPS vs. SIPP

	Primary earner transition		
	EE	EU	EN
CPS:			
Cond. prob. of spousal NE transition	6.03%	8.01%	16.79%
Cond. prob. of spousal NU transition	1.63%	5.55%	1.33%
Cond. prob. of spousal NN transition	92.34%	86.44%	81.88%
SIPP:			
Cond. prob. of spousal NE transition	2.23%	5.36%	6.28%
Cond. prob. of spousal NU transition	1.14%	4.57%	2.02%
Cond. prob. of spousal NN transition	96.63%	90.07%	91.70%

Notes: This table shows compares the probability of a spousal transition from out of the labor force conditional on primary earner transitions between the CPS and SIPP datasets.

Table B.9: Agg. Data – Joint Labor Market Transitions by Net Liquid Wealth & Age

	Primary earner transition		
	EE	EU	EN
Bottom 50% *of Net Liquid Wealth (Young):*			
Cond. prob. of spousal NE transition	10.43%	12.26%	20.50%
Cond. prob. of spousal NU transition	4.35%	11.24%	7.66%
Cond. prob. of spousal NN transition	85.22%	76.50%	71.85%
Top 50% *of Net Liquid Wealth (Young):*			
Cond. prob. of spousal NE transition	10.09%	13.82%	28.70%
Cond. prob. of spousal NU transition	2.94%	7.32%	2.61%
Cond. prob. of spousal NN transition	86.97%	78.86%	68.70%
Bottom 50% *of Net Liquid Wealth (Old):*			
Cond. prob. of spousal NE transition	4.66%	5.63%	6.93%
Cond. prob. of spousal NU transition	2.15%	7.81%	0.89%
Cond. prob. of spousal NN transition	93.19%	86.56%	92.18%
Top 50% *of Net Liquid Wealth (Old):*			
Cond. prob. of spousal NE transition	6.38%	6.09%	5.90%
Cond. prob. of spousal NU transition	1.68%	3.52%	0.92%
Cond. prob. of spousal NN transition	91.95%	90.45%	93.18%

Notes: This table shows the probability of a spousal transition from out of the labor force conditional on primary earner transitions by asset holdings and age group. "Young" refers to spouses below 40, and "Old" refers to spouses above 50. Data are aggregated to interview panel length.

Appendix C

Appendix to Chapter 3

C.1 Empirical Appendix

C.1.1 Additional Empirical Results

In addition to our main empirical analysis, we consider alternative specifications to test the robustness of our findings. More specifically, we provide results for the following variations of our main specification:

- Figure C.1 shows the analogue impulse response for monthly series.
- Figure C.2 shows the IRFs to a similar specification as in equation (3.1), but with lags for each horizon h and income group i selected independently according to the optimal selection criterion in Akaike (1974).
- In Figure C.3, we consider a different definition of household income, in which rents are subtracted from our original income variable as in Aguiar and Bils (2015).
- In Figures C.4 and C.5, we restrict our definition of consumption to respectively durable and nondurable goods.

To examine if our results are driven by households' home-ownership status, we follow Cloyne et al. (2020) and divide our sample into mortgagors and other households (renters and outright homeowners).[1] Results are displayed in

[1] Our definition of income quintiles still refers to the income distribution in the full sample, and not within housing tenure categories.

figures C.6 and C.7. A comparison between the bottom-right panels of these two figures does not reveal differences in overall consumption responses by ownership status. Focusing on the response of non-mortgagors (figure C.6) we see that response of Q1 is once again stronger than that of the other households, especially compared to Q3-5. In other words, the main takeaway of our analysis – that households at the bottom exhibit a stronger response to bank equity returns – is not driven by the response of mortgagors. In fact, figure C.7 reveals a pattern of heterogeneity that is less pronounced than in our baseline results, with the response of Q1 displaying large error bands. The sample size is particularly small for mortgagors at the bottom quintiles of the income distribution, which leads to the observed loss in precision. This is because mortgagors tend to have higher income than their counterparts. In particular, only 21 percent of households at the bottom of the income distribution are mortgagors, as opposed to 58 percent at the top quintile.

Finally, we analyse the effect of below- and above-median bank returns, plotted respectively in figures C.8 and C.9. We modify specification (3.1) by including a dummy for below-median returns interacted with r^B, and plot the coefficients corresponding to this interaction. The coefficient that multiplies r^B alone then corresponds to the effect of above-median returns. For exposition, we display a response to a *positive* shock for above median returns. The aggregate response of consumption is similar in both cases. On the other hand, in the case of below-median returns, the response of consumption for the bottom quintile is stronger – relative to the aggregate one – than in the case of above-median returns.

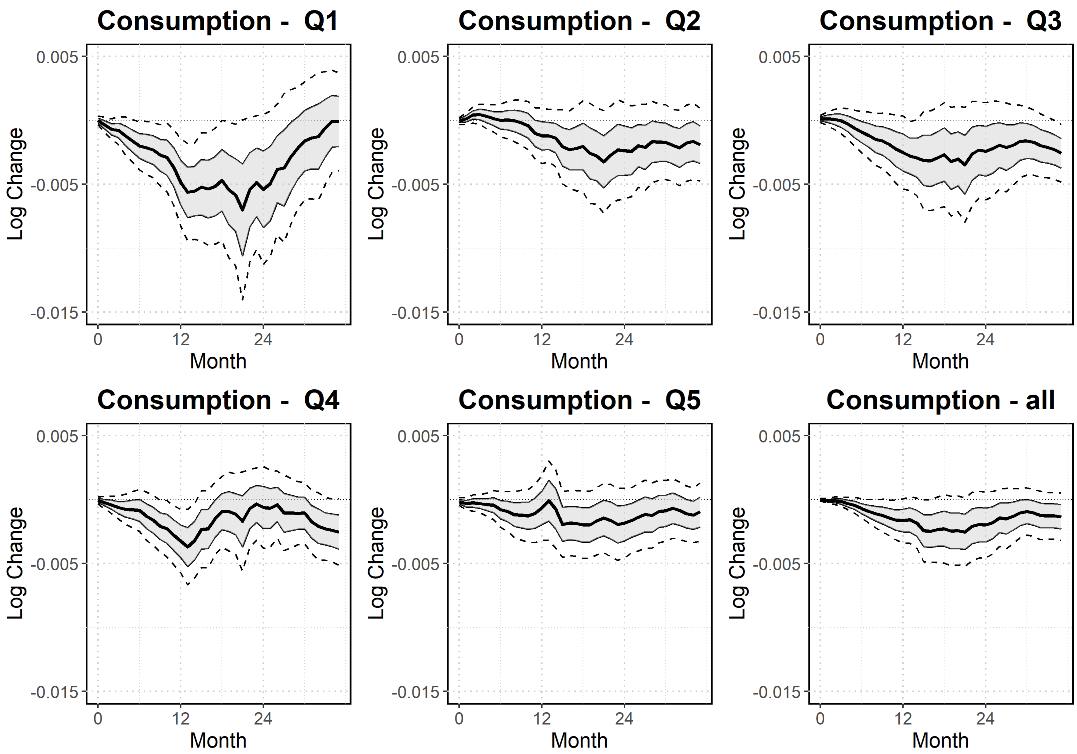


Figure C.1: Bank Equity Returns and Household Consumption—Monthly
Notes: Impulse responses of household consumption by income quintiles and aggregate, using data for 1980-2010, to a negative one standard-deviation change in r^B . The shaded areas indicate one standard-deviation confidence intervals, dashed lines are 95-percent confidence bands. Robust, Newey-West standard errors. Time (horizontal axis) in months.

C.1.2 Details on Aggregate Data Series

Data series and details on specifications for Figure 3.4:

- Top-left panel. Data series: US Bureau of Economic Analysis, Compensation of Employees, Received: Wage and Salary Disbursements [A576RC1], retrieved from FRED, Federal Reserve Bank of St. Louis; Regression specification is the same as equation 3.1, substituting consumption for the wage disbursement series adjusted by the CPI All Urban.

- Top-right panel. Data series: US Bureau of Economic Analysis, Real Gross Private Domestic Investment [GPDIC1], retrieved from FRED, Federal Reserve Bank of St. Louis; Regression specification is the same as equation 3.1, substituting consumption for the investment series.

- Bottom-left. Spread on credit card rate is obtained subtracting the 3-month T-bill rate from the the interest rate on credit cards. The regression specification is similar to equation 3.1, but substitutes consumption for the spread

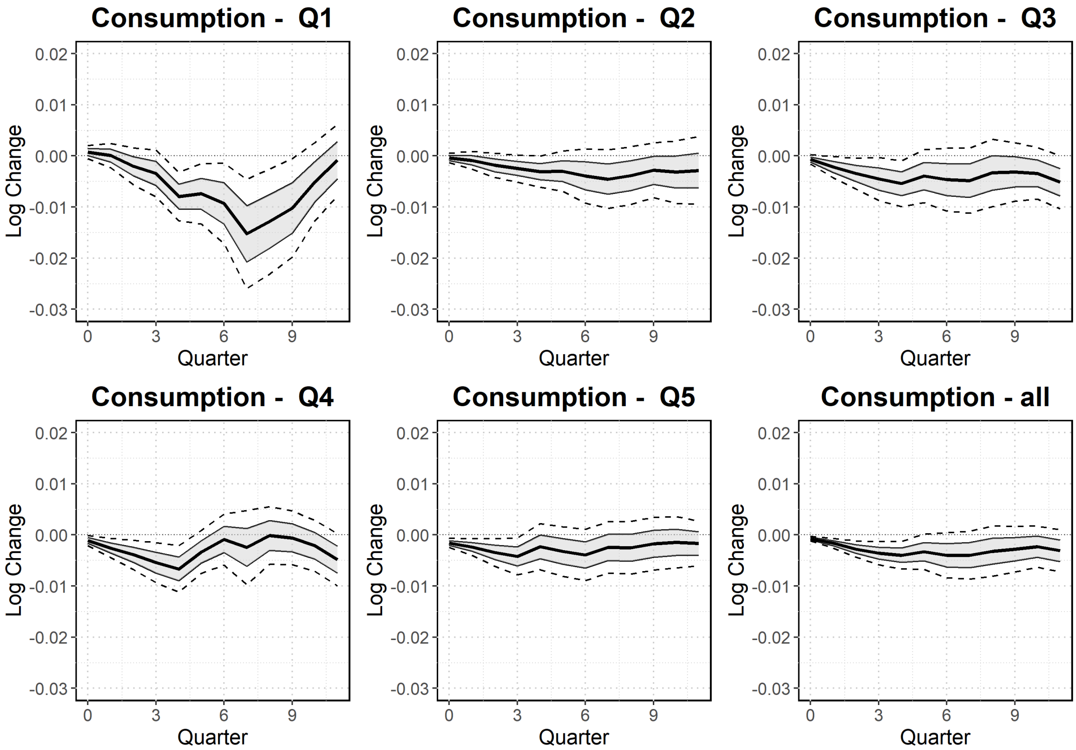


Figure C.2: Bank Equity Returns and Household Consumption—AIC
Notes: Impulse responses of household consumption by income quintiles and aggregate, using data for 1980-2010, to a negative one standard-deviation change in r^B . The shaded areas indicate one standard-deviation confidence intervals, dashed lines are 95-percent confidence bands. Robust, Newey-West standard errors. Time (horizontal axis) in quarters. Lags are selected according to Akaike (1974) optimal information criterion

series and controls for credit card charge-off rates to adjust for borrowers' default risk. Series: (i) Credit card rates: Board of Governors of the Federal Reserve System (US), Commercial Bank Interest Rate on Credit Card Plans, All Accounts [TERMCBCCALLNS], retrieved from FRED, Federal Reserve Bank of St. Louis; (ii) T-bill rates: Board of Governors of the Federal Reserve System (US), 3-Month Treasury Bill Secondary Market Rate [DTB3], retrieved from FRED, Federal Reserve Bank of St. Louis (quarterly average); (iii) Charge-off rate: Board of Governors of the Federal Reserve System (US), Charge-Off Rate on Credit Card Loans, All Commercial Banks [CORCCACBS], retrieved from FRED, Federal Reserve Bank of St. Louis.

- Bottom-right: Dow Jones Industrials Share Price Index. End-of-month indices are aggreagated at the quarterly level through simple average. The

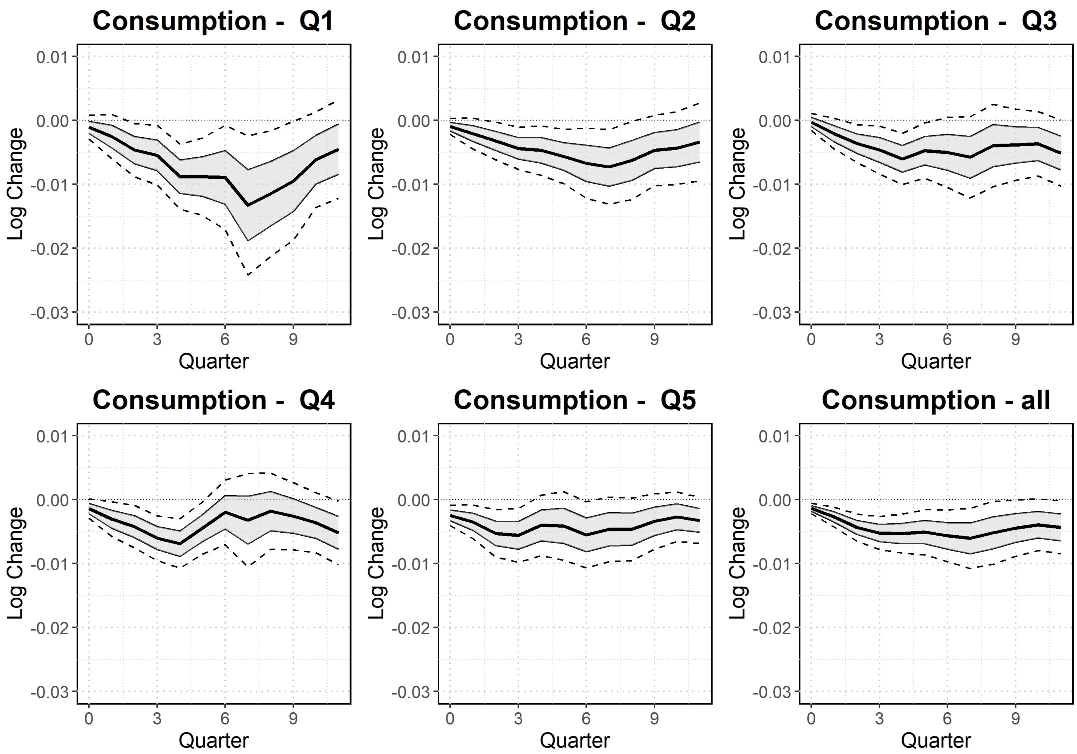


Figure C.3: Bank Equity Returns and Household Consumption—Rent Adj.

Notes: Impulse responses of household consumption by income quintiles and aggregate, using data for 1980-2010, to a negative one standard-deviation change in r^B . The shaded areas indicate one standard-deviation confidence intervals, dashed lines are 95-percent confidence bands. Robust, Newey-West standard errors. Time (horizontal axis) in quarters. Incomes are computed net of rents.

regression specification is the same as in equation (3.1), but since we control for the lagged stock market index, we exclude r^N from the set of controls.

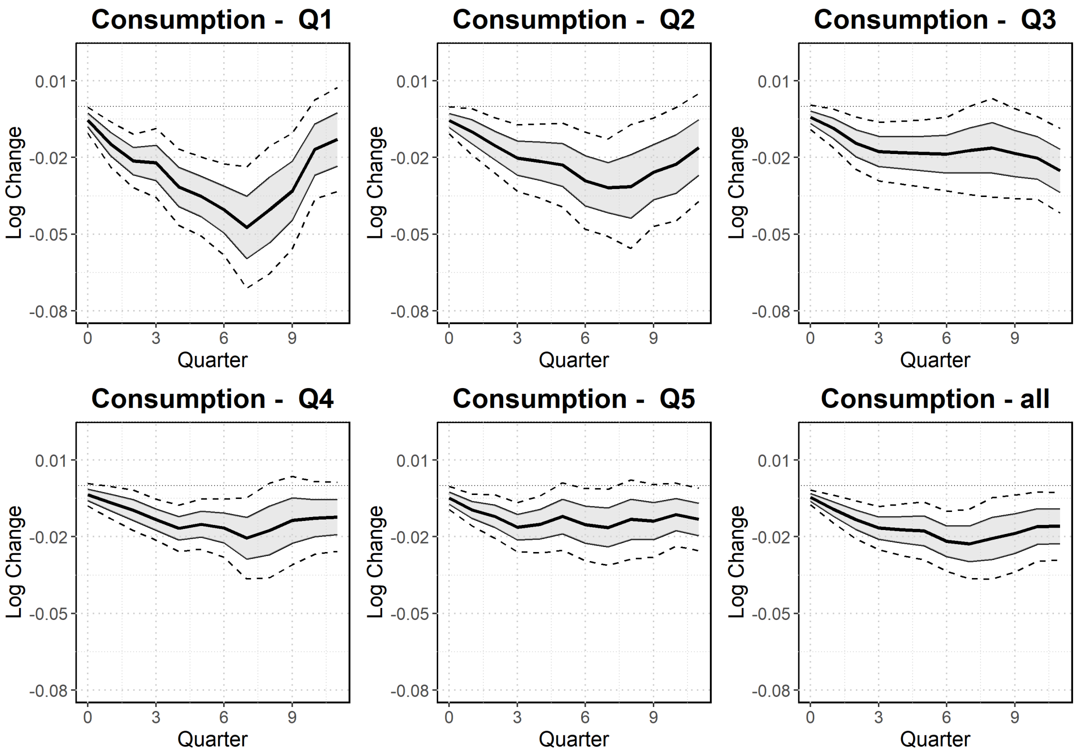


Figure C.4: Bank Equity Returns and Household Consumption—Durables

Notes: Impulse responses of household consumption by income quintiles and aggregate, using data for 1980-2010, to a negative one standard-deviation change in r^B . The shaded areas indicate one standard-deviation confidence intervals, dashed lines are 95-percent confidence bands. Robust, Newey-West standard errors. Time (horizontal axis) in quarters. Expenditures refer to durable consumption.

