Prices and Heterogeneity in Macroeconomics

Inauguraldissertation

zur Erlangung des Grades eines Doktors
der Wirtschaftswissenschaften

durch

die Rechts- und Staatswissenschaftliche Fakultät der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Maximilian Weiß

aus Bern

2023

Dekan:	Prof. Dr. Jürgen von Hagen
Erstreferent:	Prof. Dr. Christian Bayer
Zweitreferent:	Prof. Dr. Thomas Hintermaier
Tag der mündlichen Prüfung:	16. Mai 2023

Acknowledgements

I could not have written this thesis without the help of my advisor Christian Bayer. First of all, two of the papers in this thesis draw heavily on his research on heterogeneous agent business cycle models. In the talks that we had during the course of my PhD, he directed me towards interesting research questions, succinctly pointed out the critical issues I had to address in my papers and gave me very concrete guidance on how to solve them that proved invaluable to my progress. He also advised me on strategic decisions regarding my career, for which I am very grateful. In his position as the head of the research training group The Macroeconomics of Inequality, he provided us students with the opportunities to learn from excellent visiting scholars and to exchange our ideas with fellow students as well as with junior and senior faculty of the institute in Bonn.

I am likewise indebted to my second advisor Thomas Hintermaier. Thomas suggested for me to become acquainted with HANK when I did not yet know who that was (I must confess). Early on, he sketched out possible paths for my research that helped me to orient myself. Whenever I needed feedback to my work, he offered me a careful analysis from which I learned a lot. I am grateful for the attention that he paid to my work and the situation I was in.

I am grateful to my cohort at the Bonn Graduate School of Economics. Their high standard for worthwhile directions of research was inspiring, and I owe one research idea and one co-author to them. Of course, I say special thanks to my co-author Thomas Kohler for being willing to work with me on the exciting project of "surveying price-stickiness". I profited immensely from Thomas' broad knowledge of economics, and I thank him for enduring my excursiveness during our meetings.

I am grateful to my fellow students and the faculty at the Institute of Macroeconomics and Econometrics in Bonn for the inspiring conversations and presentations in reading groups, internal seminars, during lunch or in the hallway. Specifically, I would like to thank Gregor Boehl, Pavel Brendler, Joachim Jungherr, Keith Kuester, Moritz Kuhn, and Donghai Zhang for their support and interest in me and my research over the last years.

Since the completion of this thesis is a milestone in my professional life, I also want to thank a few people that I think were particularly important for me along the way. I am grateful for having attended a high school that encourages its students to be creative and pursue their individual talents, the Christian-Dietrich-Grabbe-Gymnasium in Detmold. I am thankful to my teachers Walter Hunger, Udo Mönks, and Hanne Sentker for their care and interest in my development. At the start of my undergraduate studies at the University of Bayreuth, I competed with impressive fellow students for the positions of tutor, and lost out. I am grateful to Hartmut Egger and Walter Olbricht for picking me as one of their research and teaching assistants, respectively, and inspiring confidence in me that an academic career would be a possibility.

Finally, I thank my family. My parents Christine and Rainer always were an example of professional excellence and personal integrity to me. Their unconditional support of and genuine interest in my journey are invaluable. I am grateful to my grandparents, Irmgard and Oswald, for their generous help during my undergraduate studies. I thank Oli for offering us a nice place where I could write these last few lines. I thank Elif for being by my side, and showing me so much.

Maximilian Weiß
Stuttgart, May 2023

Contents

List of Figures

List of Tables

Introduction

The present thesis consists of three self-sufficient papers, organized in the form of three chapters. Their overarching theme is the analysis of the effect of microeconomic heterogeneities on the response of macroeconomic quantities, especially asset prices and inflation, to specific shocks. Conceptually, microeconomic heterogeneities only matter for macroeconomic questions when certain simplifying assumptions are violated. Economists make these assumptions — for example, that households can trade all their assets freely at the prevailing market price, or that all firms face the same costs when adjusting their prices — in part because they seem natural, and because they allow to build tractable models that nonetheless capture the main features of the (aggregated) data well. In the first two chapters of my thesis, my co-author and myself make the case that violating these particular assumptions is necessary to explain prevalent empirical phenomena. For that, we build structural models that take account of the additional heterogeneity when the simplifying assumptions are not present, and demonstrate their ability to reproduce the observed phenomena qualitatively and quantitatively. We focus on explaining asset price fluctuations and the price dispersion of homogeneous consumer products. Market prices are relatively easily observed, but notoriously difficult to explain. In Chapter 1, which is joint work with Thomas Kohler, we additionally conduct a survey to elicit not only firms' prices, but also their motivations behind adjusting or not adjusting their prices. While we do not directly link our findings to drivers of inflation, we discuss how our structural model may be extended to explain fluctuations in the aggregate price-level in response to supply shocks in uncertain times. In Chapter 2, I explain fluctuations in the average price for public stocks within a structural macroeconomic model that additionally matches survey evidence on heterogeneous households' portfolio choices over time. Chapter 3 of this thesis is a technical contribution: allowing for heterogeneity on the microeconomic level complicates the analysis of the behavior of the economic agents in the model. I propose an approximation technique to solve such models faster. In the following, I introduce each chapter in a little more detail.

In Chapter 1, we conduct a survey among German hairdressers about their reasons for adjusting their prices during the Covid-19 pandemic. The most important reason not to increase prices is the fear of losing regular customers, while the most important reasons to increase prices refer to higher costs. We find that firms who report that their customers are more understanding of their prices are more likely to increase prices. Constructing a distribution of relative prices of hairdressers, we also find that the real price rigidity that stems from low understanding is most prevalent in the center of the distribution. We rationalize these findings in the context of a search model, where customers are uncertain about the size of an industry-wide cost shock. Firms with more understanding regular customers are more able to blame higher prices on the cost shock, which keeps their customers from searching for other firms. With heterogeneity in both quality of service and production costs, firms that charge the median price are least likely to be monopolists.

In Chapter 2, I show that news shocks about higher future capital returns can explain stock price-booms and subsequent -busts in a two-asset, heterogeneous agent New Keynesian model. The portfolio choice between liquid assets (like stocks) and illiquid capital is key, as it allows for a time-varying illiquidity premium. Upon the news, capital-rich households accept to hold more illiquid capital at a lower premium, in anticipation of higher future returns on it. This increases their consumption risk, and causes stock prices to rise. After the boom, capital-rich households trade capital for liquid assets in order to self-insure against idiosyncratic income shocks, which increases the illiquidity premium, and causes stock prices to fall. Novel evidence from survey data on portfolio choices of capital-wealthy households during stock price boom-bust cycles supports the key mechanism of the model.

In Chapter 3, I discuss the problem of the efficient computation of the second order perturbation of a state-space system with heterogeneous, forward-looking agents. I find that the nonlinear effect of the distribution constitutes the main share of the computational burden. I propose a way to split up the nonlinearity in a direct and an indirect component. Abstracting from the indirect nonlinear effect speeds up the computation and yields a suitable approximation of the system dynamics in the short run.

Chapter 1

Surveying Price Stickiness with Large Shocks*

Joint with Thomas Kohler

1.1 Introduction

What is the reason for the dispersion of prices of homogeneous products? Such price dispersion is a widespread phenomenon (Kaplan and Menzio, 2015), tends to increase with inflation[1], and has negative welfare effects under commonly made assumptions on consumer preferences and production technologies.[2] In this paper, we investigate this question by directly asking the managers of firms in a specific industry — the hairdressing business — about their reasons for setting the price for a specific service — a man's haircut. During the Covid-19 pandemic, German hairdressers were forced to shut down for several months, and had higher production

* We thank Christian Bayer, Elif Bodur-Weiß, Ricardo Duque Gabriel, Thomas Hintermaier, Matthias Kräkel, Keith Kuester, Paul Schäfer, Monika Schmitter, Fabian Schmitz, the participants of the University of Bonn Macro Lunch Seminar, and the participants of the University of Bonn Micro Theory Lab Meeting. We are grateful to the German Federal Statistical Office (Destatis) and the Institute for Macroeconomics and Econometrics in Bonn for the access to the micro-level data of the German Consumer Price Index. Maximilian Weiß gratefully acknowledges financial support from the DFG (German Research Foundation) under the RTG 2281 - The Macroeconomics of Inequality and from the ERC-CoG project with agreement ID 724204. We are grateful to all participants of our survey.

1. Sheremirov (2020) shows that in the period between 2001 to 2011 in the U.S., price dispersion in regular (non-sales) prices is positively correlated with inflation.

2. If demand falls in the relative price of a product at a decreasing rate, or firms produce at decreasing returns to scale, or both, price dispersion entails welfare losses. Demand that falls in the relative price of homogeneous goods at a decreasing rate arises when consumers substitute the goods from different firms at a constant elasticity (Dixit and Stiglitz, 1977). See Damjanovic and Nolan (2010) for a quantitative assessment of welfare losses from price dispersion.

costs due to mandatory hygiene measures. We conduct our survey right after the hairdressers could reopen their business, in March 2021.

Our survey has three main qualitative results: first, hairdressers who do not increase their price do so mainly in order to retain their regular customers. Second, hairdressers who increase their price do so mainly due to increased costs — not due to higher demand, or because their competitors' prices increased. Third, the main explanatory variable for whether a firm increases its price is customer understanding, which is a measure of whether or not customers are understanding of the firm's prices, by the manager's account.

By eliciting the hairdressers' prices, and sampling the hairdressers of various counties all over Germany, we are also able to measure the distribution of prices within and across counties, and each hairdresser's relative position on its county's price distribution. We then compare the results from our survey about the dispersion of prices with micro-level data from the German Consumer Price Index (CPI). We find that the dispersion of the price for a man's haircut within a county is sizable — the standard deviation of the price from the county's median is about 24% — and increases during the course of the pandemic. With our survey data, we can uncover one reason for the high dispersion: we find that firms in the middle of the price distribution increase their price either by less or by more than the average firm — by less, if they report to have low understanding customers, and by more, if they report that their customers are understanding of their prices. We thus find heterogeneous cost pass-through: even conditional on adjusting their prices, firms in the middle of the price distribution split in the size of the price increase by the degree of the understanding that their customers have for their prices.

In order to find an interpretation for the customer understanding-measure, we conducted a follow-up interview with the head of a local hairdresser guild, who also participated in our survey. She told us that it is common for customers to confront their hairdressers in the weeks after a price increase, asking for reasons to justify it. In their explanations, the hairdressers focused on reasons that were evident (e.g. when the news reported increased energy costs) or directly attributable to an individual treatment (e.g. increased wage costs or increased cost of dye), as these were the most effective in convincing their customers. We conclude from this that the customers' understanding of prices should be interpreted as a problem of asymmetric information: upon a price increase, the customers try to figure out if it is justified by an increase in the cost of producing the haircut, and whether that cost increase affects the whole industry or is specific to their hairdresser.

In the second part of the paper, we build a customer search model with firm heterogeneity that rationalizes these findings. In the model, firms only stay in business if they can retain their regular customer — the customer who starts his search at that firm. Searching for (and switching to) another firm entails a cost to the customer. Therefore, his expectation of his outside option — the expected value of consuming at another firm — is an important determinant of whether the firm stays in business. We impose two information asymmetries: only a share of customers — the “high understanding” customers — observe the component of marginal costs that is idiosyncratic to the firm. Additionally, in an uncertainty period, the component of marginal costs that is common to all firms is unobservable for all customers. In such a period, customers try to learn from the firm’s price about the common cost component in order to form an expectation of the value of consuming at another firm. High understanding customers are more successful in doing so, as they can disentangle the idiosyncratic and the common cost component. When the common cost increases, firms with low understanding regular customers are therefore more constrained in their price setting, as a higher price is more likely to induce their customers to search for another firm.

In order to fit the model quantitatively to our empirical findings, we postulate that firms are heterogeneous not only in production costs, but also in the quality of their product. Thereby, we take into account that some of the dispersion in the price for a man’s haircut that we observe may actually be due to differences in the quality of the service. With a perfectly competitive market, our model yields two prices for the two products that are being traded: low quality haircuts and high quality haircuts, each produced at the lowest possible cost. Relative to that benchmark, the introduction of a search cost that impedes competition actually *lowers* the standard deviation of the price distribution. The reason is that search and switching frictions allow lower productivity firms to enter the market. As monopolists, they charge a higher price than more cost-efficient firms, but a lower price than higher-quality firms. Hence, their existence “bolsters” the middle of the price distribution, and lowers its dispersion.

While differences in product quality may seem like an ad hoc-explanation for price dispersion, it is crucial to make sense of our finding that firms in the middle of the price distribution are most restricted in their price-setting by having low understanding customers. Since firms in that segment of the price distribution are less productive, they also offer a lower surplus to their customers than their competitors. Therefore, for a suitable level of search and switching costs, they can be restricted in their pricing, while more productive firms have monopoly power. This heterogeneity in restrictedness allows the model to fully account for the differences in the size of relative price increases that we see in our survey. Additionally, we demonstrate that by increasing the share of firms with low understanding customers, the model can also explain a sizable share of the standard deviation in relative price changes by firm over time that we observe for men’s haircut prices in the German CPI microdata. While we build a static model for the purpose of tractability and clarity, we think that adding dynamic strategies would further improve the fit to the pricing data, and allow for a closer comparison to leading models of price dispersion in the literature.

Related literature. Our paper contributes to several distinct strands of the literature that investigates firm pricing and price stickiness. In the empirical part, we build on the idea of Blinder et al. (1998) to ask firms about their pricing decisions. The “classical” approach is the following: firm managers are presented a list of hypotheses. Each expresses an academic theory of price stickiness in layperson’s terms. Then, the managers grade how well these hypotheses describe what they are thinking while not adjusting their firms’ prices. This approach has spurred a sizable literature that tries to replicate the findings for different countries and types of industries. In appendix 1.A, we provide the hypothesis rankings of Blinder et al. (1998) and 26 replication studies.

The hypotheses that score best overall in that literature pertain to the categories (in that order) of *implicit contracts*, *cost-based pricing*, and *coordination failure*. Implicit contracts can arise between firms and their customers if customers have an incentive to become a “regular” at a firm — due to costs that would arise from switching to another firm. Thus, the firm and the customer enter into a dynamic game: for the customer to be willing to commit to becoming a regular customer, the firm has to implicitly promise not to increase its price unduly. The customer can play a trigger strategy: if she observes a certain pricing behavior that seems excessive, she changes the firm. The literature that presented those theories includes Okun (1981), Rotemberg (2005, 2011), and Eyster, Madarász, and Michaillat (2021) (under the headline of “fair pricing”) and Nakamura and Steinsson (2011) (who analyze the implications of “internal habits”).

Cost-based pricing relates to the simple idea that firms would not increase their price if their costs did not increase. The theory of coordination failure refers to the seminal paper by Ball and Romer (1991). They show that, when adjusting one’s price is subject to a menu cost, and one firm’s price is a strategic complement to another firm’s price, then for “medium”-sized money supply shocks, indeterminancy arises: both no firm adjusting their price, and all firms adjusting their prices, are equilibria.

We deviate in two ways from the “classical” approach of asking about pricing. One prevailing criticism of the method is that the theories that are voted to the top of the rankings usually receive similar and only intermediate scores (Blanchard, 1994). Hence, while the literature has produced consistent evidence for several theories of price stickiness to be most important, the survey approach failed to conclusively select a single best theory. We think that it is unclear, however, whether such a dominant cause of price stickiness exists across different markets and types of firms. By focusing only on a single industry with relatively homogeneous firms — German hairdressers that are (in the majority) members of local guilds —, we can expect to have a better chance of identifying a single most important reason for price stickiness in this market. Second, we do not ask about hypothetical situtations, but instead query the respondents’ actual responses to recent shocks that affected all firms in a uniform way. By that, we eliminate a source of noise in the survey responses of the “classical” approach: respondents may differ in their interpretation of hypothetical situations, and the mental cost of thinking in hypotheticals may introduce a recency bias to their answers.

We find that, indeed, our approach conclusively selects the most important reasons for why firms in our survey do or do not increase prices. The possible downside of our approach is that we have to make the case for external validity: first, the firms in our survey may not even be representative of the hairdresser market, let alone of firms in other industries. Second, the firms' responses to shocks during the Covid-19 pandemic may be special to that particular episode, and may not be indicative of the firms' responses to other shocks of the same type (cost shocks), or even shocks of different types (e.g. a monetary policy shock). We address the first concern by complementing our survey data with micro-level data from the German CPI. We find that there is a clear selection bias in our survey. However, we also find that quantitative results about the divergence of relative price increases over the price distribution are consistent across the two data sources. In general, the use of the CPI micro-level data helps us to put our findings in context.

In response to the second concern about the generalizability of our results to the whole economy, we find it reassuring that two of the three theories that rank highest in the literature — customer markets/implicit contracts, and cost-based pricing — rank highest also in our survey. From the CPI micro-level data, we additionally find evidence that coordination failure may have been an important determinant of price stickiness during the first part of the pandemic, which we do not cover with our survey. In an effort to explain our results in a generalizable way, we then build a model that is not specific to the hairdresser market, and qualitatively and quantitatively matches the main findings from our survey. Our model suggests that our findings are particularly relevant in times of high uncertainty about the consequences of industry- or economy-wide shocks.

The model builds on the literature on customer (or consumer) search models. There is an established literature that explains nominal price stickiness in response to a monetary shock in search models (Bénabou, 1988, 1992; Diamond, 1993). Our paper is more closely related to papers that analyze pricing in the presence of search costs and uncertainty about aggregate cost shocks (Bénabou and Gertner, 1993; Fishman, 1996). One can view our model as an extension of the model proposed by Fishman (1996). The main difference is that we postulate a non-Bayesian, "conservative" learning rule that customers follow in order to learn from observed prices about the industry-wide cost level. While we do not microfound the rule, we show that it is "conservative" — namely, customers never overestimate the industry-wide cost level — under one critical assumption. We find it reasonable that customers follow such a rule under very rare supply-side conditions, like the ones happening during the course of a pandemic, for which they do not have prior beliefs (which makes it a case of "Knightian" uncertainty). Still, the rule allows for ample learning of the industry-wide conditions from observing the firms' prices. Due to the customers' learning, our model predicts that firms with high-understanding customers can optimally increase their prices, which leads to the heterogeneity in cost pass-throughs we observe in our survey data. By introducing this heterogeneity into a search model, we fill a gap in the literature: following the survey evidence by Blinder et al. (1998), the "fair pricing" literature assumes the existence of altruistic firms (Rotemberg, 2011) or behavioral biases of customers (Eyster, Madarász, and Michaillat, 2021) in order to explain the importance of customer markets for price stickiness. Instead, we provide a model with profit-maximizing firms and rational customers that can explain the same facts in the presence of information asymmetries.

We also relate to recent work on consumer search models with aggregate uncertainty by Janssen, Parakhonyak, and Parakhonyak (2017) and Janssen and Shelegia (2019). This strand of the search literature usually assumes that customers do not attempt to learn about other firms' prices from the first price they observe — i.e., they hold "passive beliefs" about the other firms' prices (see Anderson and Renault (2018) in Corchón and Marini (2018)). Janssen and Shelegia (2019) show that this assumption is not robust to the possibility that customers "blame" a part of an observed price increase on industry-wide conditions. Our model provides an explanation for why customers would rationally believe that a part of an observed price increase is due to higher industry-wide costs.

A large literature investigates the success of macroeconomic models with time- or state-dependent firm pricing in explaining empirically observed price responses to cost shocks. Karadi and Reiff (2019) document a flexible, but asymmetric price response to changes in value-added tax. They show that a model with menu costs can account for these observations, but failes to generate a strong output response to monetary policy shocks. In section 1.5, we discuss that our model can explain such an asymmetry without the recourse to menu costs; however, we find it more applicable to cases where the cost shock is less observable to customers, as in the case of *industry-specific* value-added tax changes (Benzarti et al., 2020). Hobijn, Nechio, and Shapiro (2021) use the Brexit as a quasi-experiment that induces a large common marginal cost shock to firms in the U.K. They find that firms with lower relative prices are more likely to increase their prices in response to the common cost shock, and increase their relative prices by more. They show that neither a model with time-dependent pricing nor a model with menu costs can account for this micro-level observation. Our findings align with those of Hobijn, Nechio, and Shapiro (2021): hairdressers on the lowest end of the relative price distribution increase their relative price by more during the time of the pandemic. Our search model can explain this finding on the intensive margin: since the cheapest firms are relatively productive, they have monopoly power. Consequently, they pass through the cost increase at an optimal rate. The same holds true for the most expensive firms. However, since the latter produce their higher-quality service at higher costs, the share of the cost increase in their overall production costs is lower. Therefore, the relative price increase declines in the initial relative price position of the firm.

Finally, our paper relates to the literature that collects general observations from micro-data about firm pricing over time, and tries to align the findings with theoretical models on pricing frictions. Nakamura and Steinsson (2008) provide facts about the frequency and size of consumer and producer price increases and decreases in the U.S., and relate them to overall inflation dynamics. Gautier et al. (2022) replicate the analysis for the euro area. Kaplan and Menzio (2015) show that price dispersion of identical goods is a prevailing feature of the data, and find evidence for search frictions as a driving force. Klenow and Willis (2016) establish facts about the behavior of nominal and *relative* price changes (relative to a sectoral price index), conditional on a price change, over time. They document a sizable standard deviation of conditional relative price changes of 14%. They conclude that this evidence implies the need for a theory of *real rigidites*, i.e. frictions that inhibit firms to change their price on the *intensive* margin, not the extensive margin. We show that low customer understanding can be such a real rigidity, and demonstrate in the model that it can explain a large share of the standard deviation in conditional relative price changes in the hairdressing market during the pandemic. The reason is that a large share of low understanding customers inhibits enough firms in their price-setting so that the price level of the median firm is low. Therefore, the price-increases of firms with monopoly power are higher relative to the median price level. This argument is similar to the one in Mongey (2021), who proposes market power in oligopoly markets as the explanation for large deviations in relative prices.

Our paper is organized as follows. In section 1.2, we outline the design of our survey and present the main quantitative findings in comparison to those from the German CPI micro-level data. In section 1.3, we present the first main qualitative result of our survey, which is the ranking of the hypotheses for why firms increase or not increase their prices. In section 1.4, we present the second main qualitative result of our survey that points towards the important role of information asymmetries, namely, how the customer understanding variable interacts with price stickiness and other firm characteristics. In section 1.5, we build a customer search model with firm heterogeneity and bring it to the data. In section 1.6, we conclude.

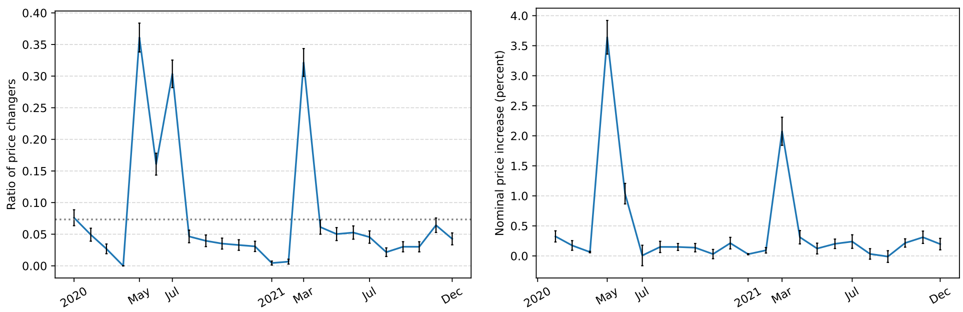
1.2 Market Description and Summary Statistics

During the course of the Covid-19 pandemic, the German government imposed two lockdowns during which any hairdressing service was forbidden in Germany. The first lockdown went from March 22 to May 4, 2020. Afterwards, hygiene rules were introduced, such as distancing rules, mandatory masking, and mandatory hair washing before any hairdressing service. The second lockdown, after which we conducted our survey, went from December 13, 2020, to February 28, 2021.

We argue that the Covid-19 pandemic and the associated lockdowns have hit all hairdressers with similar shocks: First, the firms have lost months worth of profits, paid bills from their reserves, and some had to borrow money to keep their businesses. Second, the hygiene rules that were imposed after the first lockdown became slightly stricter after the second lockdown: It got prohibited to serve walk-in

(a) Men's haircuts: price change frequency

(b) Men's haircuts: price changes



Notes: Ratio of hairdressers that changed their price for a man's haircut, monthly, from January 2020 to December 2021 (left panel). Average monthly percent change of the price for a man's haircut, from February 2020 to December 2021 (right panel). The dotted horizontal line shows the average monthly price change frequency for the median firm (left panel). Whiskers depict 68%-confidence intervals of the statistics. Source: German CPI micro-level data, N=442-461.

Figure 1.1. Men's haircut prices

customers (they needed to book appointments beforehand), the hairdresser had to wear a medical face mask and to replace it after each customer, there had to be a continuous stream of fresh air in the salon, although it was winter, and in some regions with many infections, customers had to be tested negatively. In some states, the hairdresser was allowed to conduct the test. Many of the hygiene rules stayed in place until the end of 2021. Third, the federal value-added tax changed between the two lockdowns: For the second half of 2020, the general VAT rate was reduced from 19% to 16%. Legally, price tags in Germany have to display the price including the VAT. So, firms that have passed on (some of) the tax reduction had changed their pricing lists. In our survey, few firms (14% ± 2%) reported that they passed on the VAT reduction. In the micro-data to the German CPI, we find that 23% ± 2% of firms decreased their prices from June 2020 to July 2020. Fourth, many employees in hairdressing received pay rises: January is a common month for discretionary wage increases, and the legal hourly minimum wage for hairdressing increased in several German states on January 1, 2021, by 15 Eurocent, and on July 1, 2021, by 10 Eurocent[3] Fifth, demand for haircuts was likely increased directly after the second lockdown. Several hairdressers auctioned off their first appointments for three-digit

3. In some states, there are binding collective agreements determining the minimum wage in hairdressing. This is the case in Hesse, where the collectively agreed minimum wage increased on January 1, 2021. In some states, there are collective agreements, but employers decide themselves whether to opt in. And in the other states, mostly in Eastern Germany, there are no collective agreements on the minimum wage in hairdressing. The federal minimum wage, which increased on January 1, 2021, and on July 1, 2021, applies. In fact, in these states, many employed hairdressers make the federal minimum wage.

prices (and donated them to charity).[4] The average waiting time for appointments in the online booking tool Treatwell was more than two weeks.[5] Notably, the demand later decreased to a constant level below the demand before the lockdown.[6]

We conducted our online survey on the platform *SoSci Survey* from Monday, March 08, 2021, to Friday, April 16, 2021. The questionnaire we used is in Appendix 1.B. It was necessary to possess the URL to participate. We recruited participants in two ways. First, on March 08, 2021, we contacted all local Chambers of Handicrafts (*Handwerkskammern*) because membership is mandatory for German hairdressing firms. However, the response rate was very low. Thus, second, we contacted the heads (*Obermeister*) of all local hairdressing guilds (*Friseur-Innungen*)[7], on March 15, 2021, and asked them to participate and to forward our e-mail to the other members. On April 1, we sent a reminder to the heads of the local hairdressing guilds. After deleting answers with mostly missing or contradictory answers, 281 responses remained. For comparison, in 2020, 77.166 hairdressing firms were registered in Germany (Zentralverband des Deutschen Friseurhandwerks, 2021, p. 12).

1.2.1 Nominal prices

Our measure for a hairdresser's prices is a standard man's haircut ("short back and sides, wash, cut, blow dry, 25 minutes"). We asked whether the price contained a "hygiene surcharge" and if so, what amount it is. The answers are summarized in Table 1.1. For a comparison of our results with the overall market for male haircuts in Germany, we turn to micro-data from the German CPI for the years of 2020 and 2021.[8] In figure 1.2, we show the evolution of the price for a man's haircut over time across different quantiles (left panel), as well as the evolution of the standard

4. See `https://www.rnd.de/panorama/friseurin-versteigert-ersten-termin-nach-dem-lockdown-und-spendet-erlos-TJ2JLGGO5E5DA6VHRSZ4YDCS6A.html`, last accessed June 12, 2022.

5. See `https://www.tophair.de/branche/branche-detailseite/buchungsrekord-zum-re-start`, last accessed June 12, 2022.

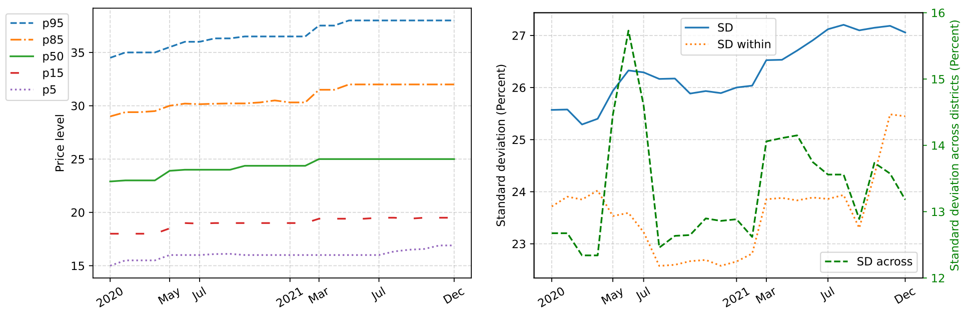
6. See `https://www.br.de/nachrichten/bayern/bayerische-friseure-leiden-weiter-unter-der-pandemie`, last accessed June 12, 2022. The hairdressers in the article conjecture that the hygiene and testing rules are the reason. Other reasons might be the fear of getting infected at the hairdresser's, the diminished importance of having a fresh or professional haircut, and the customers' smaller budgets during the recessionary period.

7. Guilds are lobby groups with voluntary membership. The local hairdressing guilds are organized on a county-level or slightly larger, and there are 247 of them in Germany in total.

8. The dataset allows us to differentiate by male and female haircuts. We only consider male haircuts. Furthermore, we only consider entries that carry the attributes "haircut" (as opposed to shaving) and "wet cut" (washing the hair before cutting became mandatory after the first lockdown, so we delete dry cut-services for all months for consistency). We use the price series "PreisBearbeitet", since it is smoother and has less missing values. We check that the results in this paper are robust to using the series "PreisErhoben" instead. Results are available upon request. We delete sales prices and observations where the quality changed over time. For each month, between 445 and 465 observations remain.

(a) Men's haircuts: price level

(b) Men's haircuts: relative price standard deviation



Notes: Price for a man's haircut at different percentiles (left panel, N=445-465). Standard deviation of price relative to the median for a man's haircut, over all hairdressers (N=445-465), on average within counties with at least 6 firms (N=121; both left axis), standard deviation across median price of counties with at least 6 firms (N=11-14; right axis) (right panel). Source: German CPI micro-level data.

Figure 1.2. Men's haircut price distribution

devation, relative to the median, overall, as well as within and across counties (right panel). We find that hairdressers in our survey charge on average about 1 Euro more for a man's haircut than the median hairdresser in Germany. Also, the price dispersion in our survey, as measured by the standard deviation, is about 2pp lower than that for all German hairdressers.

Next, we analyze the frequency of price changes in the CPI micro-level data and compare it to our survey results. The microdata allows us to measure the number of days since the last price change for each observed firm and month. We find that the price duration of the median hairdresser in March 2020, on the verge of the Covid-19 pandemic, was 409 days, or 13 and a half months. This is in line with average price durations that are commonly observed for the service sector (Gautier et al., 2022). When interpreting the duration as the average of a geometrically distributed variable, the observed duration translates to a chance of a monthly price change of 7.5%. However, as was shown by Nakamura and Steinsson (2008), the service sector exhibits a strong time interval-dependency, where many prices change after around a year. Therefore, even in normal times, we would expect to see spikes in price changing frequencies at certain points of the year, instead of uniform frequencies.

In the left panel of figure 1.1, we plot the frequency of changes in the price of a man's haircut during the height of the pandemic. Between 30% and 35% of the hairdressers changed their prices in the months directly after the two lockdowns — in May 2020 and in March 2021 — and in July 2020, where the temporary VAT decrease became effective. Since almost all price changes after the lockdowns are price increases, the plot also shows that in March 2021, a little over 30% of hairdressers increased their prices in Germany. In our survey, instead, two thirds

Table 1.1. Summary of the price related variables

Variable	*n*	Mean	SD	SD (rel.)	Min	Max
Price before the lockdown (December 2020)	281	25.93	6.22	24%	14	58
Price after the lockdown (March 2021)	281	27.35	6.48	23.7%	14	59
Hygiene surcharge	96	2.38	0.98	41%	0.5	5

Note: From our own survey. The prices and surcharges are in Euros.

of the sampled firms (64%) have increased their prices. What is more, the average increase of the price for a man's haircut in March 2021 is 2% in Germany, while it is 5.5% for firms in our survey. The differences are also there on the intensive margin: the average hairdresser who increased its price in March 2021 increased it by 7.1%, while for firms in our survey, the average conditional price increase is 12.6%. Hence, we find clear evidence that the selection of firms in our survey is biased. One possible factor is that firms who did not plan to increase their prices after the second lockdown were less interested in participating in our survey. The differences in conditional prices could imply that participants in our survey passed through less of the cost increase due to the hygiene measures up to March 2021 than the average firm. They could also be explained by the fact that the firms in our survey are predominantly members of local guilds. Guild-membership comes with duties — members commit to take on apprentices and foster the craft of hairdressing — so that guild members may be more impacted by the increase in the minimum wage on January 2021 than the average hairdresser. In fact, we find that the firms in our survey employ more hairdressers than the average firm (see below).