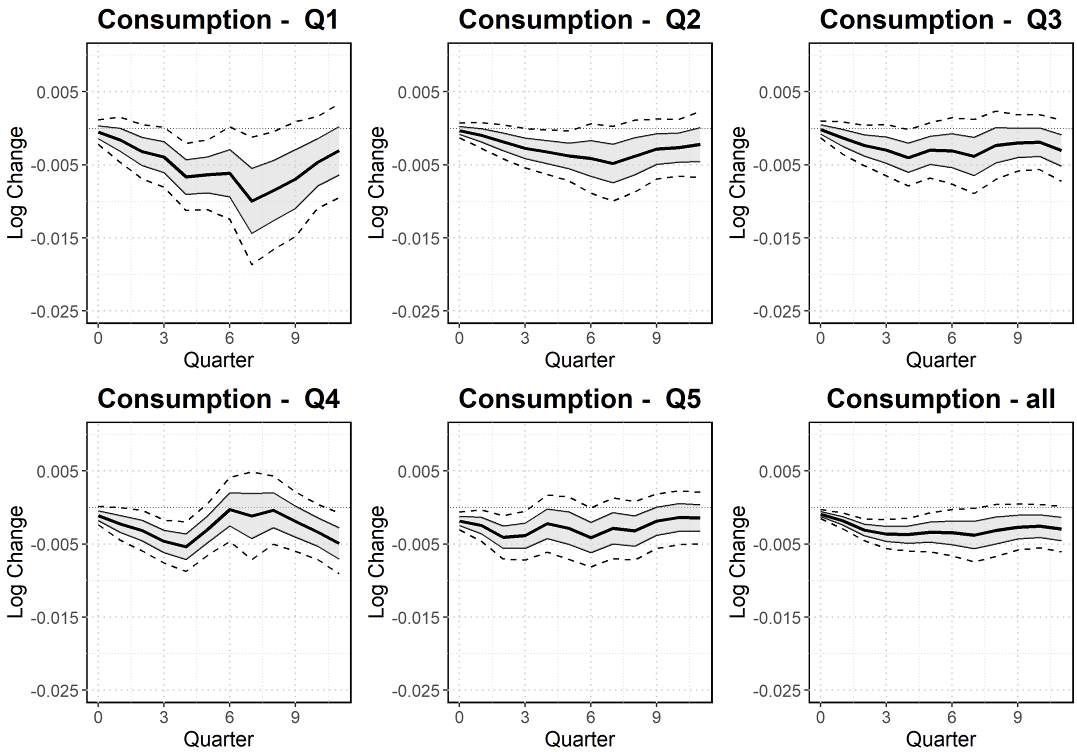


Figure C.5: Bank Equity Returns and Household Consumption—Nondurables
Notes: Impulse responses of household consumption by income quintiles and aggregate, using data for 1980-2010, to a negative one standard-deviation change in r^B . The shaded areas indicate one standard-deviation confidence interval, dashed lines are 95-percent confidence bands. Robust, Newey-West standard errors. Time (horizontal axis) in quarters. Expenditures refer to nondurable consumption.

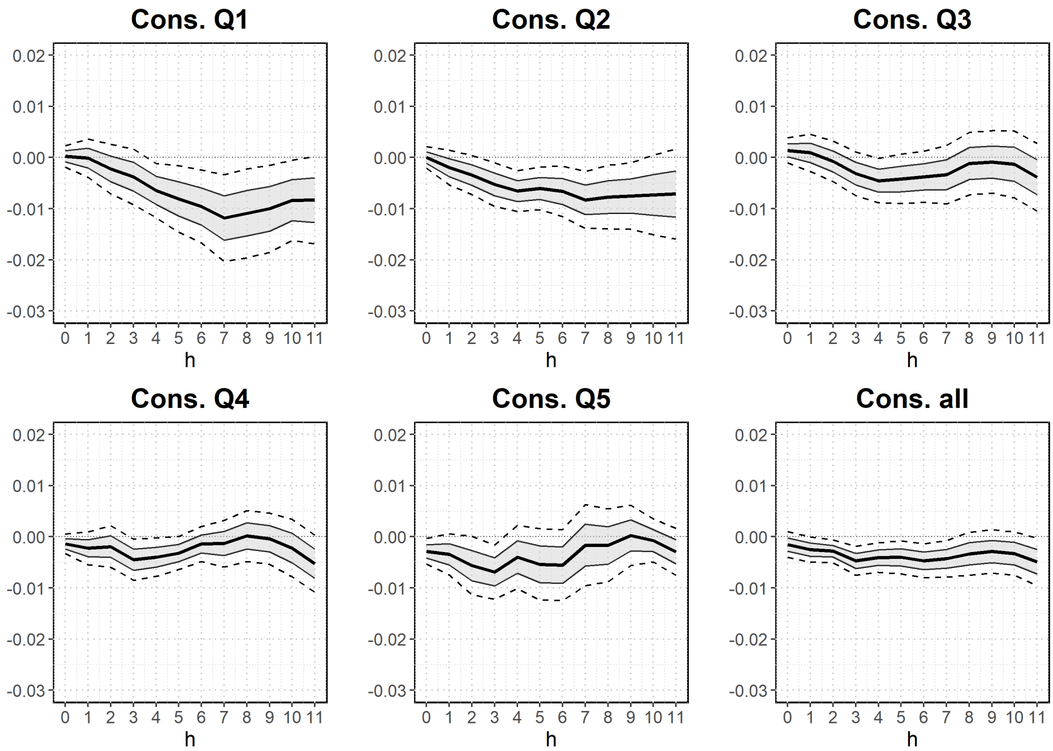


Figure C.6: Bank Equity Returns and Household Consumption—Non-Mortgagors
Notes: Impulse responses of household consumption by income quintiles and aggregate, using data for 1980-2010, to a negative one standard-deviation change in r^B . The shaded areas indicate one standard-deviation confidence intervals, dashed lines are 95-percent confidence bands. Robust, Newey-West standard errors. Time (horizontal axis) in quarters. Sample is restricted to non-mortgagors.

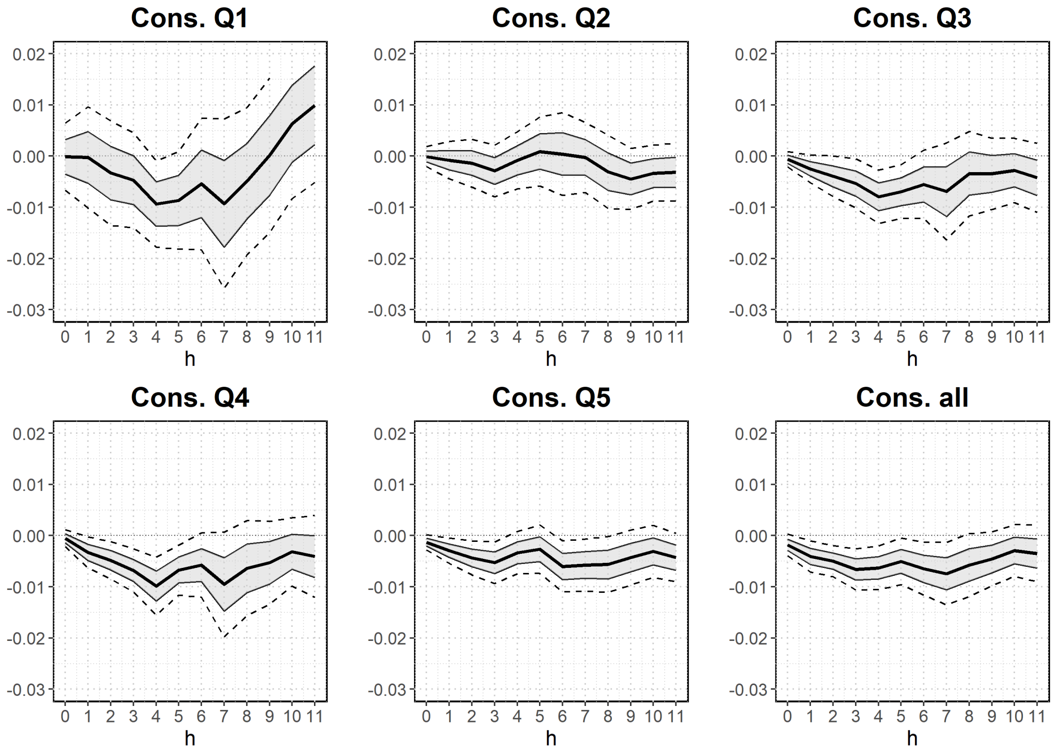


Figure C.7: Bank Equity Returns and Household Consumption—Mortgagors

Notes: Impulse responses of household consumption by income quintiles and aggregate, using data for 1980-2010, to a negative one standard-deviation change in r^B . The shaded areas indicate one standard-deviation confidence intervals, dashed lines are 95-percent confidence bands. Robust, Newey-West standard errors. Time (horizontal axis) in quarters. Sample is restricted to mortgagors.

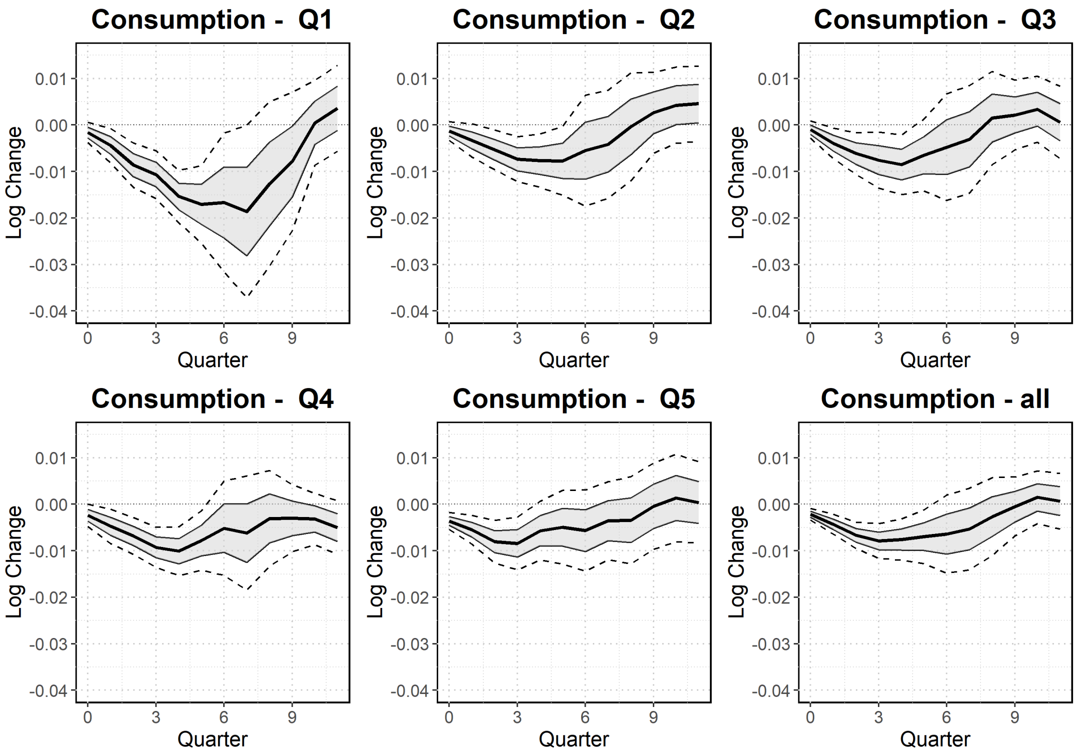


Figure C.8: Bank Equity Returns and Consumption—Below-Median Shocks
Notes: Impulse responses of household consumption by income quintiles and aggregate, using data for 1980-2010, to a negative one standard-deviation change in r^B . The shaded areas indicate one standard-deviation confidence intervals, dashed lines are 95-percent confidence bands. Robust, Newey-West standard errors. Time (horizontal axis) in quarters. Sample is restricted to below-median r^B.

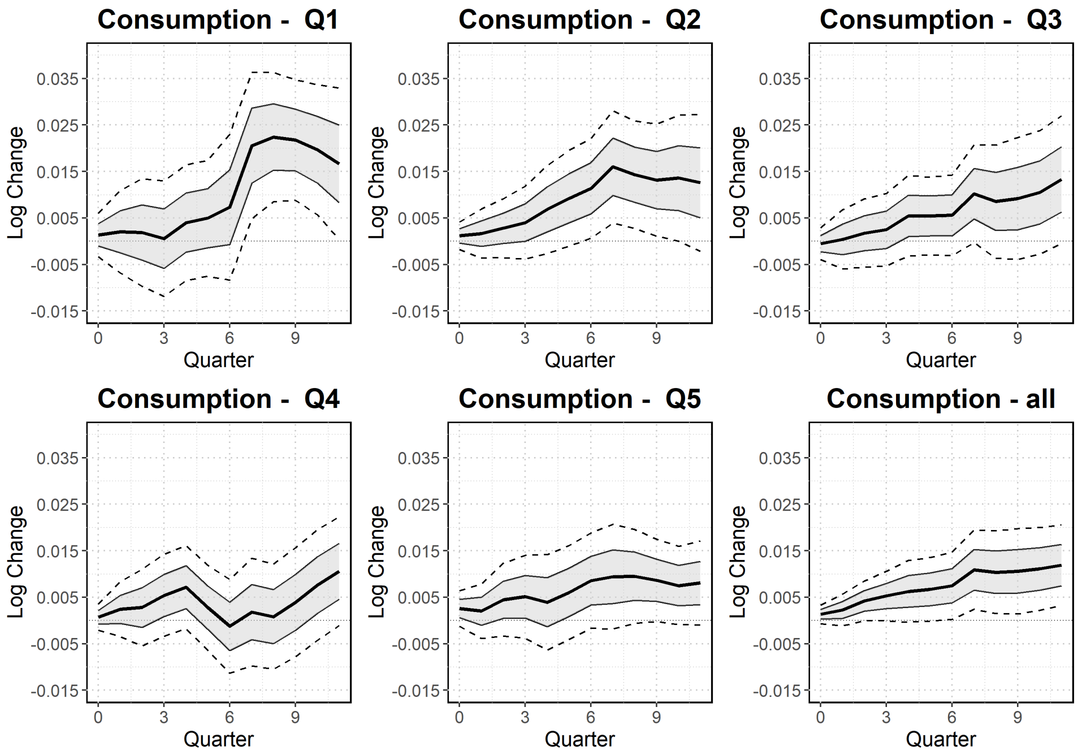


Figure C.9: Bank Equity Returns and Consumption—Above-Median Shocks

Notes: Impulse responses of household consumption by income quintiles and aggregate, using data for 1980-2010, to a **positive** one standard-deviation change in r^B . The shaded areas indicate one standard-deviation confidence intervals, dashed lines are 95-percent confidence bands. Robust, Newey-West standard errors. Time (horizontal axis) in quarters. Sample is restricted to above-median r^B.

C.2 Equilibrium Definition

An equilibrium in our model economy consists of a stream of prices $\{r_t^D, r_t^L, q_t, w_t, r_t^K\}$, stocks $\{L_t, D_t, K_t^{HH}, K_t^B\}$, flows $\{C_t, I_t, Y_t, N_t, div_t^K, div_t^B, div_t^Y\}$, value functions $\{V_t^n, V_t^a, V_t, V_t^B\}$, a measure over idiosyncratic states $\lambda_t(a_t, k_t, z_t)$, and a path of exogenous shocks $\{A_t, \xi_t^B\}$ where for initial conditions $\lambda_0(a_t, k_t, z_t)$, K_0^B, K_0^{HH}, and r_0^D, r_0^L:

1. Given prices and shocks, households and bank managers solve their problems in (3.18), (3.16), and (3.10).
2. The measure over states is induced by households' policy functions.
3. Dividends are determined by (3.7), (3.8), and (3.12).
4. K^B respects the bankers' leverage constraint (3.13) and K^{HH} respects (3.22).
5. Output Y_t is given by (3.2).
6. Bankers' equity evolves according to (3.11).
7. Loans (3.20), deposits (3.21), capital (3.23), goods (3.25), and labor (3.27) markets clear.
8. Investment is determined by (3.24).
9. The equations that jointly determine prices are: (3.3), (3.4), (3.9), (3.14), and (3.15).