To understand better the incidence of price changes between different dates, we make use of the panel dimension of the microdata. We find that the median firm changes its price once between April 2020 and December 2021, while the average price change in that time frame is 1 and a half times. The duration of the price the hairdresser charged in March 2020 is significantly negatively related to the times it changes its price from then until December 2021: Being on a higher decile on the distribution of price durations reduces the rate of price changes by 6.6pp. Hence, we can interpret the number of days the hairdresser kept the same price as of March 2020 as a measure of her idiosyncratic price stickiness. We then regress the likelihood of changing the price in March 2021 on this measure of idiosyncratic price stickiness, on an indicator variable that equals 1 if the firm reduced its price in July 2020, and on an indicator variable that equals 1 if the firm changed its price between May 2020 and December 2020. We find that the measure of idiosyncratic price stickiness significantly lowers the probability of a price change in March 2021, and that the average firm that passed through the VAT decrease in the summer of 2020 is 30pp more likely to change its price after the second lockdown. On the other hand, controlling for the other variables, the fact that the firm changed its price after the first lockdown does not have a significant effect on whether it changes its price after the second lockdown. This time-independence, together with the result that the median firm changed its price only once from March 2020 to December 2021, allows us to interpret our survey's quantitative results as the average firms' reaction to the cost shock during the whole pandemic. Therefore, we can justifiably match our model in section 1.5 to the survey data alone, even though the survey only includes price data from December 2020 to March 2021.

1.2.2 Relative prices

While a man's haircut is a fairly homogeneous product, we cannot interpret the Germany-wide standard deviation of its prices as a measure of price dispersion. The reason is that haircuts are a local good, and that the price levels across different regions in Germany might differ, for example due to differences in customers' incomes. We find that this is indeed the case in the CPI data: The right panel of figure 1.2 shows that the standard deviation across the medians of counties with at least 6 firms is at around 13% of the average median. To investigate whether the differences across counties are persistent, we regress the log of a county's median price on time-invariant county-fixed effects and the log of the average median over all counties. This simple model, which assumes that the price level of each county is a time-constant share of the average price level, explains 93% of the variation in the data (see appendix 1.C.1).

Therefore, in order to measure price dispersion of homogeneous products, we compute the prices of a man's haircut *relative* to the price level in the hairdresser's county. Throughout our analysis, we define a county's price level as its median price. We choose the median instead of the mean in order to be more robust against outliers. The panel data of the German CPI includes 11 to 14 counties where at least 6 hairdressers are sampled, depending on the month. The average number of hairdressers that are sampled within a county with at least 6 firms is 9.3, with a maximum of 25 firms. In our survey data, 21 counties include at least 6 hairdressers each. The average number of hairdressers within a county amounts to 9.2 in our survey, with a maximum of 28 firms.

The standard deviation of the price for a man's haircut *within* a county is the standard deviation of the relative prices in that county. The dotted line in the plot in the right panel in figure 1.2 shows the average of this measure across counties and over time. On average, the standard deviation within counties is at 23.6% of the county's median. This shows that the price dispersion of a fairly homogeneous product — a man's haircut in a county — is large, but still well in the range of the dispersions reported by Kaplan and Menzio (2015). The average standard deviation within counties for the December 2020 prices of our survey participants is significantly lower, at 17.7%.

Comparing the standard deviations across and within counties over time yields an interesting pattern: directly after the first lockdown, the increase in the Germany-wide standard deviation of prices for men's haircuts is driven exclusively by a sharpe increase in price devations *across* medians, while the price dispersion within counties *declines*. Conversely, after the second lockdown, and at the time of our survey, the standard deviation within countries increases by about the same magnitude as the standard devation across counties. This result seems to suggest that after the first lockdown, coordination failure played a significant role in explaining firms' pricing: while in some counties, hairdressers could coordinate to collectively increase their prices, in other counties they failed to do so. The novelty of the pandemic, as well as many distractions that could stem from new government regulations or the concern for one's own health, could have contributed to the failure to coordinate after the first lockdown. The time after the second lockdown, instead, seems to be more suitable for the study of the phenomenon of price dispersion of homogeneous goods.

Having computed the relative prices of firms, we can pool them in one common relative price distribution, that controls for the differing price levels across counties. Thereby, firms become comparable across counties by their relative price position within their respective county. We find that the relative price position of a firm at the beginning of 2020 is a significant predictor for the size of their relative price increase during the pandemic: the more expensive the firm, the lower is its relative price increase. In section 1.5, we show the concrete estimates for various measures (conditional and unconditional on price changes), and explain this finding in the context of our model. For the rest of the paper, we will analyze the relative price distribution by splitting it into tertiles. Later, we will match firms of each tertile to distinct cost- and quality-levels in the model. We find both in our survey data and in the model that firms in the second tertile are most restricted in their price setting. Evidence from the micro-level data of the CPI supports this: after winsorizing the number of days that a firm kept the same price it has in March 2020 at the 10th and 90th percentile, we find that a firm in the second relative price tertile kept its price unchanged on average 173 days longer than firms in either extremes of the distribution. The difference is statistically significant (one-sided t-test, p=4.3%, N=83). At the same time, we find that firms in the second tertile of the distribution are significantly more likely to pass through the temporary VAT reduction in July 2020 than firms in the other tertiles, where we measure the pass-through as a nominal price decrease from June 2020 to July 2020 (Chi-square-test, p=1.7%, N=115). From that, we conclude that the economy-wide VAT decrease likely was salient for many customers, so that many firms — especially those that offer a lower surplus than their competitors — were forced to reduce their prices (we further discuss the connection between asymmetric VAT pass-throughs and information asymmetries in section 1.5.1.3).

Table 1.2. Distribution of firm size in the survey

Number of employees	*n*	Frequency	continuous variable size	indicator variable size
no employees	22	8%	0	0
1 to 3 employees	105	37%	2	0
3 to 6 employees	88	31%	4.5	0
more than 6 employees	66	23%	0	1
Total	281	100%		

Notes: When using the size of a firm in regressions, we compromise between continuous variables and indicator variables. We treat the first three options as a continuous variable with the middle value of the bucket, so 0, or 2, or 4.5. The answer "more than 6 employees" is captured by an indicator variable.

1.2.3 Other firm characteristics

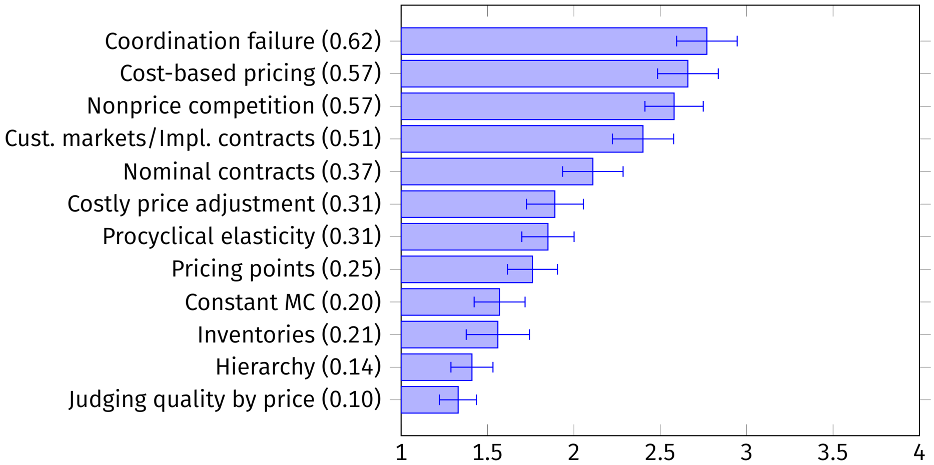
Turning to other characteristics of the firms in our survey, we find that the respondents to our survey are somewhat larger than the average hairdresser. Table 1.2 summarizes the respondents' answers to our question measuring their firms' sizes by employees. Among hairdressers in Germany, about a third most likely has no employees.[9] Of the other firms (that report their revenue to the authorities), around half of the firms has less than 5 employees, around a fifth has between 5 and 9 employees, and around a twentieth has between 10 and 19 employees (Zentralverband des Deutschen Friseurhandwerks, 2021, p. 12).

Regular customers make up the main share of demand for much of the firms in our survey. For the majority of hairdressers in our survey (60%), the share of customers who are regular is at least 80%. For only 5% of the respondents, the share of regular customers is lower than 60%.

9. For comparison with the distribution in the general market, we use data from 2018. 53.484 firms (71% of the registered firms) reported their revenues to the authorities (Zentralverband des Deutschen Friseurhandwerks, 2021, p. 12). A firm whose revenue is below the cutoff for exemption to report cannot have a single employee (paid at minimum wage) without making a loss. A firm is exempt from paying VAT if it has had a revenue of less than €17, 500 in the previous year and expects a revenue of less than €50, 000 in the current year. The federal minimum wage in 2018 was €8.84 per hour. This wage times 40 hours per week times 4.34 weeks per month times 12 months yields more than €18, 000, additional to which the employer has to pay social security contributions. It is possible that a firm is exempt from reporting its revenue because it is newly founded and expects a revenue of less than €50, 000, so it might have employees without making a loss. There are, however, few entrants to the market. The federal guild reports that 5, 867 salons—not firms—were newly registered in 2020, so this number includes existing firms moving or opening branches (p. 11).

1.3 Ranking of Hypotheses

The goal of the manager survey by Blinder et al. (1998) is to evaluate theories for price stickiness. For comparison to our own results, we report the results of their study in figure 1.3. In line with the literature that followed up on this study, the hypotheses that pertain to the categories of *coordination failure*, *cost-based pricing* and *customer markets/implicit contracts* receive top scores. In appendix 1.A, we list the results of all replication studies we are aware of and provide short descriptions of the theories underlying the hypotheses. Because many theories received similar, intermediate scores and acceptance rates,[10] Blanchard (1994) doubts the fruitfulness of the survey approach, as it seems to fail in revealing the single most important theory of price stickiness.

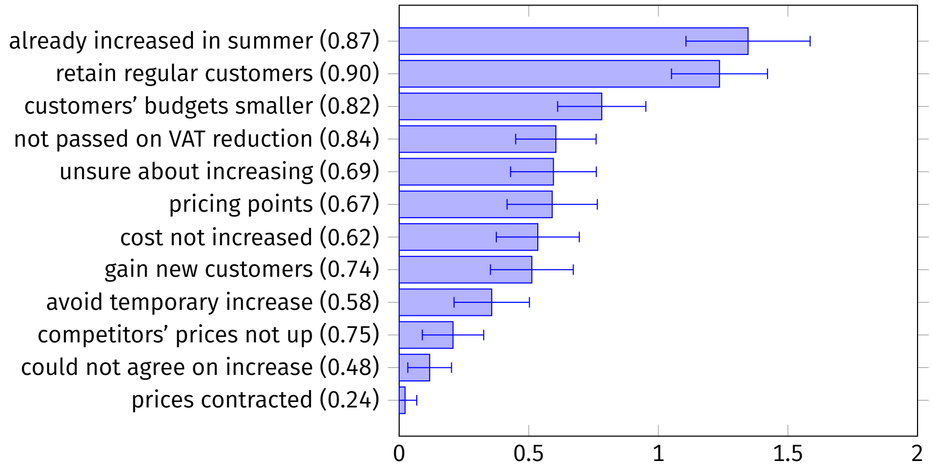


Notes: The grades go from 1 to 4. In the brackets behind the name of the hypothesis is the acceptance rate, which is defined as the share of managers that grade the respective hypothesis 3 or 4. If a firm rejected the premise of the hypothesis (e.g. the manager stated that the firm does not have inventories), Blinder et al. directly assigned the grade 1. On the *x*-axis is the average score of the hypothesis. The error bands are the 95% confidence intervals.

Figure 1.3. The original survey results by Blinder et al. (1998)

We design our survey to avoid two reasons for similar scores: The first is averaging within and across markets. As different theories are differently important in different markets, averaging leads to more intermediate scores. The same might be true for averaging within a market over small shops and multinational firms. To make the ranking clearer, we survey a single market which consists of rather homogeneous firms; made even more similar due to our sampling bias favoring guild members. The second is the noise from asking about pricing decisions in hypothetical

10. The data are from Table 5.2 in Blinder et al. (1998, p. 110). It is the finalized version of the table to which Blanchard's critique referred (Table 4.4 in Blinder (1994, p. 124)).

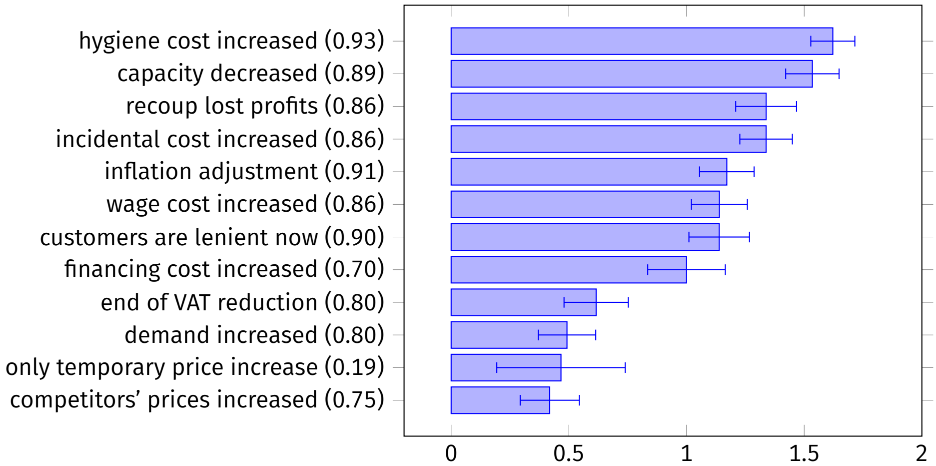


Notes: The grading scale has four items: "does not apply," "has played no role" (0), "has played some role" (1), and "has played an important role" (2). Following Blinder et al. (1998), if the respondent marked a hypothesis as not applicable and assigned no grade, we assign the lowest grade, 0. We define the acceptance rate in the bracket as the share of respondents that did not say the hypothesis does not apply, but assigned a score. On the *x*-axis is the average score with error bars showing the 95% confidence interval. N=102.

Figure 1.4. Hypthesis ranking of our survey for reasons not to increase prices

situations. The availability bias makes the managers overweight things they recently experienced when imagining to be in the hypothetical situation. The projection bias makes them misjudge what they would do if the hypothetical situation became real. To reduce this noise, we ask the hairdressers about their actual behavior in the wake of recent shocks.

The results of our survey for why firms do not increase their price show similar scores in the middle of the ranking, but two clearly winning hypotheses at the top (see figure 1.4). The presented hypotheses are based on those in Blinder et al. (1998). We asked about versions of coordination failure (competitors' prices not up), cost-based pricing (cost not up, not passed on VAT reduction), customer markets/implicit contracts (retain regular customers), nominal contracts (prices contracted), costly price adjustments (unsure about increasing, avoid temporary increase), procyclical elasticity (customers' budgets smaller), pricing points, and hierarchy (could not agree on increase). We excluded menu costs (in their literal meaning), non-price competition, inventories, constant marginal cost, and judging quality by price for different reasons. Mostly, we expected those theories to not apply or to be rated as unimportant. Because participants asked for it, we added the option "I did not increase the price because I have increased the price already after the first lockdown" a few days after the start of the survey. Therefore, only about half of our survey participants were presented the version of the survey where this was among the options. However, we also provided an open text box for the participants to re-



Notes: The grading scale has four items: "does not apply," "has played no role" (0), "has played some role" (1), and "has played an important role" (2). Following Blinder et al. (1998), if the respondent marked a hypothesis as not applicable and assigned no grade, we assign the lowest grade, 0. We define the acceptance rate in the bracket as the share of respondents that did not say the hypothesis does not apply, but assigned a score. On the *x*-axis is the average score with error bars showing the 95% confidence interval. N=179.

Figure 1.5. Hypothesis ranking of our survey for reasons to inrease prices

port other reasons for not increasing their prices. A few participants reported that they already increased prices through that channel.

We find that retaining regular customers is the winning hypothesis for not increasing one's prices among *state-dependent* reasons. Additionally, we find clear evidence for a form of time-dependent pricing: the majority of firms who did not increase their price in March 2021 report that they already increased it directly after the first lockdown. This is at odds with our finding in section 1.2 that whether hairdressers increased their price in the summer of 2020 does not significantly predict a price change in March 2021.

The results of our survey for why firms increase their price is shown in figure 1.5. Compared to the reasons not to increase prices, hairdressers agree with more of the presented hypotheses for why they increase their prices, which is reflected in higher average scores. Still, the reason that hairdressers incur higher costs of producing haircuts, reflected in direct hygiene costs and the indirect cost of a lower capacity in their salons due to distancing measures, scores significantly higher than the other hypotheses. Both hypotheses (as well as several others) are consistent with a cost-based pricing theory. Again, the theory of coordination failure (competitor's prices increased) is rejected by the firms in our survey. Interestingly, firms also report that they did not increase their prices due to higher demand. This finding might suggests that the hairdressers' production capacities are not far from constant returns to scale in the short run. Alternatively, firms with a large share of regular customers might be punished for short-term price increases (see section 1.5.1.3 for reasons why that might be the case) and therefore rather prefer to increase their capacities in the short run at higher costs.

We conclude that, for the German hairdressing market during the Covid-19 pandemic, our adaptation of the “classical” survey approach identifies clear winning theories for the reasons why firms increase or not increase their prices. We find that two theories that score highly in the literature — customer markets and cost-based pricing — are also the main explanation for price setting in our survey. In contrast, firms in our survey due not report a high role for coordination failure, which is another theory that usually scores well in surveys that ask firms about their pricing. Given our results in section 1.2, we conjecture that if we had conducted the survey directly after the first lockdown, when the standard deviation across counties spiked, the theory of coordination failure would have scored higher.

1.4 Customer Understanding: Empirical Results

In this section, we present stylized facts from our survey’s data about the role of customer understanding for firm pricing and its correlation with other firm characteristics. As a first step, we construct four variables as composites of the firm-level characteristics that we ask in our survey. The composites are sums of Likert-item scale answers to related questions and measure a common factor among these answers. By inverting some statements, answers from inattentive respondents cancel out, which increases the statistical power of this method. For comparability of the empirical results, we linearly transform the variables so that their lowest possible level is 0 and their highest possible level is 1.

The first of these variables measures the level of customers’ understanding of the firm’s prices, as reported by the firm. Table 1.3 lists the statements and shows with which sign they enter into the sum.

Table 1.3. Customer understanding variable

Sign	Statement
+	The customers express understanding for my/our prices.
–	Some customers accuse me of profiteering.
+	The reasons for price increases are understandable for customers.

Notes: Construction of a variable measuring the understanding of the hairdresser’s customers for their prices. The respondents were asked to express their agreement with the statements on a scale from 1 (totally disagree) over 3 (undecided) to 5 (totally agree). We normalize the variable by adding 3 and then dividing the whole sum by 12.

We find that the so formed *customer understanding variable* clusters into two groups: for observations up to the 40%-quantile of this variable, the values range from 0.17 to 0.75, while for observations with values above that, values range from

0.83 to 1. We denote firms with customer understanding above the 40%-quantile as firms with *high understanding customers*, while the other firms have *low understanding customers*.

The second variable measures how satisfied the owners are with their own pricing. Table 1.4 lists the statements and shows which enter positively and which negatively into the sum.

Table 1.4. Pricing satisfaction variable

Sign	Statement
+	I am satisfied with my pricing method.
+	My prices are optimal for the firm.
–	Actually, my prices should be higher.

Notes: Construction of a variable measuring how satisfied the owners are with their own pricing. The respondents were asked to express their agreement with the statements on a scale from 1 (totally disagree) over 3 (undecided) to 5 (totally agree). We normalize the variable by adding 3 and then dividing the whole sum by 12.

The third variable measures to what extent the owners see the hygiene rule mandating a hair wash before any other procedure as a pricing tool.

Table 1.5. Mandatory hair washing variable

Sign	Statement
+	The mandatory hair washing is like a price increase.
+	I profit from the mandatory hair washing.

Notes: Construction of a variable measuring how the owners view the mandatory hair washing. The respondents were asked to express their agreement with the statements on a scale from 1 (totally disagree) over 3 (undecided) to 5 (totally agree). We normalize the variable by subtracting 2 and then dividing the whole sum by 8.

The mandate could be interpreted as a price increase and, thus, deter owners from an additional price increase. Indeed, the respondents slightly agree that mandatory hair washing is like a price increase, but they slightly disagree to profiting from it. Table 1.5 summarizes the construction of the variable.

The last variable measures how pessimistic the owners are. Table 1.6 summarizes the construction of the variable.

Our data gives rise to seven stylized facts about the relation of customer understanding with a firm's pricing and other characteristics of the firm and its owner. To improve readability, all regression tables and graphs that support these findings are in Appendix 1.C. For each regression, we also estimate a specification that includes a set of control variables. The potential control variables we consider are the firm's size, the share of regular customers, the pricing satisfaction variable, the mandatory hair washing variable, the pessimism variable, and the customer understanding vari-

Table 1.6. Pessimism variable

Sign	Statement
+	There will be another lockdown this year.
–	We will be back to normal in one year.
+	The hygiene measures will stay for years.
+	Fear of infection will deter customers for a long time.
+	Customers' willingness to pay will lastingly decrease.
–	My personal financial situation will improve.

Notes: Construction of a variable measuring the owners' expectations and professional uncertainty, expressed as pessimism. The respondents were asked to express their agreement with the statements on a scale from 1 (totally disagree) over 3 (undecided) to 5 (totally agree). We normalize the variable by adding 6 and then dividing the whole sum by 24.

able. We also estimate a specification where we include the firm's relative price in December 2020 as an additional control. Since we only construct relative prices for counties where we observe at least six firms, including the relative price as a regressor reduces the sample size by a third (see section 1.2). Since we sampled most firms by hairdresser guild, we state inferences with respect to standard errors clustered at the county-level. All results are robust to not clustering. As another robustness check, we include county fixed effects into the regressions (when possible). We find that the results are robust. In light of the results in section 1.2 about higher dispersion across counties in times of higher price change-frequency, this is an important clarification: the level of customer understanding interacts with the relative position of firms *within* counties, rather than the relative position of firms *across* counties. This is crucial for our interpretation of the results within a model where firms compete on the same market (see section 1.5).

Stylized Fact 1. Among the non-increasers, the higher the understanding of a firm's customers of its prices, the less important for price stickiness is the motive of retaining its regular customers.

To support this claim, we run a logit regression where the dependent variable equals 1 if the respondent marked the hypothesis "I did not increase my prices to retain my regular customers" as applicable and 0 otherwise. Only those respondents of our survey who report that they did not increase their price between December 2020 and March 2021 are asked about this hypothesis. Therefore, we can make this statement only for firms that do not increase their nominal price (the "non-increasers"). Since only 10% of the respondents reported that the hypothesis does not apply, we cannot estimate the full regression specification with all possible controls due to singularity problems (see table 1.C.1). Still, we find the coefficient of the customer understanding variable to be negative and significant at least at the 10%-level in all specifications we consider. We thus conclude that the main reason for price stickiness in the market applies less for firms with high understanding customers.

Stylized Fact 2. Firms with high understanding customers are more likely to increase their nominal prices.

To support this claim, we run a logit regression where the dependent variable is an indicator variable for whether the respondent increased their price. Across all considered specifications, the coefficient of the customer understanding variable is positive and significant at the 5%-level (see table 1.C.2). For interpretation of the magnitude of the effect, we run another logit regression where we substitute the customer understanding variable for the high understanding customer indicator variable, and calculate the marginal effects at means (see table 1.C.3). The average firm is 25pp more likely to increase its nominal price from December 2020 to March 2021 if it has high understanding customers.

Stylized Fact 3. Firms with high understanding customers increase their nominal and relative prices by more.

We support this claim with a OLS regression where the dependent variable is the firm's nominal price increase in percent between December 2020 and March 2021. Across all considered specifications, the coefficient of the customer understanding variable is positive and significant at least at the 5%-level (see table 1.C.4). The same result obtains if the dependent variable is the firm's relative price increase, which we define as the percent increase of the relative price in December 2020 to that in March 2021 (see table 1.C.5). When we substitute the customer understanding variable for the high understanding customer indicator variable, we find that firms with high understanding customers increase their nominal price by about 2.7pp more, and increase their relative price by about 1.9pp more. However, the significance of the results is driven by the extensive margin (i.e., stylized fact 2): repeating the exercise for nominal and relative price increases *conditional* on the firm increasing its price yields positive, but insignificant results. The reason is an interaction of the relative price position of the firm with the effect of the level of understanding of its customers: we find that higher customer understanding has a significantly positive effect on the intensive margin of firm's pricing only for firms that price around the median level of the county (see figure 1.C.1). Together with the fact that the intensive margin of price increases falls in the initial relative price of the firm[11], this interaction confounds the average effect of customer understanding on the intensive margin. In section 1.5, we explain the differential effect of customer understanding on the intensive margin across the relative price distribution in the context of a search model with firm heterogeneity.

Stylized Fact 4. Firms with high understanding customers are better able to increase their profit margins.

11. In section 1.5, we show that the negative relation between the relative price position and the intensive margin is a feature also of the CPI data.

We asked the owners how their profit margins after the lockdown compare to their profit margins, first, before the pandemic and, second, before the second lockdown. The possible answers to both questions are smaller (-1), equal (0), or larger (1). We compare the average answers for firms with low understanding and high understanding customers (table 1.C.8). A two-sample *t*-test shows that firms with high understanding customers were more able to restore their profit margin to levels before the lockdown (at 10%-level), and restore the margins to levels before the pandemic (at 15%-level).

Stylized Fact 5. Firms with high understanding customers are more satisfied with their own pricing.

We support this claim with a OLS regression where the dependent variable is the owners' satisfaction with the own pricing (the pricing satisfaction variable). The stylized fact follows from the coefficient of the customer understanding variable being positive and significant at the 1%-level across all specifications that we consider (see table 1.C.9). In terms of magnitudes, the highest value of the customer understanding variable (1) predicts a value of the pricing satisfaction variable around its 75% quantile (.75). Hence, and maybe not surprisingly, owners who report that customers are more understanding of their prices on average also report to be less constrained in their price-setting.

Stylized Fact 6. Owners of firms with high understanding customers are less pessimistic.

We support this claim with a OLS regression where the dependent variable is the owners' pessimism variable. Across all specifications that we consider, the coefficient of the customer understanding variable is negative, while it is only significantly different from zero (at the 5% and 10%-level) when we do not control for the relative price (see table 1.C.10). Note that we might simply lack the statistical power to detect a significant relation in the full specification, because the sample size shrinks by one third when adding the relative price as a regressor. One can interpret the result in two ways (which we come back to in section 1.5): the owners with high understanding customers might be less pessimistic both because they expect to restore their profit margin from before the pandemic faster, and because they are more flexible in their price setting in general.

Stylized Fact 7. Firms with less employees have more understanding customers.

To support this claim, we run a logit regression where the dependent variable is the high understanding customer indicator variable. Across all specifications that we consider, the coefficients of the two variables that measure firm size by number of employees — the linear part and the indicator variable for firms with more than six employees — are negative and significantly different from zero at the 5% level (See table 1.C.11). For the average firm, the magnitude of the effect is sizable: having one more employee reduces the probability of having high understanding customers by 8.8pp, and having more than six employees reduces the probability by 41pp.

If we interpret stylized fact 7 as causal, hairsalons with many employees are more subject to the upward price rigidity described in stylized facts 2 and 3 than smaller firms. Through the lense of the search model we build to explain our empirical findings (see section 1.5), they are also less able to retain regular customers. Larger firms with many employees may therefore be more dependent on demand from occasional customers (what we call "random demand" in section 1.5). In fact, we find that larger firms have a lower share of regular customers (see table 1.C.13). In turn, firms with a lower share of regular customers (less than 80%) give a higher score to gaining new customers as a reason for not increasing their price (see table 1.C.15). For some firms, it may hence be optimal to have a business model that is more taylored towards occasional customers than towards regular customers. This can explain why owners of large firms are *less* likely to report that retaining regular customers is a reason not to increase prices (see table 1.C.1), even though they should be more constrained by the pricing friction that stems from the low understanding of their customers.

Another important prediction of our model is that more productive firms, who can set a higher price relative to their competitors because they offer a higher-quality product, are less subject to the price rigidity that stems from having low understanding customers. If for relatively unproductive, cheaper firms, the risk of having low understanding customers is a constraint on the number of employees they hire, we should observe for relatively more productive, more expensive firms a higher number of employees. Indeed, we observe that the number of employees rises in the relative price in our survey data (see table 1.C.16).

1.5 A customer search model

To rationalize our findings about the importance of customer markets and cost-based pricing for the the hairdressers' pricing during the Covid-19 pandemic, as well as the result on the differential impact of high customer understanding on real price rigidity over the price distribution, we build a customer search model. We borrow the main idea from Fishman (1996): temporary uncertainty about a general cost increase induces upward price-rigidity for firms whose regular customers perceive a better outside option. We extend their model in two directions: first, we introduce differences across firms in quality of the produced good or service, in addition to differences in production costs of firms. Together with additional assumptions on the customers' demand curve, the heterogeneity in quality allows us to explain the non-monotonic patterns over the relative price distribution that we find empirically. Second, we introduce belief heterogeneity into the search model, similar to Janssen and Shelegia (2019). We assume that customers form rational expectations. Belief heterogeneity stems from different information sets in periods of uncertainty. Customers temporarily face uncertainty about an industry-wide cost shock. Customers that observe the idiosyncratic cost-component of their "regular" firm can perfectly learn the industry-wide cost-component by observing the firm's price in equilibrium. Other customers without access to that information, using a "conservative" learning rule, believe that industry-wide costs have not changed. In order to account for the firm heterogeneity with respect to customer understanding that we find in the data, we impose that the information level is common among all regular customers of one firm. Without loss of generality, then, we consider the simple case that each firm has only one regular customer. We take the shares of the two levels of information among customers in the population as given exogenously, i.e. we do not explain how they are formed.

1.5.1 Model setup

There are three time periods, $t = 0, 1, 2$. Firms are characterized in three dimensions: costs, quality, and information type of its regular customer. There is a unit mass of firms, indexed by i, of each type.[12] Firm i produces at marginal cost C_{it} in period t that consists of two components: two possible *baseline* marginal cost levels in period t, $\underline{c_t} < \overline{c_t}$, and a time-constant, idiosyncratic cost component, ζ_i. ζ_i is drawn independently for each firm i from a continuous distribution $\mathbb{P}_\zeta$ with mean 0 and bounded support $\left[\underline{\zeta}, \overline{\zeta}\right]$. To start production in period t, firm i has to pay fixed costs F_{it}. For tractability, we choose fixed costs as a function of the other firm characteristics and such that firms without demand from regular customers have no incentive to start production (see section 1.5.1.2).

The good or service produced by firm i in period t has *quality* $q_{it} \in \{\underline{q_t}, \overline{q_t}\}$, with $\underline{q_t} < \overline{q_t}$. We assume that high baseline costs are necessary but not sufficient to produce a high-quality product or service. As a result, there are three possible baseline cost-quality-tuples each period: $(\underline{c_t}, \underline{q_t})$, $(\overline{c_t}, \underline{q_t})$, and $(\overline{c_t}, \overline{q_t})$. In appendix 1.C.10, we present evidence from our survey for differences in quality of service over the relative price distribution[13]. Together with our model, which predicts that high-quality firms charge the highest price for their product (see below), this evidence supports our assumption of heterogeneity in quality of service in the hairdresser business. The

12. In a non-cooperative, symmetric equilibrium, allowing for a finite number of firms would complicate the analysis mainly in the following way: the outside option of customers of a given firm type would (slightly) differ from that of another firm with the same customer information type, but a different cost-quality pair, as we rule out that a customer returns to the initial firm after searching with non-zero probability (see footnote 16). This would allow for more types of equilibria (see section 1.5.2.1).

13. Kohlhepp (2023) shows that hair salons in Manhattan that are more efficient in organizing their employees across several tasks offer a higher quality-service, and charge a higher price than their competitors.

type of a firm, as characterized by low or high baseline costs, low or high quality of product, and customer information type, stays constant over time. For all firms of one type, the *levels* of cost and quality of product change over time in the same way. Let p_{it} be the price that firm i charges in period t. Quality and costs are measured in the same unit as prices.

Each period and for each firm, a customer j is born that is the "regular customer" of that firm. We denote the "regular firm" of customer j by $i_t(j)$. This means that customer j starts his search in period t at firm $i_t(j)$.[14] Customers are risk-neutral. Each customer j draws firm-specific idiosyncratic preferences ξ^i_{jt}, independently for all firms i, from a uniform distribution over the support $[0,1]$. A share α of customers is of the *low understanding* type, denoted by $u(j) = 0$, which means that they do not observe the idiosyncratic cost component of its regular firm, $\zeta_{i_t(j)}$. The rest of customers, denoted by $u(j) = 1$, instead observe the idiosyncratic cost component of their regular firm. Observing equilibrium prices, all customers try to back out the new level of industry-wide baseline costs, using a "conservative" learning rule (see section 1.5.1.3). The model implies that customers who observe the idiosyncratic costs can learn about industry-wide cost changes more easily. Each firm is characterized in part by the information type of its regular customer, $u_{i_t(j)} := u(j)$. All customers are replaced by new regular customers with the same information type at the start of a new period.

1.5.1.1 The customer's problem

The customer's problem has two stages. In the second stage, customer j has decided that he considers consumption at firm i. First, the customer learns the firm-specific preference ξ^i_{jt}. If i is unequal to the initial firm $i_t(j)$, the customer also learns the firm's price p_{it}. Otherwise, the customer already observed the price of his regular firm in the first stage.[15] Then, he solves the problem whether or not he will buy the good or service:

$$\max_{d_{jt}(i)\in\{0,1\}} d_{jt}(i)(\xi^i_{jt} q_{it} - p_{it}), \tag{1.1}$$

which has the solution that the customer buys the product, $d_{jt}(i) = 1$, iff $\xi^i_{jt} \geq p_{it}/q_{it}$.

In the first stage, the customer has the choice between staying at firm $i_t(j)$, whose price he observes, or paying search cost s and searching for a different firm, whose

14. Since our data is on male haircuts, we use masculine pronouns here.

15. The assumption that firm-specific preferences are only learnt at the second stage simplifies the computations. For this assumption to be sensible, preference shocks must be more difficult to observe than prices. In the present context, one could think of the effort of making appointments with specific employees in a hair salon, who may or may not be available at a certain date. The customer only wants to make this effort once he picked the hair salon. Prices, instead, are more easily accessible on a webpage or the shopwindow.

price he has to learn. We make the following set of assumptions about the search process:

Assumption 1.5.1. (a) Each customer j searches *at most once* each period.

(b) The search is *undirected*: the allocation of the customer j to a firm after the search will be random.

(c) Customer j *cannot deliberately return* to his initial firm $i_t(j)$ after searching in that period.

We impose these assumptions for the following reasons. Random search together with firm-specific preferences generates expected demand curves that are price-elastic, even though each customer has unitary demand given his preferences. Price-elastic demand curves are necessary to have equilibrium price dispersion in a search model (Reinganum, 1979). For tractability, we impose that customers search only once, and that they cannot return to the initial firm after the search.[16] We think that this is a reasonable description of regular customers that consider switching their hairdresser: searching for different hair salons and checking up-to-date prices in shopwindows may be physically exhausting and time-consuming. Many hair salons in Germany are small and their webpage may not exist or seem unreliable. Also, we like to interpret s as including a *switching cost*: the relation of hairdressers with their clients can be intimate. Once customers decide to search for different hairdressers, they may incur the psychological cost of "cutting ties" with their old hairdresser. In fact, in our simple framework, one can interpret the cost s as a *pure* switching cost. Instead of assuming that understanding customers are more informed, which has an effect in uncertain times, we could also just assume that they have a higher switching cost. However, such an interpretation does not explain a change in price stickiness in the wake of cost shocks. Such a change is crucial for the model to generate a deviation in relative price changes (see below).

Customers know the time-constant discrete probability distribution $\mathbb{P}$ over the tuples $(c_t, q_t, u) \in \{\underline{c_t}, \overline{c_t}\} \times \{\underline{q_t}, \overline{q_t}\} \times \{0, 1\}$ that characterize all firm types, as well as the distribution of the idiosyncratic cost component $\mathbb{P}_\zeta$.[17] Given his information type $u(j)$, customer j assumes a certain baseline cost $\underline{c_t}^{u(j)}$. Conditional on this belief, he forms rational expectations about the prices of all firms that he does not already observe, $\{p_{it}^{u(j)}\}_{i \neq i_t(j)}$. In order to characterize the degree of competition that firm $i_t(j)$ is subject to, we calculate the *expected surplus* for customers with information type u of consuming at firm $i \neq i_t(j)$, which we compute as

16. "Free recall" and a costless return to the original offer generates "return demand", which has interesting implications in an ordered search model, see e.g. Armstrong (2017).

17. In line with the assumptions outlined above, customers are aware that there are no high quality firms with low costs, so they attribute zero probability mass to tuples $(\underline{c_t}, \overline{q_t}, u)$, $u \in \{0, 1\}$.

$$V_{it}^u := \int_0^1 \max\{\xi q_{it} - p_{it}^u, 0\} d\xi = \int_{p_{it}^u/q_{it}}^1 \xi q_{it} - p_{it}^u d\xi = \frac{1}{2} q_{it} - p_{it}^u - \left(\frac{p_{it}^{u\,2}}{2q_{it}} - \frac{p_{it}^{u\,2}}{q_{it}} \right)$$
$$= \frac{(q_{it} - p_{it}^u)^2}{2q_{it}}, \; p_{it}^u < q_{it}. \tag{1.2}$$

With a slight abuse of notation, we can write $V_{(c,q,u,\zeta)t}^{u(j)}$, since in equilibrium, the expected surplus for customers is the same across all firms i with the same type (c, q, u) and idiosyncratic cost component ζ (see below). Then, customer j searches in the first stage iff

$$V_{i_t(j)t} < \underbrace{\sum_{c \in \{\underline{c_t}^{u(j)}, \overline{c_t}^{u(j)}\}} \sum_{q \in \{\underline{q_t}, \overline{q_t}\}} \sum_{u \in \{0,1\}} \mathbb{P}[(c,q,u)] \int_{\underline{\zeta}}^{\overline{\zeta}} V_{(c,q,u,\zeta)t}^{u(j)} d\mathbb{P}_\zeta(\zeta)}_{=:\mathbb{E}\mathbb{V}_t^{u(j)}} - s, \tag{1.3}$$

where $V_{i_t(j)t}$ is the expected surplus for customer j of staying at his regular firm, which he can compute with the observed price of firm $i_t(j)$. In the following, we call $\mathbb{E}\mathbb{V}_t^{u_i} - s$ the *expected outside option* of the regular customer of firm i.

1.5.1.2 The firm's problem

A firm is characterized by the tuple (c_t, q_t, u, ζ). Given its marginal costs $C_{it} = c_{it} + \zeta_i$ and fixed costs $F_{it} =: F_t(C_{it}, q_{it})$, the quality of its service or good q_{it}, and the information type of its regular customer u_i, firm i chooses its price in order to maximize its expected period profits:

$$\max_{p_{it}} \mathrm{E}_j^u \left[d_{jt}(i) \right] (p_{it} - C_{it}) - F_{it}. \tag{1.4}$$

The expected demand $\mathrm{E}_j^u \left[d_{jt}(i) \right]$ is a function of the firm's price and its product's quality, as well as of the expected outside option of its regular customer, which depends on his beliefs. These, in turn, can in general be influenced by the firm's pricing. For this subsection, we assume that the firm takes the regular customer's expected outside option *after* having learnt from prices, $\mathbb{E}\mathbb{V}_t^{u_i} - s$, as given. We discuss the customer's learning from prices and additional assumptions on firms' pricing decisions in section 1.5.1.3.

In each period t, there are two possible sources of demand for each firm: the demand of its regular customer, and random demand from customers that search. We denote the mass of random demand from search expected in period t as $\mathscr{D}_t \in [0, 1]$. First, suppose that firm i's regular customer does not search, regardless of firm i's price p_{it}. Then, since the firm-specific preferences ξ^i_{jt} are independent among customers, the total expected demand of firm i is given by

$$(1+\mathscr{D}_t)\int_0^1 \mathbb{I}_{\xi\geq p_{it}/q_{it}}d\xi \;=\; (1+\mathscr{D}_t)\max\left\{\left(1-\frac{p_{it}}{q_{it}}\right),0\right\}. \tag{1.5}$$

Without competition from other firms, and under the condition $p_{it}\leq q_{it}$, firms set their monopoly price if production is profitable:

$$\begin{aligned} p^m_{it} \;:=&\arg\max_{p_{it}}\,(1+\mathscr{D}_t)(1-p_{it}/q_{it})(p_{it}-C_{it})-F_{it}\\ =&\arg\max_{p_{it}}\,-\frac{(1+\mathscr{D}_t)}{q_{it}}\left[\left(p_{it}-\frac{C_{it}+q_{it}}{2}\right)^2-\left(\frac{q_{it}-C_{it}}{2}\right)^2\right]-F_{it}\\ =&\frac{C_{it}+q_{it}}{2}\ \text{if}\ F_{it}\;\leq\;\frac{(1+\mathscr{D}_t)}{q_{it}}\left(\frac{q_{it}-C_{it}}{2}\right)^2, \end{aligned} \tag{1.6}$$

where $p^m_{it}\leq q_{it}$ holds as long as $q_{it}\geq C_{it}$.

The competition that the firm faces is characterized by inequality (1.3). If the inequality holds in period t, the firm's only expected source of profit stems from random demand. Since by assumption (1.5.1) customers search not more than once in each period, the firm could set its monopoly price if it decides to not retain its regular customer, generating expected profits

$$\pi^{\neg r}_{it} \;:=\; \frac{\mathscr{D}_t}{q_{it}}\left(\frac{q_{it}-C_{it}}{2}\right)^2-F_{it}. \tag{1.7}$$

For tractability, we assume that the firm's fixed costs $F_{it}=F_t(C_{it},q_{it})$ are such that $\pi^{\neg r}_{it}=0$ holds each period. Therefore, if the firm cannot retain its regular customer, it exits the market this period. Together with (1.6), this level of fixed costs also implies that firms that face no competition remain in business as long as production yields a positive expected surplus, i.e. as long as $q_{it}\geq C_{it}$.

The firm can retain its regular customer by lowering its price p_{it} and thereby offering a higher expected surplus V_{it}. It does so until either its offer is at least as valuable as the customer's expected outside option, $\mathbb{E}\mathbb{V}^{u_i}_t - s$, or the expected profits from retaining the customer fall below zero. Let V^*_{it} denote the expected surplus the firm offers to its customers at the threshold when expected profits are zero. In appendix 1.D.2, we show that

$$V^*_{it} \;=\; \left(1+\sqrt{1/(1+\mathscr{D}_t)}\right)^2\frac{F_{it}}{2\mathscr{D}_t}. \tag{1.8}$$

Intuitively, the higher the fixed costs are relative to expected random demand, the higher is the expected surplus that the firm is willing to offer its regular customer in order to retain him.

If the regular customer does not search at the firm's monopoly price, it implies that the offered surplus at the monopoly price, defined as V^m_{it}, exceeds his expected outside option. The firm will never offer a lower expected surplus than V^m_{it}. In sum, firm i exits the market in period t if $q_{it} < C_{it}$ or $\mathbb{E}\mathbb{V}^{u_i}_t - s > V^*_{it}$ hold, and otherwise offers the expected surplus

$$V_{it} \;=\; \max\{\mathbb{E}\mathbb{V}^{u_i}_t - s, V^m_{it}\} \tag{1.9}$$

while retaining its regular customer. It is easily shown that the optimal price of firm i is in the interval $p_{it} \in [C_{it}, q_{it})$.[18] Then, using the result in (1.2), it is a function of the optimal expected surplus that firm i offers in period t:

$$p_{it} \;=\; p(V_{it}, q_{it}) \;:=\; q_{it} - \sqrt{V_{it} \cdot 2q_{it}}. \tag{1.10}$$

1.5.1.3 Customers' learning from prices

We now describe how customers learn from observed prices about industry-wide costs. In principle, firms can have an incentive to adjust their prices in order to manipulate the customer's belief about the industry-wide cost level, thereby changing the level of competition they are subject to. This could imply a deviation from the above description of optimal firm pricing. The following assumption, together with assumptions we make about the learning behavior of customers, is enough to ensure that this will not be the case in our setting.

Assumption 1.5.2. Firms do not charge a higher price than their monopoly price, i.e. for any firm i in any period t, $p_{it} \leq p^m_{it}$.

This assumption is only binding in periods where customers face uncertainty about the firms' industry-wide baseline marginal cost. As in Fishman (1996), we assume that any uncertainty period t is preceded and succeded by certainty periods. Any customer j with information type $u(j)$ will observe the price $p_{i_t(j)t}$ of his regular firm $i_t(j)$ if it is in business in period t (otherwise, he will directly search for a new firm). The customer enters the period with last period's belief about his outside option, $\mathbb{E}\mathbb{V}^{u(j)}_{t-1} - s$, that is shared among customers with his information type, and that is consistent with his knowledge of last period's industry-wide baseline marginal cost level $\underline{c}^{u(j)}_{t-1}$.

If the customer is of the type $u(j) = 1$, he observes the idiosyncratic cost component $\zeta_{i_t(j)}$ of the firm. Together with his knowledge of the last period's industry-wide baseline cost, he can calculate the expected monopoly price of the firm, using (1.6). If the observed price lies at or below the expected monopoly price, the customer does not attempt to update beliefs. If the observed price lies above the expected

18. Let p^*_{it} denote the price that conforms to V^*_{it}. For $\mathscr{D}_t \to 0$, $p^*_{it} \to C_{it}$ by equations (1.8) and (1.2). Hence, $C_{it} \leq p^*_{it} \leq p_{it} \leq p^m_{it} < q_{it}$, where the last inequality follows by $C_{it} < q_{it}$, which holds for all firms that stay in the market.

monopoly price, the customer concludes that industry-wide baseline costs must have increased, and backs out the new cost level $\tilde{\underline{c}}^1_{i_t(j)t}$ under the assumption that the observed price is the firm's new monopoly price. Also, he updates his expected outside option $\tilde{\mathbb{EV}}^1_{i_t(j)t} - s$ as implied by the equilibrium that obtains with the new cost level. If the offered expected surplus implied by the observed price $p_{i_t(j)t}$ is at least as large as the updated outside option, and if the price is smaller than the observed quality $q_{i_t(j)t}$, the customer maintains his updated belief (and stays at firm $i_t(j)$): $\underline{c}^1_{i_t(j)t} = \tilde{\underline{c}}^1_{i_t(j)t}$, $\mathbb{EV}^1_{i_t(j)t} - s = \tilde{\mathbb{EV}}^1_{i_t(j)t} - s$. Otherwise, the customer goes back to his old beliefs about the industry-wide baseline cost, $\underline{c}^1_{i_t(j)t} = \underline{c}^1_{t-1}$, and updates his expected outside option accordingly, $\mathbb{EV}^1_{i_t(j)t} - s = \mathbb{EV}^1_t(\underline{c}^1_{t-1}) - s$ (and starts to search).[19]

Proposition 1.5.1. Given assumption 1.5.2 holds, a high understanding customer of firm i learns a fraction γ^1_i of increases in the industry-wide marginal baseline cost in uncertainty period t:

$$\underline{c}^1_{it} = \underline{c}_{t-1} + \gamma^1_i(\underline{c}_t - \underline{c}_{t-1}),$$

where $\gamma^1_i \in [0,1]$. Any firm i with a high understanding regular customer behaves as described in section 1.5.1.2, with the customer's expected outside option $\mathbb{EV}^1_{it} - s$ as the one that obtains after he learnt from its price p_{it}.

Proof: Firm i only has an incentive to signal higher industry-wide costs if it is restricted by its competition, in the sense that the expected surplus it offers as monopolist, V^m_{it}, is below its customer's expected outside option as implied by his *non-updated beliefs*, $\mathbb{EV}^1_{t-1} - s$. Let $\underline{c}_t$ denote the true industry-wide marginal baseline cost in period t, which is at least as high as its level in period $t-1$, $\underline{c}_{t-1}$. $\mathbb{EV}_t - s$ denotes the expected outside option that would prevail if the customer knew the industry-wide cost (his true outside option). The customer's initial expected price is $p^{m,e}_{it} = \frac{c_i(\underline{c}^1_{t-1}) + \zeta_i + q_{it}}{2}$. In order to trigger a belief update by the customer, the firm must set a price that is the monopoly price of a firm subject to baseline costs $\tilde{\underline{c}_{it}}^1$ larger than $\underline{c}^1_{t-1}$. On the other hand, by assumption 1.5.2, the firm never sets a price higher than its monopoly price, so that $\tilde{\underline{c}_{it}}^1 \leq \underline{c}_t$.

If the firm is a monopolist under the true outside option of the customer, $V^m_{it} \geq \mathbb{EV}_t - s$, it optimally sets its price to p^m_{it}, which triggers the customer to learn the true industry-wide costs ($\gamma^1_i = 1$). Otherwise, the firm will set a price that signals the largest industry-wide cost $\tilde{\underline{c}_{it}}^1$ such that the implied expected surplus offered by the firm, $V_{it}(\tilde{\underline{c}_{it}}^1)$, equals the implied expected outside option, $\tilde{\mathbb{EV}}^1_{it} - s$. The

19. We leave the possibility open that other shocks can happen simultaneously in the uncertainty period, so that the outside option in general differs from period $t-1$ to t even if the customer does not change his belief about the industry-wide baseline costs. For the experiment we consider, however, customers that do not learn about the cost increase will just revert to their last period's expected outside option.

learnt expected outside option is at least as large as $\mathbb{EV}_t - s$, since the expected surpluses that the firm's competitors offer are weakly decreasing in (expected) baseline costs.[20] Since the firm cannot lower the level of its (expected) competition below the one that obtains under perfect information, and since customers do not update their beliefs when the observed price exceeds the observed quality of the product, the conditions for staying in business remain as in section 1.5.1.2, subject to the updated belief $\mathbb{EV}^1_{it} - s$. □

If customer j is of the low understanding type, $u(j) = 0$, he does not observe the idiosyncratic cost component. However, given some belief about the industry-wide baseline costs, $\underline{c}^e$, and the knowledge of the idiosyncratic cost distribution, he can calculate the maximum price he expects to be charged by firm $i_t(j)$:

$$p^{max}_{i_t(j)t}(\underline{c}^e) \; := \; \frac{q_{i_t(j)t} + c_i(\underline{c}^e) + \overline{\zeta}}{2}. \tag{1.11}$$

If the customer observes a price $p_{i_t(j)t}$ at or below $p^{max}_{i_t(j)t}(\underline{c}^0_{t-1})$, he does not update his beliefs. If the observed price $p_{i_t(j)t}$ is higher than the level of the maximum expected price, he considers the possibility that industry-wide baseline costs have increased to the level $\tilde{\underline{c}}^0_{it}$ such that $p^{max}_{i_t(j)t}(\tilde{\underline{c}}^0_{it}) = p_{i_t(j)t}$. He also updates the belief about his outside option that would obtain in equilibrium, $\mathbb{E}\tilde{\mathbb{V}}^0_{it} - s$. If the offered expected surplus implied by the observed price $p_{i_t(j)t}$ is at least as large as the updated expected outside option, and if the price is smaller than the observed quality $q_{i_t(j)t}$, the customer maintains his updated belief (and stays at firm $i_t(j)$): $\underline{c}^0_{it} = \tilde{\underline{c}}^0_{it}$, $\mathbb{EV}^0_{it} - s = \mathbb{E}\tilde{\mathbb{V}}^0_{it} - s$; otherwise he only updates his belief about his outside option, $\mathbb{EV}^0_{it} - s = \mathbb{EV}^0_t(\underline{c}^0_{t-1}) - s$.

Corollary 1.5.1. Given assumption 1.5.2 holds, a low understanding customer of firm i learns a fraction $\gamma^0_i \leq \gamma^1_i$ of increases in the industry-wide marginal baseline cost in uncertainty period t:

$$\underline{c}^0_{it} \; = \; \underline{c}_{t-1} + \gamma^0_i(\underline{c}_t - \underline{c}_{t-1}),$$

where $\gamma^0_i \in [0,1]$. Any firm i with a low understanding regular customer behaves as described in section 1.5.1.2, with its customer's outside option $\mathbb{EV}^0_{it} - s$ as the one that obtains after he learnt from its price p_{it}. If the firm's marginal cost change $C_{it} - C_{it-1}$ is bounded above by $\overline{\zeta} - \zeta_i$, it cannot signal any industry-wide cost increase to its customers ($\gamma^0_i = 0$).

20. We assume that in each period t, at least one firm type is a monopolist in all possible equilibria, so that the outside option does not become indeterminate. Also, it could be optimal for the firm to offer an expected surplus V_{it} strictly higher than its customer's outside option, if the expected industry value $\mathbb{EV}_t$ was discontinuous in expected baseline costs $\underline{c}_t$. Possible discontinuity points are at the threshold values V^*_{it}, which correspond to threshold idiosyncratic costs $\overline{\zeta}^*_{it}(\underline{c}_t)$ at which firms of a given type exit the market. Since the expected industry value integrates over the continuous distribution of idiosyncratic costs, these discontinuities in $\underline{c}_t$ smoothen out in the aggregate.

Proof: Firms with low understanding customers can only signal industry-wide costs that are upper bounded by what firms with high understanding customers can signal: $\tilde{\underline{c}}_{it}^{\,0} \leq \tilde{\underline{c}}_{it}^{\,1}$. Then, the proof of proposition 1.5.1 goes through, with $\gamma_i^0 \leq \gamma_i^1$. For the last statement, we use that the baseline cost expected by the customers ex-ante, $\underline{c}_{t-1}^0$, equals the true baseline cost that obtained in period $t-1$. Therefore, $C_{it} - C_{it-1} \leq \overline{\zeta} - \zeta_i$ implies $C_{it} \leq c_i(\underline{c}_{t-1}^0) + \overline{\zeta}$, so that firm i cannot trigger a belief update by its customer with a low understanding type unless it sets a price that exceeds its monopoly price, which is ruled out by assumption 1.5.2. □

We want to add two remarks to our description of the customers' learning from prices. First, while we set the learning rules ad-hoc, we think that its properties are justifiable: given assumption 1.5.2, customers following the rules cannot be fooled by firms into overestimating the industry-wide cost increase, and in that sense act *conservatively*. This might be rational if they want to minimize the risk of sticking with a firm when they should have searched.[21] At the same time, customers are interested in learning: when industry-wide conditions worsen and firms have to exit the market as they cannot credibly blame aggregate shocks, customers pay unnecessary search (and switching) costs.

In consumer search markets with uncertainty, demand can increase in the posted price, which was analyzed by Janssen, Parakhonyak, and Parakhonyak (2017). Assumption 1.5.2 rules out "extreme" instances of this phenomenon, and is crucial to obtain the above result. To justify the assumption, we make recourse to typical properties of dynamic pricing that we abstract from in our model: in the presence of nominal pricing frictions, firms set the price near their long run price target at the cost of forfeiting higher profits in the short run. The only reason why a firm would increase its price above its monopoly price is to signal higher than realized industry-wide costs to its regular customer, which is only effective in an uncertainty period. When the change in industry-wide costs is expected to be more persistent than the customers' uncertainty about it, the firm's long run price target therefore is upper bound by its monopoly price in the uncertainty period.[22] What is more, our setting with long-lived regular customers lends itself to a micro-foundation for nominal price stickiness: customers could follow a *dynamic* learning rule, where they punish temporary price increases during uncertainty periods by subsequently leaving the firm. Knowing this threat, firms abstain from signaling industry-wide costs that are too high. In turn, customers are willing to learn from the firm's prices. This argument is close to the results in Nakamura and Steinsson (2011): nominal price

21. We think this is a reasonable assumption when the change in industry-wide costs is the result of a rare event, like a pandemic, where customers cannot draw on prior knowledge about probabilities of cost shocks. A "conservative" estimation strategy to deal with such Knightian uncertainty can be micro-founded using robust control theory (Hansen and Sargent, 2022).

22. More than 80% of the respondents to our survey (N=257) in the spring of 2021 agree with the prediction that mandatory hygiene measures to prevent the spread of the virus will remain in place for years to come. Less than 3% of the respondents disagree.

rigidities can be a *commitment device* that helps firms to achieve more favourable equilibria in customer markets with information asymmetry.

As a second remark, we only consider cost *increases* when describing the learning rules, since this is the relevant case for explaining our survey evidence. However, our analysis also uncovers an asymmetry between cost increases and decreases that may be of interest on its own. At a first glance, learning about industry-wide cost *decreases* might appear to be an easier problem: firms have no incentive to signal *lower* than realized cost decreases, as this would only increase their customer's expected outside option. Hence, high understanding customers could attempt to learn about cost decreases using a symmetric version of the rule described above, but without the need for an analogue to assumption 1.5.2. However, some firms may have an incentive to exploit the *sluggishness* of the resulting learning rule, which stems from the assumption of adaptive expectations. Suppose that industry-wide costs decrease in an uncertainty period, and consider a firm whose offered expected surplus at its monopoly price before the cost decrease was exactly equal to its customer's outside option then. If the firm does not change its price, even though its monopoly price now is lower, it does not trigger a belief update by its regular customer. The firm prefers this strategy if the *expected marginal increase* in the customer's expected outside option from signaling lower industry-wide costs outpaces the expected marginal gain to the customer from the firm's price decrease. This is most likely to be the case for firms who offer a lower surplus than their competitors.[23] Different from the case of cost increases, the neglect to pass through lower costs is not easily detectable via patterns in nominal prices: firms may just fix their nominal price until the uncertainty subsides. This scheme is thus also uninhibited by nominal price rigidities.[24]

Our theory thus predicts that industry-wide cost decreases are more difficult to learn from prices than cost increases, especially for regular customers of less profitable firms. This may be a possible explanation for the evidence of asymmetric incidence of tax changes: Benzarti et al. (2020) show for the case of the Finnish hairdressing business, among others, that the pass-through of an industry-targeted decrease in value-added taxes is only half of the pass-through of the subsequent increase in value-added taxes. The effect is driven by firms with low profit margins.

23. This can be shown with expected surpluses of monopolists, V_{it}^m. Combining equations (1.2) and (1.6), it holds that $V_{it}^m = \frac{((q_{it}-C_{it})/2)^2}{2q_{it}}$. The derivative of this surplus by marginal costs, $\partial V_{it}^m/\partial C_{it} = -\frac{1-C_{it}/q_{it}}{4}$, falls in the quality-cost ratio q_{it}/C_{it} of the firm, and thereby in its productivity/profitability.

24. Note that if customers could credibly commit not to learn about industry-wide conditions from the firm's price decreases, this would be a Pareto improvement: customers would benefit from more pass-through of lower costs, while firms could freely set their lower monopoly price. A low but positive inflation rate of the overall consumption-basket could be such a commitment (or *obfuscation*) device: by keeping their nominal prices constant, firms could decrease their real prices over time, which may go unnoticed by customers who pay little attention to low inflation rates (Coibion and Gorodnichenko, 2015).

1.5.1.4 Equilibrium

Each period t, the price of firm i is determined by its type (c_{it}, q_{it}, u_i) and its idiosyncratic cost component ζ_i. Each customer j of with information type $u(j)$ accounts for this and computes the expected price of firm i in period t as a function of the tuple (C_{it}, q_{it}, u_i). In periods without uncertainty, there are three firm types: low-cost, low-quality firms, high-cost, low-quality firms, and high-cost, high-quality firms. We denote the *median* equilibrium prices of these types as $\underline{\underline{p_{t,u}}}$, $\overline{\underline{p_{t,u}}}$, and $\overline{\overline{p_{t,u}}}$, respectively, where the firm's prices may also differ by their regular customer's information type u in periods with uncertainty, and the median is over the distribution of the idiosyncratic cost component.[25] We denote the corresponding expected customer surpluses of consuming at firms with these types that arise in equilibrium as $\underline{\underline{V_{t,u}}}$, $\overline{\underline{V_{t,u}}}$, and $\overline{\overline{V_{t,u}}}$.

In periods with uncertainty, the expected prices and surpluses $p_{it}^{u(j)}$ and $V_{it}^{u(j)}$ by customer j of information type $u(j)$ do not generally coincide with the equilibrium prices and surpluses. However, by the assumption of rational expectations, all customers' expectations must be consistent with *some* equilibrium. We will consider an equilibrium where customers who observe the idiosyncratic cost component of their regular firm can identify the industry-wide cost increase that induces the true equilibrium ($\gamma_i^1 = 1$ for all firms i with $u_i = 1$), while some customers without this information have the counterfactual belief that costs did not increase, and form expectations consistent with an alternative equilibrium ($\gamma_i^0 = 0$ for some firms i with $u_i = 0$). In order for the counterfactual belief to be rational, the support of the idiosyncratic cost component, $\left[\underline{\zeta}, \overline{\zeta}\right]$, has to be wide enough, which we assume (see below). For any customer j with information type $u(j) = 1$, it then holds that $p_{it}^1 = p_{it}$ and $V_{it}^1 = V_{it}$ for all firms i. The customers calculate their expected outside option in period t as the integral over all possible expected surpluses that they expect to obtain in equilibrium, as in (1.3). To be consistent, both the true as well as the counterfactual surpluses fulfill condition (1.9) for all firm types. The median of idiosyncratic prices (1.10) over $\mathbb{P}_\zeta$ yields the median prices that obtain in equilibrium from the expected customer surpluses of each firm type in the true equilibrium.

The expected mass of searching customers in equilibrium, $\mathscr{D}_t$, is equal to the expected mass of firms that do not retain their regular customers:

$$\mathscr{D}_t^{\tilde{u}} = \sum_{c\in\{\underline{c_t}^{\tilde{u}}, \overline{c_t}^{\tilde{u}}\}} \sum_{q\in\{\underline{q_t}, \overline{q_t}\}} \sum_{u\in\{0,1\}} \mathbb{P}[(c,q,u)] \int_{\underline{\zeta}}^{\overline{\zeta}} \mathbb{I}_{\left\{V^{*,\tilde{u}}_{(c,q,u,\zeta)t} < \mathbb{E}\mathbb{V}_t^u - s \text{ or } c+\zeta > q\right\}} d\mathbb{P}_\zeta(\zeta). \quad (1.12)$$

25. Taking the median instead of the mean helps for the comparison with the data, where we take the median in order to be robust to outliers on the observed price distribution.

In general, customers who form expectations consistent with the alternative equilibrium, where industry-wide costs did not increase, will expect a counterfactual mass of searching customers, denoted by $\mathscr{D}_t^0$. Firms on the other hand all observe the true cost shock, and hence all know the expected random demand of the actual equilibrium, $\mathscr{D}_t^1 = \mathscr{D}_t$. Customers do not observe which firms stay in the market or exit the market, so that they cannot learn about the true level of random demand, and by extension about industry-wide costs, from firms' decisions whether to produce. If their regular firm stays in business, customers with the low information type will in general assume that its idiosyncratic cost component is higher than it actually is.[26]

1.5.2 Model experiment

In the periods $t = 0$ and $t = 2$, all customers are perfectly informed about the baseline levels of the firms' costs, $\underline{c_t}$ and $\overline{c_t}$. We consider the experiment where the baseline production costs increase over time by a fixed amount $\kappa > 0$: $\underline{c_0} =: \underline{c} < \underline{c}' = \underline{c} + \kappa := \underline{c_1}$ and $\overline{c_0} =: \overline{c} < \overline{c}' = \overline{c} + \kappa := \overline{c_1}$. We assume that the customers know that any possible cost-increase is a fixed amount κ that is added to low or high baseline costs. Hence, customers who learn about the new level of the low baseline cost $\underline{c}'$ also learn about the new level of the high baseline cost, $\overline{c}'$. In the period $t = 1$, while baseline costs have already increased to $\underline{c}'$, low understanding customers instead believe that they are still at the level $\underline{c}$: $\underline{c}_{i1}^0 = \underline{c}$ for some firms i. In period $t = 2$, all low understanding customers have learned the higher baseline cost-levels.[27]

With this industry-wide cost-shock, we aim to capture the adverse effect of the pandemic on the hairdressing-business in Germany between the years 2020 and 2021: first, mandatory hygiene- and distancing-measures that were in place during that time increased the marginal and fixed costs of producing haircuts for all kinds of hairdressers. Second, the two mandatory shutdowns that lasted several months also increased the ex-post fixed cost of running a hairdresser-business. Third, the federal minimum wage in Germany increased at the first of January 2021, and several federal states who have an independent minimum wage for the hairdresser industry increased it at that time as well.

1.5.2.1 Solving for an equilibrium

We look for an equilibrium of the model where the relative price distribution is as disperse as in the data, and the information type of the firm's customer imposes

26. Note that there is no dynamic learning by the assumption that customers are replaced each period.

27. Modeling the dynamic learning of customers is beyond the scope of this paper. We could imagine that the customers learn over time by occasionally observing the prices of random firm types, or that customers with the lower information type learn over time from customers with the high information type.

a constraint on the price-setting of at least some firm types in period $t = 1$. For measuring price dispersion in the model, we look at the dispersion of the median equilibrium prices across types, across firms that stay in business in that period. The three firm types that obtain in the model in periods without uncertainty induce a *fundamentals-based* order of relative prices, which can be derived from the monopoly prices in (1.6): low cost firms are the relatively cheapest, while high quality firms are the relatively most expensive. We match this fundamentals-based order to tertiles of the empirical relative price distribution. In the data, we observe relatively little transitions of firms across the tertiles over time.[28] Therefore, as an additional requirement, we only consider equilibria where the relative prices of firms of a given firm type conform to the fundamentals-based ordering of firm types.

Only a narrow set of equilibria fulfills these conditions: In period $t = 1$, the low baseline cost and the high quality firm types set their monopoly price, while the high baseline cost, low quality firm type is split into two: the median firm with low understanding customers that stays in business sets a price lower than its monopoly price in order to retain its regular customer, while the median firm with understanding customers that stays in business either sets its monopoly price, or a lower price, which is however still higher than that of the firm which is subject to less understanding customers.