C.3 Additional Quantitative Results

C.3.1 Baseline Results - Additional Figures and Tables

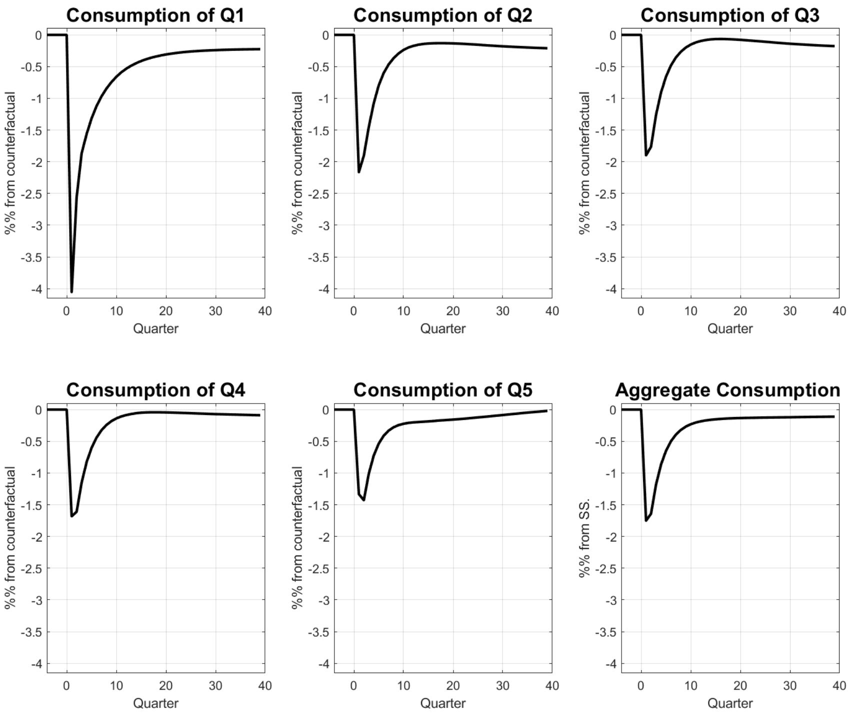


Figure C.1: Consumption Responses by Quintile of Net Worth
Note: Impulse responses are displayed relative to the counterfactual evolution of consumption for each group in the absence of the shock. Quintiles based on total net worth in the steady state.

Table C.1: Household Characteristics by Quintile of Welfare Change

Average Income	Q1	Q2	Q3	Q4	Q5
	1.02	0.70	0.92	1.07	1.28
Average Capital	Q1	Q2	Q3	Q4	Q5
	0.50	0.22	0.39	0.70	3.16
Average Networth	Q1	Q2	Q3	Q4	Q5
	0.77	0.23	0.36	0.64	2.98

Note: Lowest quintile corresponds to largest welfare losses. Characteristics represented as multiple of economy-wide average.

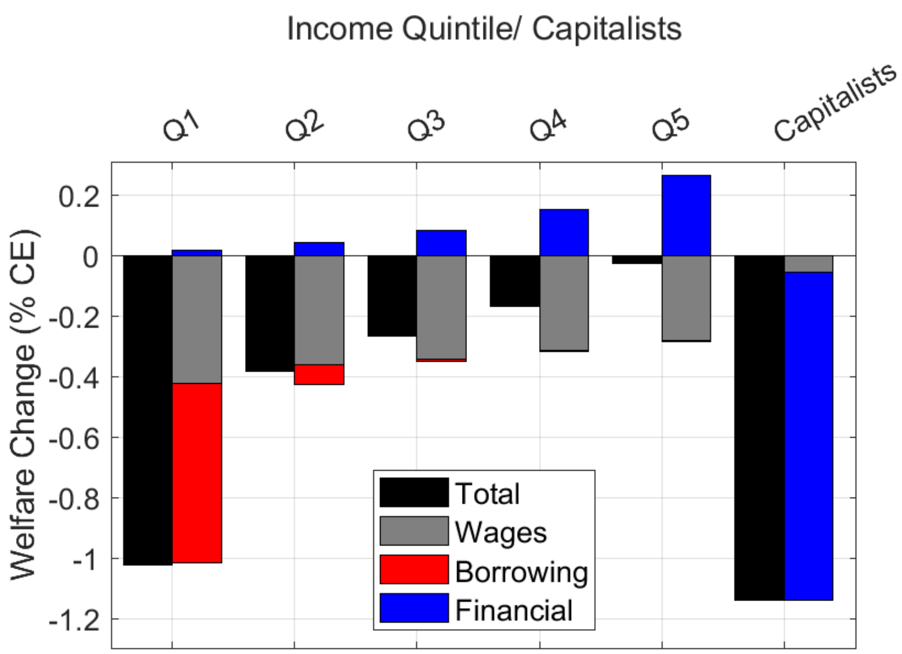


Figure C.2: Decomposition of Welfare Changes by Income Quintile

Note: Decomposition of welfare changes due to wages $\{w_t\}_{t=0}^T$, the lending rate $\{r_t^L\}_{t=0}^T$, and financial variables (jointly $\{r_t^D, r^K, q_t, div_t\}_{t=0}^T$). The black bars represent the general equilibrium welfare changes. Each of the gray and colored bars is obtained by simulating the economy in response to the general equilibrium path of one variable (or all four, in the case of financial variables).

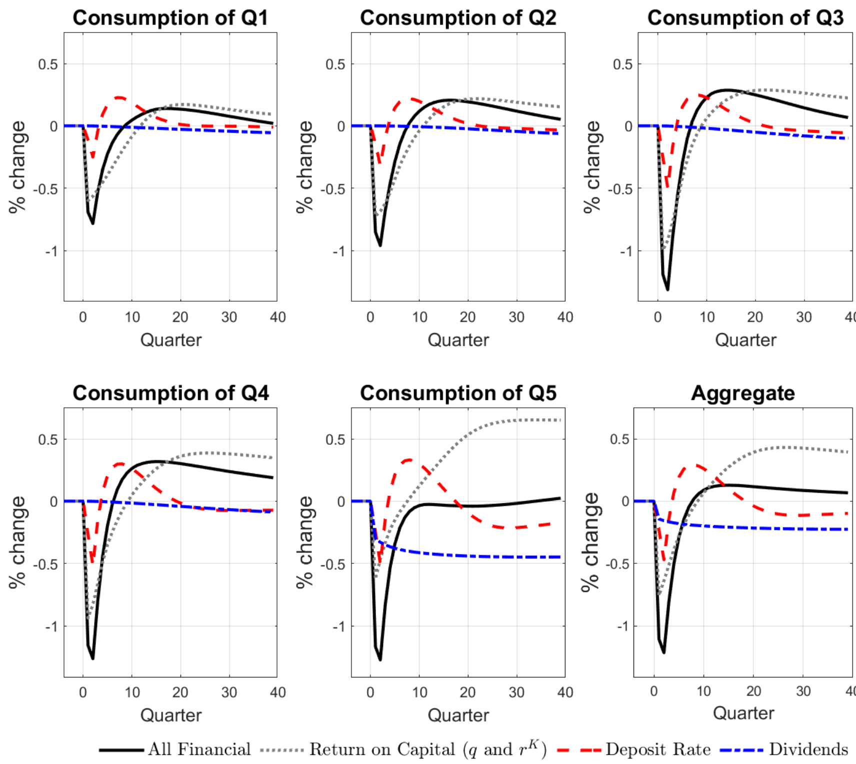


Figure C.3: Consumption Decomposition—Financial Variables

Note: Model-implied consumption responses to changes in *financial variables*. Income quintiles sorted based on total income in steady state, including earnings, interest received, and dividends. Impulse responses are displayed relative to the counterfactual evolution of consumption for each group in the absence of any price variation. Consumption responses are decomposed into partial-equilibrium effects of return on capital $\{q_t, r_t^k\}_{t=0}^T$, the deposit rate $\{r_t^D\}_{t=0}^T$, dividends $\{div_t\}_{t=0}^T$, and financial variables (jointly $\{r_t^D, r^K, q_t, div_t\}_{t=0}^T$).

C.3.2 Alternative Shock – A Direct Bank Equity Loss

We now consider an alternative shock to induce bank losses: A direct bank equity shock. This is a reduced form way to capture unexpected losses in bank equity, without (directly) affecting production. In this sense, this shock provides a "clean" exercise in which only beginning-of-period bank equity is affected, but the consequences of the shock will be felt throughout the economy due to general equilibrium effects. The loss to bank equity is, however, a deadweight loss to the economy's resource constraint and could be interpreted as the banking sector incurring some depreciation on external assets not affecting the economy directly. The presence of this shock changes two expressions in the model. Equation (3.11) becomes:

$$E_t = \underbrace{(1+r_t^L)L_t}_{\text{repayment from borrowing HHs}} + \underbrace{((1-\delta)q_t + \xi_t^B r_t^K)K_t^B}_{\text{returns of holding capital}} - \underbrace{(1+r_t^D)D_t}_{\text{repaying depositors}} - \underbrace{\varepsilon_t}_{\text{Equity Shock}}$$

and equation (3.26) is now:

$$\Xi_t = \frac{\phi_K}{2}\left(\frac{I_t}{K_{ss}} - \delta\right)^2 K_{ss} + \tau L_t + \varepsilon_t.$$

Together with the bank equity shock, we consider a demand externality as in Krueger et al. (2016), which makes output partially demand-driven and enables its endogenous response on impact.[2] Namely, equation (3.2) becomes:

$$Y_t = \widehat{A_t} K_t^{\alpha} N_t^{1-\alpha},$$

where $\widehat{A_t}$ is total factor productivity A_t adjusted for an externality from aggregate demand (consumption C plus investment I) such that:

$$\widehat{A_t} = A_t \left(\frac{C_t + I_t}{C^{SS} + I^{SS}}\right)^{\phi_Y}.$$

Finally, factor payments (equations (3.3) and (3.4)) are adjusted accordingly.

[2]The presence of the demand externality does not impact the distributive results. Simulations without it are available upon request. See Bai et al. (2019) for microfoundations via search for quantities in the goods market. A possible interpretation is the following: there are some sectors in the economy (especially services, but also e.g. customized investment goods) that are unable to produce for inventory and hence require immediate demand for input factors to be productive. Cooper and Ejarque (2000) introduce a similar externality by assuming complementarity of the output of multiple firms based on the work of Baxter and King (1991).

We calibrate the magnitude and persistence of the bank equity shock to again match a 20 percent decline in bank equity on impact and a 12-quarter cumulative response of 8.6 percent. We reproduce all figures and tables from Section 3.5 below (Figures C.4, C.5, C.6, C.7, C.8, C.9, C.10, C.11, C.12, C.13, and C.14, together with Tables C.2 and C.3).

Overall, the conclusions from Section 3.5 remain intact: Consumption and welfare responses are heterogeneous, low-income households along with claimants to dividends are the biggest losers, and high-income individuals take advantage of movements in financial markets, with some (7 percent) standing to gain. This is because the transmission channels are very similar to the ones following the bank capital productivity shock.

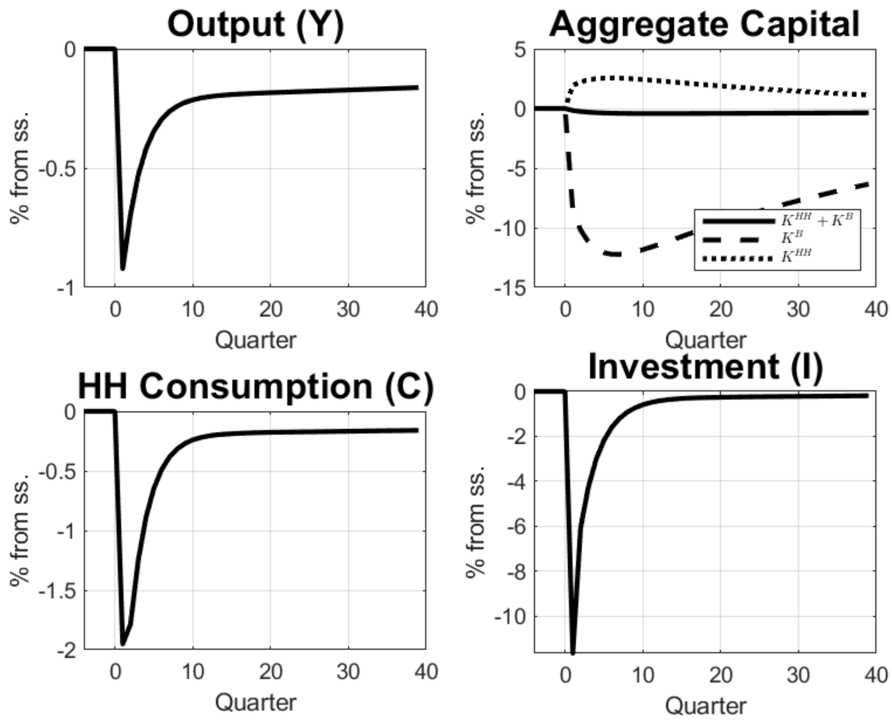


Figure C.4: Dynamics of Macroeconomics Aggregates
Note: Responses of macroeconomic aggregates to the bank equity shock. All variables reported in percentage deviation from their respective steady state levels.

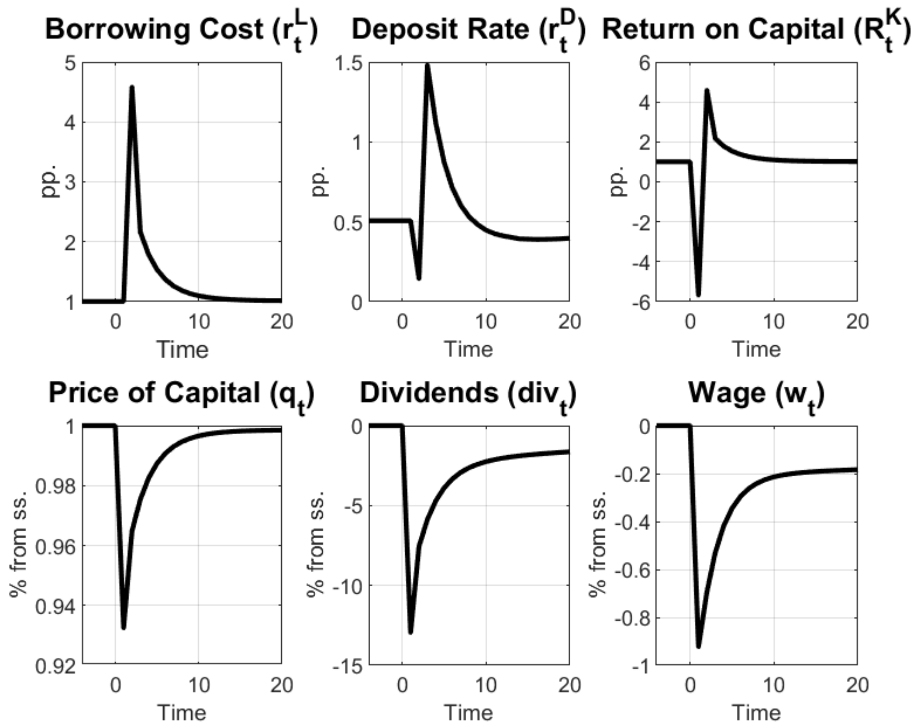


Figure C.5: General Equilibrium Price Responses

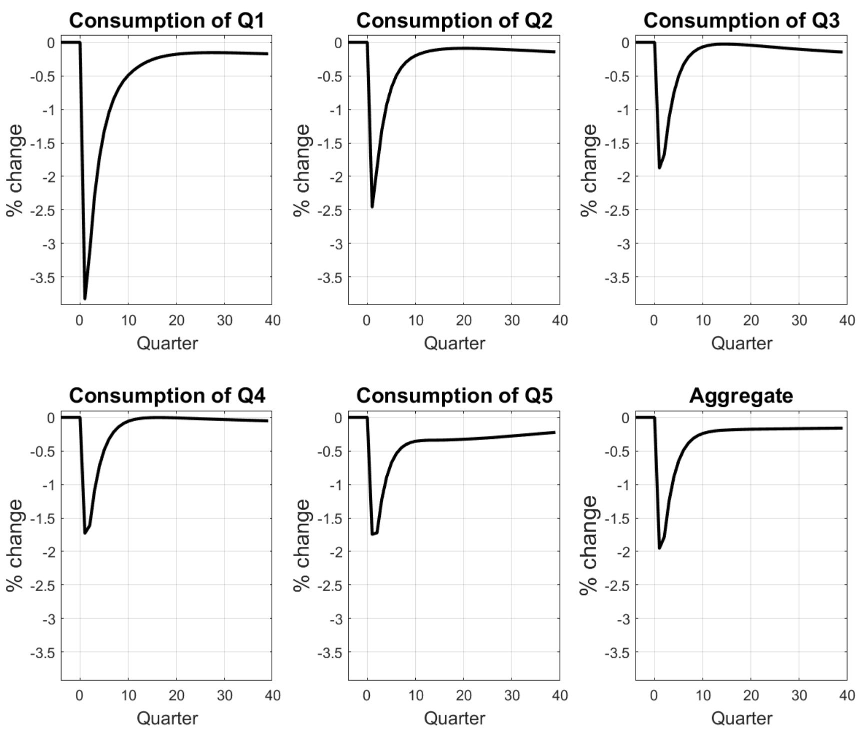
Note: Model-implied general equilibrium response of prices to the bank equity shock. The top three panels consist of rates. The three bottom panels consist of percent deviations from their respective steady-state values. The return on capital is defined as $R_t^k \equiv \frac{(r_t^K + (1-\delta)q_t}{q_{t-1}} - 1$

Table C.2: Characteristics of Gainers and Losers from the Bank Equity Shock

	Negative CE	Positive CE
Average liquid assets	0.52	7.22
Average capital holdings	0.56	3.9
Average Earnings	0.97	1.41
Average (total) income	0.95	1.65
Average Portfolio Liquidity	0.94	1.61
Dependence on labor income	92%	65%

Note: "Dependence on labor income" refers to the average share of earnings in households' total income. With the exception of the last row, numbers are displayed as a multiple of economy-wide averages.

Figure C.6: Consumption Responses by Income Quintile



Note: Model-implied consumption responses to the bank equity shock. Income quintiles are sorted based on total income in the steady state, including earnings, interest received, and dividends. Impulse responses are displayed relative to the counterfactual evolution of consumption for each group in the absence of the shock.

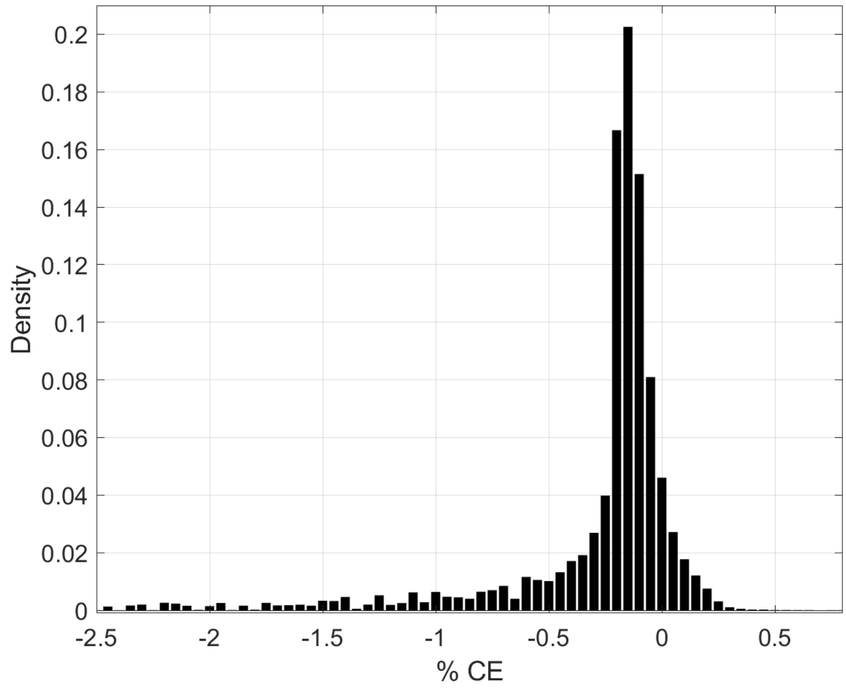


Figure C.7: Distribution of Welfare Changes - BE
Note: Distribution of welfare changes due to the bank equity shocl, measured in consumption equivalent units, as in equation 3.28.

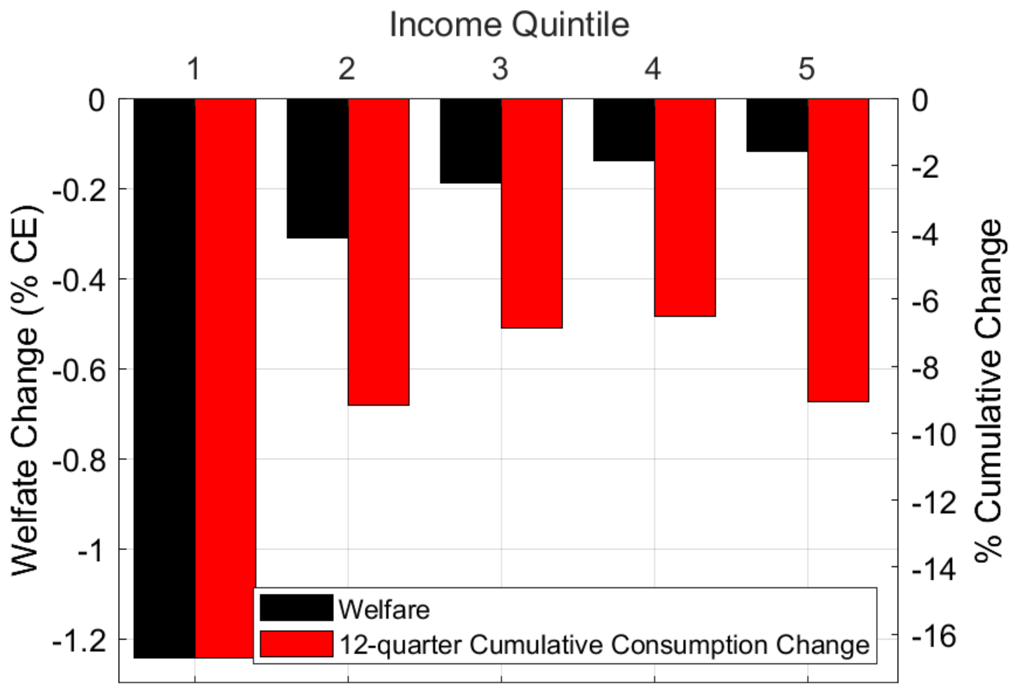


Figure C.8: Welfare and Consumption Changes - Bank Equity Shock
Note: Welfare changes, whose scale is on the left y-axis, are computed according to equation (3.28) and aggregated within each income quintile. Cumulative consumption changes are measured on the right y-axis.

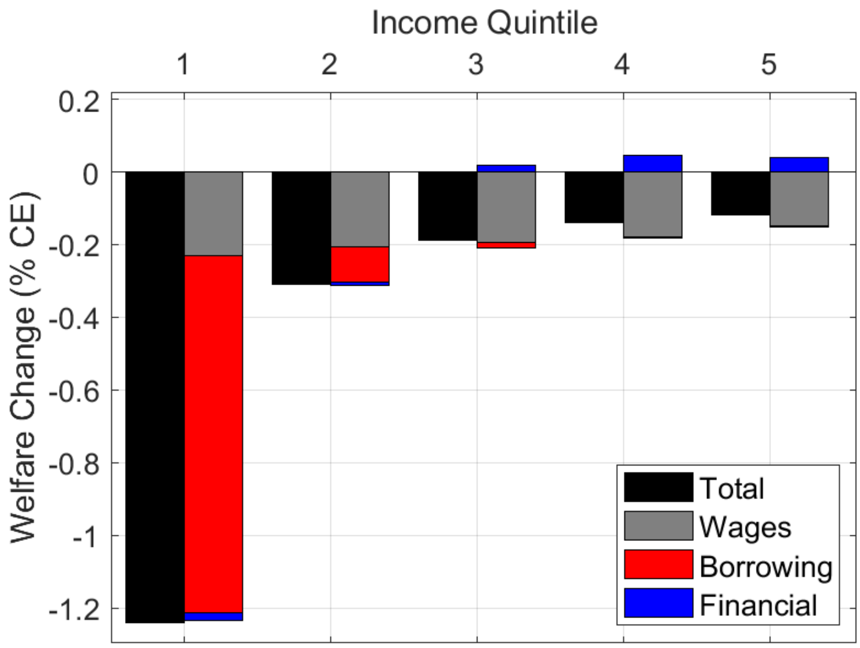


Figure C.9: Decomposition of Welfare Changes by Income Quintile

Note: Decomposition of welfare changes due to wages $\{w_t\}_{t=0}^{T}$, the lending rate $\{r_t^L\}_{t=0}^{T}$, and financial variables (jointly $\{r_t^D, r^K, q_t, div_t\}_{t=0}^{T}$) in response to the bank equity shock. The black bar represents the general equilibrium welfare changes, replicating figure C.8. Each of the gray and colored bars is obtained by simulating the economy in response to the general equilibrium path of one variable (or all four, in the case of financial variables).

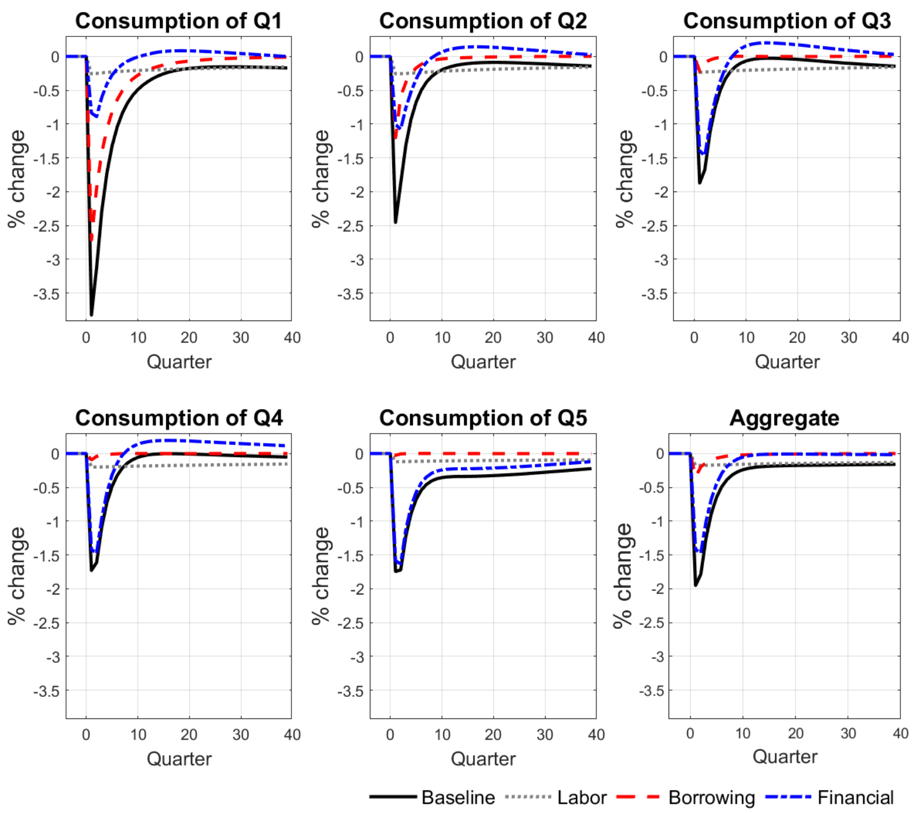


Figure C.10: Consumption Decomposition

Note: Model-implied consumption responses to the bank equity shock. Income quintiles are sorted based on total income in steady-stat, including earnings, interest received, and dividends. Impulse responses are displayed relative to the counterfactual evolution of consumption for each group in the absence of any price variation. Consumption responses are decomposed into partial equilibrium effects of wages $\{w_t\}_{t=0}^{T}$, the lending rate $\{r_t^L\}_{t=0}^{T}$, and financial variables (jointly $\{r_t^D, r^K, q_t, div_t\}_{t=0}^{T}$).

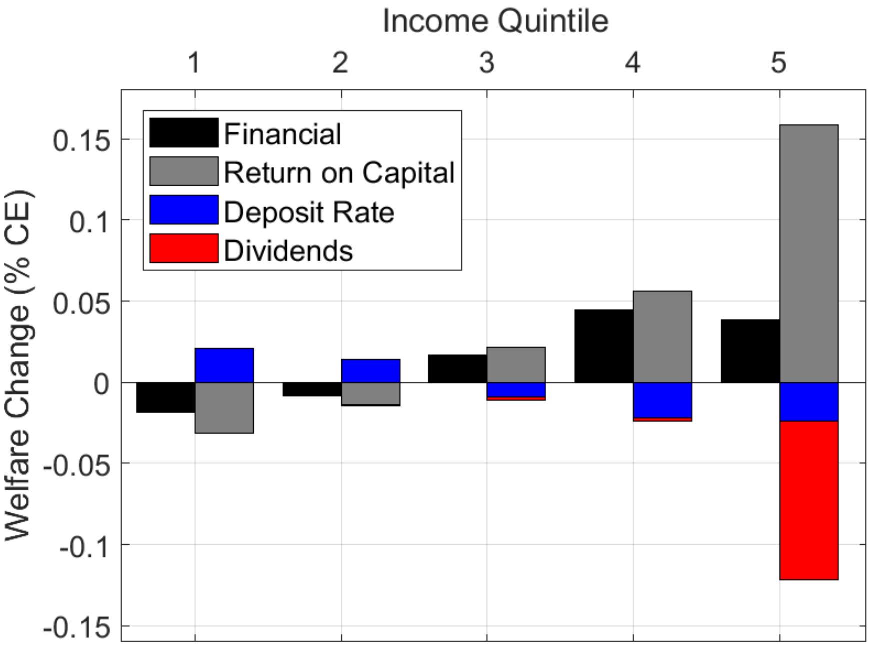


Figure C.11: Decomposition of Welfare Changes by Income Quintile
Note: Decomposition of welfare changes (in response to the bank equity shock) due to wages financial variables (jointly $\{r_t^D, R_t^K, div_t\}_{t=0}^T$, in the black bar) and each of its separate components (gray and colored bars). Each of the gray and colored bars is obtained by simulating the economy in response to the partial-equilibrium path of one variable (or all four, in the case of the black bar).

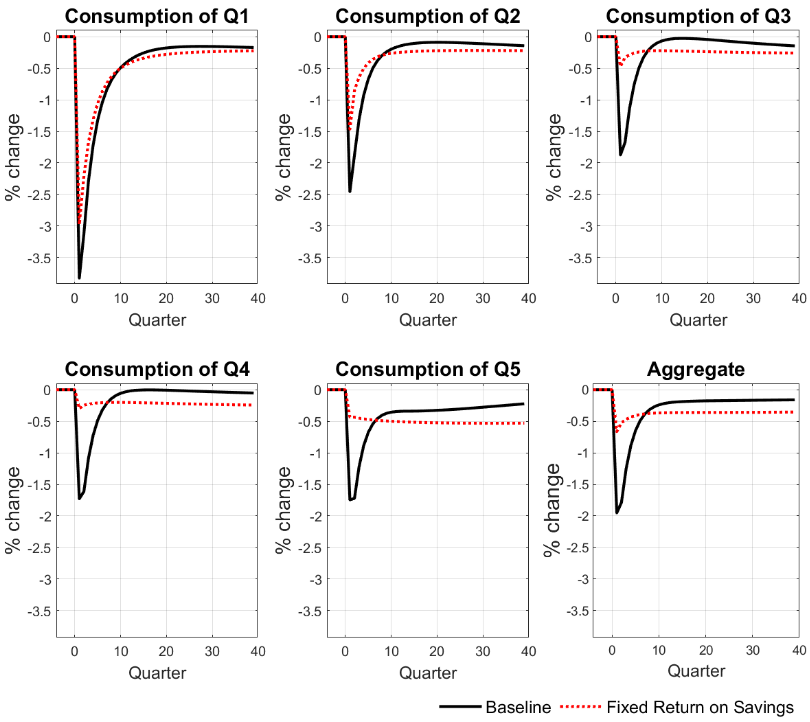


Figure C.12: Consumption Decomposition - The Role of Savings Returns

Note: Model-implied consumption responses to the bank equity shock in general equilibrium (solid line) and partial equilibrium (dotted line). Income quintiles sorted based on total income in steady state, including earnings, interest received, and dividends. Impulse responses are displayed relative to the counterfactual evolution of consumption for each group in the absence of any price variation. The dotted line shows the partial-equilibrium response to changes *only in* wages, the lending rate, and dividends ($\{w_t, r_t^L, div_t\}_{t=0}^T$).

Table C.3: Welfare Changes due to Bank Equity Shock - Heterogeneity

Quintile	Q1	Q2	Q3	Q4	Q5
by Income	-1.242	-0.311	-0.189	-0.139	-0.117
by Net Worth	-1.315	-0.239	-0.184	-0.144	-0.119

Notes: Changes in welfare measured in consumption equivalent units, as in equation 3.28.