The intuition for this result is the following: The search cost s is the main model parameter that we can vary to select equilibria. Trivially, with s large enough, all firm types can charge their monopoly price, which differs across cost-quality pairs, but not customer information types. Lowering s, the first firm type that is subject to real price rigidity is the one that offers the lowest expected surplus to customers relative to its competitors. Naturally, this is the case for the firms producing low quality goods or services at high baseline costs, and, in periods with uncertainty, those with less understanding customers. If s is low enough such that more than two firm types are restricted by their competition, the price distribution starts to collapse: all firms that are restricted by competition, and whose customers' outside options are the same, set the same price. While different customer information levels imply different expected outside options, and the model in principle allows for infinitely many information levels, the learning rules from section 1.5.1.3 imply that information levels above some firm type-specific threshold $\underline{\gamma}_i$ allow firms to set their monopoly price, while for prices below that, customers remain at the lowest information level $\gamma = 0$.[29] As a consequence, only up to two firm types can be restricted

28. We find that the majority of firms (63% ±5%) remain in the same relative price tertile for all months between March 2020 and December 2021 (CPI micro-level data, N=89).

29. We restrict ourselves to equilibria where $\gamma_i^1 = 1$ and $\gamma_i^0 \in \{0\} \vee [\underline{E}, \overline{E}], \underline{E}, \overline{E} \in (0, 1)$ for all firms i; see appendix 1.D.

and still set different prices.[30] Unless firms transition across tertiles, which we rule out, not more than two firm types can thus be restricted by their competition for the price distribution to remain dispersed.[31]

In order to solve for an equilibrium that fulfills the above criteria, we choose the search cost s between two bounds, which are determined by the equilibrium conditions of the high baseline cost, low quality, low understanding firm type in period $t = 1$, characterized as $(\overline{c_1}, \underline{q_1}, 0)$. The upper bound on s binds when the median firm of this type can set its price at the monopoly price level. At that point, differences in customers' beliefs do not imply differences in (median) prices, which is contrary to our interpretation of the data. The lower bound on s binds when *from the perspective of* low understanding customers firms of this type with the highest idiosyncratic cost level are on the brink of stopping production. For lower s, low understanding customers would rationally expect to only observe firms of that type in period t with idiosyncratic costs up to a bound smaller than $\overline{\zeta}$. Then, they would follow a different learning rule than the one we describe in section 1.5.1.3. We choose to set s equal to this lower bound.

In appendix 1.D, we describe the numerical algorithm we use to solve the model for each period and numerically check the requirements on the equilibrium. In particular, we check that for our calibration, the chosen s fulfills two requirements: in uncertainty period $t = 1$, the low understanding customers of firms that are search-restricted in their price setting rationally expect firms with *high* understanding customers to be search-restricted as well, since they believe that industry-wide conditions have not changed from last period. At the same time, all firms with high understanding customers in fact have monopoly power in that period, unless their costs exceed their quality, at which point they exit the market. By proposition 1.5.1, this ensures that all customers of the high information type learn the true industry-wide cost increase in the uncertainty period ($\gamma_i^1 = 1$). At the same time, we check that firms with high costs, low quality, and low understanding customers, whose idiosyncratic costs ζ_i are in the interval $[\underline{\zeta}_1^*, \overline{\zeta}_1^*]$, are restricted to set their price equal to the outside option of their customer. Firms with even lower costs, $\zeta_i < \underline{\zeta}_1^*$, are able to charge their monopoly price, while their low understanding customers still do not learn about the cost increase. Firms with higher costs $\zeta_i > \overline{\zeta}_1^*$ either exit the market or (at another threshold, $\zeta_i > \overline{\zeta}_1^E$) signal higher industry-wide costs by charging their monopoly prices. We calibrate the model to the targeted data moments by numer-

30. This result could be attenuated by considering finite firm size, which would induce different outside options for firms with different cost-quality pairs.

31. With transitions across tertiles, one could for example have an equilibrium where the median high quality firm with low understanding customers is search-restricted, while the median high quality firm with high understanding customers can charge its monopoly price. In this situation, low quality firms with high baseline costs and high understanding customers would charge more than some firms who offer a high quality product, so that the fundamentals-based order of firm types on the relative price distribution would be violated.

Table 1.7. Calibration of model parameters

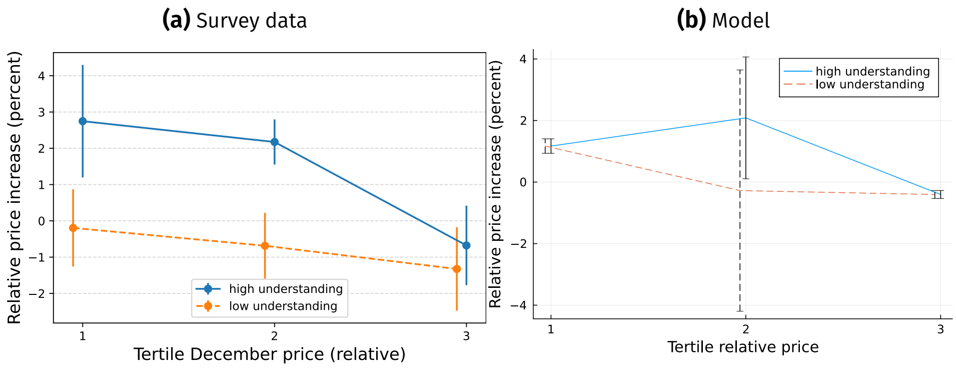
Parameter	Value	Matched data moment
$\underline{c}$	1	- (normalization)
$\bar{c}$	1.55	relative price dispersion December
$\underline{q}$	1.99	relative price dispersion December
$\bar{q}$	2.53	relative price dispersion December
κ	0.18	relative price increases March
$\bar{\zeta}$	0.21	relative price gap March
α	0.45	survey evidence
s	2.88%	choice of equilibrium

ically minimizing the sum of squared deviations from the targeted data moments over the parameter space.

1.5.2.2 Calibration and Results

We have two main calibration targets: the dispersion of the relative price distribution for the December 2020-prices of our survey participants, and the heterogeneous relative price increases from December 2020 to March 2021 across the relative price distribution. For these data moments, we construct the relative prices of hairdressers in our survey as described in section 1.2. Then, we split the relative price distribution of December 2020, pooled across all surveyed counties with at least 6 firms, into tertiles. We find that the median price of firms in the first tertile is 14.7pp lower than the median price of the overall distribution, while the median price of firms in the third tertile is 16.7pp higher than the median price. Our calibrated model matches this dispersion quite well: the median firm in the first relative price-tertile, which is of the low cost-type, prices 14.8pp below the median price, while the median firm in the third tertile, which is of the high quality-type, prices 16.3pp above the median price.

Next, we calibrate the changes in costs and quality to the observed relative price changes from December 2020 to March 2021, which we define as the percentage change in relative prices between these two periods. We average these price changes over firms in the tertiles of the relative price distribution of December 2020. In the data, we find a significantly lower relative price change from the second to the third tertile among firms with high understanding customers (see the left panel of figure 1.6), where the firms in the middle tertile increase their relative price by about 2.5pp more. We find a similar, statistically significant gap between the relative price changes of firms with high understanding and firms with low understanding customers within the second tertile. In other tertiles, differences in the level of customer understanding do not lead to significantly different relative price changes. Our calibrated model likewise generates gaps of 2.5pp between the second and third tertile and 2.4pp between firms with high and low understanding firms in the second ter-



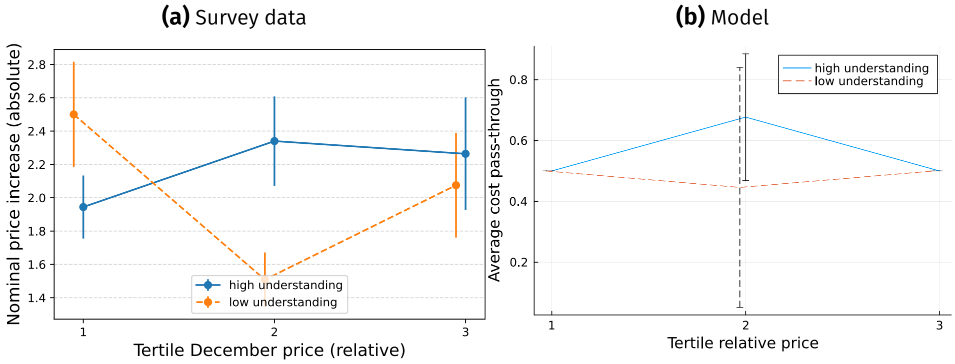
Notes: Relative price increases over the relative price distribution and by understanding-type. For the survey data, the increase is from December 2020 to March 2021. The difference between the two understanding-types is only statistically significant for the second tertile (two-sample t-test, standard errors clustered at the county level; p=2%, $N_L = 16$ (13 cluster), $N_H = 25$ (15 cluster)). The whiskers denote 68% confidence intervals (left) and 68% coverage intervals (right).

Figure 1.6. Relative price increases: data and model

tile (right panel of figure 1.6). The relative price increase declines over the relative price distribution, since for expensive firms in the upper tertile, the industry-wide cost increase by the fixed amount κ makes up a lower share of their higher baseline cost than for the cheaper firms in the lower tertile. The reason for the gap between firms who have customers with different information types is as discussed above: customers who observe the idiosyncratic costs can learn about the industry-wide cost increase, and as consequence their firms can charge their monopoly prices, while customers who do not have this information learn about the cost increase to a much lesser extent, and force their regular firms to either increase their price by less, or to exit the market.[32]

An important parameter of the model is the share of customers of the high information type, α. We estimate it from the share of firms who report a high understanding of prices by their customers in the third tertile. We do not estimate the share from all tertiles, since our model predicts that firms in the second tertile with low understanding customers disproportionally exit the market, which would downward bias our estimate. The dispersion of idiosyncratic costs, which are distributed uniformly and symmetrically around 0, is determined by the distribution's upper bound, $\overline{\zeta}$. We calibrate it so that the real upward price rigidity from having a low understanding customer is as strong as in the data. Comparing the model's coverage intervals (right panel of figure 1.6) with the confidence intervals of the mean relative price increases in the data (left panel) shows that the model-implied dispersion

32. Our model predicts that 2.9% of firms exit the market during the uncertainty period $t = 1$, while no firms exit in the periods $t = 0$ and $t = 2$.



Notes: Nominal price increase conditional on a price change in Euros (left) and cost pass-through (right) over the relative price distribution and by understanding-type. For the survey data, the increase is from December 2020 to March 2021. The difference between the two understanding-types is only statistically significant for the second tertile (two-sample t-test, standard errors clustered at the county level; p=2%, $N_L = 9$ (9 cluster), $N_H = 20$ (12 cluster)). The cost pass-through is defined as $(p_1 - p_0)/(C_1 - C_0)$. The whiskers denote 68% confidence intervals (left) and 68% coverage intervals (right).

Figure 1.7. Nominal price increases: data and model

of relative price increases is too large. This is mainly a consequence of our simplifying assumption that the idiosyncratic costs are uniformly distributed. With such a one-parameter distribution, only a large support can prevent low understanding customers from learning about the industry-wide cost increase. Allowing for a distribution with long tailes would give the model another degree of freedom to match the observed heterogeneity in price increases as well. Here, we aim to match only the average gap between firms with high and low understanding customers.

In terms of nominal prices, the model predicts an average nominal price increase of 5.5%, which is in line with the 5.7% (± 0.4%, N=281) nominal increase of the firms in our survey, even though it was not a target of the calibration exercise. The model also predicts that, as in the data[33], no firm decreases its nominal price. Nominal price stickiness is small, however: only 1.8% of firms keep their price at the same level, which is an order of magnitude lower than in the survey data. The reason is that there is not a lot of overlap on the idiosyncratic cost distribution of firms that are search-restricted in their price-setting in both periods $t = 0$ and $t = 1$. Firms with relatively high idiosyncratic costs are search-restricted in period $t = 0$, but in period $t = 1$, if the firms have low understanding customers, they either have to exit the market, or can set a monopoly price that is high enough so that their customers will learn (some) of the industry-wide cost increase.

33. One out of 282 firms in our survey reported a nominal price decrease from December 2020 to March 2021.

Figure 1.7 shows the absolute nominal price increase, conditional on a price change, of the hairdressers in our survey over the relative price distribution and by customer understanding (left panel). The absolute price increase is a scaled measure of cost pass-through under the maintained assumption that the cost increased by the same fixed amount for all firms. Under that assumption, the model matches the data qualitatively quite well (see right panel of figure 1.7). Quantitatively, we can use our survey participants' report on the hygiene surcharge as a measure for the average pass-through of the increase in marginal costs due to the hygiene measures (see table 1.1), which lies at 2.38 Euro ± 10 Eurocents. We find that firms in the extremes of the price distribution increase their prices by about the same nominal amount in absolute terms, about 2.20 Euro. This is consistent with the prediction by the model that firms with monopoly power increase their prices by the same absolute amount, as they are subject to the same cost-shock. In the model, the cost pass-through diverges for less productive firms in the middle of the price distribution for two reasons: as their customers' outside option falls with the industry-wide cost increase, firms with high understanding customers that were search-restricted in their pricing in period $t = 0$ now can set their monopoly price; hence, they increase the markup on their product. Meanwhile, firms with low understanding customers that were monopolists in period $t = 0$ are restricted by the fact that their customers do not update their beliefs about their outside option; hence, they reduce their markup by more than monopolists. In the survey data, we do not see the former effect — that firms with high understanding customers in the second tertile charge a higher markup — but we find evidence that firms with low understanding customers in the middle of the price distribution reduce their markup, given the assumption of a common cost increase.

Table 1.8 shows the effects of a change in the share of low understanding customers, α, on the average relative price changes by tertile. It is clear that with more low understanding customers, the real price rigidity affects more firms in the middle of the relative price distribution, lowering the average relative price increase of those firms compared to firms in the other tertiles. We also calculate the standard deviation of relative price changes *conditional* on price changes for each firm over time, denoted as $\sigma(\Delta p)$, as in Klenow and Willis (2016). In the context of the model, we treat the relative prices of firms in period $t = 0$ as resulting from a price change. Therefore, even for firms who do not change their prices from periods $t = 1$ to $t = 2$, the standard deviation is well defined. Instead, among high cost, low quality firms with low understanding customers, firms who are constrained in their pricing in period $t = 1$ change their price also from period $t = 1$ to $t = 2$, as period $t = 2$ is a certainty period when their customers learn about the higher industry-wide cost. Table 1.8 shows that this measure of real rigidities varies quite strongly with different incidences of customer understanding. When 90% of firms in the market have low understanding customers, we can explain 39% of the fluctuations of conditional relative price changes over time of firms in the CPI data. In comparison,

Table 1.8. Relative price changes in model and data

Source	α	Relative price changes over tertiles	$\sigma(\Delta_1 p)$	$\sigma(p_1)$
Model	0.0	(0.2%, 1.11%, −1.38%)	0.8%	11.3%
Model	0.45	(1.17%, 1.15%, −0.40%)	1.1%	11.4%
Model	0.9	(6%, 4.86%, 4.43%)	2.9%	12%
Survey (unc.)	-	(1.95% ±1.0%, 0.88% ±1.2%, -0.91% ±0.8%)	-	17.7%
Survey (con.)	-	(4.97% ±1.1%, 2.99% ±1.5%, 2.33% ±1.4%)	-	-
CPI (unc.)	-	(-2.21% ±1.4%, -4.11% ±1.4%, -5.85% ±1.7%)	-	23.1%
CPI (con.)	-	(6.55% ±2.1%, 3.33% ±1.4%, 4.03% ±1.4%)	7.5%	-

Notes: Average relative price changes over the relative price distribution and the standard deviation of conditional relative price changes in model and data. For the distribution, the position in the tuple represents the tertile-number. For the survey, the relative price distribution is from December 2020, while for the CPI-micro data, it is from March 2020. "unc." and "con." refer to unconditional relative price changes or changes conditional on firms adjusting their nominal price, respectively. $\sigma(\Delta p)$ refers to the SD of *conditional price changes* relative to the median, while $\sigma(p)$ refers to the SD of *prices* relative to the median. Standard errors of survey data are clustered at the county-level, and sample sizes are $N = 189$ (unc.) and $N = 121$ (con.). The unconditional CPI-results denote averages from March 2020 to March 2021, with sample size $N = 103$. The conditional CPI-results denote the average montly change from January 2020 to December 2021, conditional on a nominal price adjustment, and control for county fixed effects and a measure of nominal price-stickiness, with sample size $N = 86$.

the standard deviation of relative prices (not price *changes*), denoted as $\sigma(p)$, does not change that much with the share of firms with low understanding customers. This shows that the restrictedness of unproductive firms in an uncertainty period changes the relative prices of more productive firms more strongly *over time* than within the period. The reason is that the firms with low understanding customers cause the median price to fluctuate over time[34]. The fact that some firms with low understanding customers also change their prices from period $t = 1$ to $t = 2$, which can in principle increase the standard deviation of price changes further, only has a negligible impact on the result.

The model explains only about half of the relative price dispersion, $\sigma(p)$, that we see in the data. The reason is twofold: by assuming a uniform distribution of idiosyncratic costs, we abstract from prices in the tails of the empirical price distribution. At the same time, the search model is able to generate a higher price dispersion in period 0: $\sigma(p_0) = 13.1\%$. Hence, the model predicts a *decline* of price dispersion in the uncertainty period. This is the case because cheap and expensive firms in the model pass through the industry-wide cost increase at the same rate, so that the relative gap between the two extremes of the price distribution shrinks.

34. The results are much subdued if relative prices are instead calculated with respect to the mean. An increase in the share of firms with low understanding customers from $\alpha = 0.45$ to $\alpha = 0.9$ only increases $\sigma(\Delta p)$ by 17% in that case.

In the remaining rows of table 1.8, we report the distribution of relative price changes from different data sources, and measured unconditionally or conditionally on a nominal price adjustment. While we match our model to the unconditional relative price changes of firms in our survey, conditional relative price changes as proposed in the literature (Klenow and Willis, 2016) offer a way to disentangle real from nominal price stickiness. Notably, the gap between unconditional and conditional relative price changes is larger in the CPI than in the survey data. This confirms the result from section 1.2 that the respondents in our survey are subject to less nominal price stickiness than the firms sampled by the German statistical offices. Still, the pattern that the size of relative price increases falls in the initial relative price position of the firm is consistent and significant across all measures. Interpreted as the response to the common cost-increases during the Covid-19 pandemic, this empirical finding is in line with Hobijn, Nechio, and Shapiro (2021), who call it the "mean reversion of magnitude" of price changes. Our model explains this pattern for cheap and expensive firms by virtue of their identical (monopolistic) cost pass-through, and for firms in the middle of the price distribution if the share of firms with low understanding customers is sufficiently large.

1.6 Conclusion

In this paper, we make the case that customer markets with information asymmetries are an important source of price dispersion and relative price fluctuations in firm pricing. First, we conduct a survey among German hairdressers during the Covid-19 pandemic, asking about their reasons for setting their prices. Then, we take the hairdressers at their word and build a model that matches the results of our survey both qualitatively and quantitatively.

Our interpretation of the data is that the Covid-19 pandemic induced uncertainty about firms' production costs on the side of the customer. We believe that this is the reason why the firms in our survey reported that retaining their regular customers is the most important reason not to increase their prices, and it explains the importance of the customers' understanding for the firm's price setting. With the supply-chain disruptions during the pandemic and the energy crisis as a consequence of the war in the Ukraine, there is reason to believe that uncertainty about supply-side conditions can explain a sizable share of recent price dispersion in other industries as well, given our model. The fact that our survey results are in line with the numerous papers that ask managers about their pricing decisions makes us confident about the external validity of the mechanism we propose. We view as one of our main contributions to the literature that we explain the heterogeneity in price rigidity due to low customer understanding in a model with rational customers and profit-maximizing firms that fits the data quantitatively quite well.

The main factor that we do not microfound in our model is why firms have low or high understanding customers. However, we provide evidence from our survey data that firms with more employees are more likely to have low understanding customers. As a next step, we would therefore like to investigate the properties of our model in a dynamic setting with firm investment. In the context of the search model, this would necessitate to introduce directed search into the theory. Then, less productive firms would trade-off building a customer base by setting low prices (see Foster, Haltiwanger, and Syverson (2016) for empirical evidence for this) with growing large, and being possibly more at risk of future price rigidity due to low customer understanding.

Appendix 1.A Details on the Other Surveys on Price Stickiness

After Blinder et al. (1998), the Inflation Persistence Network of the European Central Bank has conducted similar surveys in many European countries: Austria (Kwapil, Baumgartner, and Scharler, 2005), Belgium (Aucremanne and Druant, 2005), France (Loupias and Ricart, 2004), Germany (Stahl, 2005), Italy (Fabiani, Gattulli, and Sabbatini, 2004), Luxembourg (Lünnemann and Mathä, 2006), the Netherlands (Hoeberichts and Stokman, 2006), Portugal (Martins, 2005), and Spain (Álvarez and Hernando, 2005). Their results are summarized in the meta study by Fabiani et al. (2006). Independent researchers have also conducted similar studies in other countries: the United Kingdom (Hall, Walsh, and Yates (2000) and Greenslade and Parker (2012)), Japan (Nakagawa, Hattori, and Takagawa, 2000), Canada (Amirault, Kwan, and Wilkinson, 2006), Sweden (Apel, Friberg, and Hallsten, 2005), Norway (Langbraaten, Nordbø, and Wulfsberg, 2008), Romania (Copaciu, Neagu, and Braun-Erdei, 2010), Estonia (Dabušinskas and Randveer, 2006), Turkey (Sahinoz and Saraçoğlu, 2008), Pakistan (Malik, Satti, and Saghir, 2008), Poland (Jankiewicz and Kolodziejczyk, 2008), Iceland (Ólafsson, Pétursdóttir, and Vignisdóttir, 2011), Lithuania (Virbickas, 2011), New Zealand (Parker, 2014), Brazil (Correa, Petrassi, and Santos, 2018), Tanzania (Kimolo, 2018), and Vietnam (Pham, Nguyen, and Nguyen, 2019).

The following tables summarize the results of these studies. Because both the selection of hypotheses and their number differ across the studies, we categorized the hypotheses in 8 categories and color-coded the rankings to make them better comparable. We also added short descriptions of the theories underlying the hypotheses. The tables list the authors of a study, where the results are published, when and where the survey was conducted, how many managers responded, what kind of survey it is, the scale on which hypotheses are rated, and their ranking. We interpreted all hypotheses that are referred to as “kinked-demand curve” as coordination failure, and we excluded the hypotheses for why prices are increased (instead of why prices are sticky) in Loupias and Ricart (2004).

Authors (Source)	Blinder, Canetti, Lebow, Rudd (1998 Monography)	Hall, Walsh, Yates (2000 Oxford Economic Papers)	Nakagawa, Hattori, Takagawa (2000 Bank of Japan Working Paper)
Country (Year of Survey)	United States (1990-1992)	United Kingdom (1995)	Japan (2000)
Responses	200	654	630
Small firms excluded?	> $10 million revenue	unclear	In First Section of Tokyo Stock Exchange
Kind of survey	structured interview	paper questionnaire after agreeing to participate	paper questionnaire
Hypotheses and Ranking	Scale: 1 to 4	Scale: 7 to 1	Scale: 5 to 1
1	Coordination failure (2.77)	Explicit contracts (2.2)	Coordination failure (2.86)
2	Cost-based pricing (2.66)	Cost-based pricing (2.3)	Implicit contracts (2.86)
3	Nonprice competition (2.58)	Coordination failure (2.5)	Explicit contracts (3.10)
4	Implicit contracts (2.40)	Pricing points (2.8)	Pricing points (3.60)
5	Explicit contracts (2.11)	Implicit contracts (2.9)	Nonprice competition (3.61)
6	Costly price adjustment (1.89)	Constant MC (3.1)	Procyclical elasticity (3.99)
7	Procyclical elasticity (1.85)	Inventories (3.1)	Menu costs (4.18)
8	Pricing points (1.76)	Nonprice competition (3.3)	Judging quality by price (4.23)
9	Constant MC (1.57)	Procyclical elasticity (3.3)	Delivery lags/service (4.35)
10	Inventories (1.56)	Judging quality by price (3.6)	
11	Hierarchy (1.41)	Physical menu costs (3.8)	
12	Judging quality by price (1.33)		
13			
14			
15			

Authors (Source)	Amirault, Kwan, Wilkinson (2006 Bank of Canada Working Paper)	Apel, Friberg, Hallsten (2005 Journal of Money, Credit and Banking)	Langbraaten, Nordbø, Wulfsberg (2008 Norges Bank Economic Bulletin)
Country (Year of Survey)	Canada (2002-2003)	Sweden (2000)	Norway (2007)
Responses	170	48.7% of 1285	725
Small firms excluded?	> 20 employees	> 5 employees	no
Kind of survey	structured interview	paper questionnaire	paper questionnaire
Hypotheses and Ranking	Scale: 0 or 1	Scale: 1 to 4	Scale: 1 to 4 (scores not reported)
1	Cost-based pricing (67.1%)	Implicit contracts (3.00)	Explicit contracts
2	Customer relations (55.3%)	Cost-based pricing and constant MC (2.45)	Coordination failure
3	Explicit contracts (45.3%)	Explicit contracts (2.27)	Customer relationship
4	Nonprice competition (44.1%)	Kinked demand curve (coordination failure) (2.17)	Pricing points
5	Coordination failure upwards (41.2%)	Countercyclical cost of finance (2.08)	Costly information Menu costs
6	Low inflation (33.5%)	Liquidity constraints (1.85)	
7	Implicit contracts (31.8%)	Pricing points (1.85)	
8	Coordination failure downwards (31.2%)	Procyclical elasticity (1.75)	
9	Factor stability (31.2%)	Deviation from collusion (1.68)	
10	Menu costs (21.2%)	Thick market (supply side) (1.60)	
11	Sticky information (13.5%)	Physical menu costs (1.54)	
12		Thick market (demand side) (1.50)	
13		Information costs (1.40)	
14			
15			

Authors (Source)	Inflation Persistence Network Meta-Study (2006 International Journal of Central Banking)	Kwapil, Scharler, Baumgartner (2005 European Central Bank Working Paper)	
Country (Year of Survey)	EU (2003-2004)	Austria (2004)	
Responses	more than 11000	873	
Small firms excluded?	differs across countries	In WIFO Business Cycle Survey	
Kind of survey	differs across countries	paper questionnaire	
Hypotheses and Ranking	Scale: 1 to 4 (unweighted averages of country scores)	Scale: 1 to 4 Increases only	Scale: 1 to 4 Decreases only
1	Implicit contracts (2.7)	Implicit contracts (3.04)	Implicit contracts (3.04)
2	Explicit contracts (2.6)	Explicit contracts (3.02)	Explicit contracts (2.94)
3	Cost-based pricing (2.6)	Cost-based pricing (2.72)	Kinked demand curve (2.69)
4	Coordination failure (2.4)	Kinked demand curve (2.69)	Cost-based pricing (2.49)
5	Judging quality by price (2.1)	Coordination failure (2.47)	Coordination failure (2.13)
6	Temporary shocks (2.0)	Information costs (1.61)	Nonprice competition (1.98)
7	Nonprice competition (1.7)	Menu costs (1.52)	Judging quality by price (1.88)
8	Menu costs (1.6)	Nonprice competition (1.49)	Temporary shocks (1.470
9	Costly information (1.6)	Temporary shocks (1.42)	Information costs (1.61)
10	Pricing points (1.6)	Pricing points (1.32)	Menu costs (1.52)
11		Judging quality by price (not applicable)	Pricing points (1.24)
12			
13			
14			
15			

Authors (Source)	Aucremanne, Druant (2005 European Central Bank Working Paper)	Loupias, Ricart (2004 European Central Bank Working Paper)	
Country (Year of Survey)	Belgium (2004)	France (2003-2004)	
Responses	1979	1662	
Small firms excluded?	In monthly survey of the National Bank of Belgium	unclear	
Kind of survey	paper questionnaire	face-to-face, phone, mail	
Hypotheses and Ranking	Scale: 1 to 4	Scale: 1 to 4 Increases only	Scale: 1 to 4 Decreases only
1	Implicit contracts (2.5)	Cost-based pricing (3.0 commodity prices) (2.5 labor cost) (1.8 productivity)	Explicit contracts (2.5)
2	Explicit contracts (2.4)	coordination failure (3.0 others don’t change) (2.3 match others’ prices)	Cost-based pricing (2.6 commodity prices) (1.9 labor cost) (2.2 productivity)
3	Cost-based pricing and constant MC (2.4)	Explicit contracts (2.7)	coordination failure (2.8 match others’ prices) (2.1 others don’t change)
4	Liquidity constraints (2.2)	Implicit contracts (2.2)	Negative demand shock (2.3)
5	Kinked demand curve (2.2)	Temporary shocks (2.1)	Temporary shocks (2.1)
6	Procyclical elasticity (2.1)	Positive demand shock (2.0)	Implicit contracts (2.0)
7	Thick market (demand side) (2.0)	Fewer competitors (1.8)	More competitors (2.0)
8	Judging quality by price (1.9)	Pricing points (1.7)	Pricing points (1.6)
9	Thick market (supply side) (1.8)	Inventory + delay (1.4)	Inventory + delay (1.6)
10	Temporary shock (1.8)	Physical menu costs (1.4)	Physical menu costs (1.4)
11	Nonprice competition (1.7)		
12	Countercyclical cost of finance (1.7)		
13	Pricing points (1.7)		
14	Information costs + bureaucracy (1.6)		
15	Menu costs (1.5)		

Authors (Source)	Fabiani, Gattulli, Sabbatini (2004 European Central Bank Working Paper)	Stahl (2005 European Central Bank Working Paper)	
Country (Year of Survey)	Italy (2003)	Germany (2004)	
Responses	333	1200	
Small firms excluded?	> 50 employees	In survey of Ifo institute	
Kind of survey	paper questionnaire	paper questionnaire	
Hypotheses and Ranking	Scale: 1 to 4	Scale: 1 to 4 Increases only	Scale: 1 to 4 Decreases only!
1	Explicit contracts (2.64)	Coordination failure (2.6)	Explicit contracts (2.4)
2	Coordination failure (2.59)	Explicit contracts (2.4)	Coordination failure (2.2)
3	Temporary shocks (1.97)	High elasticity for increases (2.1)	Low elasticity for decreases (2.1)
4	Menu costs (1.58)	Regular date (2.0)	Temporary shock (2.0)
5	Pricing points (1.43)	Regular interval (1.9)	Regular date (2.0)
6	Bureaucratic costs 1.30	Temporary shock (1.8)	Regular interval (1.9)
7		Sluggish costs (1.8)	Sluggish costs (1.8)
8		Menu costs (1.4)	Menu costs (1.4)
9		„Other“ (1.1)	„Other“ (1.1)
10			
11			
12			
13			
14			
15			

Authors (Source)	Lünnemann, Mathä (2006 European Central Bank Working Paper)	Alvarez, Hernando (2005 European Central Bank Working Paper)	
Country (Year of Survey)	Luxemburg (2004)	Spain (2004)	
Responses	367	2008	
Small firms excluded?	> 5 employees	> 5 employees	
Kind of survey	paper questionnaire	paper questionnaire	
Hypotheses and Ranking	Scale: 1 to 4 (scores not reported)	Scale: 1 to 4 Increases only	Scale: 1 to 4 Decreases only
1	Implicit contracts	Implicit contracts (2.56)	Coordination failure (2.21)
2	Constant MC	Coordination failure (2.42)	Explicit contracts (2.09)
3	Explicit contracts	Explicit contracts (2.25)	Temporary shocks (1.82)
4	Procyclical elasticity	Temporary shocks (1.82)	Judging quality by price (1.82)
5	Thick markets (demand)	Pricing points (1.49)	Pricing points (1.42)
6	Liquidity constraints	Menu costs (1.43)	Menu costs (1.39)
7	Judging quality by price	Nonprice competition (1.34)	Nonprice competition (1.34)
8	Thick markets (supply)	Costly information (1.33)	Costly information (1.30)
9	Coordination failure	Judging quality by price (not applicable)	Implicit contracts (not asked)
10	Pricing points		
11	Temporary shock		
12	Countercyclical cost of finance		
13	Menu cost		
14	Nonprice competition		
15	Costly information		

Authors (Source)	Hoeberichts, Stokman (2006 European Central Bank Working Paper)	Martins (2005 European Central Bank Working Paper)	Copaciu, Neagu, Braun-Erdei (2010 Managerial and Decision Economics)
Country (Year of Survey)	Netherlands (2004)	Portugal (2004)	Romania (2006)
Responses	1246	1370	377
Small firms excluded?	no	> 20 employees	> 10 employees
Kind of survey	email	paper questionnaire	unclear
Hypotheses and Ranking	Scale: 1 to 4	Scale: 1 to 4	Scale: 1 to 4
1	Implicit contracts (2.66)	Implicit contracts (3.14)	Implicit contracts (3.12)
2	Explicit contracts (2.57)	Coordination failure (2.84)	Explicit contracts (3.10)
3	Judging quality by price (2.34)	High fixed costs (=liquidity constraints) (2.80)	Judging quality by price (2.19)
4	Temporary shocks (2.34)	Constant MC (2.70)	Price readjustments (2.15)
5	Coordination failure (2.22)	Explicit contracts (2.63)	Coordination failure (1.97)
6	Nonprice competition (2.07)	Procyclical elasticity (2.61)	Costly information (1.74)
7	Pricing points (1.80)	Temporary shock (2.46)	Menu costs (1.62)
8	Menu costs (1.71)	Bureaucratic delays (2.45)	
9		Judging quality by price (2.28)	
10		Menu costs (1.89)	
11		Pricing points (1.78)	
12		Costly information (1.70)	
13			
14			
15			

Authors (Source)	Dabušinskas, Randveer (2006 Bank of Estonia Working Paper)		Sahinoz, Saraçoğlu (2008 Developing Economies)
Country (Year of Survey)	Estonia (2005)		Turkey (2005)
Responses	208		999
Small firms excluded?	no		unclear
Kind of survey	internet survey		unclear
Hypotheses and Ranking	Scale: 1 to 4 Increases only (scores not reported)	Scale: 1 to 4 Decreases only (scores not reported)	Scale: 0 to 3 and *100
1	Implicit contracts	Cost-based pricing	Constant markup (44.8)
2	Explicit contracts	Implicit contracts	Temporary shocks (40.6)
3	Cost-based pricing	Judging quality by price	Explicit contracts (37.1)
4	Coordination failure	Coordination failure	Implicit contracts (36.9)
5	Pricing points	Explicit contracts	Coordination failure (30.8)
6	Nonprice competition	Nonprice competition	Constant MC (22.6)
7	Costly information	Pricing points	
8	Menu costs	Menu costs	
9		Costly information	
10			
11			
12			
13			
14			
15			

Authors (Source)	Greenslade, Parker (2012 Economic Journal)		Malik, Satti, Saghir (2008 Pakistan Development Review)
Country (Year of Survey)	United Kingdom (unclear)		Pakistan (2008)
Responses	693		343
Small firms excluded?	unclear		> 10 employees
Kind of survey	unclear		structured interview
Hypotheses and Ranking	Scale: 0 to 1 Increases only	Scale: 0 to 1 Decreases only	Scale: 1 to 4
1	Coordination failure (60%)	Coordination failure (35%)	Implicit contracts (framed as „customers prefer stable prices“) (2.66)
2	It would anger customers (56%)	Explicit contracts (35%)	Explicit contracts (2.41)
3	Explicit contracts (47%)	Implicit contracts (29%)	Coordination failure (2.35)
4	Implicit contracts (38%)	Temporary shocks (28%)	Temporary shocks (1.84)
5	Temporary shocks (32%)	Constant MC (26%)	Judging quality by price (1.84)
6	Constant MC (31%)	It would anger customers (25%)	Costly information (1.62)
7	Pricing points (24%)	Pricing points (15%)	Menu costs (1.59)
8	Menu costs (10%)	Menu costs (9%)	
9			
10			
11			
12			
13			
14			
15			

Authors (Source)	Jankiewicz, Kolodziejczyk (2008 Bank i Kredyt)		Ólafsson, Pétursdóttir, Vignisdóttir (2011 Central Bank of Iceland Working Paper)
Country (Year of Survey)	Poland (2006)		Iceland (2008)
Responses	752		580
Small firms excluded?	unclear		> 3 employees
Kind of survey	unclear		structured interview
Hypotheses and Ranking	Frequency top two answers Increases only	Frequency top two answers Decreases only	Assign 100 to one and 50 to another hypothesis (some not reported)
1	Coordination failure (53.4%)	Temporary shocks (33.5%)	Implicit contracts (34.1)
2	Explicit contracts (40.5%)	Explicit contracts (30.8%)	Explicit contracts (31.0)
3	Temporary shocks (22.0%)	None (29.0%)	Temporary shocks (28.8)
4	None (17.1%)	Other (22.1%)	Coordination failure (26.1)
5	Other (15.7%)	Judging quality by price (19.1%)	Pricing points (15.0)
6	Formal and legal difficulties (7.3%)	Pricing points (5.3%)	Menu costs
7	Pricing points (misreported) 3.8??	Formal and legal difficulties (3.2%)	Nonprice competition
8	Menu costs (1.4%)	Menu costs (2.0%)	Judging quality by price
9			
10			
11			
12			
13			
14			
15			

Authors (Source)	Virbickas (2011 Bank of Lithuania Working Paper)		Parker (2014 Reserve Bank of New Zealand Discussion Paper)
Country (Year of Survey)	Lithuania (2008)		New Zealand (2010)
Responses	343		5369
Small firms excluded?	> 5 employees		> NZD 30,000 revenue and > 5 employees
Kind of survey	unclear		unclear
Hypotheses and Ranking	Scale: 1 to 4 and then frequency of 3 and 4 Increases only	Scale: 1 to 4 and then frequency of 3 and 4 Decreases only	Scale: 0 to 1 (scores not reported)
1	Cost-based pricing (74.2%)	Cost-based pricing (61.7%)	Explicit contracts
2	Explicit contracts (63.2%)	Explicit contracts (51.1%)	Implicit contracts
3	Implicit contracts (50.9%)	Temporary shocks (50.9%)	Coordination failure
4	Coordination failure (41.1%)	Judging quality by price (48.1%)	Temporary shocks
5	Costly information (40.5%)	Coordination failure (37.8%)	Pricing points
6	Temporary shocks (33.4%)	Costly information (30.2%)	Nonprice competition
7	Pricing points (21.5%)	Nonprice competition (27.4%)	Menu costs
8	Nonprice competition (18.3%)	Pricing points (19.6%)	
9	Menu costs (17.0%)	Menu costs (16.4%)	
10			
11			
12			
13			
14			
15			

Authors (Source)	Correa, Petrassi, Santos (2018 Journal of Business Cycle Research)	Kimolo (2018 Journal of Economics and Sustainable Development)	Pham, Nguyen, Nguyen (2019 working paper)
Country (Year of Survey)	Brazil (2011-2012)	Tanzania (2014)	Vietnam (2014)
Responses	7002	79	1560
Small firms excluded?	unclear	> 10 employees and > 7 years old	unclear
Kind of survey	unclear	structured interview	unclear
Hypotheses and Ranking	Unclear (Authors' ranking)	Scale: 5 to 1	not reported
1	Menu cost and costly information (46.7%)	Implicit contracts (2.00)	
2	Cost-based pricing (79.4%)	Explicit contracts (2.70)	
3	Explicit contracts (20%[sic])	Pricing points (2.94)	
4	Implicit contracts and not antagonizing customers (79%)	Judging quality by price (3.14)	
5	Coordination failure (67%)	Coordination failure (3.28)	
6	Non-price competition (75% - 54%)	Nonprice competition (3.33)	
7	Judging quality by price (38.3%)	Menu Costs (3.53)	
8		Temporary shocks (3.68)	
9			
10			
11			
12			
13			
14			
15			

List of Categories and Their Hypotheses

	There is no reason to change the prices	
1	Constant MC	The supply is perfectly elastic in the relevant range.
2	Factor stability	Nothing changes, so there is no reason to change the prices.
3	Low inflation	The price leves does not change, so there is no reason to change the prices.
	Rules (of thumb) how prices are set	
4	Regular date	Prices are only changed on specific dates.
5	Regular interval	Prices are only changed after specific (potentially stochastic) intervals
6	Cost-based pricing	Price = Piece cost + markup
7	Pricing points	Exploit the leading digit bias of the consumers (e.g. let prices end in .99).
8	Constant markup	Change the price only if the (real) mark-up falls out of a pre-specified range.
9	Explicit contracts	Long-term contracts with customers fix the prices (potentially pegged to inflation measures).
	Customer goodwill	
10	Implicit contracts	Invisible handshake: Customers want stable prices to reduce uncertainty and to be regular customers.
11	Customer relations	Don't want to lose customers' goodwill.
	Market environment changes in cycles	
12	Countercyclical cost of finance	In recessions, financing costs are larger, so prices are not reduced.
13	Liquidity constraints	Firms have to recoup their fixed costs, so they cannot reduce prices too much in recessions.
14	Procyclical elasticity	The mark-up changes over the cycle because the elasticity changes (e.g. in recessions only loyal people buy).
15	Thick market (demand side)	In booms, people buy more, so searching for cheaper prices becomes worthwhile.
16	Thick market (supply side)	In booms, firms can reach customers easier and get more demand by not increasing their prices.
	Competition	
17	Coordination failure (upwards)	The first firm to increase the price gets punished by the customers' leaving.
18	Coordination failure (downwards)	Decreasing the price starts a price war.
19	Deviation from collusion	Decreasing the price breaks the collusion and leads to punishment.
	Adjustment costs	
20	(Physical) Menu costs	Changing the price incurs costs directly.
21	Costly information	Gathering information and making decisions is costly.
22	Hierarchy	Within the firm, consensus has to be reached.
23	Formal and legal difficulties	Price changes might have to be justified.
24	Temporary shocks	To save adjustment costs, the price might not be changed if the optimal price will revert soon.

	Adjust other things than price	
25	Nonprice competition	Change other things than the price. E.g. increase delivery lags instead of increasing the price.
26	Inventories	Keep a stock to satisfy excess demand and build up a stock if demand is low.
	Asymmetric information	
27	Judging quality by price	If the price goes down, people think that the quality went down.

Appendix 1.B Questionnaire

The following is the translation of our survey into English. Below the translation is the German original.

English Translation of Our Questionnaire

Page 1
Dear Sir or Madam,

on March 1, you were finally allowed to open up again. For our dissertations in economics at the University of Bonn, we investigate how the pandemic and the lockdown in Germany affect the hairdressers and the prices for haircuts.

We kindly ask you to take 10 to 15 minutes to fill out our survey. If you have less time at your proposal, we would also be happy for partially filled out forms (all answers are optional). You can also save your progress and continue the survey later; to do so, please click on "save progress" on the bottom of the page.

The survey is anonymous. We do not ask for or save any personal data. Your answers will be treated confidentially and only used for scientific purposes.

Thank you very much for your support!

Thomas Kohler and Maximilian Weiß

Page 2

First, we would like to get to know you and your firm better.

1. What is your role in the firm?
() I am the owner.
() I am a franchisee.
() I am an employed manager.
() I am an employee.
() Other: [free text field]
() not applicable

2. Are you involved in the pricing in your firm?
() Yes, I set the prices.
() Yes, I suggest prices to my superior.
() Yes, I set the prices in accordance with my franchising contract.
() Yes, my associates and I set the prices together.
() No

() Other: [free text field]

3. How many branches does your firm have? (In case of franchises, please for the franchisee)
() no branch (mobile hairdresser)
() one branch
() two branches
() three to five branches
() more than five branches
() can't or won't say

4. How many employees does your firm have? (In case of franchises, please for the franchisee)
() none
() one to three
() three to six
() more than six
() can't or won't say
Comment: [free text field]

5. Which share of your customers are regulars?
() 0 % to 19 %
() 20 % to 39 %
() 40 % to 59 %
() 60 % to 79 %
() 80 % to 100 %
() can't or won't say

Page 3

On this page, we'll ask you some questions about the price of a man's haircut in your firm. If you do not offer this haircut, please indicate so (You will then receive questions about the price of a woman's haircut).

6. What is the price of the following man's haircut in your firm?
short back and sides, wash, cut, blow dry, 25 minutes

Please fill in the price including a possible hygiene surcharge.
Please fill in the base price if you charge other surcharges (e.g. for Mondays, late appointments, new customers, or other).

Before this lockdown (until December 16, 2020): [free text field] Euros
() can’t or won’t say

First week of March 2021: [free text field] Euros
() can’t or won’t say

[Planned] April 2021: [free text field] Euros
() can’t or won’t say

() I don’t offer this kind of haircut (in this case, please indicate "can’t or won’t say" everywhere in this question, ignore the rest of the page, and click on "Continue").

7. Had you lowered your prices because of the VAT reduction in the second half-year of 2020?
() yes
() no
() can’t or won’t say

8. Pricing parts (begin of March 2021)
If the price you filled in (for begin of March 2021) contains a hygiene surcharge, please indicate what it is. If you charge different hygiene surcharges for different services, please indicate the hygiene surcharge for the haircut described above.
If new customers pay more than regular customers, please indicate the price difference.
If you charge a surcharge for late appointments, Monday appointments or weekend appointments, please indicate the surcharge.

hygiene surcharge: [free text field] Euros
new customer surcharge: [free text field] Euros
surcharge for late appointments: [free text field] Euros
surcharge for Monday appointments: [free text field] Euros
surcharge for weekend appointments: [free text field] Euros
() can’t or won’t say

9. Do you make more or less profit per customer with the haircut described above compared to before the pandemic (February 2020)?
() today less
() same
() today more
() can’t or won’t say

10. Do you make more or less profit per customer with the haircut described above compared to before the last lockdown (December 2020)?
() today less
() same
() today more
() can't or won't say

Page 4 [only if indicated that the reference man's haircut is not offered]

On this page, we'll ask you some questions about the price of a woman's haircut in your firm.

11. What is the price of the following woman's haircut in your firm?
Length is to the shoulders; wash, cut, brush, blow dry. Total time around 45 minutes. No dying or highlights or similar.

Please fill in the price including a possible hygiene surcharge.
Please fill in the base price if you charge other surcharges (e.g. for Mondays, late appointments, new customers, or other).

Before this lockdown (until December 16, 2020): [free text field] Euros
() can't or won't say

First week of March 2021: [free text field] Euros
() can't or won't say

[Planned] April 2021: [free text field] Euros
() can't or won't say

12. Had you lowered your prices because of the VAT reduction in the second half-year of 2020?
() yes
() no
() can't or won't say

13. Pricing parts (begin of March 2021)
If the price you filled in (for begin of March 2021) contains a hygiene surcharge, please indicate what it is.
If you charge different hygiene surcharges for different services, please indicate the hygiene surcharge for the haircut described above.
If new customers pay more than regular customers, please indicate the price difference.

If you charge a surcharge for late appointments, Monday appointments or weekend appointments, please indicate the surcharge.

hygiene surcharge: [free text field] Euros
new customer surcharge: [free text field] Euros
surcharge for late appointments: [free text field] Euros
surcharge for Monday appointments: [free text field] Euros
surcharge for weekend appointments: [free text field] Euros
() can't or won't say

14. Do you make more or less profit per customer with the haircut described above compared to before the pandemic (February 2020)?
() today less
() same
() today more
() can't or won't say

15. Do you make more or less profit per customer with the haircut described above compared to before the last lockdown (December 2020)?
() today less
() same
() today more
() can't or won't say

Page 5 [only if the indicated price for March strictly larger than the price for December]

16. Why have you increased your prices since December?
You have indicated that at least one of your prices was larger in March 2021 than in December 2020. Which role did the following factors play in your increasing the prices?
Reduced capacity due to distancing rules
() no role
() a small role
() a big role
() does not apply
() can't or won't say

Recoup lost revenue / reduced reserves due to lockdown
() no role
() a small role
() a big role

() does not apply
() can’t or won’t say

Increased demand
() no role
() a small role
() a big role
() does not apply
() can’t or won’t say

Increased financing cost (for example because of new loans)
() no role
() a small role
() a big role
() does not apply
() can’t or won’t say

Adjustment to the general price level
() no role
() a small role
() a big role
() does not apply
() can’t or won’t say

Increased wage cost
() no role
() a small role
() a big role
() does not apply
() can’t or won’t say

The price increase is only temporary
() no role
() a small role
() a big role
() does not apply
() can’t or won’t say

Increased incidental cost
() no role
() a small role
() a big role

() does not apply
() can’t or won’t say

Increased hygiene cost (masks, disinfection, time)
() no role
() a small role
() a big role
() does not apply
() can’t or won’t say

Expectation that the customers understand the price increases
() no role
() a small role
() a big role
() does not apply
() can’t or won’t say

Competitors have increased their prices
() no role
() a small role
() a big role
() does not apply
() can’t or won’t say

End of the VAT reduction
() no role
() a small role
() a big role
() does not apply
() can’t or won’t say

Other important factors:
[free text field]
[free text field]
[free text field]

17. To what extent do you agree with these statements about your experiences with your customers?

The customers express understanding for my/our prices.
() totally disagree
() somewhat disagree

() undecided
() somewhat agree
() totally agree
() can’t or won’t say

The customers complain to me about their own financial situation.
() totally disagree
() somewhat disagree
() undecided
() somewhat agree
() totally agree
() can’t or won’t say

Some customers accuse me of profiteering.
() totally disagree
() somewhat disagree
() undecided
() somewhat agree
() totally agree
() can’t or won’t say

The customers tip more.
() totally disagree
() somewhat disagree
() undecided
() somewhat agree
() totally agree
() can’t or won’t say

The customers tip less.
() totally disagree
() somewhat disagree
() undecided
() somewhat agree
() totally agree
() can’t or won’t say

page 6 [only if the indicated price for March is not larger than the price for December]

18. Why have you not increased your prices since last December?

You have indicated that at least one of your prices is not larger in March 2021 than in December 2020.
Which role did the following factors play in your decision to not increase the price?
The prices are contracted [in the ranking table: prices contracted]
() no role
() a small role
() a big role
() does not apply
() can't or won't say

Within the firm, we could not agree on a price increase [in the ranking table: could not agree on increase]
() no role
() a small role
() a big role
() does not apply
() can't or won't say

I am not sure whether increased prices would be better for the firm [in the ranking table: unsure about increasing]
() no role
() a small role
() a big role
() does not apply
() can't or won't say

A price increase would seem larger than it actually is [in the ranking table: pricing points]
() no role
() a small role
() a big role
() does not apply
() can't or won't say

Increase the market share / gain new customers [in the ranking table: gain new customers]
() no role
() a small role
() a big role
() does not apply
() can't or won't say

The prices were already increased after the first lockdown (spring 2020) [not in the ranking table]
() no role
() a small role
() a big role
() does not apply
() can't or won't say

The customers' budgets are smaller during the pandemic [in the ranking table: customers' budgets smaller]
() no role
() a small role
() a big role
() does not apply
() can't or won't say

VAT reduction was not passed on in the second half-year of 2020 [in the ranking table: not passed on VAT reduction]
() no role
() a small role
() a big role
() does not apply
() can't or won't say

The competitors have not increased their prices [in the ranking table: competitors' prices not up]
() no role
() a small role
() a big role
() does not apply
() can't or won't say

The prices were not increased, so they don't have to be decreased again soon [in the ranking table: avoid temporary increase]
() no role
() a small role
() a big role
() does not apply
() can't or won't say

The costs have not increased [in the ranking table: cost not increased]
() no role

() a small role
() a big role
() does not apply
() can't or won't say

Retaining regular customers [in the ranking table: retain regular customers]
() no role
() a small role
() a big role
() does not apply
() can't or won't say

Other important factors:
[free text field]
[free text field]
[free text field]

19. To what extent do you agree with these statements about your experiences with your customers?

The customers express understanding for my/our prices.
() totally disagree
() somewhat disagree
() undecided
() somewhat agree
() totally agree
() can't or won't say

The customers complain to me about their own financial situation.
() totally disagree
() somewhat disagree
() undecided
() somewhat agree
() totally agree
() can't or won't say

Some customers accuse me of profiteering.
() totally disagree
() somewhat disagree
() undecided
() somewhat agree
() totally agree

() can't or won't say

The customers tip more.
() totally disagree
() somewhat disagree
() undecided
() somewhat agree
() totally agree
() can't or won't say

The customers tip less.
() totally disagree
() somewhat disagree
() undecided
() somewhat agree
() totally agree
() can't or won't say

Page 7

On this page we ask you questions about how your company is dealing with the political measures and how you assess future developments.
20. If you received more requests for appointments for the beginning of March than you could satisfy: how did you deal with it?

Multiple answers are possible.

[ ] preferential treatment of new customers
[ ] hire more employees to offer more appointments
[ ] preferential treatment of customers that had appointments canceled in the past months
[ ] preferential treatment of regular customers
[ ] first come, first served
[ ] extend the opening hours to offer more appointments
[ ] charge a surcharge for new customers
() does not apply
() can't or won't say

21. To what extent do you agree with these statements about the mandate to wash the customers' hair?

I feel safer when I wash the customers’ hair before the treatment.
() totally disagree
() somewhat disagree
() undecided
() somewhat agree
() totally agree
() can’t or won’t say

The mandatory hair washing is like a price increase.
() totally disagree
() somewhat disagree
() undecided
() somewhat agree
() totally agree
() can’t or won’t say

The customers find the mandatory hair washing acceptable.
() totally disagree
() somewhat disagree
() undecided
() somewhat agree
() totally agree
() can’t or won’t say

I profit from the mandatory hair washing.
() totally disagree
() somewhat disagree
() undecided
() somewhat agree
() totally agree
() can’t or won’t say

22. How accurate do you think the following predictions are?

We will be back to normal in one year.
() not at all
() rather not
() unclear
() rather
() very
() can’t or won’t say

The hygiene measures will stay for years.
() not at all
() rather not
() unclear
() rather
() very
() can’t or won’t say

Fear will deter customers for a long time.
() not at all
() rather not
() unclear
() rather
() very
() can’t or won’t say

My personal financial situation will improve (compared to today).
() not at all
() rather not
() unclear
() rather
() very
() can’t or won’t say

Due to (fighting) the pandemic, the customers’ willingness to pay will lastingly decrease.
() not at all
() rather not
() unclear
() rather
() very
() can’t or won’t say

There will be another lockdown this year.
() not at all
() rather not
() unclear
() rather
() very
() can’t or won’t say

23. How unsure are you about your own professional future?
() not at all
() barely
() somewhat
() a lot
() can’t or won’t say

Page 8
On this page, we ask general questions about pricing in your firm.

24. In general, what do you pay most attention to when setting prices?
Multiple answers are possible.

[ ] Costs
[ ] The competitors’ prices
[ ] The quality of my offer
[ ] Customer satisfaction
[ ] Adjustment to the general price level
[ ] Something else: [free text field]
() can’t or won’t say

25. To what extent do you agree with these statements about your pricing?

I am satisfied with my pricing method.
() totally disagree
() somewhat disagree
() undecided
() somewhat agree
() totally agree
() can’t or won’t say

My prices are optimal for the firm.
() totally disagree
() somewhat disagree
() undecided
() somewhat agree
() totally agree
() can’t or won’t say

Actually, my prices should be higher.
() totally disagree
() somewhat disagree

() undecided
() somewhat agree
() totally agree
() can't or won't say

The reasons for price increases are understandable for customers.
() totally disagree
() somewhat disagree
() undecided
() somewhat agree
() totally agree
() can't or won't say

Page 9
Thank you very much for participating in our study!
26. If you want to tell us anything, you can do so anonymously here (note: this answer will be saved together with the other answers, but without any personal information).
If you have a question that you would like an answer to, please feel free to email us.
[free text field]

Last page

Thank you again for participating!
Your answers have been saved, you may close the browser window now.

oeffnung_der_friseure → base 19.04.2021, 11:38

Seite 01

Startseite

Sehr geehrte Damen und Herren,

am 01. März durften Sie endlich wieder öffnen. Im Rahmen unserer Doktorarbeiten in VWL an der Universität Bonn untersuchen wir, wie sich die Pandemie und der Lockdown in Deutschland auf die Friseur/innen und die Preise für Haarschnitte auswirken.

Wir bitten Sie, sich 10 bis 15 Minuten Zeit zu nehmen, um unseren Fragebogen auszufüllen. Sollten Sie weniger Zeit zur Verfügung haben, würden wir uns auch über teilweise ausgefüllte Bögen freuen (alle Antworten sind optional). Sie können auch Ihren zwischenzeitlichen Fortschritt abspeichern und die Befragung zu einem späteren Zeitpunkt an der Stelle fortsetzen; dazu klicken Sie bitte auf "Fortschritt speichern" am unteren Rand der Seite.

Die Befragung ist anonym. Es werden keinerlei personenbezogene Daten erhoben oder gespeichert. Ihre Angaben werden vertraulich behandelt und nur für wissenschaftliche Zwecke verwendet.

Herzlichen Dank für Ihre Unterstützung!
Thomas Kohler und Maximilian Weiß

PHP-Code

```
$pageNr = 1;
replace('%ownPageNumber%',$pageNr);
option('progress',round(100*$pageNr/7));
option('progress.last','KO');
```

Seite 02

PHP-Code

```
$pageNr = 2;
replace('%ownPageNumber%',$pageNr);
option('progress',round(100*$pageNr/7));
```

Teil 1 Allgemein

Zunächst möchten wir etwas über Sie und Ihr Unternehmen erfahren.

AI03

1. Was ist Ihre Rolle in Ihrem Unternehmen?

- ○ Ich bin der/die Besitzer/in
- ○ Ich bin Franchise- oder Lizenznehmer/in
- ○ Ich bin angestelle/r Betriebsleiter/in
- ○ Ich bin Angestellte/r
- ○ Anderes: ______

- ○ Nicht zutreffend

AI02

2. Sind Sie an der Preissetzung in Ihrem Unternehmen beteiligt?

- ○ Ja, ich bestimme die Preise selbst
- ○ Ja, ich schlage meiner/m Vorgesetzten Preise vor
- ○ Ja, ich wähle die Preise im Rahmen meines Franchise-Vertrags
- ○ Ja, mein/e Geschäftspartner/in und ich wählen die Preise gemeinsam
- ○ Nein
- ○ Anderes: ______

AI04

3. Wie viele Filialen hat Ihr Unternehmen? (Bei Franchises bitte für das Franchise-nehmende Unternehmen)

- ○ keine Filiale (mobiler Friseur)
- ○ eine Filiale
- ○ zwei Filialen
- ○ drei bis fünf Filialen
- ○ mehr als fünf Filialen

- ○ Kann / Möchte ich nicht sagen

AI05

4. Wie viele Angestellte hat Ihr Unternehmen? (Bei Franchises bitte für das Franchise-nehmende Unternehmen)

- ○ keine
- ○ eine/n bis drei
- ○ drei bis sechs
- ○ mehr als sechs

- ○ Kann / Möchte ich nicht sagen

AI08

Anmerkung:

AI01

5. Welcher Anteil Ihrer Kunden sind Stammkunden?

- ○ 0 % bis 19 %
- ○ 20 % bis 39 %
- ○ 40 % bis 59 %
- ○ 60 % bis 79 %
- ○ 80 % bis 100 %

- ○ Kann / Möchte ich nicht sagen

Seite 03

PHP-Code

```
$pageNr = 3;
replace('%ownPageNumber%',$pageNr);
option('progress',round(100*$pageNr/7));
```

Teil 2 Preise Haarschnitt 1

Auf dieser Seite stellen wir Ihnen einige Fragen zum Preis eines Herren-Haarschnitts in Ihrem Unternehmen. Falls Sie diesen Haarschnitt nicht anbieten, markieren Sie dies bitte (Sie erhalten dann Fragen zum Preis eines Damen-Haarschnitts).

PL01

6. Wie viel kostet der folgende Herren-Haarschnitt in Ihrem Unternehmen?

Klassischer Fassonschnitt. Waschen, Schneiden, Föhnen. Gesamtdauer etwa 25 Minuten.

Bitte geben Sie den Preis inklusive einer eventuellen Hygienepauschale an.

Bitte geben Sie den Grundpreis an, falls Sie andere Zuschläge (z.B. montags, späte Termine, für Neukunden oder ähnliches) erheben.

Vor diesem Lockdown (bis zum 16. Dezember 2020)		Euro	☐ Kann / Möchte ich nicht sagen
Erste Märzwoche 2021		Euro	☐ Kann / Möchte ich nicht sagen
April 2021		Euro	☐ Kann / Möchte ich nicht sagen

PL14

☐ Ich biete diese Art Haarschnitt nicht an (Bitte kreuzen Sie in diesem Fall bei dieser Frage überall „Kann ich nicht sagen" an und ignorieren Sie bitte den Rest dieser Seite und klicken auf „Weiter".)

PL16

7. Hatten Sie aufgrund der Mehrwertsteuersenkung im zweiten Halbjahr 2020 Ihre Preise gesenkt?

○	○	○
ja	nein	Kann / Möchte ich nicht sagen

PL05

8. Preisbestandteile (Anfang März 2021)

Falls der angegebene Preis (Anfang März 2021) eine Hygienepauschale beinhaltet, geben Sie bitte an, wie hoch diese ist. Falls Sie eine unterschiedlich hohe Hygienezuschläge für unterschiedliche Dienstleistungen erheben, geben Sie bitte den Hygienezuschlag für den oben beschriebenen Haarschnitt an.

Falls Neukunden mehr zahlen als Stammkunden, geben Sie bitte den Preisunterschied an.

Falls Sie einen Zuschlag für späte Termine, für Termine am Montag oder für Termine am Wochenende erheben, geben Sie bitte die Höhe des Zuschlags an.

☐ Hygienepauschale: ____ Euro

☐ Neukunden-Zuschlag: ____ Euro

☐ Zuschlag für späte Termine: ____ Euro

☐ Zuschlag für Termine am Montag: ____ Euro

☐ Zuschlag für Termine am Wochenende: ____ Euro

☐ Kann / Möchte ich nicht sagen

9. Machen Sie mit dem oben beschriebenen Haarschnitt pro Kunde heute mehr oder weniger Gewinn als vor der Pandemie (Februar 2020)?

heute weniger	gleich viel	heute mehr	Kann / Möchte ich nicht sagen
○	○	○	○

PL10

10. Machen Sie mit dem oben beschriebenen Haarschnitt pro Kunde heute mehr oder weniger Gewinn als vor dem letzten Lockdown (Dezember 2020)?

heute weniger	gleich viel	heute mehr	Kann / Möchte ich nicht sagen
○	○	○	○

Seite 04

PHP-Code

```
if (value('PL14_01')==1){
goToPage('PH');
}
$pageNr = 3;
replace('%ownPageNumber%',$pageNr);
option('progress',round(100*$pageNr/7));
```

Teil 2 Preise Haarschnitt 2

Auf dieser Seite stellen wir Ihnen einige Fragen zum Preis eines Damen-Haarschnitts in Ihrem Unternehmen.

PL02

11. Wie viel kostet der folgende Damen-Haarschnitt in Ihrem Unternehmen?

Haarlänge: etwa schulterlang

Waschen, Schneiden, Kämmen, Föhnen. Gesamtdauer etwa 45 Minuten.

Keine Farbe, Strähnchen oder ähnliches.

Bitte geben Sie den Preis inklusive einer eventuellen Hygienepauschale an.

Bitte geben Sie den Grundpreis an, falls Sie andere Zuschläge (z.B. montags, späte Termine, für Neukunden oder ähnliches) erheben.

Vor diesem Lockdown (bis zum 16. Dezember 2020)		Euro	☐ Kann / Möchte ich nicht sagen
Erste Märzwoche 2021		Euro	☐ Kann / Möchte ich nicht sagen
April 2021		Euro	☐ Kann / Möchte ich nicht sagen

PL17

12. Hatten Sie aufgrund der Mehrwertsteuersenkung im zweiten Halbjahr 2020 Ihre Preise gesenkt?

○ ja	○ nein	○ Kann / Möchte ich nicht sagen

PL13

13. Preisbestandteile (Anfang März 2021)

Falls der angegebene Preis (Anfang März 2021) eine Hygienepauschale beinhaltet, geben Sie bitte an, wie hoch diese ist. Falls Sie eine unterschiedlich hohe Hygienezuschläge für unterschiedliche Dienstleistungen erheben, geben Sie bitte den Hygienezuschlag für für den oben beschriebenen Haarschnitt an.

Falls Neukunden mehr zahlen als Stammkunden, geben Sie bitte den Preisunterschied an.

Falls Sie einen Zuschlag für späte Termine, für Termine am Montag oder für Termine am Wochenende erheben, geben Sie bitte die Höhe des Zuschlags an.

☐ Hygienepauschale: ____ Euro

☐ Neukunden-Zuschlag: ____ Euro

☐ Zuschlag für späte Termine: ____ Euro

☐ Zuschlag für Termine am Montag: ____ Euro

☐ Zuschlag für Termine am Wochenende: ____ Euro

☐ Kann / Möchte ich nicht sagen

14. Machen Sie mit dem oben beschriebenen Haarschnitt pro Kunde heute mehr oder weniger Gewinn als vor der Pandemie (Februar 2020)?

heute weniger	gleich viel	heute mehr	Kann / Möchte ich nicht sagen
○	○	○	○

PL12

15. Machen Sie mit dem oben beschriebenen Haarschnitt pro Kunde heute mehr oder weniger Gewinn als vor dem letzten Lockdown (Dezember 2020)?

heute weniger	gleich viel	heute mehr	Kann / Möchte ich nicht sagen
○	○	○	○

Seite 05

PH

PHP-Code

```
if (
(
(value('PL14_01') == 1) and (value('PL01_02') <= value('PL01_01'))
)
or
(
(value('PL14_01') == 2) and (value('PL02_02') <= value('PL02_01'))
)
){
goToPage('PG');
}
$pageNr = 4;
replace('%ownPageNumber%',$pageNr);
option('progress',round(100*$pageNr/7));
```

PL03

16. Weshalb haben sich Ihre Preise seit letztem Dezember erhöht?

Sie haben angegeben, dass mindestens einer Ihrer Preise im März 2021 höher ist als er im Dezember 2020 war. Welche Rolle haben die folgenden Faktoren bei der Preiserhöhung gespielt?