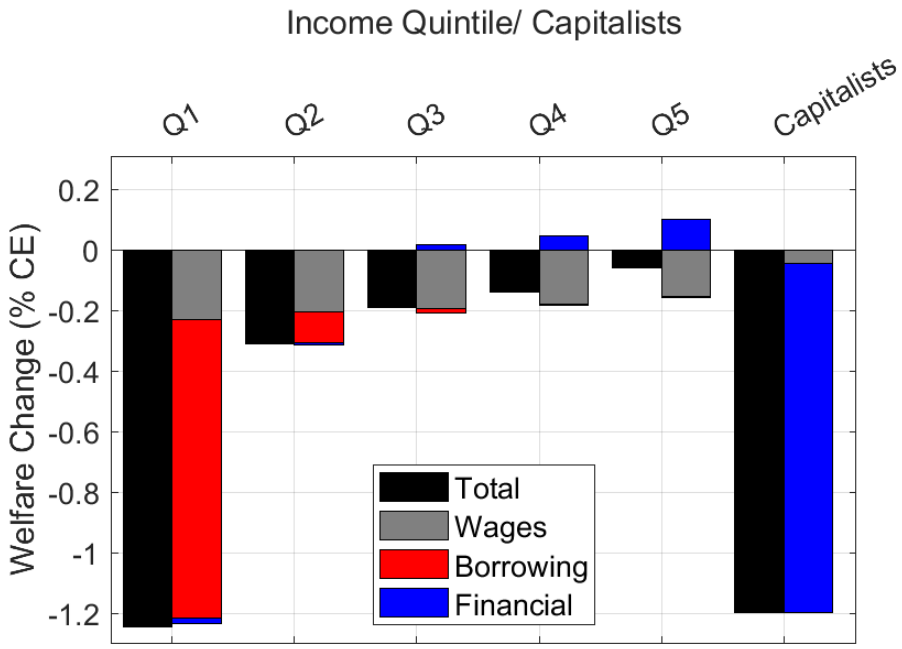


Figure C.13: Decomposition of Welfare Changes by Income Quintile

Note: Decomposition of welfare changes (due to the bank equity shock) due to wages $\{w_t\}_{t=0}^T$, the lending rate $\{r_t^L\}_{t=0}^T$, and financial variables (jointly $\{r_t^D, r^K, q_t, div_t\}_{t=0}^T$). The black bars represent the general-equilibrium welfare changes. Each of the gray and colored bars is obtained by simulating the economy in response to the general-equilibrium path of one variable (or all four, in the case of financial variables).

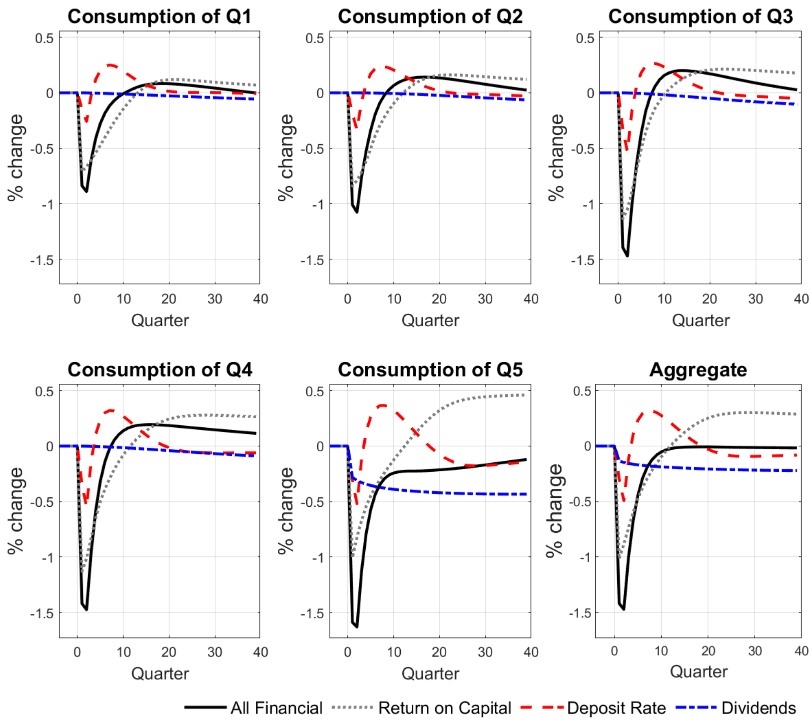


Figure C.14: Consumption Decomposition - Bank Equity Shock -Financial Variables
Note: Model-implied consumption responses to the bank equity shock to changes in *financial* variables. Income quintiles are sorted based on total income in steady state, including earnings, interest received, and dividends. Impulse responses are displayed relative to the counterfactual evolution of consumption for each group in the absence of any price variation. Consumption responses are decomposed into partial-equilibrium effects of return on capital $\{q_t, r_t^k\}_{t=0}^T$, deposit rates $\{r_t^D\}_{t=0}^T$, dividends $\{div_t\}_{t=0}^T$, and financial variables (jointly $\{r_t^D, r^K, q_t, div_t\}_{t=0}^T$).

C.3.3 Credit Policies - Additional Details

Description of Alternative Policy Interventions

Lump-Sum Rebate. In the case of the lump-sum rebate, we introduce a transfer T_t^{ls} to the right-hand side of households' budgets. Since there is a unit measure of households, $T_t^{ls} = \mathcal{T}_t$.

Proportional Rebate. For the proportional rebate, we introduce the following term to the right-hand-side of consumers' budgets:

$$T_t^p(a,k,z) = \eta_t \left[r_t^D a_t \mathbb{I}(a_t \geq 0) + w_t z + \mathbb{I}(z = z^*) div_t + (r_t^k - (1-\delta) q_t) k \right]$$

Budget balance requires $\mathcal{T}_t = \sum_{a,k,z} T_t^p(a,k,z) d\lambda(a,k,z)$, which is achieved by selecting the adequate sequence of η_t.

Rebate to Banks. In this case, equation (3.11) is modified to:

$$E_t = \underbrace{(1 + r_t^L) L_t}_{} + \underbrace{((1-\delta) q_t + \xi_t^B r_t^K) K_t^B}_{} - \underbrace{(1 + r_t^D) D_t}_{} + \mathcal{T}_t$$

Baseline Credit Policy - Additional Figures

Below we compare the transmission mechanisms (prices, interest rates, dividends) with and without the baseline credit policy intervention (lump sum rebate, $\xi^G = 1$).

Alternative Credit Policies - Results

Table C.4 below compares the welfare impacts of the credit policy proposed in section 3.6 under distinct assumptions regarding the productivity of capital financed by the government and how the proceeds from intermediation are rebated. In all cases, the credit policy improves welfare, with gains concentrated at the bottom of the income distribution.

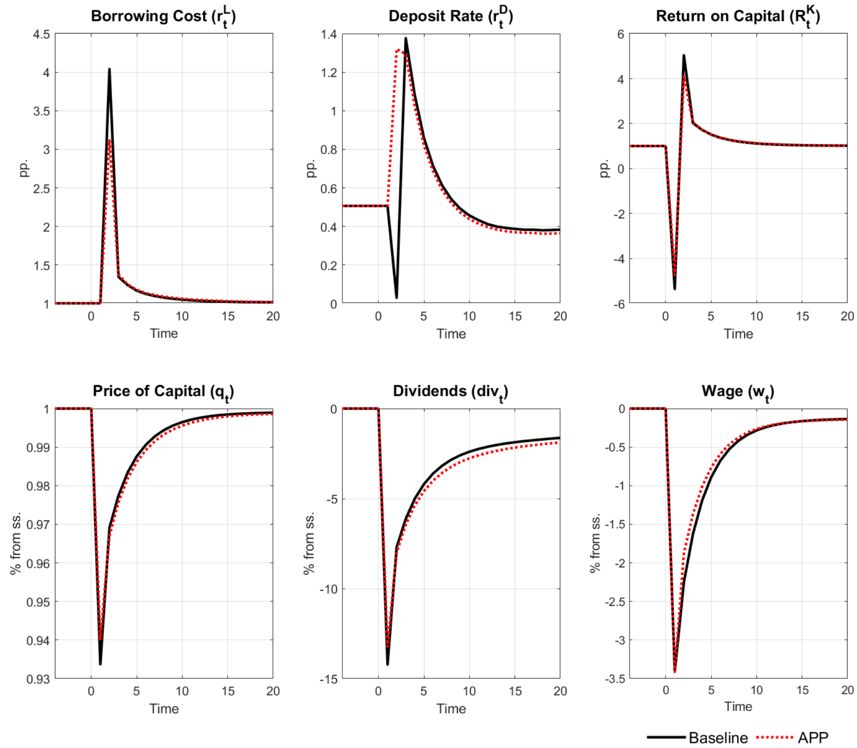


Figure C.15: General Equilibrium Price Responses

Note: Model-implied general-equilibrium responses of prices to baseline shock in the presence (red dotted line) and absence (solid black line) of the credit policy described in 3.6. The top three panels consist of rates. The three bottom panels consist of percent deviations from their respective steady-state values. The return on capital is defined as $R_t^k \equiv \frac{(r_t^K+(1-\delta)q_t}{q_{t-1}} - 1$

Table C.4: Welfare Responses - Credit Policies

	Q1	Q2	Q3	Q4	Q5	Aggregate	Capitalists
Lump sum, $\xi^G = 1$	0.2718	0.0758	0.0459	0.0349	0.0312	0.0927	-0.0155
Lump sum, $\xi^G = \xi^B$	0.2284	0.0560	0.0324	0.0265	0.0277	0.0749	-0.0208
Prop. Tax, $\xi^G = 1$	0.1956	0.0569	0.0386	0.0337	0.0357	0.0727	0.0063
Prop. Tax, $\xi^G = \xi^B$	0.1855	0.0454	0.0283	0.0258	0.0302	0.0636	-0.0090
Banks, $\xi^G = 1$	0.1861	0.0465	0.0279	0.0245	0.0297	0.0635	0.0207
Banks, $\xi^G = \xi^B$	0.1802	0.0396	0.0224	0.0206	0.0269	0.0586	-0.0008

Note: Change in welfare when credit policy is available, compared to the baseline shock. Welfare is measured according to equation 3.28. Different rebating schemes are described above. $\xi^G = \xi^B$ denotes the case when the productivity of government-intermediated capital equals that of the bank-intermediate capital.

C.3.4 Distributive Consequences of TFP Shocks

We now compare the distributive consequences of our baseline shock that *only* affects the banking sector with a recession of the same magnitude, but induced by a decline in aggregate productivity (A_t). This allows us to understand how large the distributive consequences of recessions uniquely originated in the banking sector are, relative to a disruption that affects all sectors in the economy equally. In other words, it gives us an idea of whether the bank loss channel amplifies or dampens the distributive impact of business cycles as a whole. For comparison, we calibrate the magnitude and the persistence of the TFP shock to match the same on-impact and 12-quarter-cumulative declines in aggregate consumption as in our baseline specification in Section 3.5.

Figure C.16 compares the two consumption responses across income quintiles. The TFP shock has substantially less impact for households at the bottom, with their on-impact consumption decline reduced by 0.67 percentage points, or 24 percent. For the other quintiles, differences are smaller. Figure C.17 however shows that welfare changes are more evenly distributed in the case of a TFP shock: For quintiles 1-2, the TFP shock is less harmful than the baseline, whereas the opposite is true for Q3-Q5.

Even though the transmission mechanisms described in Section 3.5 are still operative – the bank also suffers from the decline in aggregate productivity – the increases in the spread and in the future returns on capital are not as strong (Figure C.18). This is because the banks' losses in net worth associated with the TFP shock are much smaller (Figure C.19). The less severe consequences for the banking sector lead to a smaller decline in welfare for low-income households. In contrast, high-income individuals cannot benefit as much from movements in financial variables as in the case of the bank capital shock (Figure C.20) and hence face a larger decline in their welfare.

Taken together, these results suggest that even though the consumption responses to an aggregate TFP shock are similar to those in response to banking sector losses, this masks differences in welfare inequality due to the underlying transmission mechanisms.

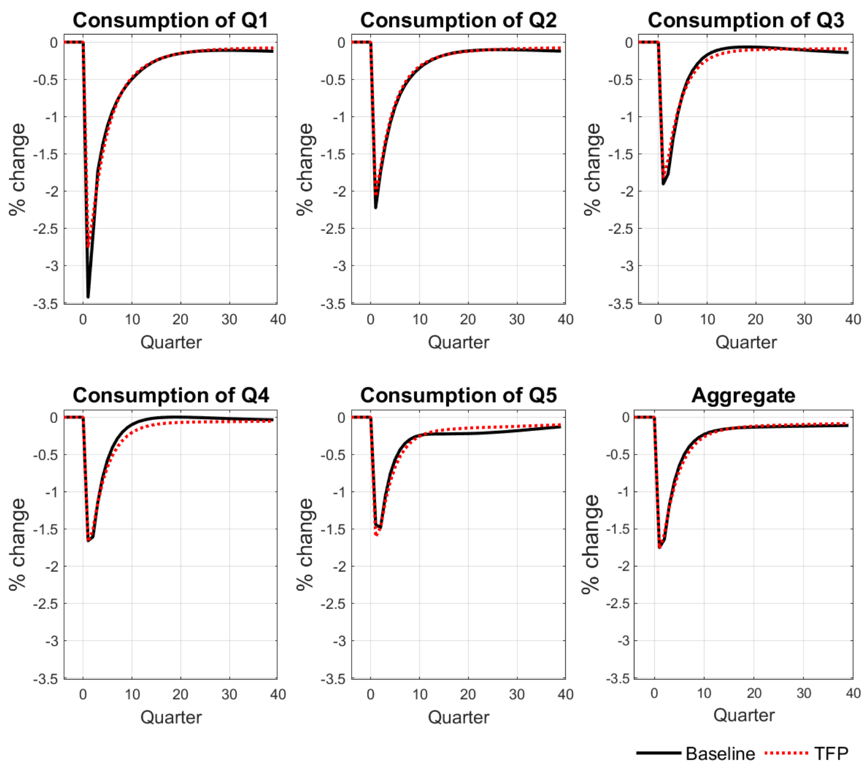


Figure C.16: Consumption Responses - TFP Shock
Note: Model-implied consumption responses to the TFP shock. Income quintiles are sorted based on total income in the steady state, including earnings, interest received, and dividends. Impulse responses are displayed relative to the counterfactual evolution of consumption for each group in the absence of the shock.

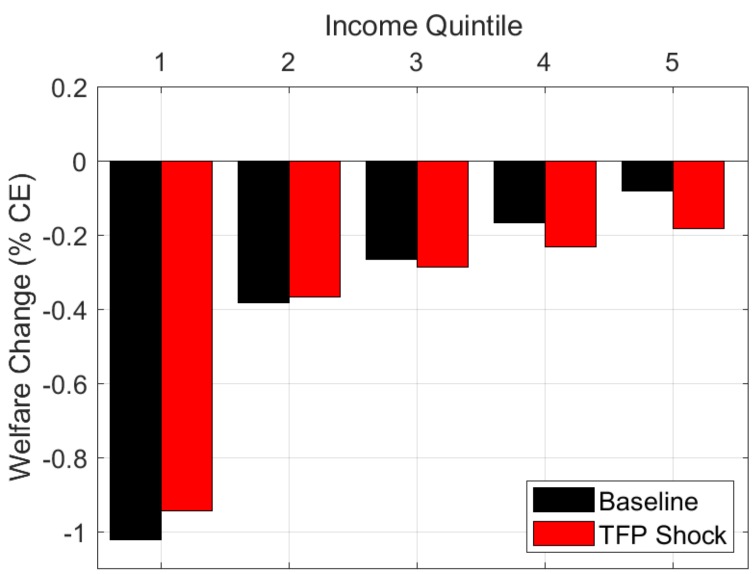


Figure C.17: Welfare Changes - Baseline vs. TFP Shock
Note: Welfare changes, computed according to equation (3.28) and aggregated within which income quintile, for baseline and TFP shocks.