	Keine Rolle	Eine kleine Rolle	Eine große Rolle	Trifft nicht zu	Kann / Möchte ich nicht sagen
verringerte Kapazität durch Abstandsregelungen	○	○	○	○	○
Ausgleich des entgangenen Umsatzes / des Rücklagenabbaus durch den Lockdown	○	○	○	○	○
höhere Nachfrage	○	○	○	○	○
gestiegene Finanzierungskosten (zum Beispiel wegen Kreditaufnahme)	○	○	○	○	○
Anpassung an das allgemeine Preisniveau	○	○	○	○	○
gestiegene Lohnkosten	○	○	○	○	○
Die Preiserhöhung ist nur kurzfristig.	○	○	○	○	○
gestiegene Nebenkosten	○	○	○	○	○
gestiegener Hygieneaufwand (Masken, Desinfektionsmittel und Zeit)	○	○	○	○	○
Erwartung, dass Kunden für Preiserhöhung Verständnis haben	○	○	○	○	○
gestiegene Preise der Konkurrenz	○	○	○	○	○
Ende der Mehrwertsteuersenkung	○	○	○	○	○

PL07

Sonstige wichtige Faktoren:

PL15

17. Inwiefern stimmen Sie diesen Aussagen über Ihre Erfahrungen mit Ihren Kunden zu?

	stimme gar nicht zu	stimme eher nicht zu	unent-schieden	stimme eher zu	stimme voll zu	Kann / Möchte ich nicht sagen
Die Kunden äußern Verständnis für meine/unsere Preise.	○	○	○	○	○	○
Die Kunden beklagen sich aufgrund ihrer eigenen finanziellen Situation über die Preise.	○	○	○	○	○	○
Einzelne Kunden haben mir vorgeworfen von der Krise profitieren zu wollen.	○	○	○	○	○	○
Die Kunden geben mehr Trinkgeld.	○	○	○	○	○	○
Die Kunden geben weniger Trinkgeld.	○	○	○	○	○	○

Seite 06

PG

PHP-Code

```
if ((value('PL01_02') > value('PL01_01')) or (value('PL02_02') > value('PL02_01'))) {
goToPage('RA');
}
$pageNr = 4;
replace('%ownPageNumber%',$pageNr);
option('progress',round(100*$pageNr/7));
```

PL04

18. Weshalb haben sich Ihre Preise seit letztem Dezember nicht erhöht?

Sie haben angegeben, dass mindestens einer Ihrer Preise im März 2021 nicht höher ist als er im Dezember 2020 war. Welche Rolle haben die folgenden Faktoren bei der Entscheidung, den Preis nicht zu erhöhen, für Sie gespielt?

	keine Rolle	eine kleine Rolle	eine große Rolle	Trifft nicht zu	Kann / Möchte ich nicht sagen
Die Preise sind vertraglich festgelegt.	○	○	○	○	○
Innerhalb des Unternehmens konnten wir uns nicht auf Preissteigerungen einigen.	○	○	○	○	○
Ich weiß nicht, ob höhere Preise besser für das Unternehmen wären.	○	○	○	○	○
Eine Preiserhöhung würde größer scheinen als sie wirklich ist.	○	○	○	○	○
Erhöhung des Marktanteils / neue Kunden gewinnen	○	○	○	○	○
Die Preise wurden bereits nach dem 1. Lockdown (Frühjahr 2020) erhöht.	○	○	○	○	○
Zahlungskraft der Kunden ist in der Pandemie geringer	○	○	○	○	○
Mehrwertsteuersenkung im zweiten Halbjahr 2020 wurde nicht weitergegeben	○	○	○	○	○
Die Konkurrenz hat ihre Preise nicht erhöht.	○	○	○	○	○
Die Preise wurden nicht erhöht, um sie nicht in absehbarer Zeit wieder senken zu müssen.	○	○	○	○	○
Die Kosten sind nicht gestiegen.	○	○	○	○	○
Erhalt der Stammkunden	○	○	○	○	○

PL08

Sonstige wichtige Faktoren:

19. Inwiefern stimmen Sie diesen Aussagen über Ihre Erfahrungen mit Ihren Kunden zu? PL15

	stimme gar nicht zu	stimme eher nicht zu	unent-schieden	stimme eher zu	stimme voll zu	Kann / Möchte ich nicht sagen
Die Kunden äußern Verständnis für meine/unsere Preise.	○	○	○	○	○	○
Die Kunden beklagen sich aufgrund ihrer eigenen finanziellen Situation über die Preise.	○	○	○	○	○	○
Einzelne Kunden haben mir vorgeworfen von der Krise profitieren zu wollen.	○	○	○	○	○	○
Die Kunden geben mehr Trinkgeld.	○	○	○	○	○	○
Die Kunden geben weniger Trinkgeld.	○	○	○	○	○	○

Seite 07

RA

PHP-Code

```
$pageNr = 5;
replace('%ownPageNumber%',$pageNr);
option('progress',round(100*$pageNr/7));
```

Teil 3 Zustand nach Lockdown

Auf dieser Seite stellen wir Ihnen Fragen dazu, wie Ihr Unternehmen mit den politischen Maßnahmen umgeht, und wie Sie die zukünftige Entwicklung einschätzen.

LO04

20. Falls Sie für Anfang März mehr Terminanfragen erhalten haben, als Sie Termine zu vergeben hatten: wie sind Sie damit umgegangen?

Mehrfachantworten sind möglich

- ☐ Bevorzugung von Neukunden
- ☐ Anstellung von Mitarbeitern, um mehr Termine anbieten zu können
- ☐ Bevorzugung von Kunden, deren Termine in den letzten Monaten abgesagt werden mussten
- ☐ Bevorzugung von Stammkunden
- ☐ Wer zuerst angefragt hat, hat Termine bekommen
- ☐ Ausweitung der Öffnungszeiten, um mehr Termine anbieten zu können
- ☐ Erhebung eines Zuschlags für Neukunden

- ☐ Trifft nicht zu
- ☐ Kann / Möchte ich nicht sagen

LO01

21. Inwiefern stimmen Sie diesen Aussagen über die Pflicht zum Haarewaschen zu?

	Stimme gar nicht zu	Stimme eher nicht zu	Unent-schieden	Stimme eher zu	Stimme voll zu	Kann / Möchte ich nicht sagen
Ich fühle mich sicherer, wenn die Haare der Kunden vor der Behandlung gewaschen werden.	○	○	○	○	○	○
Die Pflicht zum Haarewaschen ist wie eine Preiserhöhung.	○	○	○	○	○	○
Die Kunden finden die Pflicht zum Haarewaschen akzeptabel.	○	○	○	○	○	○
Ich profitiere finanziell von der Pflicht zum Haarewaschen.	○	○	○	○	○	○

LO07

22. Für wie zutreffend halten Sie die folgenden Vorhersagen?

	gar nicht	eher nicht	unklar	eher ja	sehr	Kann / Möchte ich nicht sagen
In einem Jahr werden wir wieder den Zustand von vor der Pandemie haben.	○	○	○	○	○	○
Infektionsschutzmaßnahmen werden noch für Jahre vorgeschrieben bleiben.	○	○	○	○	○	○
Die Angst vor dem Virus wird manche Menschen noch lange Zeit von einem Friseurbesuch abhalten.	○	○	○	○	○	○
Meine persönliche finanzielle Situation wird sich längerfristig verbessern (verglichen zu heute).	○	○	○	○	○	○
Infolge der Pandemie(bekämpfung) wird die Zahlungsbereitschaft meiner/unserer Kunden nachhaltig sinken.	○	○	○	○	○	○
Es wird dieses Jahr einen weiteren Lockdown geben, in dem Friseurläden wieder schließen müssen.	○	○	○	○	○	○

LO08

23. Wie unsicher sind Sie sich über Ihre berufliche Zukunft?

gar nicht	kaum	etwas	sehr	Kann / Möchte ich nicht sagen
○	○	○	○	○

Seite 08

PHP-Code

```
$pageNr = 6;
replace('%ownPageNumber%',$pageNr);
option('progress',round(100*$pageNr/7));
```

Teil 4 Preissetzung allgemein

Auf dieser Seite stellen wir Ihnen allgemeine Fragen zur Preissetzung in Ihrem Unternehmen.

PA01

24. Im Allgemeinen, worauf achten Sie am meisten bei der Preissetzung?

Mehrfachantworten sind möglich

- ☐ Kosten
- ☐ Preise der Konkurrenz
- ☐ Qualität meines Angebots
- ☐ Kundenzufriedenheit
- ☐ Anpassung an das allgemeine Preislevel
- ☐ Anderes: ____________

- ☐ Kann / Möchte ich nicht sagen

PA02

25. Inwiefern stimmen Sie diesen Aussagen über Ihre Preissetzung zu?

	stimme gar nicht zu	stimme eher nicht zu	unent-schieden	stimme eher zu	stimme voll zu	Kann / Möchte ich nicht sagen
Ich bin zufrieden mit der Art wie ich/wir Preise setze/n.	○	○	○	○	○	○
Die Preise sind optimal für das Unternehmen gewählt.	○	○	○	○	○	○
Eigentlich sollten die Preise höher sein.	○	○	○	○	○	○
Die Gründe für Preiserhöhungen sind für die Kunden nachvollziehbar.	○	○	○	○	○	○

Seite 09

KO

PHP-Code

```
$pageNr = 7;
replace('%ownPageNumber%',$pageNr);
option('progress',round(100*$pageNr/7));
```

Danke

Vielen Dank für Ihre Teilnahme an unserer Studie!

S001

26. Wenn Sie uns etwas mitteilen möchten, können Sie dies hier anonym tun

Anmerkung: Diese Antwort wird zusammen mit Ihren anderen Antworten, aber ohne personenbezogene Informationen gespeichert.

Sollten Sie eine Frage haben, auf die Sie eine Antwort wünschen, können Sie uns gerne eine E-Mail schreiben.

Letzte Seite

Nochmals vielen Dank für Ihre Teilnahme!

Ihre Antworten wurden gespeichert, Sie können das Browser-Fenster nun schließen.

Thomas Kohler und Maximilian Weiß, Bonn Graduate School of Economics

Rheinische Friedrich-Wilhelms Universität Bonn – 2021

Impressum:
Maximilian Weiß
Kaiserplatz 7-9, 4. Stock
53113 Bonn
Tel.: +49 (0)228-73 3925

Appendix 1.C Regression Tables and Supplemental Figures

1.C.1 CPI evidence: stable differences across counties

Dep. Variable:	Median County	**R-squared:**	0.931
Model:	OLS	**Adj. R-squared:**	0.926
Method:	Least Squares	**Log-Likelihood:**	576.35
No. Observations:	304	**AIC:**	-1115.
Df Residuals:	285	**BIC:**	-1044.
Df Model:	18		
Covariance Type:	HAC		

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
average median	0.8894	0.071	12.493	0.000	0.750	1.029
16	0.3732	0.230	1.621	0.105	-0.078	0.824
34	0.1972	0.231	0.854	0.393	-0.255	0.650
64	0.3339	0.232	1.439	0.150	-0.121	0.789
81	0.2993	0.232	1.292	0.196	-0.155	0.753
138	0.4430	0.231	1.914	0.056	-0.011	0.897
163	0.0860	0.233	0.370	0.712	-0.370	0.542
210	0.5870	0.232	2.534	0.011	0.133	1.041
223	0.3361	0.230	1.464	0.143	-0.114	0.786
370	0.4360	0.232	1.881	0.060	-0.018	0.890
384	0.5442	0.228	2.383	0.017	0.097	0.992
444	0.5065	0.231	2.196	0.028	0.054	0.958
564	0.4531	0.232	1.954	0.051	-0.001	0.908
565	0.3059	0.231	1.327	0.184	-0.146	0.758
867	0.2269	0.229	0.990	0.322	-0.222	0.676
967	0.4629	0.230	2.016	0.044	0.013	0.913
1150	0.3198	0.230	1.394	0.163	-0.130	0.770
1238	0.4634	0.231	2.004	0.045	0.010	0.917
1518	0.3828	0.229	1.673	0.094	-0.066	0.831

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 1 lags and without small sample correction
[2] The regressors 16, 34, ..., 1518 are the fixed effects of the counties in the panel data. Source: German CPI micro-level data.

1.C.2 Retaining Regulars Applies Less Often

Stylized Fact 1. *Among the non-increasers, the higher the understanding of a firm's customers of its prices, the less important for price stickiness is the motive of retaining its regular customers.*

Table 1.C.1. Importance of retaining regular customers and customer understanding

	(1)	(2)	(3)
Dummy for retain regulars applies			
Cust. understand prices	-7.956**	-20.82*	-20.61*
	(4.035)	(12.43)	(10.77)
Employees (linear part)		-7.377***	-9.426***
		(0.822)	(0.993)
Dummy for many employees=1		-32.12***	-40.27***
		(3.128)	(4.482)
Satisfaction with pricing		2.814**	
		(1.309)	
Hairwashing		-0.934	
		(1.054)	
Pessimism		-4.867	
		(3.837)	
Share of regular customers		-0.431	
		(0.831)	
Rel. price December			-3.753***
			(1.432)
Constant	8.848**	55.51***	64.37***
	(3.575)	(17.74)	(12.30)
Observations	81	74	52
Pseudo R2	0.134	0.585	0.543

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Logit regression. The dependent variable is whether the respondent marked the hypothesis "Retain regular customers" as applicable or not.

1.C.3 More Likely to Increase Prices

Stylized Fact 2. *Firms with high understanding customers are more likely to increase their nominal prices.*

Table 1.C.2. Nominal price increase and customer understanding: extensive margin

	(1)	(2)	(3)
Price increased during the lockdown?			
Cust. understand prices	2.593***	3.553***	3.566**
	(0.710)	(1.205)	(1.562)
Employees (linear part)		0.0790	0.0953
		(0.113)	(0.139)
Dummy for many employees=1		0.443	0.207
		(0.446)	(0.590)
More than one salon=1		-0.560	-0.400
		(0.492)	(0.617)
Satisfaction with pricing		-1.504*	-1.589*
		(0.813)	(0.959)
Hairwashing		0.201	0.0528
		(0.537)	(0.671)
Pessimism		-0.101	-0.739
		(1.557)	(2.046)
Share of regular customers		0.0280	0.150
		(0.234)	(0.277)
Rel. price December			-1.411*
			(0.765)
Constant	-1.485**	-1.697	-0.189
	(0.580)	(1.411)	(1.934)
Observations	237	207	137
Pseudo R2	0.0343	0.0521	0.0836

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Logit regression. The dependent variable is a dummy indicating whether the respondent increased the price during the lockdown or not.

Table 1.C.3. Probability of nominal price increase and customer understanding: marginal effects

	(1)
Employees (linear part)	0.0212
	(0.0308)
Dummy for many employees=1	0.0875
	(0.120)
More than one salon=1	-0.126
	(0.161)
High understanding customers=1	0.237***
	(0.0888)
Satisfaction with pricing	-0.201
	(0.173)
Hairwashing	0.0397
	(0.146)
Pessimism	-0.205
	(0.437)
Rel. price December	-0.312*
	(0.164)
Share of regular customers	0.0590
	(0.0661)
N	138

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Logit regression with high understanding customers-indicator. Marginal effects at means.

1.C.4 Higher Nominal and Relative Price Increase

Stylized Fact 3. *Firms with high understanding customers increase their nominal and relative prices by more.*

Table 1.C.4. Nominal price increase and customer understanding: intensive margin

	(1)	(2)	(3)
Cust. understand prices	6.757***	8.909**	9.748**
	(2.074)	(3.450)	(4.198)
Employees (linear part)		-0.183	-0.290
		(0.328)	(0.402)
Dummy for many employees=1		0.393	-1.307
		(1.475)	(1.439)
More than one salon=1		-1.964	-1.102
		(1.384)	(1.610)
Satisfaction with pricing		-3.502*	-4.373**
		(1.931)	(1.973)
Hairwashing		0.532	-0.336
		(1.459)	(1.267)
Pessimism		2.098	3.504
		(4.642)	(5.564)
Share of regular customers		-0.491	-0.430
		(0.630)	(0.704)
Rel. price December			-4.460***
			(1.259)
Constant	0.179	1.978	6.262
	(1.628)	(4.628)	(6.166)
Observations	237	207	137
R2	0.0361	0.0576	0.146

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: OLS regression. The dependent variable is the percent increase of the firm's nominal price from December 2020 to March 2021.

Table 1.C.5. Relative price increase and customer understanding

	(1)	(2)	(3)
Cust. understand prices	6.946***	8.862**	8.917**
	(2.049)	(3.533)	(3.558)
Employees (linear part)		0.00928	0.198
		(0.312)	(0.304)
Dummy for many employees=1		-0.396	0.500
		(1.004)	(1.167)
More than one salon=1		-2.272	-1.891
		(2.227)	(1.805)
Satisfaction with pricing		-4.442*	-4.784**
		(2.256)	(2.050)
Hairwashing		0.517	0.740
		(1.665)	(1.729)
Pessimism		-1.865	-1.582
		(3.649)	(3.505)
Share of regular customers		-0.463	-0.337
		(0.720)	(0.729)
Rel. price December			-6.039***
			(1.225)
Constant	-4.888**	-0.473	4.429
	(1.759)	(4.879)	(5.591)
Observations	157	137	137
R2	0.0523	0.0927	0.169

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: OLS regression. The dependent variable is the percent increase of the firm's relative price from December 2020 to March 2021. The relative price is defined as the nominal price divided by the county's median price.

Table 1.C.6. Nominal price increase and customer understanding: intensive margin, binary regressor

	(1)	(2)	(3)
High understanding customers=1	1.997***	2.469**	2.714**
	(0.703)	(0.971)	(1.159)
Employees (linear part)		-0.149	-0.293
		(0.320)	(0.387)
Dummy for many employees=1		0.707	-0.849
		(1.454)	(1.417)
More than one salon=1		-1.982	-1.383
		(1.382)	(1.560)
Satisfaction with pricing		-1.786	-2.515
		(1.713)	(1.514)
Hairwashing		0.786	0.00207
		(1.426)	(1.227)
Pessimism		1.692	2.814
		(4.660)	(5.640)
Share of regular customers		-0.263	-0.156
		(0.640)	(0.785)
Rel. price December			-4.512***
			(1.317)
Constant	4.325***	5.485	10.17
	(0.557)	(4.165)	(6.090)
Observations	281	209	138
R2	0.0234	0.0438	0.122

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: OLS regression. The dependent variable is the percent increase of the firm's nominal price from December 2020 to March 2021.

Table 1.C.7. Relative price increase and customer understanding: binary regressor

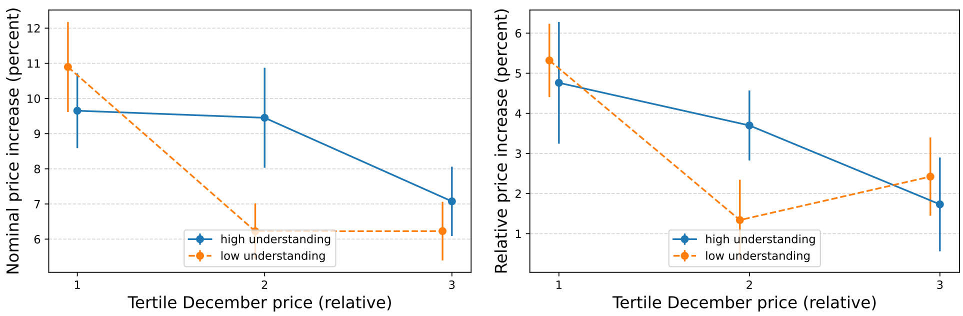
	(1)	(2)	(3)
High understanding customers=1	1.671**	1.897**	1.911*
	(0.620)	(0.896)	(0.935)
Employees (linear part)		-0.0303	0.157
		(0.307)	(0.302)
Dummy for many employees=1		-0.154	0.738
		(0.955)	(1.123)
More than one salon=1		-2.550	-2.175
		(2.244)	(1.892)
Satisfaction with pricing		-2.325	-2.673
		(1.908)	(1.703)
Hairwashing		0.796	1.020
		(1.637)	(1.725)
Pessimism		-2.872	-2.545
		(3.867)	(3.640)
Share of regular customers		-0.221	-0.100
		(0.759)	(0.743)
Rel. price December			-6.090***
			(1.337)
Constant	-0.477	3.480	8.482
	(0.748)	(5.068)	(5.664)
Observations	186	138	138
R2	0.0204	0.0573	0.134

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: OLS regression. The dependent variable is the percent increase of the firm's relative price from December 2020 to March 2021. The relative price is defined as the nominal price divided by the county's median price.

(a) Conditional, nominal

(b) Conditional, relative



Notes: Nominal and relative price increases conditional on price change, over the relative price distribution and by understanding-type. The increase is from December 2020 to March 2021. For the nominal price increase (left panel), the difference between the two understanding-types is only statistically significant for the second tertile (two-sample t-test, standard errors clustered at the county level; p=6%, $N_L = 9$ (9 cluster), $N_H = 20$ (12 cluster)). For the relative price increase (right panel), the difference between the two understanding-types is only statistically significant for the second tertile (two-sample t-test, standard errors clustered at the county level; p=9%, $N_L = 9$ (9 cluster), $N_H = 20$ (12 cluster)). The whiskers denote 68% confidence intervals.

Figure 1.C.1. Conditional price increases and customer understanding

1.C.5 Rather Restored Profit Margin

Stylized Fact 4. *Firms with high understanding customers are better able to increase their profit margins.*

Table 1.C.8. Profit margins and customer understanding

	(1) mean	(2) mean	(3) t
Profit margin before pandemic	-0.14	-0.04	(-1.04)
Profit margin before lockdown	-0.20	-0.09	(-1.33)
Observations	84	173	257

Notes: Means of the answers to the question about the profit margin compared to before the pandemic and before the lockdown for firms with low understanding (first column) and high understanding (second column) customers. The third column is the t statistic of a two-sample t-test for the null hypothesis that the means are equal. The null can be rejected for the one-sided alternative that high understanding customers have a higher mean at the 10%-level for the profit margin before lockdown (second row), while it cannot be rejected for the profit margin before the pandemic (first row, at p=15%).

1.C.6 More Satisfied with Own Pricing

Stylized Fact 5. *Firms with high understanding customers are more satisfied with their own pricing.*

Table 1.C.9. Pricing satisfaction and customer understanding

	(1)	(2)	(3)
Cust. understand prices	0.531***	0.512***	0.503***
	(0.0549)	(0.0624)	(0.0753)
Employees (linear part)		-0.00425	0.00255
		(0.00956)	(0.0134)
Dummy for many employees=1		0.0116	0.0464
		(0.0336)	(0.0451)
More than one salon=1		0.0381	0.0599
		(0.0352)	(0.0418)
Hairwashing		0.0518	0.109**
		(0.0351)	(0.0425)
Pessimism		-0.0939	-0.0818
		(0.0821)	(0.119)
Share of regular customers		-0.0194	-0.00533
		(0.0250)	(0.0343)
Rel. price December			-0.0286
			(0.0407)
Constant	0.193***	0.329**	0.252
	(0.0467)	(0.127)	(0.173)
Observations	224	207	137
R2	0.228	0.239	0.247

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: OLS regression. The dependent variable is the summary variable of the respondent's satisfaction with the own pricing method.

1.C.7 Are Less Pessimistic

Stylized Fact 6. *Owners of firms with high understanding customers are less pessimistic.*

Table 1.C.10. Pessimism and customer understanding

	(1)	(2)	(3)
Cust. understand prices	-0.165***	-0.106*	-0.0832
	(0.0439)	(0.0624)	(0.0844)
Employees (linear part)		-0.00674	-0.00381
		(0.00605)	(0.00696)
Dummy for many employees=1		-0.0321	-0.0270
		(0.0233)	(0.0337)
More than one salon=1		0.0435	0.0564
		(0.0267)	(0.0454)
Satisfaction with pricing		-0.0498	-0.0365
		(0.0464)	(0.0568)
Hairwashing		0.0116	0.00289
		(0.0283)	(0.0385)
Share of regular customers		-0.0285**	-0.0270
		(0.0138)	(0.0190)
Rel. price December			0.0106
			(0.0331)
Constant	0.710***	0.836***	0.796***
	(0.0338)	(0.0712)	(0.102)
Observations	228	207	137
R2	0.0481	0.0639	0.0598

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: OLS regression. The dependent variable is the summary variable of the respondent's pessimism concerning the firm's and the owner's professional future.

1.C.8 Are Smaller

Stylized Fact 7. *Firms with less employees have more understanding customers.*

Table 1.C.11. Customer understanding and firm size

	(1)	(2)	(3)
High understanding customers			
Employees (linear part)	-0.216**	-0.296**	-0.370***
	(0.0944)	(0.116)	(0.134)
Dummy for many employees=1	-0.803*	-1.478***	-1.752***
	(0.417)	(0.526)	(0.652)
More than one salon=1	-0.281	0.0958	-0.271
	(0.533)	(0.682)	(1.144)
Pessimism		-2.734**	-2.180
		(1.254)	(1.613)
Satisfaction with pricing		3.996***	3.280***
		(1.033)	(1.125)
Hairwashing		-0.614	-0.367
		(0.490)	(0.593)
Share of regular customers		0.0196	-0.206
		(0.302)	(0.376)
Rel. price December			0.0117
			(0.531)
Constant	1.441***	0.977	1.999
	(0.347)	(2.198)	(2.648)
Observations	276	209	138
Pseudo R2	0.0167	0.132	0.120

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Logit regression. The dependent variable is high understanding customers indicator variable.

Table 1.C.12. Customer understanding and firm size: marginal effects

	(1)
Pessimism	-0.521
	(0.390)
Employees (linear part)	-0.0884***
	(0.0317)
Dummy for many employees=1	-0.412***
	(0.135)
More than one salon=1	-0.0662
	(0.283)
Satisfaction with pricing	0.784***
	(0.260)
Hairwashing	-0.0878
	(0.142)
Rel. price December	0.00280
	(0.127)
Share of regular customers	-0.0492
	(0.0905)
N	138

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Marginal effects at means.

1.C.9 Large Firms and Occasional Customers

Table 1.C.13. Share of regular customers and firm size

	(1)	(2)	(3)
Low share of regular customers			
Employees (linear part)	0.199**	0.247*	0.290
	(0.0934)	(0.133)	(0.195)
Dummy for many employees=1	1.646***	1.662***	1.879***
	(0.376)	(0.488)	(0.660)
More than one salon=1		0.0927	1.501**
		(0.417)	(0.683)
Cust. understand prices		-0.929	-0.109
		(1.121)	(1.564)
Satisfaction with pricing		0.958	0.0788
		(1.029)	(1.226)
Hairwashing		0.0143	0.119
		(0.415)	(0.507)
Pessimism		2.827**	2.917
		(1.325)	(2.015)
Rel. price December			-0.844
			(0.626)
Constant	-1.278***	-2.997**	-2.701
	(0.317)	(1.243)	(1.897)
Observations	280	207	137
Pseudo R2	0.0481	0.0740	0.120

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Logit regression. The dependent variable is an indicator variable that equals 1 if less than 80% of the firm's customers are regular customers, and 0 otherwise.

Table 1.C.14. Share of regular customers and firm size: marginal effects

	(1)
Employees (linear part)	0.0653
	(0.0429)
Dummy for many employees=1	0.435^{***}
	(0.136)
More than one salon=1	0.358^{**}
	(0.149)
Cust. understand prices	-0.0246
	(0.353)
Satisfaction with pricing	0.0178
	(0.277)
Hairwashing	0.0267
	(0.115)
Pessimism	0.658
	(0.455)
Rel. price December	-0.190
	(0.143)
N	137

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Marginal effects at means.

Table 1.C.15. Gain customers and share of regular customers

	(1)	(2)	(3)
Low share of regular customers=1	0.550***	0.560***	0.605**
	(0.151)	(0.157)	(0.225)
Employees (linear part)		-0.0281	-0.0729
		(0.0572)	(0.0692)
Dummy for many employees=1		0.398	0.325
		(0.381)	(0.530)
More than one salon=1		-0.338	-0.739
		(0.442)	(0.462)
Cust. understand prices		-0.929	-0.828
		(0.645)	(1.091)
Satisfaction with pricing		0.0371	-0.175
		(0.478)	(0.757)
Hairwashing		-0.250	-0.481
		(0.252)	(0.341)
Pessimism		0.530	1.714*
		(0.808)	(0.981)
Rel. price December			-0.0999
			(0.201)
Constant	0.255***	0.760	0.493
	(0.0806)	(0.836)	(1.108)
Observations	83	67	43
Pseudo R2			

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: OLS regression. The dependent variable is the grade to the hypothesis "Price not increased: gain new customers". We allocate the score 0 if the respondent answered "Does not apply".

Table 1.C.16. Number of employees and relative price position

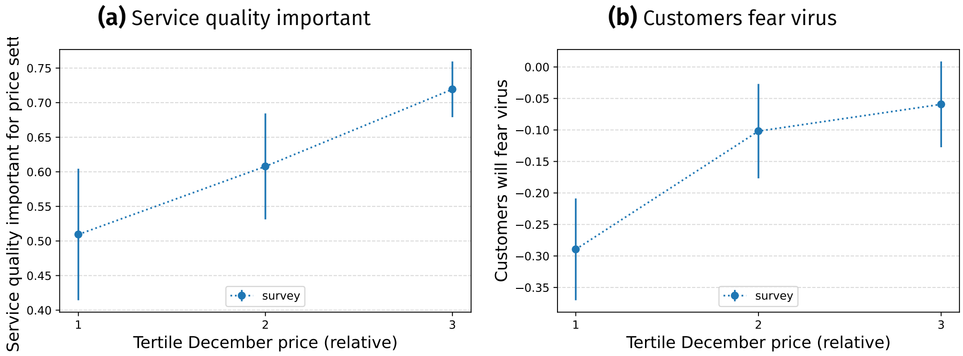
	(1)	(2)
Number of employees		
Second rel. price tertile	-0.0683	0.343
	(0.335)	(0.405)
Third rel. price tertile	0.872***	1.132***
	(0.336)	(0.385)
More than one salon=1		3.548***
		(0.631)
Cust. understand prices		-1.506
		(1.029)
Satisfaction with pricing		0.938
		(0.783)
Hairwashing		-0.507
		(0.522)
Pessimism		-1.292
		(1.483)
Low share of regular customers=1		0.968***
		(0.355)
/		
cut1	-2.310***	-3.625**
	(0.374)	(1.696)
cut2	0.160	-0.837
	(0.265)	(1.618)
cut3	1.521***	0.808
	(0.294)	(1.617)
Observations	186	137
Pseudo R2	0.0220	0.130

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Ordered logistic regression. The dependent variable is an ordered variable for the number of employees, with the values 1:no employees, 2:one to three employees, 3:four to six employees, 4:more than six employees.

1.C.10 Quality difference among hairdressers

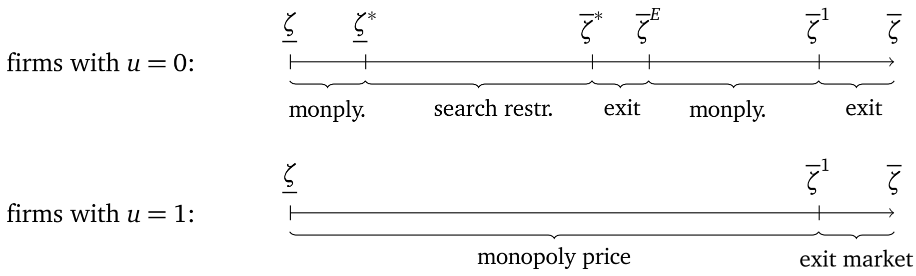
Figure 1.C.2 presents evidence for the claim that relatively more expensive hairdressers in our survey produce a higher-quality haircut. We ask the participants in our survey how important the quality of their service is for their price-setting. The plot in the left panel shows the average answer to this question over the relative price distribution. Relatively more expensive firms place a higher importance on the quality of their service when setting their price, and the difference between the first and the third tertile is statistically significant. In the context of our model, this suggests that more expensive firm offer a higher-quality service, and therefore can set a higher price.



Notes: Survey responses over the relative price distribution. The responses are scaled between -1 (completely disagree) and 1 (completely agree). For the importance of service quality for price setting (left panel), the difference between the first and the third tertile is statistically significant (two-sample t-test, standard errors clustered at the county level; p=5.2%, $N_1 = 53$ (21 cluster), $N_3 = 57$ (20 cluster)). For the expectation that customers will fear infection with the coronavirus (right panel), the difference between the first and the third tertile is statistically significant (two-sample t-test, standard errors clustered at the county level; p=3.6%, $N_1 = 53$ (21 cluster), $N_3 = 59$ (20 cluster)). The whiskers denote 68% confidence intervals.

Figure 1.C.2. Quality differences over relative price distribution

The plot in the right panel shows the expectation of the hairdressers about whether the fear of the virus will deter their customers from demanding their service in the future. It shows that relatively more expensive firms are significantly more pessimistic. We conjecture that the added value of a higher-quality haircut has a large social component: in higher-quality hairsalons, customers like to spend more time, and maybe enjoy the interaction with the hairdresser and other clients more. Spending more time in a social setting, however, also increases the risk of getting infected with a virus. Therefore, we interpret the evidence that more expensive hairdressers report more fearful customers as supportive of the claim that more expensive hairdressers offer a higher-quality, namely a more socially enjoyable, service.



Notes: Firm pricing decisions during the uncertainty period by customer information type u and idiosyncratic cost component ζ_i.

Figure 1.D.1. Model: idiosyncratic cost component cut-offs

Appendix 1.D Solving the model

Figure 1.D.1 shows the cutoffs on the distribution of idiosyncratic costs at which firms with high baseline costs and low product quality change their optimal pricing strategy in the uncertainty period $t = 1$. We choose an equilibrium such that firms with high understanding customers ($u = 1$) with idiosyncratic costs $\zeta_i < \overline{\zeta}^1$ can charge the monopoly price, as it allows their customers to learn the new industry-wide marginal costs (by proposition 1.5.1), which lowers their expected outside option. At the threshold $\overline{\zeta}^1$, the firms' costs exceed their product's expected quality, so that they exit the market.

The situation is more complex for firms with high baseline costs and low product quality whose regular customer is of the low information type ($u = 0$). Charging the monopoly price does not guarantee that customers learn the true industry-wide costs, since customers cannot observe the idiosyncratic costs of the firm. However, customers know the upper bound of the distribution of idiosyncratic costs, and therefore will learn of the industry-wide cost increase if a firm has a large enough idiosyncratic cost and charges its monopoly price (see corollary 1.5.1). As a result, whether firms are search-restricted or not is related in a *non-monotonic* way to their idiosyncratic costs. Firms whose idiosyncratic marginal cost component is below $\underline{\zeta}^*$ offer, by charging their monopoly price, a higher expected surplus to their customer than their expected outside option, even though customers do not understand that industry-wide costs increased. Firms whose costs lie between $\underline{\zeta}^*$ and $\overline{\zeta}^E$ will not induce their customer to learn of a higher industry-wide cost, since if they set their highest price (the monopoly price), their customer would leave (in terms of corollary 1.5.1, the γ^0 is not high enough). In that region, firms either lower their price below their monopoly price in order to compete with the customer's outside option (costs below $\overline{\zeta}^*$) or exit the market (costs above $\overline{\zeta}^*$), which happens when the expected profits at the search-restricted price are at most zero (see equation (1.8)). With costs above $\overline{\zeta}^{E}$, firms can set their monopoly price, as their customer learns enough about the industry-wide cost increase to estimate that searching would yield a lower expected surplus than staying at the firm.[35] Firms with the highest idiosyncratic costs, $\zeta_i > \overline{\zeta}^{1}$, exit the market, for the same reason as firms with high understanding customers.

1.D.1 Thresholds for different information levels

In the uncertainty period $t = 1$, we have to solve an equilibrium for the three different information levels of customers. We are only concerned with customers of high cost, low quality firms, since all other firms are never restricted in their pricing by their competition. While customers of the information *type* $u = 1$ only have information level $\gamma^1 = 1$ in the equilibrium we consider, customers of the information type $u = 0$ can have the information levels $\gamma^0 \in \{0\} \vee [\underline{E}, \overline{E}]$, where $\underline{E}, \overline{E} \in (0, 1)$ are the information levels that obtain at idiosyncratic costs $\overline{\zeta}^{E}$ and $\overline{\zeta}^{1}$, respectively.

We denote the thresholds of the *true* equilibrium, i.e. the one implied by the correct level of industry-wide costs in period $t = 1$, $\underline{c}'$, as $\underline{\zeta}^{*}, \overline{\zeta}^{*}, \overline{\zeta}^{E}$, and $\overline{\zeta}^{1}$. Naturally, this is the equilibrium that is expected by customers with information level $\gamma^1 = 1$. For the equilibrium that is expected by customers who do not learn, $\gamma^0 = 0$, we denote the respective thresholds as $\underline{\zeta}^{*,0}, \overline{\zeta}^{*,0}, \overline{\zeta}^{E,0}$, and $\overline{\zeta}^{1,0}$, while for the one expected by customers who learn up to the level $\gamma^0 = \underline{E}$, we write $\underline{\zeta}^{*,E}, \overline{\zeta}^{*,E}, \overline{\zeta}^{E,E}$, and $\overline{\zeta}^{1,E}$.

We calibrate the model such that in period $t = 0$, there is no firm who produces at a lower quality than its marginal costs, i.e. $\bar{c} + \overline{\zeta} < \underline{q}$. Therefore, since customers with information level $\gamma^0 = 0$ assume that industry-wide costs did not change, $\overline{\zeta}^{1,0} = \overline{\zeta}$. Also, since they already believe that costs did not change, they do not think that any firm could charge a monopoly price that would indicate higher industry-wide costs, i.e. $\overline{\zeta}^{E,0} = \overline{\zeta}$. We solve for the thresholds $\underline{\zeta}^{*,0}$ and $\overline{\zeta}^{*,0}$, which pose a fixed point problem, numerically.

At information level $\underline{E}$, the customer with information type $u = 0$ estimates that only the firm with the highest idiosyncratic costs, $\overline{\zeta}$, is able to convince low-understanding customers of an industry-wide cost increases, hence $\overline{\zeta}^{E,E} = \overline{\zeta}$. Likewise, he estimates that the industry-wide cost increase is so low that even the least productive can survive, i.e. $\overline{\zeta}^{1,E} = \overline{\zeta}$. We solve for the thresholds $\underline{\zeta}^{*,E}$ and $\overline{\zeta}^{*,E}$ numerically.

35. We make use of the result that, for unproductive firms who offer a relatively low surplus, learning of an increase in industry-wide costs reduces the expected value from searching by more than it reduces the expected value of staying at the unproductive firm (see footnote 23). Hence, if charging the monopoly price at idiosyncratic costs $\overline{\zeta}^{E}$ is favourable to the firm, it will be favourable at all higher idiosyncratic costs, as the level of industry-wide costs learnt by the firm's customer increases in its charged price.

Finally, we solve for $\underline{\zeta}^*, \overline{\zeta}^*, \overline{\zeta}^E$, and $\overline{\zeta}^1$ numerically, taking all the other solutions into account. To make this practicable, we use a "guess and verify"-approach: we impose the equilibrium that we want to solve for, i.e. the one where low-cost firms and high-quality firms are always monopolists, choose the search cost s accordingly (see the main text), and then check for a given parameterization if the equilibrium conditions are fulfilled for all types of firms.

1.D.2 Derivation of zero profit threshold

It is to show that

$$V_{it}^* = \left(1 + \sqrt{1/(1+\mathscr{D}_t)}\right)^2 \frac{F_{it}}{2\mathscr{D}_t}.$$

Proof: V_{it}^* is the expected surplus the firm offers at which expected profits are zero. Let p_{it}^* denote the according price the firm sets. We assume that $p_{it}^* < q_{it}$. Then, the expected revenue at that price, given that the firm can hold its regular customer, is

$$(1+\mathscr{D}_t)(1-p_{it}^*/q_{it})p_{it}^*$$

We now subtract the expected marginal costs and the fixed costs and set the difference to zero, where we can rewrite the first line as in equation 1.6:

$$(1+\mathscr{D}_t)(1-p_{it}^*/q_{it})(p_{it}^*-C_{it})-F_{it} = 0 \tag{1.D.1}$$

$$\Leftrightarrow -\left(p_{it}^*-(C_{it}+q_{it})/2\right)^2 = F_{it}q_{it}/(1+\mathscr{D}_t)-\left((q_{it}-C_{it})/2\right)^2 \tag{1.D.2}$$

Substituting $F_{it} = \frac{\mathscr{D}_t}{q_{it}}\left(\frac{q_{it}-C_{it}}{2}\right)^2$ yields

$$\left(p_{it}^*-(C_{it}+q_{it})/2\right)^2 = \left((q_{it}-C_{it})/2\right)^2(1-\mathscr{D}_t/(1+\mathscr{D}_t)) \tag{1.D.3}$$

which has the solution

$$p_{it}^* = (C_{it}+q_{it})/2+\sqrt{1/(1+\mathscr{D}_t)}(C_{it}-q_{it})/2$$

Inserting this into formula 1.2 yields

$$V_{it}^* = (q_{it}-p_{it}^*)^2/(2q_{it}) \tag{1.D.4}$$

$$= \frac{\left((q_{it}-C_{it})/2-\sqrt{1/(1+\mathscr{D}_t)}(C_{it}-q_{it})/2\right)^2}{2q_{it}} \tag{1.D.5}$$

$$= \frac{\left((q_{it}-C_{it})/2\right)^2}{q_{it}}\left(1+\sqrt{1/(1+\mathscr{D}_t)}\right)^2/2 \tag{1.D.6}$$

$$= \left(1+\sqrt{1/(1+\mathscr{D}_t)}\right)^2\frac{F_{it}}{2\mathscr{D}_t} \tag{1.D.7}$$

References

Álvarez, Luis J., and Ignacio Hernando. 2005. "The Price Setting Behaviour of Spanish Firms: Evidence from Survey Data." *European Central Bank Working Paper Series* No. 538. [46]

Amirault, David, Carolyn Kwan, and Gordon Wilkinson. 2006. "Survey of Price-Setting Behaviour of Canadian Companies." *Bank of Canada Working Paper* No. 2006-35. [46]

Anderson, Simon P., and Régis Renault. 2018. "Chapter 8: Firm pricing with consumer search." In *Handbook of Game Theory and Industrial Organization, Volume II.* Cheltenham, UK: Edward Elgar Publishing. [8]

Apel, Mikael, Richard Friberg, and Kerstin Hallsten. 2005. "Microfoundations of Macroeconomic Price Adjustment: Survey Evidence from Swedish Firms." *Journal of Money, Credit and Banking* 37 (2): 313–38. [46]

Armstrong, Mark. 2017. "Ordered Consumer Search." *Journal of the European Economic Association* 15 (5): 989–1024. [28]

Aucremanne, Luc, and Martine Druant. 2005. "Price-Setting in Belgium: What Can Be Learned from an ad hoc Study?" *European Central Bank Working Paper Series* No. 448. [46]

Ball, Laurence, and David Romer. 1991. "Sticky Prices as Coordination Failure." *American Economic Review* 81 (3): 539–52. Accessed February 27, 2023. [6]

Bénabou, Roland. 1988. "Search, Price Setting and Inflation." *Review of Economic Studies* 55 (3): 353–76. [7]

Bénabou, Roland. 1992. "Inflation and Efficiency in Search Markets." *Review of Economic Studies* 59 (2): 299–329. [7]

Bénabou, Roland, and Robert Gertner. 1993. "Search with Learning from Prices: Does Increased Inflationary Uncertainty Lead to Higher Markups." *Review of Economic Studies* 60 (1): 69–93. Accessed February 28, 2023. [7]

Benzarti, Youssef, Dorian Carloni, Jarkko Harju, and Tuomas Kosonen. 2020. "What goes up may not come down: asymmetric incidence of value-added taxes." *Journal of Political Economy* 128 (12). [8, 35]

Blanchard, Olivier J. 1994. "Comment." In *Monetary Policy,* edited by N. Gregory Mankiw, 150–54. The University of Chicago Press. [6, 17]

Blinder, Alan S. 1994. "On Sticky Prices: Academic Theories Meet the Real World." In *Monetary Policy,* edited by N. Gregory Mankiw, 117–54. The University of Chicago Press. [17]

Blinder, Alan S., Elie R. D. Canetti, David E. Lebow, and Jeremy B. Rudd. 1998. *Asking About Prices: A New Approach to Understanding Price Stickiness.* New York: Russell Sage Foundation. [6, 8, 17–19, 46]

Coibion, Olivier, and Yuriy Gorodnichenko. 2015. "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts." *American Economic Review* 105 (8): 2644–78. [35]

Copaciu, Mihai, Florian Neagu, and Horia Braun-Erdei. 2010. "Survey Evidence on Price-Setting Patterns of Romanian Firms." *Managerial and Decision Economics* 31 (2-3): 235–47. [46]

Corchón, Luis C., and Marco A. Marini. 2018. *Handbook of Game Theory and Industrial Organization, Volume II: Applications.* Cheltenham, UK: Edward Elgar Publishing. [8]

Correa, Arnildo, Myrian Petrassi, and Rafael Santos. 2018. "Price-Setting Behavior in Brazil: Survey Evidence." *Journal of Business Cycle Research* 14 (2): 283–310. [46]

Dabušinskas, Aurelijus, and Martti Randveer. 2006. "Comparison of pricing behavior of firms in the euro area and Estonia." *Bank of Estonia Working Paper Series* No. 8/2006. [46]

Damjanovic, Tatiana, and Charles Nolan. 2010. “Relative Price Distortions and Inflation Persistence*.” *Economic Journal* 120 (547): 1080–99. [3]

Diamond, Peter A. 1993. “Search, Sticky Prices, and Inflation.” *Review of Economic Studies* 60 (1): 53–68. Accessed February 28, 2023. [7]

Dixit, Avinash K., and Joseph E. Stiglitz. 1977. “Monopolistic Competition and Optimum Product Diversity.” *American Economic Review* 67 (3): 297–308. Accessed February 27, 2023. [3]

Eyster, Erik, Kristóf Madarász, and Pascal Michaillat. 2021. “Pricing Under Fairness Concerns.” *Journal of the European Economic Association* 19 (3): 1853–98. [6, 8]

Fabiani, Silvia, Martine Druant, Ignacio Hernando, Claudia Kwapil, Bettina Landau, Claire Loupias, Fernando Martins, et al. 2006. “What Firms’ Surveys Tell Us about Price-Setting Behavior in the Euro Area.” *International Journal of Central Banking* 2 (3): 3–47. [46]

Fabiani, Silvia, Angela Gattulli, and Roberto Sabbatini. 2004. “The Pricing Behaviour of Italian Firms: New Survey Evidence on Price Stickiness.” *European Central Bank Working Paper Series* No. 333. [46]

Fishman, Arthur. 1996. “Search with Learning and Price Adjustment Dynamics*.” *Quarterly Journal of Economics* 111 (1): 253–68. [7, 25, 31]

Foster, Lucia, John Haltiwanger, and Chad Syverson. 2016. “The Slow Growth of New Plants: Learning about Demand?” *Economica* 83 (329): 91–129. [46]

Gautier, Erwan, Cristina Conflitti, Riemer P. Faber, Brian Fabo, Ludmila Fadejeva, Valentin Jouvanceau, Jan-Oliver Menz, et al. 2022. “New facts on consumer price rigidity in the Euro Area.” Temi di discussione (Economic working papers) 1375. Bank of Italy, Economic Research and International Relations Area. [9, 12]

Greenslade, Jennifer V., and Miles Parker. 2012. “New Insights into Price-Setting Behaviour in the UK: Introduction and Survey Results.” *Economic Journal* 122 (558): F1–F15. [46]

Hall, Simon, Mark Walsh, and Anthony Yates. 2000. “Are UK Companies’ Prices Sticky?” *Oxford Economic Papers* 52 (3): 425–46. [46]

Hansen, Lars Peter, and Thomas J. Sargent. 2022. “Risk, Ambiguity, and Misspecification: Decision Theory, Robust Control, and Statistics.” Working Paper, University of Chicago, Becker Friedman Institute for Economics 2022-157. University of Chicago, Becker Friedman Institute for Economics. [34]

Hobijn, Bart, Fernanda Nechio, and Adam Hale Shapiro. 2021. “Using Brexit to identify the nature of price rigidities.” *Journal of International Economics* 130: 103448. NBER International Seminar on Macroeconomics 2020. [8, 45]

Hoeberichts, Marco M., and Ad C. J. Stokman. 2006. “Price Setting Behaviour in the Netherlands: Results of a Survey.” *European Central Bank Working Paper Series* No. 607. [46]

Jankiewicz, Zofia, and Danuta Kolodziejczyk. 2008. “The Price-Setting Behaviour of Polish Firms: Comparison between the Euro Area and Poland.” *Bank i Kredyt* 39 (2): 19–42. [46]

Janssen, Maarten, and Sandro Shelegia. 2019. “Beliefs and Consumer Search in a Vertical Industry.” *Journal of the European Economic Association* 18 (5): 2359–93. [8, 25]

Janssen, Maarten C.W., Alexei Parakhonyak, and Anastasia Parakhonyak. 2017. “Non-reservation price equilibria and consumer search.” *Journal of Economic Theory* 172: 120–62. [8, 34]

Kaplan, Greg, and Guido Menzio. 2015. “The morphology of price dispersion.” *International Economic Review* 56 (4): 1165–206. [3, 9, 14]

Karadi, Peter, and Adam Reiff. 2019. “Menu Costs, Aggregate Fluctuations, and Large Shocks.” *American Economic Journal: Macroeconomics* 11 (3): 111–46. [8]

Kimolo, Deogratius Wenceslaus. 2018. "Price Adjustment Behaviour of Manufacturing and Service Sector Firms in Tanzania: A Survey Evidence of Price Stickiness." *Journal of Economics and Sustainable Development* 9 (24): 128–58. [46]

Klenow, Peter J., and Jonathan L. Willis. 2016. "Real Rigidities and Nominal Price Changes." *Economica* 83 (331): 443–72. [9, 43, 45]

Kohlhepp, Jacob. 2023. "The Inner Beauty of Firms." Mimeo. University of California, Los Angeles. [26]

Kwapil, Claudia, Josef Baumgartner, and Johann Scharler. 2005. "The Price-Setting Behavior of Austrian Firms." *European Central Bank Working Paper Series* No. 464. [46]

Langbraaten, Nina, Einar W. Nordbø, and Fredrik Wulfsberg. 2008. "Price-Setting Behaviour of Norwegian Firms – Results of a Survey." *Norges Bank Economic Bulletin* 79 (2): 13–34. [46]

Loupias, Claire, and Roland Ricart. 2004. "Price Setting in France: New Evidence from Survey Data." *European Central Bank Working Paper Series* No. 423. [46]

Lünnemann, Patrick, and Thomas Y. Mathä. 2006. "New Survey Evidence on the Pricing Behaviour of Luxembourg Firms." *European Central Bank Working Paper* No. 617. [46]

Malik, Wasim Shahid, Ahsan ul Haq Satti, and Ghulam Saghir. 2008. "Price Setting Behaviour of Pakistani Firms: Evidence from Four Industrial Cities of Punjab." *Pakistan Development Review* 47 (3): 247–66. [46]

Martins, Fernando. 2005. "The Price Setting Behaviour of Portuguese Firms: Evidence from Survey Data." *European Central Bank Working Paper Series* No. 562. [46]

Mongey, Simon. 2021. "Market Structure and Monetary Non-neutrality." NBER Working Papers 29233. National Bureau of Economic Research, Inc. [9]

Nakagawa, Shinobu, Ryota Hattori, and Izumi Takagawa. 2000. "Price-Setting Behavior of Japanese Companies." *Mimeo.* [46]

Nakamura, Emi, and Jón Steinsson. 2008. "Five Facts about Prices: A Reevaluation of Menu Cost Models*." *Quarterly Journal of Economics* 123 (4): 1415–64. [8, 12]

Nakamura, Emi, and Jón Steinsson. 2011. "Price Setting in Forward-Looking Customer Markets." *Journal of Monetary Economics* 58 (3): 220–33. [6, 34]

Okun, Arthur M. 1981. *Prices And Quantities: A Macroeconomic Analysis.* Washington: The Brookings Institution. [6]

Ólafsson, Thorvardur Tjörvi, Ásgerdur Pétursdóttir, and Karen Á. Vignisdóttir. 2011. "Price Setting in Turbulent Times: Survey Evidence from Icelandic Firms." *Central Bank of Iceland Working Paper Series* No. 54. [46]

Parker, Miles. 2014. "Price-Setting Behaviour in New Zealand." *Reserve Bank of New Zealand Discussion Paper* 2014/04. [46]

Pham, Anh The, Hoang Huy Nguyen, and Hung Duc Nguyen. 2019. "Price Adjustment Lags and Their Asymmetries in Vietnam." *Mimeo.* [46]

Reinganum, Jennifer F. 1979. "A Simple Model of Equilibrium Price Dispersion." *Journal of Political Economy* 87 (4): 851–58. Accessed January 5, 2023. [28]

Rotemberg, Julio J. 2005. "Customer Anger at Price Increases, Changes in the Frequency of Price Adjustment and Monetary Policy." *Journal of Monetary Economics* 52 (4): 829–52. [6]

Rotemberg, Julio J. 2011. "Fair Pricing." *Journal of the European Economic Association* 9 (5): 952–81. [6, 8]

Sahinoz, Saygin, and Bedriye Saraçoğlu. 2008. "Price–Setting Behavior In Turkish Industries: Evidence From Survey Data." *Developing Economies* 46 (4): 363–85. [46]

Sheremirov, Viacheslav. 2020. "Price dispersion and inflation: New facts and theoretical implications." *Journal of Monetary Economics* 114: 59–70. [3]

Stahl, Harald. 2005. "Price Setting in German Manufacturing: New Evidence from New Survey Data." *European Central Bank Working Paper Series* No. 561. [46]

Virbickas, Ernestas. 2011. "Price Setting in Lithuania: More Evidence from the Survey of Firms." *Bank of Lithuania Working Paper Series* No. 11. [46]

Zentralverband des Deutschen Friseurhandwerks. 2021. "Die wirtschaftliche Lage des Friseurhandwerks." *Spotlight Friseurhandwerk 2020/2021,* 8–17. [11, 16]

Chapter 2

Fundamental Stock Price Cycles*

2.1 Introduction

Why does the stock market predictably yield lower returns after a boom? Campbell and Shiller (1988) show that, due to the return predictability of the stock market, most of the fluctuations in aggregate stock prices can be explained by expected movements in the discount factor: if the future return is expected to be low, future dividend payments are expected to be more valuable — they are discounted by less — , which appreciates the value of the stock asset today. Fluctuations in future dividend growth, instead, explain only a smaller part of the variance in aggregate stock prices. This finding has been reiterated by Cochrane (2011) for the post-war U.S. economy, and has been reproduced by Kuvshinov (2022) for 17 advanced economies since 1870.[1] Thus, the main driver of stock price fluctuations is return predictability — the pattern of boom-bust cycles — that is unexplained by movements in dividend growth. In this paper, I offer a novel explanation for this main empirical pattern of the stock market. I model the stock market within a business cycle model. Thereby, the theory is also able to explain the positive correlation of stock price booms with business cycle booms that I find in the data.

* I am grateful to Christian Bayer, Marco Bassetto, Jonathan Heathcote, Thomas Hintermaier, Rustam Jamilov, Joachim Jungherr, Keith Kuester, Jochen Mankart, Edouard Schaal, Mirko Wiederholt, and Donghai Zhang for their helpful comments and suggestions. I likewise thank seminar participants at the University of Bonn and at the Deutsche Bundesbank, participants at the 2021 Warwick PhD Conference, at the ECONtribute Rhineland Workshop 2021, at the 15th RGS Doctoral Conference, at the YES 2022, and at the Konstanz Doctoral Workshop 2022. I am grateful to my discussants Matthias Kaldorf and Tobias König. I thank Alina Bartscher for her generous help. I gratefully acknowledge financial support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the RTG 2281 - The Macroeconomics of Inequality and from the ERC-CoG project with agreement ID 724204.

1. Kuvshinov (2022) finds that in the sample he considers, the discount rate news and the cash flow news each explain about half of the price dividend ratio.

There are three main factors that allow the theory to explain return predictability of stocks. First, markets are incomplete. As Constantinides and Duffie (1996) show theoretically, movements in the stochastic discount factor of investors that are large enough to explain the observed stock price fluctuations can be explained by risk in the investors' incomes that is uninsured. If investors instead were only subject to aggregate risk, aggregate consumption, in order to yield similar movements in the discount factor of the representative investor, would have to fluctuate by several orders of magnitude more than empirically observed, or risk aversion would have to be unrealistically high. My theory adds to this that households' incomes in general come from various sources: households receive labor income, but also asset incomes. Since asset incomes have a procyclical pattern, this addition to the theory is crucial for explaining not just a high fluctuation in stock prices, but the structure of boom-bust cycles.

Second, households can save both in liquid and illiquid assets. Illiquid assets can only be traded infrequently. If all assets were liquid, wealthy households could never be liquidity-constrained, and uninsurable income risk could only affect households without any savings. With two asset classes, households can be wealthy in the illiquid asset and still be liquidity-constrained, so that the income they receive from holding their assets can influence their consumption growth, and thus their stochastic discount factor. The notion of infrequently traded assets is established in the literature about the importance of incomplete markets and portfolio choice for macroeconomics (see, e.g. Kaplan, Moll, and Violante (2018)). The innovation of this paper with respect to this literature is to divide the assets that allow households to hold a share of the profits that accrue in the production process, that is, claims to equity, in a liquid and an illiquid category.[2] In combination with idiosyncratic income risk, this leads to public (liquid) equity being less risky than private (illiquid) equity, similar to the incomplete markets-model by Angeletos (2007). Sorensen, Wang, and Yang (2014) find that the illiquidity of private equity is an important component of its return risk. Sagi (2020) finds the same for investments in real estate, and rationalizes the illiquidity of the market within a search and matching model.

Third, households anticipate future changes in technology and productivity. Examples for this are the anticipation of the adoption of the internet in firm-customer relationships during the 1990s stock market boom, or the anticipation of the development of a vaccine during the Covid-19 pandemic. I find that such "news shocks" are much more consequential when households choose between liquid and illiquid assets in their portfolios, than in models with only liquid assets. When investing in

2. In Alves et al. (2020), the authors analyze the implications of partly liquid profits for the transmission of monetary policy shocks in the two-asset HANK model. However, the liquid profits are not traded in that model, but accrue to households proportional to their idiosyncratic productivity. The contribution of the present paper is to analyze the valuation of *traded* liquid profits in response to news shocks.

an illiquid asset, households expect to not be able to trade it for several years. As a consequence, when households receive the information that illiquid assets are expected to yield a higher return some time in the future, they attempt to invest *early*, in order to reduce the risk of not being able to invest before the higher returns materialize. News about future higher productivity thus induces some households to shift their portfolio towards illiquid assets. This comes at the cost of higher *idiosyncratic* risk for these households.

These three pillars of the theory work together to explain a typical boom-bust cycle on the stock market, that is, as in the data, caused by time-varying discount rates. It starts with the news about a temporary increase in productivity growth some time in the future. The stock market appreciates at the onset of the news, as higher productivity implies higher dividend growth in the future. However, stock prices continue to rise during the following years, when the higher productivity has not yet materialized (the "anticipation phase"). The reason is that the equilibrium return on liquid assets is high during that time: households are less willing to save in liquid assets, since they expect higher future incomes, which lowers their precautionary-savings motive. The growth in labor and asset incomes occurs already during the anticipation phase, as the wealthy households' shift from liquid assets to illiquid assets in their portfolios causes an investment-driven business cycle boom. Since stocks are liquid, a higher return on liquid assets implies, for a given dividend-stream, a gradual growth in stock prices.[3]

Once the temporary acceleration in productivity growth materializes, the return on illiquid assets peaks and recedes back to its steady state value. Therefore, households who hold most of their wealth in the illiquid asset and thus are subject to high idiosyncratic risk at the same time face falling incomes — due to their declining asset income. Consequently, they demand more liquid assets for self-insurance, which depresses the return on liquid assets in equilibrium. This implies that stock prices, for a given dividend-stream, persistently fall after the productivity growth has peaked. The theory thus identifies the marginal trader of the stock price cycle: households with high illiquid wealth, who face the largest consumption fluctuations due to changes in illiquid asset returns, which cause large fluctuations (in absolute terms) in their asset income. The appreciation of the stock market during the anticipation phase is in part due to higher expected dividend growth, and in part due to the expected movements in the equilibrium return on liquid assets — the rate at which stock dividends are discounted.

The theory can be understood as proposing a time-varying *illiquidity premium*, rather than a time-varying aggregate risk premium (Campbell and Cochrane, 1999;

3. It is important to note that the higher return on liquid assets is *expected* by households. As long as the return on liquid assets is expected to remain high, the demand for stocks is reduced, due to a no-arbitrage condition. The closer one gets to the moment where the return falls, the higher is the demand for stocks, and the higher is the stock price.

Bansal and Yaron, 2004), as the main explanation for stock price fluctuations. I define the (ex-ante) illiquidity premium as the (expected) difference between the return on illiquid assets and the return on liquid assets.[4] The illiquidity premium varies due to the time-varying propensity to bear consumption risk at the *individual* level. The expectation of higher future returns on illiquid assets induces wealthy households to bear more consumption risk, by holding more illiquid assets, in the anticipation phase. Thus, the illiquidity premium is low in the anticipation phase. Once the investment and stock price boom subsides, the illiquidity premium rises above its steady state value, since the marginal traders have more illiquid portfolios and face falling incomes, so that liquid assets become more valuable. Since stocks are liquid, the growth in stock prices correlates negatively with the illiquidity premium. Kuvshinov (2022) compares the risk factor that causes fluctuations in stock prices with the risk factors that drive fluctuations in returns to housing and corporate bonds. He finds that the risk factors do not comove across asset classes. This finding is inconsistent with theories that hinge on aggregate risk, which affects all those assets, as the main cause of asset price fluctuations, while my theory can accomodate this evidence: stocks differ from housing and corporate bonds in their higher liquidity. In this paper, I analyze only the effect of the illiquidity premium on the stock price cycle, and abstract from an aggregate risk premium.[5]

In my model, stocks are claims to a share of the profits of the monopolistically competitive firms in the economy. Stock-supply is time-invariant (normalized to one), so that I abstract from financing decisions of firms. The behavior of stock prices is solely explained by the households' demand for stocks. The value of the stock asset is determined by two properties: the expected dividend stream, and its liquidity. Since stocks are liquid, households compare it to other liquid assets, like government bonds.[6] A consumption-based explanation of stock prices is in line with the empirical finding that among all commonly traded financial assets, public eq-

4. In the literature, this is also called the liquidity premium, see e.g. Bayer et al. (2019).

5. I solve the quantitative model up to first-order (the news about technologial progress is an unexpected "MIT-shock"). I conjecture that solving the model non-linearly would not diminish the role of the time-varying illiquidity premium for explaining rising stock prices: in stock price booms, the share of wealth that is held in stocks rises (mechanically, but also by active stock-investment; section 2.5 provides evidence for this). If stocks are risky, this increases the riskiness of households' portfolios, which in turn increases the risk premium households are willing to pay, and puts *downward* pressure on stock prices. This mechanism is well-known for consumption-based asset pricing models with aggregate risk. Chien, Cole, and Lustig (2012) generate rising stock price booms with risky stocks by having "Mertonian" investors, who price the asset, *sell* their risky shares to intermittently rebalancing investors during a boom, which circumvents the problem. For the housing boom of the early 2000s, Favilukis, Ludvigson, and Van Nieuwerburgh (2017) conclude that relaxed financing constraints, that is, an institutional change that makes housing an individually less risky asset, is needed to model a simultaneous house price boom and rising share of housing equity in households' portfolios.

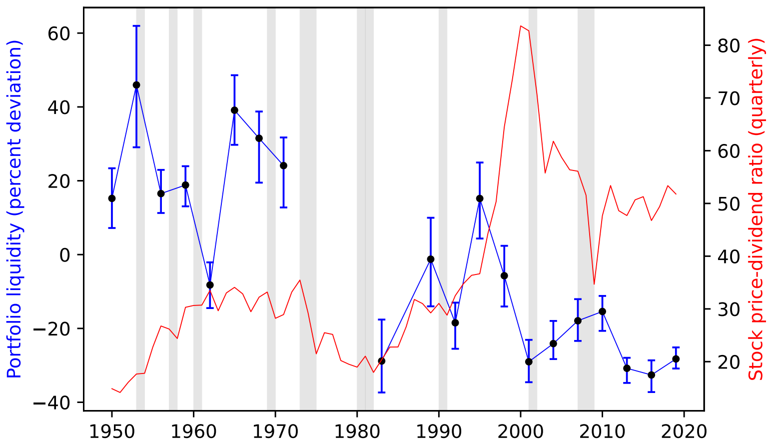
6. Such an arbitrage condition between stocks and government bonds is assumed, e.g., in Caballero and Simsek (2020).

uity is directly held by households the most, followed by treasury bonds (Haddad and Muir, 2021). The authors also find that excess returns on stocks and treasury bonds are the least predictable — with a coefficient close to 0 — by the health of intermediaries in the economy. Hence, a theory that explains price fluctuations with frictions on the level of financial intermediares, or institutional frictions, is less convincing for the case of stocks.

Liquid assets, like stocks or bonds, do not enter into the production function of the representative firm. Since I abstract from the firm's financing choices and frictions in the model, a stock price boom also does not ease the firm's financing constraints. Illiquid assets, instead, aggregate to the capital stock in the economy, which is the most important production factor for the firm. In that sense, stocks in the model are "unproductive". However, the only reason why liquid assets have value in the economy is a financial friction on the household side, namely, that capital is illiquid. When the returns on liquid assets like stocks rise during a stock price boom, households can afford to shift more wealth to the illiquid asset. Hence, a growing stock market is *indirectly* productive in the model by the virtue of households who use the additional liquid asset income to invest more in productive capital.[7] I calibrate capital in the model to fixed assets in the U.S. Bureau of Economic Analysis tables. A third of fixed assets is housing. At the same time, Asker, Farre-Mensa, and Ljungqvist (2014) estimate that about half of investment in the non-residential sector is carried out by noncorporate private firms. In sum, around two thirds of productive capital in the U.S. economy is not financed by stock issuance. A part of the valuation of fixed assets is also due to (unproductive) market power by firms, which is passed through exclusively to stock-owners in the model. For these reasons, I view the modeling of stocks as unproductive assets as a useful approximation to the data.

The mechanism I propose to explain stock price boom-bust cycles has several testable implications. First, the theory hinges on a no-arbitrage condition between government bonds and stocks. I document that in U.S.-data, stock price-growth is positively associated with higher returns on government bonds, and yearly moving averages of stock returns and bond returns also comove (correlations are around 0.2). The low correlation is to be expected, since I use realized returns, which contain surprise shocks that add noise. Second, falling stock prices should coincide with falling capital rents, since the latter cause the fall in asset incomes of the marginal traders in the model. I find a positive correlation of about 0.26 in the data. Finally, employing the extended Survey of Consumer Finances-dataset provided by Kuhn, Schularick, and Steins (2020) that ranges from 1950 to 2016, I provide evidence

7. Since there is a negative borrowing limit, the effect of government bond interest rates on investment is non-monotone — when the rates are expected to grow too high, households who borrow in the liquid asset will have to pay too much on their debt, which deters them from investing into illiquid assets. I revisit this case when discussing the effects of monetary and fiscal poliy.