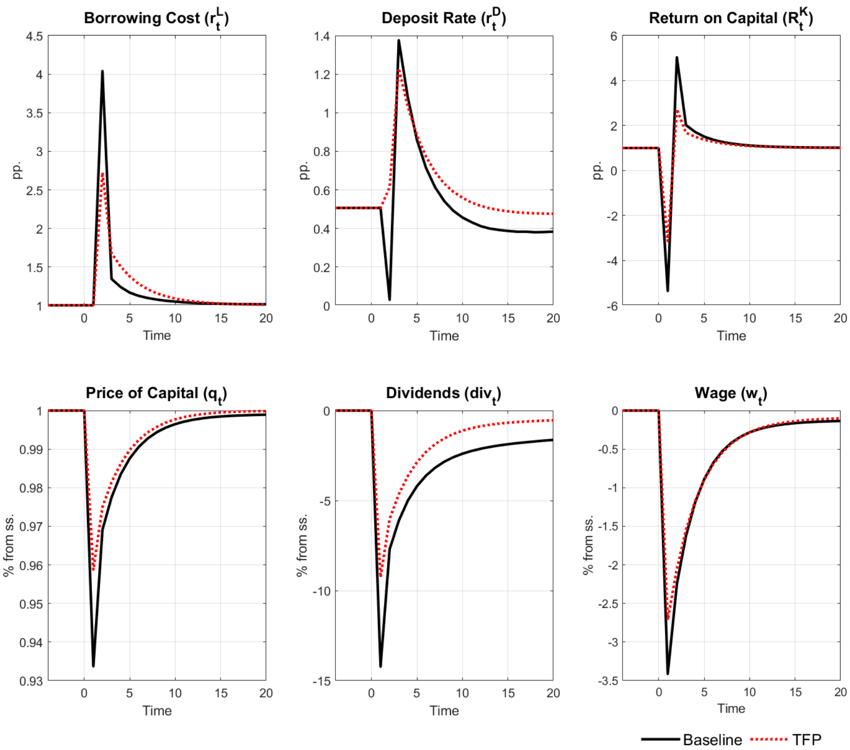


Figure C.18: General Equilibrium Price Responses

Note: Model-implied general-equilibrium response of prices to baseline and TFP shocks. The top three panels consist of rates. The three bottom panels consist of percent deviations from their respective steady-state values. The return on capital is defined as $R_t^k \equiv \frac{(r_t^K + (1-\delta)q_t}{q_{t-1}} - 1$

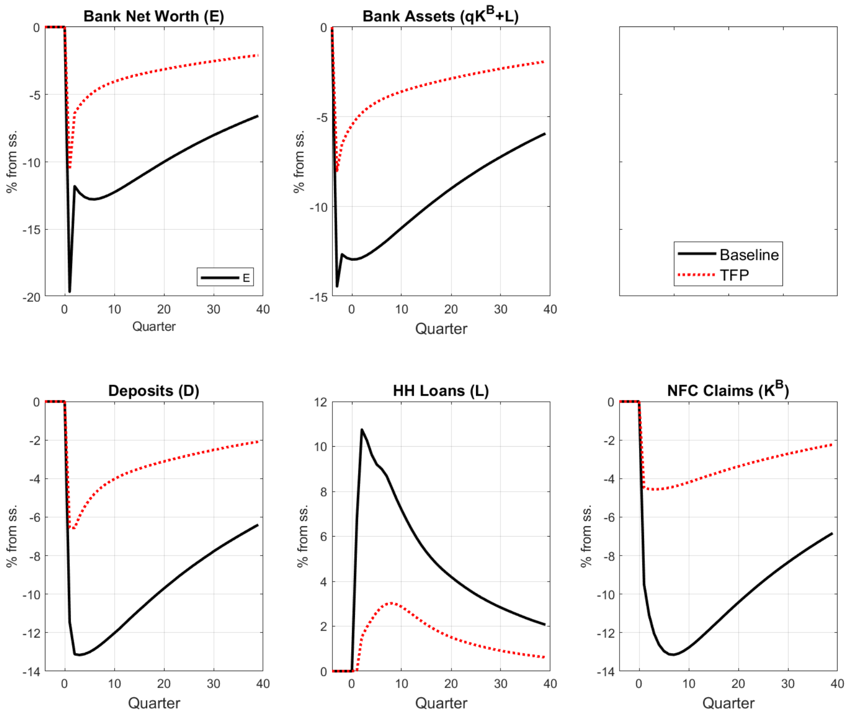


Figure C.19: Evolution of Banks' Balance Sheet Components
Note: Responses of components of the banks' balance sheets to the baseline and the TFP shock, itself represented in the top-right panel.

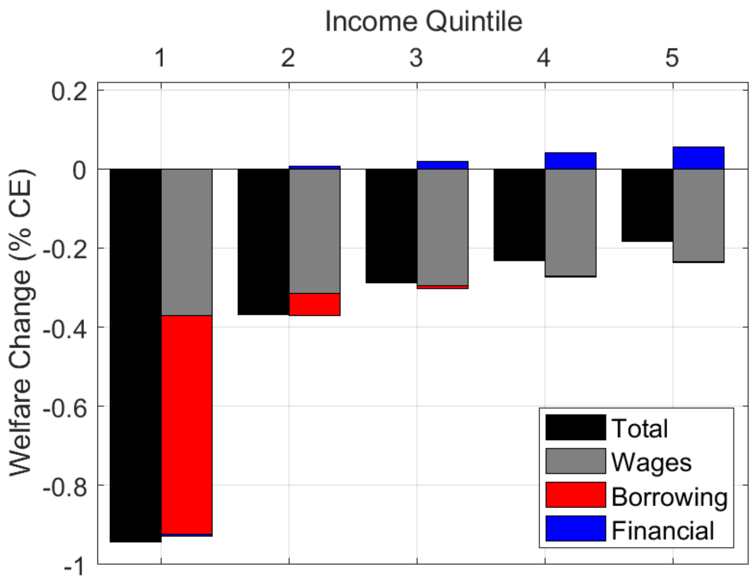


Figure C.20: Decomposition of Welfare Changes by Income Quintile - TFP Shock

Note: Decomposition of welfare changes due to wages $\{w_t\}_{t=0}^T$, the lending rate $\{r_t^L\}_{t=0}^T$, and financial variables (jointly $\{r_t^D, r^K, q_t, div_t\}_{t=0}^T$) in response to a TFP shock. The black bar represents the general-equilibrium welfare changes. Each of the gray and colored bars is obtained by simulating the economy in response to the general-equilibrium path of one variable (or all four, in the case of financial variables).

Appendix D

Appendix to Chapter 4

D.1 Empirical Observations

D.1.1 Michigan Survey of Consumers

The Michigan Survey of Consumers (MSC) is one of the most established sources of data on households' expectations. Compared to our main data source it has a disadvantage in that it does not provide comprehensive data on the wealth of participants. It only reports the current value of individuals' stock market portfolios. We use this value as a proxy for financial wealth and repeat part of the analysis on DHS data for the Michigan Survey.

An advantage of the MSC is the long time series for which consistent data are available. Data on inflation expectations and stock investment are continuously provided since September 1998. Furthermore, the data is available at monthly frequency. This does not only increase the number of observations along the time dimension, but also allows for a more precise computation of the forecast error as we can pin down the exact month of the observation. Applying the same approach as discussed above for the DHS data, we assign observations to investment quintile groups based on their position in the stock portfolio distribution in the month of their observation. We compute the expectation error as the reported forecast minus the realized inflation rate in the 12 months following the month of observation.

Figure D.1 reports the within quintile group standard deviation of expectation errors as well as the mean absolute forecast error by quintile group. Similar to

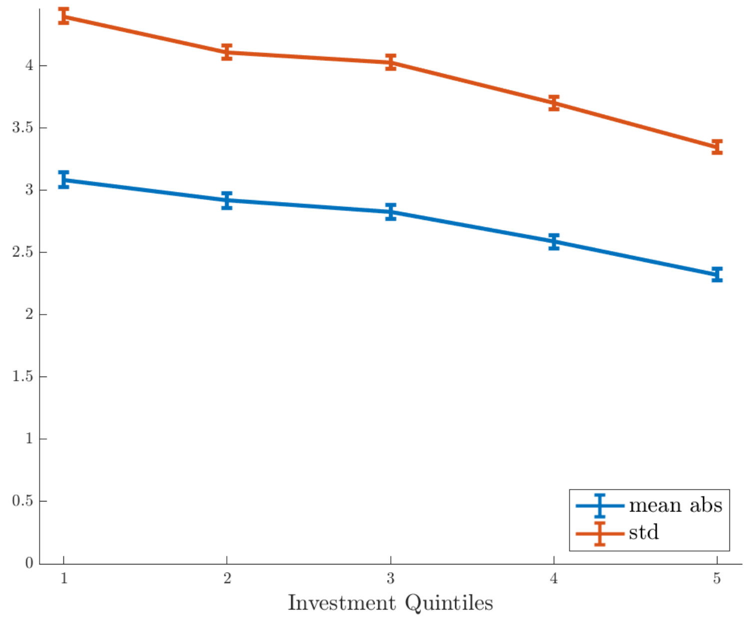
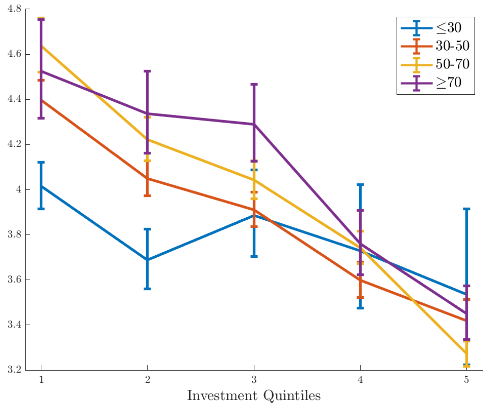


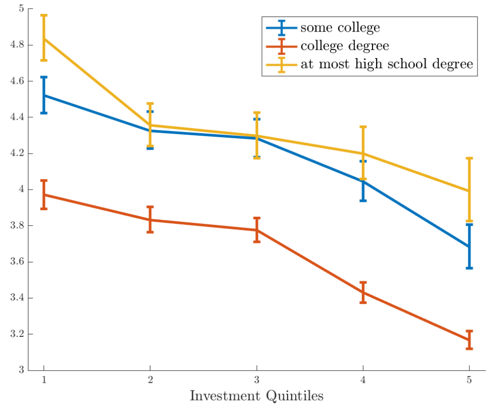
Figure D.1: Expectation Errors by Investment Quintiles (Michigan Data)

The figure plots the within quintile group standard deviation of errors and the mean absolute forecast error. Bars provide confidence bands at the 95% level. Expectation errors are ex-ante point forecasts minus ex-post realizations. Data from Michigan Survey of Consumers waves 09/1998-04/2018.

the DHS data both are declining in investment value, a pattern that is statistically significant. Note that the first quintile now begins at zero investment as naturally there are no observations reporting a negative value of their stock market portfolio. Hence, we cannot observe any drop for negative wealth levels. Interestingly, we also cannot observe a flattening out of the decline for high levels of stock investment. Again, the pattern is robust to controlling for age or education. Figure D.2 shows that the standard deviation of errors split by education and age groups. As in the Dutch data college education reduces disagreement about future inflation rates. The findings are similar for mean absolute errors, as presented in Figure D.3.



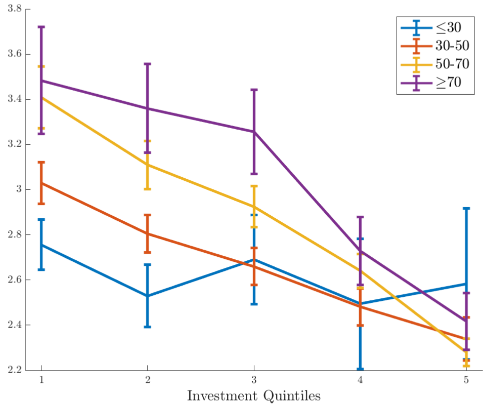
(a) by age



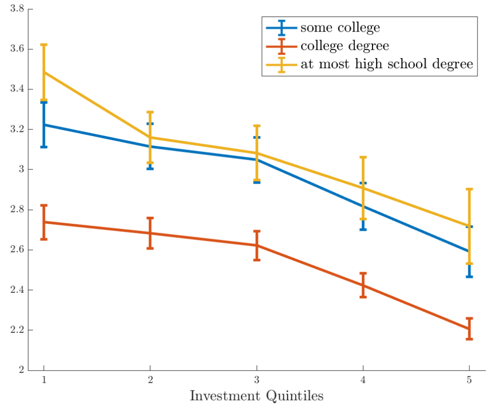
(b) by education

Figure D.2: Standard Deviation of Expectation Errors by Investment Quintiles (Michigan Data) – Controls

The figure plots the within quintile group standard deviation of errors by age (a) and education groups (b). Bars provide confidence bands at the 95% level. Expectation errors are ex-ante point forecasts minus ex-post realizations. Data from Michigan Survey of Consumers waves 09/1998-04/2018.



(a) by age



(b) by education

Figure D.3: Mean Absolute Expectation Error by Investment Quintiles (Michigan Data) – Controls

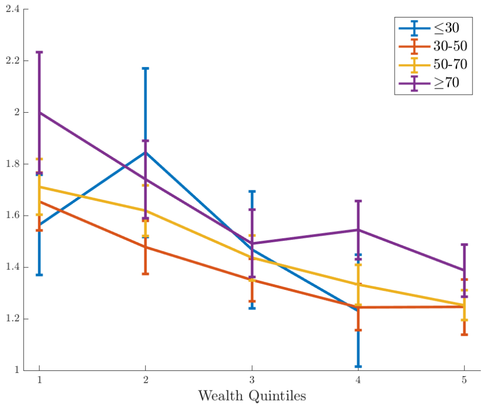
The figure plots the mean absolute forecast error for each investment quintile group by age (a) and education groups (b). Bars provide confidence bands at the 95% level. Expectation errors are ex-ante point forecasts minus ex-post realizations. Data from Michigan Survey of Consumers waves 09/1998-04/2018.

D.1.2 Additional Empirical Results

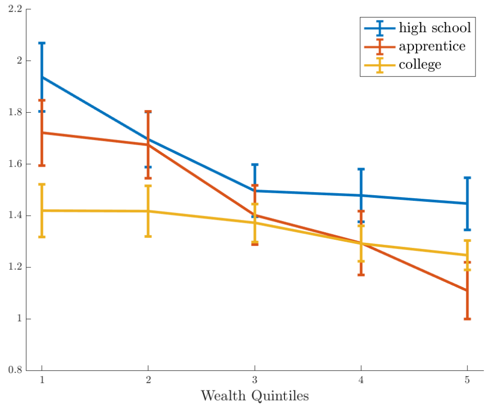
Table D.1: Net Financial Wealth Decile Groups – Summary

				Forecast Errors			
	Decile	**N**	**Mean Assets**	**Mean**	**Sd**	**Mean abs**	**N missing**
deciles by wave	1	1,272	-27,513	1.18	1.84	1.58	149
	2	1,267	-54	1.47	2.11	1.84	170
	3	1,262	1,599	1.27	1.90	1.63	128
	4	1,264	4,574	1.24	1.80	1.59	123
	5	1,261	9,670	1.04	1.70	1.47	122
	6	1,270	17,134	0.99	1.57	1.37	110
	7	1,260	27,175	1.02	1.57	1.39	104
	8	1,265	45,373	0.97	1.56	1.33	84
	9	1,264	84,787	0.86	1.49	1.29	66
	10	1,260	289,130	0.91	1.48	1.30	53
deciles pooled sample	1	1,268	-27,640	1.17	1.83	1.57	142
	2	1,261	-33	1.47	2.10	1.83	172
	3	1,265	1,519	1.30	1.92	1.67	121
	4	1,264	4,497	1.25	1.80	1.60	130
	5	1,266	9,469	1.10	1.74	1.51	118
	6	1,267	16,938	0.97	1.53	1.34	113
	7	1,261	27,115	0.96	1.55	1.36	98
	8	1,268	45,120	0.97	1.56	1.33	96
	9	1,261	84,871	0.85	1.50	1.27	64
	10	1,264	289,088	0.90	1.48	1.29	55
	Total	12,645	45,061	1.09	1.72	1.47	1,109

Data from DNB Household Survey waves 2010-2018. Summary statistics by net financial wealth decile groups. Net financial wealth refers to net wealth ex housing, mortgages, businesses and vehicles. The first block sorts households into deciles by year of observations and then pools deciles across waves. The second block pools all observations and computes deciles based on the full sample.



(a) by age



(b) by education

Figure D.4: Mean Absolute Expectation Error by Wealth Quintiles – Controls

The figure plots the mean absolute forecast error for each wealth quintile group by age (a) and education groups (b). Bars provide confidence bands at the 95% level. Expectation errors are ex-ante point forecasts minus ex-post realizations. Data from DNB Household Survey waves 2010-2018. Combination of youngest age and highest wealth quintile omitted due to lack of observations.

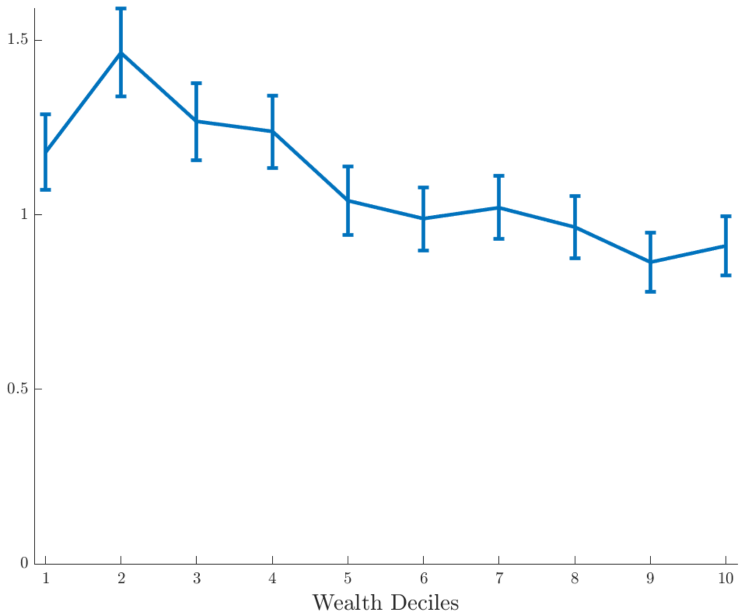


Figure D.5: Expectation Errors by Wealth Decile Groups – Mean

The figure plots the average expectation error by net financial wealth decile group. Bars provide confidence bands at the 95% level. Expectation errors are ex-ante point forecasts minus ex-post realizations. Data from DNB Household Survey waves 2010-2018.

Table D.2: Individual Absolute Forecast Errors

	$abs(err^i_{t+1})$
net financial wealth decile 2	0.208**
	(0.092)
net financial wealth decile 3	0.027
	(0.083)
net financial wealth decile 4	-0.021
	(0.079)
net financial wealth decile 5	-0.109
	(0.075)
net financial wealth decile 6	-0.205***
	(0.074)
net financial wealth decile 7	-0.195**
	(0.077)
net financial wealth decile 8	-0.229***
	(0.074)
net financial wealth decile 9	-0.272***
	(0.073)
net financial wealth decile 10	-0.253***
	(0.076)
high school	-0.012
	(0.130)
apprenticeship	-0.126
	(0.131)
college	-0.219*
	(0.127)
age 30-50	-0.089
	(0.083)
age 50-70	-0.062
	(0.084)
age >70	0.075
	(0.088)
home owner	-0.103**
	(0.045)
constant	1.817***
	(0.147)
Observations	11532
Adjusted R^2	0.0215

Data from DNB Household Survey waves 2010-2018. Household level regression of absolute forecast errors on households' wealth decile, education and age of the household head and an indicator for owning the primary residence. Standard errors (in parentheses) clustered at the household level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

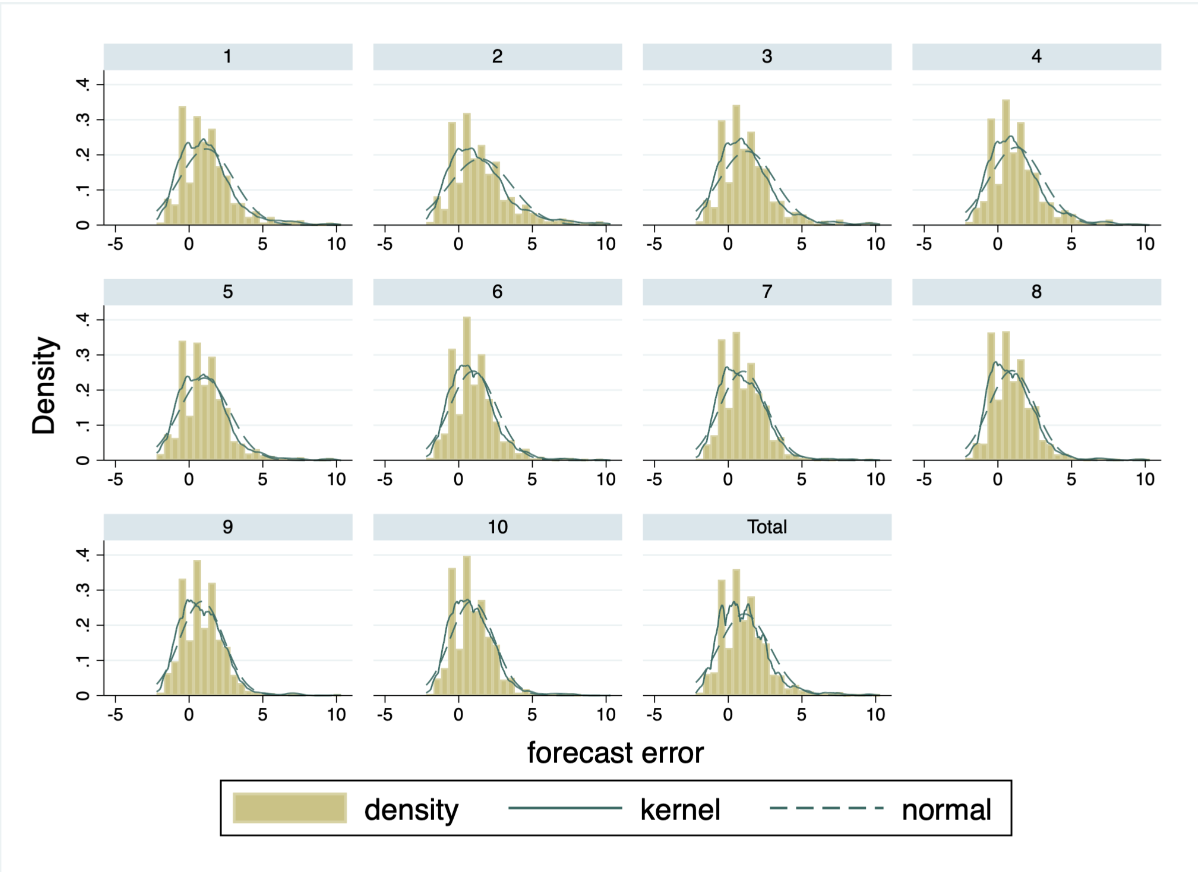


Figure D.6: Distribution of Errors by Wealth Decile Groups

The figure plots histograms of the distribution of expectation errors by wealth decile group along with fitted kernel densities and normal densities with identical mean and standard deviation as a reference point. Expectation errors are ex-ante point forecasts minus ex-post realizations. Data from DNB Household Survey waves 2010-2018.

D.2 Theoretical Framework

D.2.1 Extension - Fundamental Disagreement

In our baseline model of expectation formation we abstract from fundamental disagreement about the underlying model of inflation and its parameters or any other exogenously imposed heterogeneity in beliefs to focus on endogenous expectation formation. Nevertheless, additional sources of disagreement among households can be important to capture moments of the data our baseline model fails to explain, such as e.g. a positive mean error or a positive error covariance.[1] We therefore extend our empirical analysis in order to evaluate the potential impact of other sources of heterogeneity in expectations on our findings.

To test for robustness towards including fundamental disagreement, we adjust our baseline model of expectation formation to incorporate heterogeneity in beliefs about the long run mean of inflation μ. Household i's belief about μ is denoted μ^i and assumed to be distributed normally among households. Furthermore, we assume $\mu^i \perp\!\!\!\perp s_t^i \; \forall i, t$. With all other notation as before, household i's inflation expectation and expectation error are now given as

$$\mathbb{E}_t^i[\pi_{t+1}|\hat{e}_{t+1}^i, n_t^i] = (1-\rho)\mu^i + \rho\pi_t + \omega_{t+1}^i(n_t^i)\hat{e}_{t+1}^i \tag{D.1}$$

$$\begin{aligned} err_{t+1}^i &= \mathbb{E}_t^i[\pi_{t+1}|\hat{e}_{t+1}^i, n_t^i] - \pi_{t+1} \\ &= (1-\rho)(\mu^i - \mu) + (\omega_{t+1}^i(n_t^i)s_{t+1}^i - (1-\omega_{t+1}^i(n_t^i))e_{t+1}). \end{aligned} \tag{D.2}$$

The error now includes an additional term accounting for households' misperception of the long run mean. Denote the average belief about the long term mean of a group g of households as $\bar{\mu}^g$ and its variance as σ_μ^{g2}. Assuming, as before, that households in group g exert the same effort $\bar{n}_t^g$, the variance of errors across households in group g and over time becomes

$$\begin{aligned} \text{Var}^g(err_{t+1}^i) &= (1-\rho)^2\text{Var}(\mu^i) + (\omega_{t+1}^g(\bar{n}_t^g))^2\sigma_s^2(\bar{n}_t^g) + (1-\omega_{t+1}^g(\bar{n}_t^g))^2\sigma_e^2 \\ &= (1-\rho)^2\sigma_\mu^{g2} + \frac{\sigma_e^2\sigma_s^2(\bar{n}_t^g)}{\sigma_e^2 + \sigma_s^2(\bar{n}_t^g)} = (1-\rho)^2\sigma_\mu^{g2} + \overline{SU}_{t+1}^{g2}, \end{aligned} \tag{D.3}$$

[1] See Figure D.5 and D.1.

where now the endogenous subjective uncertainty term $\overline{SU}^{g2}_{t+1}$ is adjusted by the within-group fundamental disagreement about μ. Disagreement among households can hence be decomposed into disagreement about the long run mean and households' subjective uncertainty. We can also compute the covariance of the ex-post errors across time. This is given as

$$\text{Cov}^g(err^i_{t+1}, err^i_t) = (1-\rho)^2\mathbb{E}[(\mu^i-\mu)^2] - (1-\rho)^2(E[(\mu^i-\mu)])^2 = (1-\rho)^2\sigma^{g2}_{\mu}. \tag{D.4}$$

Together, (D.3) and (D.4) allow us to identify the endogenous component of error dispersion in the presence of fundamental disagreement from the difference between variance and covariance of forecast errors as

$$\overline{SU}^g_{t+1} = \sqrt{\text{Var}^g(err^i_{t+1}) - \text{Cov}^g(err^i_{t+1}, err^i_t)}. \tag{D.5}$$

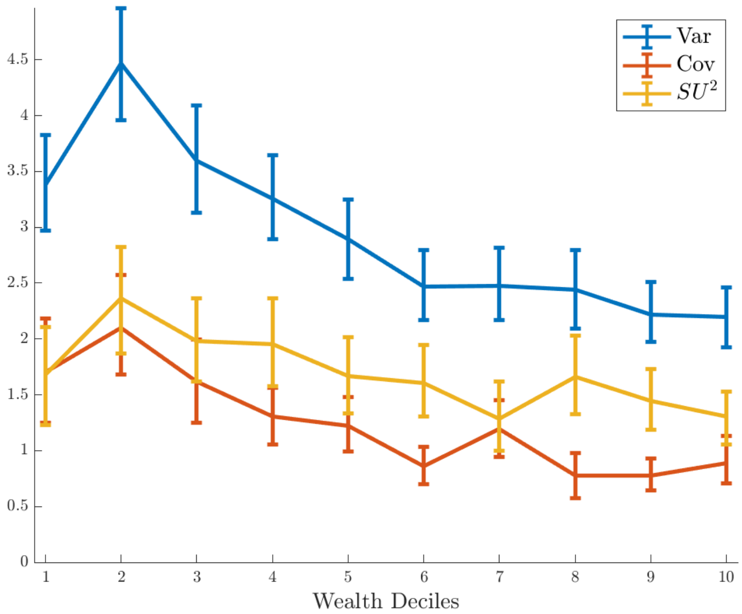


Figure D.1: Expectation Error Variance – Decomposition

The figure decomposes the variance of expectation errors across households (Var) by wealth decile groups into error covariance (Cov) and the square of subjective uncertainty (SU^2) as in (D.5). Data from DNB Household Survey waves 2010-2018. Bootstrapped 95% confidence intervals.

Intuitively, the covariance of errors is a sufficient statistic to measure heterogeneity in beliefs about the long run mean as we assume noise to be uncorrelated over time. Persistent beliefs about misreporting in current inflation (household i assuming actual inflation $\tilde{\pi}^i_t = \pi_t + \bar{\pi}^i$), as well as dispersion in beliefs about the mean of the signal s or the shock e can be treated similarly as long as they are constant over time at the household level.

We apply equation (D.5) to our data and compute the implied subjective uncertainty of households by subtracting for each wealth decile group the error covariance over time from the within group variance. The result is presented in Figure D.1. The implied subjective uncertainty exhibits a similar pattern as our benchmark results. It is slightly increasing between the first and second decile group and broadly decreasing for further increases in wealth. The covariance, which according to the extended model is driven by the dispersion of beliefs about the long-run mean, is equally higher among households with lower wealth and decreasing alongside subjective uncertainty. Both the decreases in subjective uncertainty and covariance between their respective peaks and lowest points are significant at the 95% level . Of the overall drop in the variance of expectation errors across households, about half is attributable to the fall in exogenous disagreement about the long run mean (covariance) and half to a fall in endogenous subjective uncertainty.

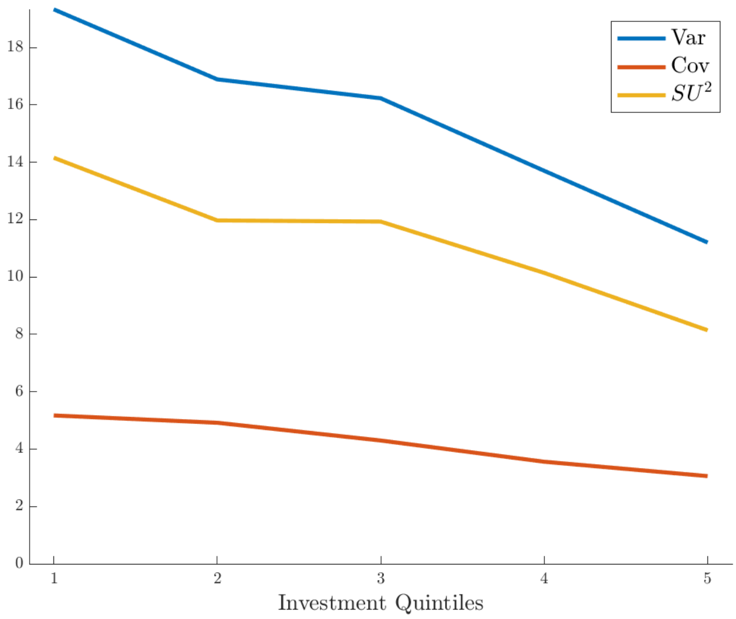


Figure D.2: Expectation Error Variance – Decomposition (Michigan Data)

The figure decomposes the cross-sectional variance of expectation errors (Var) by investment quintile groups into error covariance (Cov) and subjective uncertainty (SU) as in (D.5). Data from the Michigan Survey of Consumers waves 09/1998-04/2018.

Figure D.2 provides the decomposition of error variance across households by investment quintile group into covariance and subjective uncertainty as of equation (D.5) in the MSC. Even after allowing for fundamental disagreement almost all of the decline of within quintile group error variance is attributed to a decline in subjective uncertainty, while error covariance declines only modestly.

We take these findings as evidence that existence of the mechanism in our benchmark model is robust to incorporating fundamental disagreement.

D.2.2 Dynamic Budget Constraint - From Nominal to Real

Starting with nominal assets $\hat{a}$

$$
\begin{aligned}
Pc + \hat{a}' &= (1+r^n)\hat{a} + y - P\mathcal{F}(n,i) \\
c + \frac{\hat{a}'}{P} &= (1+r^n)\frac{\hat{a}}{P} + y - \mathcal{F}(n,i).
\end{aligned}
$$

Define $a' = \frac{\hat{a}'}{P}$, i.e. tomorrow's nominal assets in today's real consumption, and inflation rate $1+\pi = \frac{P}{P_{-1}}$

$$
\begin{aligned}
c + a' &= (1+r^n)\frac{P_{-1}}{P}a + y - \mathcal{F}(n,i) \\
c + a' &= \frac{1+r^n}{1+\pi}a + y - \mathcal{F}(n,i).
\end{aligned}
$$

D.3 Endogenous Expectations in a Two Period Model

To highlight the mechanism through which households' wealth levels impact their expectation formation, it is instructive to analyze the properties of a two period model. In the interest of a simpler exposition, we abstract from inflation entirely and focus directly on risk to the real interest rate. This is without loss of generality, since fluctuations in inflation translate into fluctuations in the real interest rate as long as nominal rates are not assumed to adjust one-for-one with inflation. Furthermore, their impact on real interest rates is the only channel through which fluctuations in inflation are relevant to the household's problem as long as additional (labour) income is assumed to be in real terms. These are the same assumptions we impose in the dynamic model where we consider inflation explicitly, making the two approaches comparable.

D.3.1 A Two Period Model

A household lives for two periods and maximizes utility by choosing consumption in both periods (c_1 and c_2) as well as savings a between periods. In both periods he receives a deterministic and constant income y. Additionally, at the beginning of the first period the household receives initial assets A. Preferences of the household are recursive, following Epstein and Zin (1989).

The real interest rate r between the two periods is stochastic. Before choosing savings in period 1, the household receives a noisy signal $\hat{r}$ about the interest rate. The distribution of the interest rate and the signal are given as

$$r \sim \mathcal{N}(\bar{r}, \sigma_r^2) \qquad \hat{r} = r + s \qquad s \sim \mathcal{N}(0, \sigma_s^2(n)), \tag{D.6}$$

where s is pure noise. Before receiving the signal, the household can influence the variance of the noise by exerting some effort n, for which he has to incur a monetary cost $\mathcal{F}(n)$. Based on the signal, the household forms a Bayesian posterior belief about the true interest rate r, attaching weight $\omega(n)$ to the signal received. Hence, conditional on n and $\hat{r}$, the posterior distribution is given as

$$\begin{aligned} r_{|n,\hat{r}} \sim \mathcal{N}((1-\omega(n))\bar{r} + \omega(n)\hat{r}, \omega(n)\sigma_s^2(n)) \\ \omega(n) = \frac{\sigma_r^2}{\sigma_r^2 + \sigma_s^2(n)}. \end{aligned} \tag{D.7}$$

We will refer to the standard deviation of a household's posterior belief about r (given by $\sqrt{\omega(n)\sigma_s^2(n)}$) as his subjective uncertainty about the future interest rate.

The household's effort choice problem is then given as

$$\tilde{V}(A) = \max_n \mathbb{E}_{\hat{r}}[V(A, n, \hat{r})|n]. \tag{D.8}$$

Conditional on having exerted effort n and receiving signal $\hat{r}$, the consumption-savings problem is given by

$$\begin{aligned} V(A,n,\hat{r}) = & \max_a \left(c_1^{1-\gamma} + \beta \left(\mathbb{E}_r[c_2^{1-\alpha}|\hat{r},n] \right)^{\frac{1-\gamma}{1-\alpha}} \right)^{\frac{1}{1-\gamma}} \\ & c_1 = A + y - a - \mathcal{F}(n) \\ & c_2 = (1+r)a + y \quad \forall\, r \end{aligned} \tag{D.9}$$

For the cost of effort and the relationship between effort and noise in the signal we assume functional forms

$$\sigma_s(n) = \frac{\chi}{1+n} \qquad \mathcal{F}(n) = (\theta n)^{\phi}. \tag{D.10}$$

These choices yield convex cost of and convex gains from exerting effort.[2] Note that with these functional forms χ is the variation in the noise if zero effort is exerted, i.e. the maximum variation possible, and that zero effort implies zero cost.

To highlight some properties of the proposed mechanism, we calibrate the model outlined above. The calibration is ad-hoc and for instructive purposes only. It is provided in Table D.1.

Table D.1: Two Period Model – Calibration

Parameter	Value
γ	2
α	2
β	0.98
y	4
$\bar{r}$	0.02
σ_r	0.01
ϕ	2
θ	0.005
χ	0.03

Calibration for the two period model. Values are ad-hoc and only for instructive purpose.

[2] $\sigma_s'(n) < 0,\ \sigma_s''(n) > 0$ and $\mathcal{F}'(n) > 0,\ \mathcal{F}''(n) \geq 0,$ iff $\phi \geq 1$.

D.3.2 Information Incentives

To study households' incentives to form precise expectations, we begin by taking the effort choice n as exogenously given. In order to do so, we drop the $\max$-operator in (D.8) and set the cost in (D.9) to $\mathcal{F}(n) = 0 \; \forall \; n$. After solving the households' problem for given n we can compute a certainty equivalence consumption level cec_n, satisfying

$$\tilde{V}_n(A) = \left(cec_n^{1-\gamma} + \beta \left(cec_n^{1-\alpha} \right)^{\frac{1-\gamma}{1-\alpha}} \right)^{\frac{1}{1-\gamma}}, \tag{D.11}$$

where $\tilde{V}_n(A)$ is the value of (D.8) for exogenously given n and zero cost of effort. We use this certainty equivalent to construct a measure of the benefit of decreasing the noise in the signal as

$$\Delta cec_n = \frac{cec_n}{cec_0} - 1, \tag{D.12}$$

which is the percentage change in the certainty equivalence consumption level if effort is increased from 0 to n, and hence the standard deviation of the noise is decreased from χ to $\sigma_s(n)$.

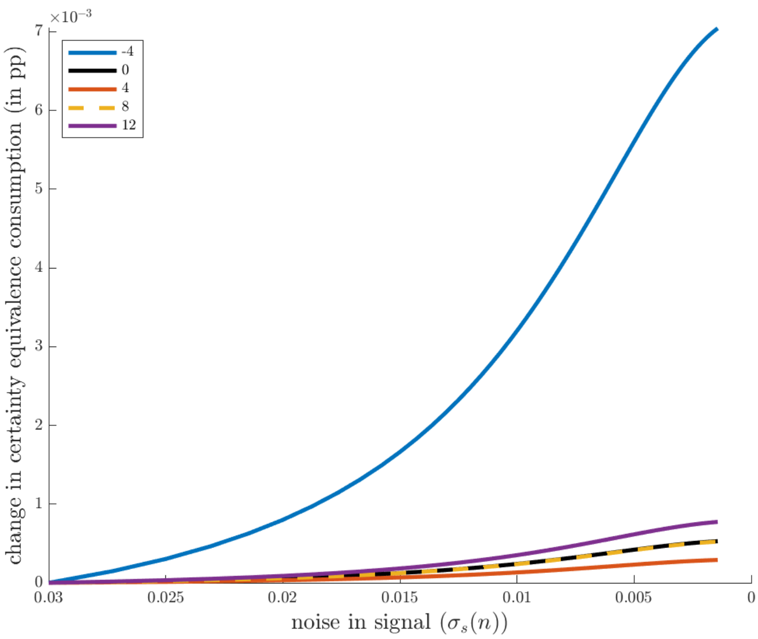


Figure D.1: Change in Certainty Equivalence Consumption

The figure plots the percentage gain in certainty equivalence consumption (cec_n, as defined in (D.11)) of decreasing the standard deviation of the noise in the signal from χ to $\sigma_s(n)$. Each line represents a different initial asset level A.

Figure D.1 plots results from the calibrated model for a range of initial asset values A. The gain from decreasing the variation in noise is highest for households starting with debt. It decreases as initial asset levels increase towards zero and modestly positive values for A and increases again once A becomes substantially positive. Note that the gains of decreasing the variation in the noise are small, for the given calibration below 0.01% of the certainty equivalent consumption level. This is evidence that already small cost of forming precise expectations might deter households from doing so.

The pattern of noise in wealth can be explained by two forces, governing households' incentives to form precise expectations: Exposure and absolute risk aversion. Exposure is given by the absolute value of a household's savings or borrowing between the two periods. It determines the relevance of the risk for a household. The higher absolute savings, the larger are expected fluctuations in period 2 consumption due to fluctuations in the interest rate. In the presence of risk aversion, fluctuations in future consumption reduce expected utility. Hence households with larger fluctuations in their future consumption due to the risk have stronger incentives to reduce the perceived risk and form more precise expectations. The exposure effect is therefore higher for households with either higher initial debt or higher (positive) initial assets, who engage in borrowing/saving between periods, but low for households with A close to zero, as these households save/borrow little between $t = 1$ and $t = 2$. Absolute risk aversion, as usual, implies that any absolute fluctuation in consumption has higher cost in terms of expected utility to households with a lower average consumption level. This effect is hence highest for households with higher debt (A substantially negative), as these households have the lowest consumption levels, and decreases as A increases.

To highlight the two effects on the change in certainty equivalence consumption, we conduct two quantitative experiments. For the first, we eliminate differences in the absolute risk aversion of households with different A to focus solely on exposure. This is achieved by compensating each household to obtain the same average consumption level as a benchmark household, which we chose to be a household with initial assets $A = -4$. More specifically, we fix the savings choice of a household at the optimal choice without any compensation.

Conditional on the exogenously set effort n and the signal received $\hat{r}$, each household receives a deterministic transfer for both periods, satisfying

$$\begin{aligned}\Delta c_1(A,n,\hat{r}) &= c_1(-4,n,\hat{r}) - c_1(A,n,\hat{r}) \\ \Delta c_2(A,n,\hat{r}) &= \mathbb{E}_r[c_2(-4,n,\hat{r})|n,\hat{r}] - \mathbb{E}_r[c_2(A,n,\hat{r})|n,\hat{r}].\end{aligned} \tag{D.13}$$

As this equalizes consumption levels across households, any difference in the remaining effect on the certainty equivalence consumption should be due to different exposure.

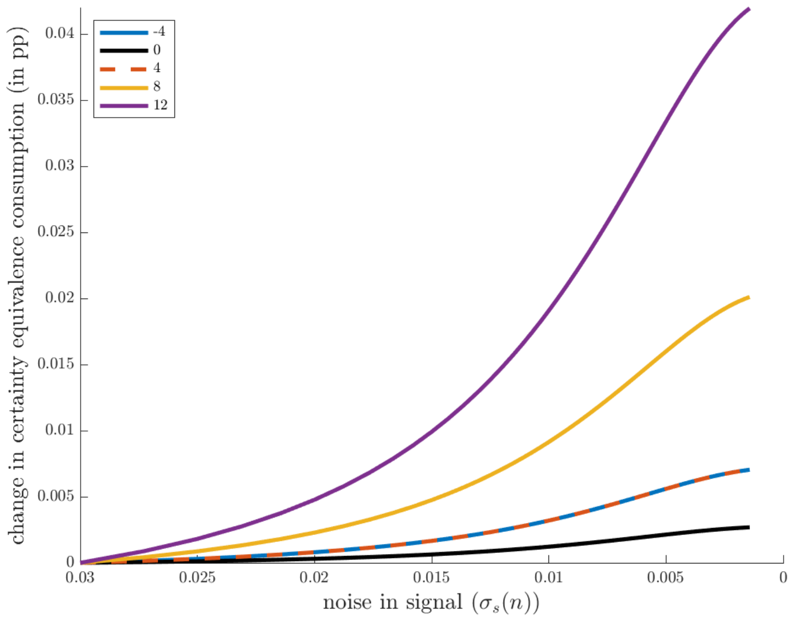


Figure D.2: Change in Certainty Equivalence Consumption – Exposure

The figure plots the adjusted percentage gain in certainty equivalence consumption (cec_n, as defined in (D.11)) of decreasing the standard deviation of the noise in the signal from χ to $\sigma_s(n)$. Adjustment equalizes average consumption levels across households to the level of a households with $A = -4$, as given in (D.13), while leaving the savings choice unchanged. Each line represents a different initial asset level A.

Figure D.2 plots the quantitative results. As expected, the change in the certainty equivalence consumption level is monotonically increasing in the absolute value of A, which is directly related to the absolute value of households' savings between periods. Note that the effect of decreasing the variation in noise is almost identical for households with $A = 4$ and $A = -4$. This reflects their, in absolute values and on average across signals, almost identical savings choices, implying a similar exposure to interest rate risk.

To control for the exposure effect and highlight the influence of absolute risk aversion, we can conduct a similar experiment by normalizing households savings choice. We assign every households the savings choice of a household with $A = 10$ (i.e. $s(10, n, \hat{r})$), controlling for n and $\hat{r}$. We additionally assign transfers, such that the household has the same average consumption level as before. These are given as

$$\begin{aligned} \tilde{\Delta} c_1(A, n, \hat{r}) &= s(10, n, \hat{r}) - s(A, n, \hat{r}) \\ \tilde{\Delta} c_2(A, n, \hat{r}) &= \mathbb{E}_r[c_2(A, n, \hat{r})|n, \hat{r}] - \mathbb{E}_r[c_2(10, n, \hat{r})|n, \hat{r}]. \end{aligned} \tag{D.14}$$

The results can be interpreted as the gain from decreasing the variation in noise for households with identical savings choice but varying consumption levels.

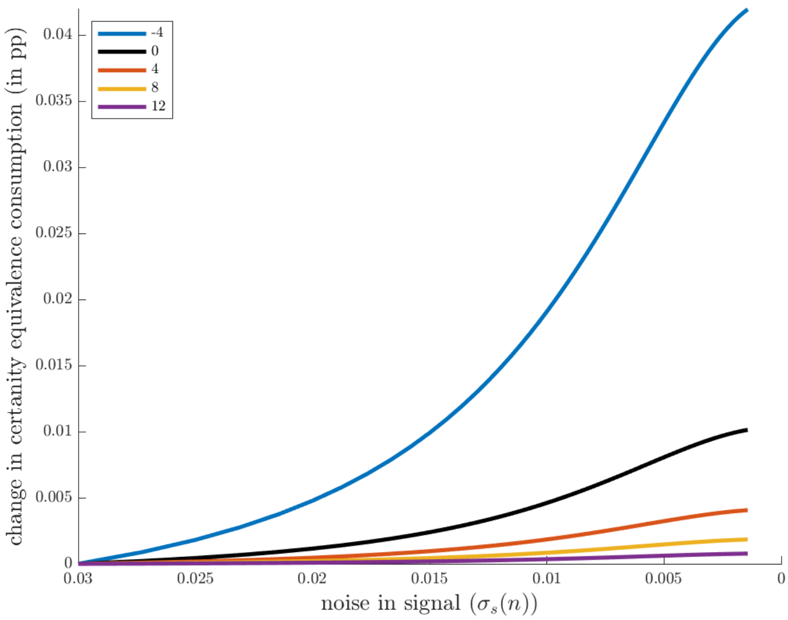


Figure D.3: Change in Certainty Equivalence Consumption – Absolute Risk Aversion

The figure plots the adjusted percentage change in certainty equivalence consumption (cec_n, as defined in (D.11)) of decreasing the standard deviation of the noise in the signal from χ to $\sigma_s(n)$. Adjustment equalizes savings across households to the level of a households with $A = 10$ as given in (D.14) while leaving the average consumption level of the household unchanged. Each line represents a different initial asset level A.

Figure D.3 plots the quantitative results. Unsurprisingly, when controlling for the savings choice, households with lower consumption level (and hence higher absolute risk aversion) profit more from a reduction of uncertainty. The gain from increasing n / reducing $\sigma_s(n)$ is decreasing in A.

D.3.3 Information Choice

We can summarize the findings above to make predictions about how households decide on effort n when the choice is endogenous. The exposure effect is increasing in households absolute initial wealth, as their future absolute savings will be equally increasing. This implies, that starting at a wealth level of zero, the further away we move in any direction along the wealth distribution the more effort households should want to exert due to the exposure effect. This effect is almost symmetric for positive and negative values of initial assets A. Absolute risk aversion is, however, monotonically decreasing in wealth. It reinforces the exposure effect, but more so for negative asset levels. The effect of absolute risk aversion is hence asymmetric in positive/negative wealth. We should hence expect the chosen noise in the signal to peak around zero wealth, decline as we move away from zero wealth in any direction, but decline steeper for negative wealth than for positive wealth. All discussion above assumes that effort is equally costly for all households. With the specification for effort to have monetary cost, this is not true in utility terms, as the same monetary costs transmit into higher utility cost for households with lower consumption levels. This adds an additional dimension of heterogeneity in incentives.

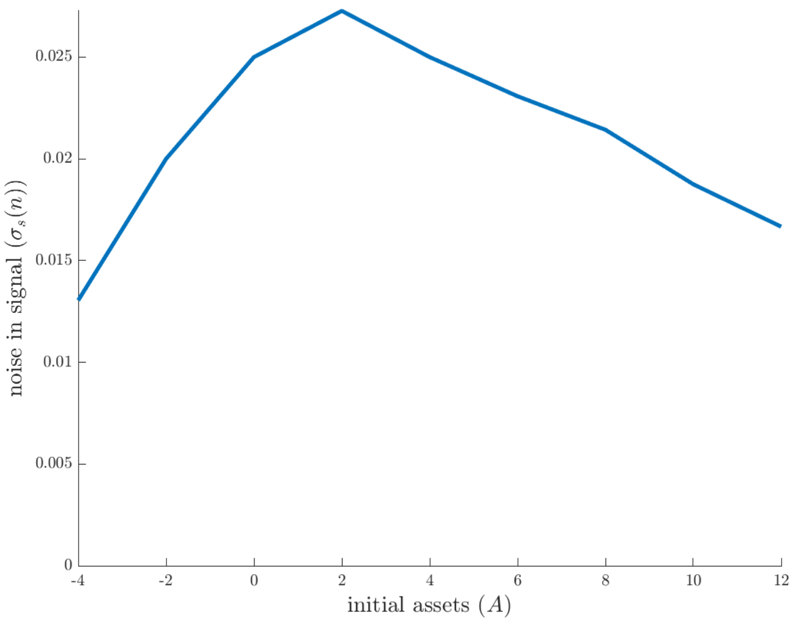


Figure D.4: Endogenous Effort – Chosen Standard Deviation of Noise

The figure plots the standard deviation of the noise implied by the endogenous choice for n, solving (D.8) for given initial asset level A.

We confirm the predictions of our exercise by moving on to an endogenous choice of effort according to (D.8) and (D.9), subject to the cost function and return to effort as outlined in (D.10). The calibration remains the same as before. Figure D.4 plots the standard deviation of the noise implied by effort choice $n(A)$ across a range of initial asset level A. The findings confirm our conjecture from Section D.3.2. With increasing absolute wealth level (positive or negative), households decide to exert more effort to reduce the noise in the signal, driven by the exposure effect. Additionally, households with negative wealth choose to exert more effort (reduce the noise further) than households with similar positive wealth. This is due to the asymmetric impact of absolute risk aversion which has equally been discussed above.