Notes: Survey evidence from SCF+ (Kuhn, Schularick, and Steins, 2020), stock market data from S&P500 (Robert Shiller), recession years (grey areas) by NBER. Portfolio liquidity is defined as the ratio of liquid assets by total wealth. Left axis shows the relative deviation of portfolio liquidity of households whose main share of income (>75%) is capital income, from portfolio liquidity of the top 10% of wealth distribution. Whiskers are 68%-confidence intervals.

Figure 2.1. Portfolio liquidity of "rentiers" and the stock market

(see figure 2.1) that households whose income mainly stems from capital income — the "rentiers" in the economy, who I identify as the marginal traders of the stock market — decrease their portfolio liquidity in stock price-booms, and increase it in stock price-busts, like the model predicts.[8]

In section 2.2, I demonstrate the mechanism in a highly stylized, but tractable model, that only considers the capital-wealthy subset of households that are the marginal traders of the stock market. In section 2.3, I explore the mechanism quantitatively in a heterogeneous agent New Keynesian ("HANK") model with two assets. My analysis of stock price cycles in a general equilibrium setting uncovers the dependence of the stock price cycle on the elasticity of liquid asset supply. Since government bonds are liquid assets, they are in less demand once the news about higher productivity arrives: at the onset of the news, households immediately shift from bonds to stocks, since the discounted sum of future dividends increases. During the anticipation phase, wealthier households additionally shift from liquid to illiquid assets in their portfolios. Thus, the fiscal authority faces a pressure to reduce its balance sheet. However, the government can also induce higher inflation and allow for higher output gaps late in the anticipation phase, thereby raising (inflation) taxes. These policies lower asset incomes early and decrease profits late in the anticipation

8. Specifically, the change in the relative portfolio liquidity of the "rentiers" over subsequently sampled years, and the growth in the stock price-dividend ratio, are negatively correlated at about −0.27. See section 2.5 for a discussion of the evidence.

phase, which cuts into the income of wealthy households. Since they have the highest marginal propensity to invest, the wish to substitute liquid assets (like bonds) for capital weakens in the aggregate. While this equilibrates the bond market, it prevents the wealthier households from generating an investment-driven boom during the anticipation phase. Conversely, a policy that stabilizes inflation and "smoothes out" the increase in the real rate on liquid assets over the anticipation phase, is only consistent with a strong reduction of the aggregate liquid asset supply.[9] Fundamentally, the real interest rate is "smoothed out" over the anticipation phase by allowing for the crowding out of unproductive liquid assets by productive capital, so that households' incomes increase long before the productivity boost. This comes at a cost of increased consumption risk for wealthy households, which they are willing to trade off against the anticipated higher future return on their wealth.[10]

I show that a sizable share of stock price fluctuations can be quantitatively accounted for by two alternative news shocks about two kinds of fundamentals: accelerated growth in total factor productivity (TFP), or a higher capital share in the production process. Importantly, however, these fundamental changes should be (expected to be) *temporary*, since only then investment in the anticipation phase is urgent enough to drive the business cycle. The anticipation of a temporary productivity boost can be motivated by the 1990s "dot-com" boom in the U.S., which was a R&D-investment boom (Brown, Fazzari, and Petersen, 2009), and is thought of by many as an anticipation-driven boom (Jermann and Quadrini, 2007; Ben Zeev, 2018). Since R&D capital, like other "intangible" capital, depreciates relatively fast (due to technological obsolescence and increased competition, c.f. Li and Hall (2020)), the households expect the future productivity acceleration to be temporary.[11] Alternatively, Karabarbounis and Neiman (2014) find that the decline in the relative price of IT investment goods lowered the labor share in recent decades. In Karabarbounis and Neiman (2019), they argue that the most plausible explanation for the "excess" value added is an increase in capital rents, rather than an increase in firm's profits (or markups), or a large share of unmeasured, intangible capital. The present paper

9. Domínguez-Díaz (2021) analyzes a HANK-model with portfolio choice, where the main provider of liquidity is the banking system. His analysis shows that, if the banks are subject to a moral hazard-problem, and are at their borrowing constraint, the supply of liquidity rises in the illiquidity premium, since the banks' profitability increases with the spread between capital returns and returns on deposits. Hence, in an environment with constrained banks, the low illiquidity premium during the anticipation phase of a news-induced boom would lower the supply of liquidity also through that channel, independent of the fiscal side.

10. The effectiveness of investment in amplifying booms in HANK models with portfolio choice, where capital is illiquid, is highlighted in Auclert, Rognlie, and Straub (2020). Luetticke (2021) shows the importance of the redistribution towards high marginal propensity to invest (MPI) households for the transmission of monetary policy shocks.

11. Bianchi, Kung, and Morales (2019) also interpret the 1990s boom as driven by R&D investment which provides spillover effects, and interpret the bust after 2000 as a shock to equity financing, as the value of pledgable capital falls. I will discuss disappointed expectations in section 2.5.

relates to their analysis in two ways: on the one hand, it provides a rationale for a time-varying wedge between the real interest rate of government bonds, and the capital rents necessary to account for the excess value added — the illiquidity premium. Additionally, Karabarbounis and Neiman (2019) provide evidence that the share of value added attributable to IT capital declined after 2000, lending credence to the idea that the 1990s boom was driven by the expectation of a *temporary* increase in capital returns. Smith et al. (2019) document a rising share of value added that accrues to business owners who pass through firm profits to their own (capital) income, making them the top earners in the economy. My findings suggest that these households are the marginal traders of the stock market.

Related literature. Some of the channels through which time-varying idiosyncratic risk and heterogeneous portfolios affect equilibrium prices in this paper are also present in the more stylized model by Fernández-Villaverde, Hurtado, and Nuño (2022): There, a representative financial expert rents capital to firms, financed via the issuance of risk-free bonds, and consumes nothing but the returns from her investment. At the same time, households that are subject to idiosyncratic income risk can save in the bonds issued by the financial expert, but are not able to access the capital market on their own. The critical financial friction is that the financial expert cannot share her capital risk with the other households. Similarly to my model, the dominance of capital income in the financial expert's budget provides a strong incentive to increase their investment, by issuing more debt, upon the expectation of higher excess returns on capital. Fernández-Villaverde, Hurtado, and Nuño (2022) solve the model globally and show that in the "high leverage"-stochastic steady state, recessions caused by negative aggregate capital shocks are more severe than in the "low leverage"-stochastic steady state. While that paper focuses on "supercycles", where the state of the economy fluctuates between these two stochastic steady states, I analyze the effect of these channels at the business cycle-frequency, with a special focus on anticipation and the valuation of liquid assets, and in a richer general equilibrium-setting.

Papers within the consumption-based asset pricing framework have shown to be able to generate stock return-predictability by imposing special preferences (Campbell and Cochrane, 1999), or special stochastic processes that households face (Bansal and Yaron, 2004), among others.[12] Kekre and Lenel (2022) explain the stock market response to a monetary policy shock through its effect on the risk premium within a HANK model with portfolio choice, and calibrate it using the Survey of Consumer Finances, as I do in this paper. They assume heterogeneity in risk aversion, which allows for the comovement of investment and stock prices: when a shock redistributes towards households with lower risk aversion, investment rises, while the risk premium falls. In my model, the heterogeneity in the *individual* riskiness of

12. In many models, a reduced-form "discount rate shock" is introduced instead; see for a discussion of the leading asset pricing models also Gormsen (2021).

portfolios, by means of their liquidity, comes about endogenously (as an outcome of optimal portfolio choices in response to idiosyncratic and aggregate shocks).

A burgeoning literature challenges the assumption of rational expectations in explaining asset price fluctuations, and especially asset price "puzzles", on the basis of survey data (Adam, Marcet, and Beutel, 2017; Bordalo et al., 2020; Beutel and Weber, 2022). This literature finds evidence for irrational optimism about future stock returns during stock price booms. As solution, variants of subjective expectations that have an extrapolative component, and may be rational for forecasting future stock prices under information asymmetry, are proposed by many authors. However, in this forecasting exercise, households would form only partial equilibrium expectations, instead of forming conditional expectations, given the observed positive correlation of stock price- with business cycle-booms. Adam and Merkel (2019) develop a model with learning where surprise productivity shocks can trigger an endogenous belief propagation that gives rise to boom-bust cycles in stock prices and investment. The mechanism I propose abstracts from learning, as the anticipation of higher future returns on capital is modelled as an exogenous news shock. Through the lense of my model, and in contrast to the results in Adam and Merkel (2019), the expectations about temporarily higher future productivity are accurate on average (I discuss noise shocks in section 2.5). While my model fails to generate observed irrational swings in expectations[13], it matches observed patterns in households portfolios over the stock price cycle.

Following Krusell, Mukoyama, and Smith (2011), the analytical literature on asset pricing in heterogeneous agent models often makes critical simplifications (e.g. Ravn and Sterk (2017) and Broer et al. (2019)): the rate on the liquid asset, which is in zero net supply, is such that the marginal trader (or "marginal saver") optimally holds no assets. Since the impact of aggregate risk on households' budgets is small, the marginal trader can be identified from the stochastic process of idiosyncratic endowments. Often, a dichotomy between "capitalists" and "workers" is introduced, where only the latter are subject to idiosyncratic shocks, so that the worker with the highest income today prices the liquid asset each period. In the analytical HANK-model of Bilbiie (2019, 2020), the roles are reversed: households that receive the returns on capital in the economy price the liquid asset, while the other households do not have access to markets and just consume labor income and transfers. Households switch roles stochastically.

The setting of Bilbiie is closer to my results from the numerical HANK model: at the peak of the stock price cycle, the liquid rate is set by (capital)-wealthy households who want to self-insure. The main difference from the (analytically tractable) model of Bilbiie is that there is a heterogeneity among households in the unconstrained state that plays a role over the cycle: Households choose to dissave their

13. However, Bordalo et al. (2020) gives empirical support to the importance of long-run expectations about fundamentals during stock price-booms.

liquid asset holdings, as they want to hold on to their capital stock, in anticipation of higher returns. Therefore, more households end up closer to the constrained state, which makes them more susceptible to income risk. Since stock price fluctuations are an aggregate phenomenon, it appears reasonable that an explanation for co-varying returns on liquid assets hinges on an aggregate component of income (i.e., the dynamics of capital rents). News about a temporary increase of this income *endogenously* generates time-varying idiosyncratic risk, i.e. exposure to idiosyncratic risk that varies with the stock price cycle, by virtue of the optimal portfolio choices of households.[14] In sum, I find that time-varying idiosyncratic risk, which has been shown to generate amplification of business cycles when poor households price the asset (Ravn and Sterk, 2017), can also yield amplification when a certain subset of wealthy households prices the asset, at which point a change in capital income, instead of labor income, becomes the decisive factor.

Finally, this paper relates to the question of what drives the business cycle. There is a long-standing literature on news-driven business cycles, starting with Beaudry and Portier (2004, 2006), who employ stock prices to empirically identify news shocks.[15] Christiano et al. (2010) show that the New Keynesian model can generate booms from news shocks when monetary policy follows a naive Taylor rule. The reason is twofold: higher future productivity anchors inflation expectations at a level below steady state, and sufficiently high price stickiness lowers prices already in the anticipation phase. As a consequence, the policy rate falls, which boosts demand. Since it is a (inefficiently) low interest rate that causes the boom, the one-asset New Keynesian model does not account for the positive correlation of real rates and stock price growth in the data (see section 2.5). The low real rate is inefficient, since a positive news shock, which increases consumption of households in the future, raises the natural rate today. In the model with heterogeneous agents, instead, the business cycle boom coincides with a high real rate. Liquid savings are not held down by an inefficient monetary policy; instead, households want to save less, and consume more, due to higher incomes in the anticipation phase. The economy is more productive ahead of the exogenous technology shock, as households increase their capital stock early. Households are willing to have more illiquid assets in their portfolios — at a lower premium, and a higher consumption risk — since they expect higher future returns on their wealth. For satisfying the higher demand for capital goods, output has to rise. The resource constraint of the economy is partly satisfied by the crowding out of government expenditure, and partly by higher labor

14. In a related paper, Bilbiie, Känzig, and Surico (2022) place emphasis on the fact that redistributing capital income to constrained households amplifies demand shocks, as capital income is procyclical. They model the redistribution exogenously (via fiscal policy), while in the present model, anticipation generates the same kind of "redistribution" endogenously, only in reverse: households with a large share of capital income choose to become more constrained.

15. Beaudry and Portier (2014) give a comprehensive summary. The news are typically about long-run productivity in that literature, while I consider news about a temporary productivity boost.

supply of workers, who earn higher real wages as markups fall (the standard New Keynesian mechanism).[16]

The structure of my paper is as follows: In section 2.2, I illustrate the main mechanism to generate a stock price cycle from anticipation in a simple, tractable heterogeneous agent model, making use of the stylized framework developed by Challe and Ragot (2016). In section 2.3, I describe the full quantitative HANK model, which is taken from the literature[17] and ammended to include liquid stocks and news shocks. In section 2.4, I show that technology news — either about TFP or factor share shifts — generate a stock price and business cycle boom in this model, and analyze the importance of the hetereogeneous agent and two-asset structure (liquid and illiquid assets) for obtaining the results. In section 2.5, I document that aggregate data on asset returns, as well as survey data of households' portfolio choices over time, are consistent with the mechanism I propose, and I investigate the quantitative success of the mechanism as the main driver of stock price fluctuations in a simulation exercise under different specifications of dividend cyclicality and news accuracy. Section 2.6 concludes.

2.2 Illustration of the stock price cycle

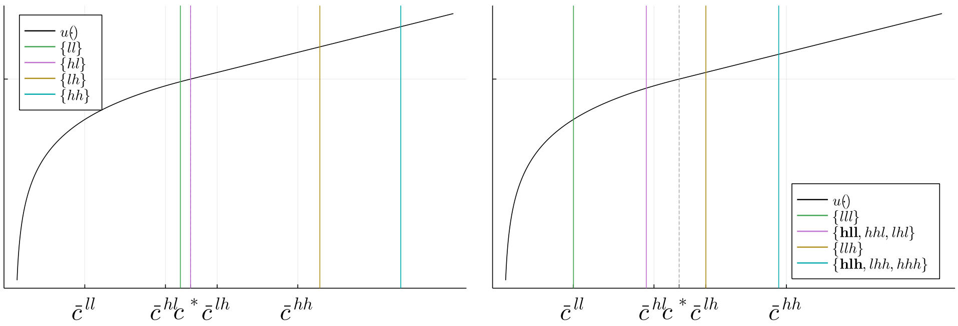
In this section, I illustrate the mechanism how wealthy hand-to-mouth households can drive down the equilibrium return on liquid assets. I abstract, however, from portfolio choice between liquid and illiquid assets. I analyze a situation in which all households hold little liquid wealth relative to their income risk, i.e. they are poorly insured, while their illiquid wealth is high. In the full model, this situation applies to a small subset of households, as a result of their portfolio choice, at the end of the anticipation phase. In addition to the technology news, in this simplified setting agents are also subject to a shortage of liquidity in the anticipation phase.[18] I apply

16. Görtz, Tsoukalas, and Zanetti (2022) build a RANK model with financial frictions and show that a financial accelerator enables news to cause a business cycle boom. In an estimation exercise, they find that news shocks account for about half of the fluctuations in real business cycle variables. Instead of the time-varying markups of the New Keynesian model framework, one could adopt other explanations for rising labor hours during the anticipation phase of a news-induced boom. McGrattan and Prescott (2010) argue that the 1990's increase in labor hours *preceding* higher wages can be explained by workers investing "sweat capital". In a similar vein, the notion of illiquidity could be widened to include (a part of) human capital, which workers would be willing to invest into more when the expected returns are high.

17. I am building on the HANK model with portfolio choice by Bayer, Born, and Luetticke (2022) which is estimated using U.S. business cycle and inequality data from 1954 onwards.

18. This is necessary to bring the market for liquidity into equilibrium: the news about future productivity lowers the demand for liquid savings. The real rate is bounded above by the inverse of the time preference rate, and therefore cannot rise enough to fully offset the lack of demand for liquid assets.

(a) Capital rate increase ($t = 1$) **(b)** Capital rate decrease ($t = 2$)



Notes: $\{...y_{t-1}y_t\}$ denotes the history of income shocks at time t, with $y_t \in \{l, h\}$, $l < h$. $\bar{c}^{yy'}$ denote the optimal consumption levels at all possible states $\{yy'\}$ of the ergodic wealth and income distribution.

Figure 2.2. Optimal consumption levels

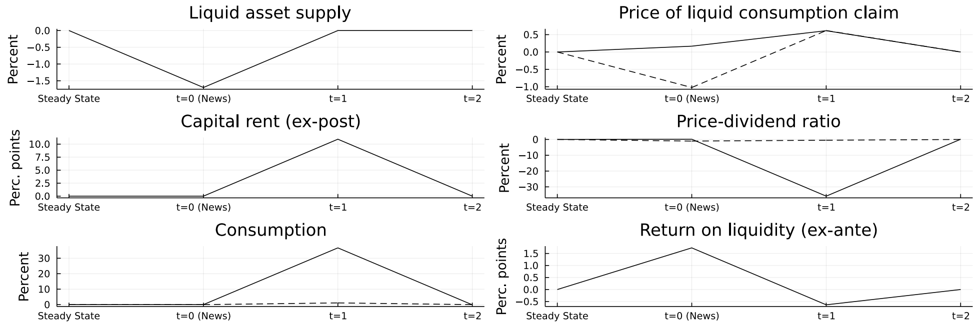
the technique by Challe and Ragot (2016) to make heterogeneous agent models with a non-degenerate wealth distribution tractable.

Consider a unit mass of households who hold two assets, a liquid asset and a fixed amount of illiquid capital. They can borrow in the liquid asset up to the constraint $\underline{L} < 0$. Their income encompasses interest on the assets they hold, and idiosyncratic income $y \in \{l, h\}$, $l < h$, which follows a stochastic Markov process. They derive utility each period from consumption c, where the utility function $u(\cdot)$ is concave up to point c^*, and has a constant slope afterwards.

The steady state is calibrated[19] such that all households that receive the low income, l, consume at a level below c^*, which is so low that they like to borrow more than $\underline{L}$. On the other hand, all households that receive income h consume at a level above c^*. They like to self-insurance against the risk of receiving the low income, and hence save $\tilde{b}$ liquid assets. Since they consume at the linear segment of the utility function, their marginal utilites are all identical, so that $\tilde{b}$ is the optimal saving for all households with high income. The economy has a liquid outside asset at the positive net supply $L = \pi^l \underline{L} + (1 - \pi^l)\tilde{b}$, where π^l is the unconditional probability of receiving a low income.

The grey lines in figure 2.2 show the steady state consumption allocation in the model. Since all households hold the same (positive) amount of fixed capital, the joint distribution over income and liquid asset wealth has only four mass points in steady state: $(l, \underline{L})$, $(l, \tilde{b})$, $(h, \underline{L})$, and $(h, \tilde{b})$. In a first step, I consider a surprise, one-period increase of the capital rent. I choose a rent increase such that households who change from the high to the low state, (hl), now optimally consume c^* and save a positive amount b' for self-insurance. In other words, they become unconstrained

19. Table 2.A.1 in the appendix shows the values of the parameters.



Notes: Responses to a news shock about higher future capital rents, and a *simultaneous* surprise drop in the asset supply, at $t = 0$. Dashed lines are for the case with a share of $\alpha = 97\%$ perfectly insured, capital-poor households (see section 2.2.2).

Figure 2.3. Impulse responses

due to the higher capital income, but since they face lower capital income again in the future, they want to save part of their income gains. Since the liquid asset supply is constant, the households who receive high income today have to save less than $\tilde{b}$ this period for the bond market to clear. Equilibrium is obtained with a falling return on the liquid asset. For simplicity, I assume the income process to be symmetric[20], so that high-income households will also save the amount $b' < \tilde{b}$. As a result, next period, those households that were lifted out of the constrained state due to the higher capital income are at higher consumption levels than in steady state, while households that received high incomes last period consume slightly less (see figure 2.2b).

In a second step, I consider the case where the capital rent increase is anticipated one period in advance. To keep the solution tractable, I require that the optimal consumption and liquid asset choices stay the same as above, once capital rents change. This implies that unconstrained households decide to fully insure themselves upon the news (since next period, even if they get low income, they will be unconstrained due to higher capital income). Therefore, the equilibrium return on liquid assets has to increase to $1/\beta - 1$ (β being the time discount factor). For this to be an equilibrium outcome, bond supply has to be depressed in the period of the news shock.

Figure 2.3 shows the responses of the return on liquid assets (ex-ante), the price of a liquid consumption claim (i.e. the "stock" price), and its price-dividend ratio, to this experiment. The price of the consumption claim appreciates at the onset of the news. It is also higher than steady state in period $t = 1$, due to the lower liquid asset return then. The price-dividend ratio also increases upon the news. However, the increase in the dividends, once the capital rent rises in the subsequent period,

20. I choose the conditional probabilites of losing a high income (e.g. job separation) and gaining a high income (e.g. job finding) to sum to 100%.

has a larger effect in this calibration. Still, the result illustrates how anticipation can generate a stock price cycle as seen in the data, i.e. high stock prices followed by low returns.

2.2.1 Equilibrium prices from household optimization

Unlike Challe and Ragot (2016), I consider an equilibrium where the household optimization determines the return on liquid savings endogenously. I abstract from risk in aggregate variables. For all households i in the economy, it has to hold that

$$u'(c_t^i) \geq \beta R_t \mathrm{E}_t^i \left[u'(c_{t+1}^i) \right], \tag{2.1}$$

where equation (2.1) holds with equality for all unconstrained households (i.e. households with a high income realization in steady state), and R_t denotes the *ex ante* gross return on liquid savings. In terms of stochastic discount factors $SDF_{t+1}^i := \beta \frac{u'(c_{t+1}^i)}{u'(c_t^i)}$, the equilibrium condition can be written as

$$\frac{1}{R_t} \geq \mathrm{E}_t^i \left[SDF_{t+1}^i \right] \; \forall i. \tag{2.2}$$

The necessary optimality conditions for a pattern of higher than steady state liquid asset returns in $t = 0$, followed by lower than steady state liquid asset returns in $t = 1$, are thus:

- When liquid asset returns are above steady state, $R_0 > \overline{R}$, it must hold that $\mathrm{E}_0^i \left[SDF_1^i \right] < \overline{ESDF}^i$ *for all* households i[21]
- When liquid asset returns are below steady state, $R_1 < \overline{R}$, there *exists* a household j where $\mathrm{E}_1^j \left[SDF_2^j \right] > \overline{ESDF}^j$. j must be unconstrained by the borrowing limit on liquid savings.

The last condition on household j follows, as the level of the liquid asset return is always determined in equilibrium by the households where equation (2.2) holds with equality, i.e. by unconstrained households.

In the example above, both conditions are fulfilled: since in period $t = 1$, households at all wealth and income-positions consume more than in steady state (see figure 2.2a), all households discount the future by more upon the news in period $t = 0$. In period $t = 1$, there are three unconstrained household-types: those with income histories $\{hl\}$, $\{lh\}$, and $\{hh\}$, who all save the amount $b' > \underline{L}$. Since they all have the same expected marginal utility of consumption in period $t = 2$ (under the assumption of the symmetric income process), and the same marginal utility of consumption today (as they consume at the linear segment of the utility function), their

21. $\overline{ESDF}^i$ denotes the steady state expected stochastic discount factor of household i.

expected stochastic discount factor is the same. It is given by (in terms of households with a high income realization today)

$$\mathrm{E}_1^h\left[SDF_2^h\right] = \frac{\beta}{\gamma}\left(\pi^{hl}u'(c_2^{hl}) + (1-\pi^{hl})\gamma\right), \tag{2.3}$$

where $\gamma := u'(c)\ \forall\, c \geq c^*$ is the slope at the linear part of the utility function, and π^{hl} denotes the conditional probability of falling to the low income level from the high income level. The condition $\mathrm{E}_1^h\left[SDF_2^h\right] > \overline{ESDF}^h$ is then equivalent to $c_2^{hl} - \bar{c}^{hl} = R_1 b' - \overline{R}\tilde{b}$ being strictly negative. This is the case, as $R_1 < \overline{R}$ and $b' < \tilde{b}$. Intuitively, the additional income from the illiquid asset holding in period $t = 1$ allows more households to purchase consumption claims for period $t = 2$, which, by goods market clearing, implies that the high-income households expect to consume less, and are therefore willing to save at a lower rate.

2.2.2 Extension: segmented markets

The fluctuation of aggregate consumption in this model economy, where all households hold a large amount of illiquid capital they cannot use to smoothe income shocks, while they are close to the borrowing constraint in liquid assets, is two orders of magnitudes too large compared to quarterly consumption fluctuations in U.S.-data. This is by design: the example was built to illustrate the consumption risk that these "wealthy hand-to-mouth" households are exposed to, who are then pricing the liquid asset. The idea that incomplete markets can generate realistic price fluctuations, while aggregate consumption remains flat, follows the seminal work by Constantinides and Duffie (1996). In order to clarify this contribution of my paper, I now insert more households into the model economy.

The newly introduced households insure themselves perfectly against idiosyncratic income shocks (trading Arrow securities amongst themselves), but cannot access "outside" financial markets, in the sense that they cannot hold capital, and cannot issue debt to households that hold capital (segmented markets). The optimality condition with respect to their liquid asset holding is then

$$\frac{1}{R_t} \geq E_t^{\alpha}\left[SDF_{t+1}^{\alpha}\right] = \beta, \tag{2.4}$$

where α denotes capital-poor households, who do not have consumption risk and thus have a constant discount factor β. By complementary slackness, their saving in liquid assets is zero if inequality (2.4) is strict. If they are indifferent (when $1/R_t = \beta$), I assume that they decide to stay at the borrowing constraint, i.e. $b_t^{\alpha} = 0$. Note that, since R_t peaks at $1/\beta$ in period $t = 0$ in the above experiment, the optimality condition (2.4) is always fulfilled, and the liquid asset is still priced by the uninsured households.

Let α denote the share of perfectly insured households that do not hold capital in the economy. Since they have no income besides l or h, they consume the constant $c_t^\alpha = \pi^l l + (1-\pi^l)h =: \bar{y}$. Households who hold capital, but cannot trade Arrow securities to insure themselves against income shocks, consume $\tilde{c}_t := \sum_{j\in J} \pi^j c_t^j$, where J encompasses all possible income histories $\{lll\}, \{hll\}, ..., \{hhh\}$, and π^j is the probability weight of these histories. The aggregate consumption is then given by $c_t = \alpha\bar{y} + (1-\alpha)\tilde{c}_t$. Choosing α high enough such that the consumption of insured, capital-poor households makes up more than 90% of aggregate consumption in steady state then yields an attenuation of aggregate consumption fluctuations by almost two orders of magnitude[22], while the fluctuation in the returns to the liquid asset remain unchanged (see the dashed lines in figure 2.3). The movements in the price-dividend ratio are attenuated; in the quantitative model, stocks are only claims to a fraction of output, and dividend payments are smoothed out, so that return volatility will have a bigger impact on the price-dividend ratio.

2.2.3 Interpretation

The liquid asset can be thought of as incorporating both, a share of a publicly traded firm, and government bonds. Let B/L denote the aggregate share of government bonds within the liquid asset class. The share of the publicly traded firm yields the return $(q_t^\Pi + c_t)/q_{t-1}^\Pi$, where q^Π denotes the share price, and consumption c is the dividend that the publicly traded firm pays. The analysis above can be thought of as the limit case $B \to L$, since it abstracts from the income effect of the jump of the share price upon the positive news. Still, since both government bonds and stocks are liquid assets, and there is no aggregate risk (the news shock is unexpected), the sequence of prices q^Π is determined through the no-arbitrage condition on the ex-ante returns on stocks: $\mathrm{E}_t(q_{t+1}^\Pi + c_{t+1})/q_t^\Pi = \mathrm{E}_t r_{t+1}^b$, where r^b denotes the real gross return on bonds. This condition arises from the Euler equation with respect to the liquid asset from household optimization. The expected increase in the future dividend appreciates, ceteris paribus, today's stock price. This leads to a "front-loading" of the future expected return of the liquid asset. However, if also the expected future returns on bonds change, the response of the stock price is altered. In the quantitative model, where the news horizon is longer, the initial increase in the stock price due to the news shock is attenuated by an increase in the return to bonds during the subsequent anticipation phase. This comes about through a decrease in the stochastic discount factor of households: the investment boom lets incomes rise, so that households want to save less in the liquid asset. In order for the real rate not to increase too much, government bond supply has to fall in the anticipation period:

22. Choosing $\alpha = 0.97$, which corresponds to the 2.7% of households whose income is dominated by capital income in the full model (see below), yields a peak-increase of aggregate consumption of about 1%, a factor 35 reduction from the case without insured households.

$B_0 < B_{SS}$. In the full model, a fiscal rule determines the bond supply endogenously, reacting to inflation by lowering the supply of bonds. Once the higher capital rent has materialized, households' precautionary savings motive depresses the return to bonds, which increases the preceding stock price.

The capital, on the other hand, can be thought of as a share in a private firm, which is illiquid (alternatively, it can be thought of as a financial asset with a long maturity, like a share in a pension fund, or a physical asset, like a house, that can only be traded infrequently/at a high cost). In this simple example, the return to capital increases exogenously. In the full model, while the capital rent increases due to an exogenous increase in productivity, capital gains increase endogenously: poorer households want to hold the illiquid asset after the stock price-boom, when the illiquidity premium increases. However, these anticipated high returns are *not* front-loaded via intertemporal arbitrage, as for the liquid asset. The reason is the illiquidity of capital. In this section, capital was fixed. In the full model, capital can only be traded each period with some probability. Therefore, in the anticipation period, households do not want to realize possible capital gains of their illiquid asset, since by selling capital, they might forfeit the chance to hold the asset once the capital returns increase.

This is, thus, one fundamental reason why the illiquidity premium falls upon the news of higher future productivity: the higher future returns on liquid assets obtain already in anticipation, while the higher future returns on illiquid assets do not. The other fundamental reason is that illiquid assets are productive; hence, households that hold onto them increase the productivity of the economy, and thereby cause a boom, which raises the return on liquid assets in the anticipation phase.

For the rest of the paper, I solve the response to technology news in a HANK model with portfolio choice, which is calibrated to match micro data on labor income processes and wealth inequality.

2.3 A HANK model of the stock market

The model economy consists of heterogeneous households, who are subject to idiosyncratic income shocks and stochastic (illiquid) capital market access, a production sector with intermediate goods producers, who hire workers and rent capital, and final goods producers, who set prices subject to price adjustment costs, and a government sector, where a monetary and a fiscal authority react to business cycle conditions by setting the nominal interest rate and the bond supply according to fixed rules. In the following, I describe each sector individually, before stating the market clearing conditions and giving the definition of the equilibrium of the model.[23] The model is partly calibrated to aggregate data of the U.S. economy from 1954 to 2015, and partly estimated by Bayesian methods (see Bayer, Born, and Luetticke (2022)). One period denotes one quarter. $\bar{X}$ denotes the steady state value of variable X, and $\hat{X}$ the relative deviation of X from $\bar{X}$.

2.3.1 Households

There is a unit mass of ex-ante identical households, indexed by i, who are infinitely lived, discount the future with the factor β, and optimize their (time-separable) preferences of the Constant Relative Risk Aversion (CRRA) type, $u(x) = \frac{1}{1-\xi}x^{1-\xi}$, over consumption, c_{it}, and leisure. Each period t, they choose consumption, labor supply n_{it}, future holdings of liquid assets, b_{it+1}, and non-negative illiquid/capital assets, k_{it+1}, subject to their budget constraint, the debt limit $\underline{B}$, and the ability of market access to the illiquid asset. Their budget is composed of (after tax) labor income, $w_t h_{it} n_{it}$, profit incomes Π_t^F (final goods firms' rents) and Π_t^U (labor union rents), and asset returns. While w_t denotes the aggregate wage rate, their individual productivity h_{it} is determined stochastically according to

$$h_{it} = \frac{\tilde{h}_{it}}{\int \tilde{h}_{it} di}, \tag{2.5}$$

$$\tilde{h}_{it} = \begin{cases} \exp(\rho_h \log \tilde{h}_{it-1} + \epsilon_{it}^h) & \text{with probability } 1-\zeta \text{ if } \tilde{h}_{it-1} \neq 0, \\ 1 & \text{with probability } \iota \text{ if } \tilde{h}_{it-1} = 0, \\ 0 & \text{else.} \end{cases}$$

$\tilde{h}$ follows a log-AR(1) process, with $\epsilon_{it}^h \sim \mathcal{N}(0, \sigma_{h,t}^2)$, for the times when the household is a worker. Its volatility moves endogenously in response to aggregate hours: $\sigma_{h,t}^2 = \bar{\sigma}_h^2 \exp(\hat{s}_t)$, $\hat{s}_{t+1} = \rho_s \hat{s}_t + \Sigma_Y \hat{N}_{t+1}$. ζ is the probability of becoming an entrepreneur. Entrepreneurs have no labor income ($h_{it} = 0$), but gain a share of the profits of the final goods firms, Π_t^F, and raise funds by emitting stock (see section 2.3.2). With probability ι, they return to being a worker with mean productivity. The average of individual productivity h is normalized to 1. In addition to their wages, workers also receive a lump-sum share of the labor union rent, Π_t^U. The existence of entrepreneurs solves the problem of the allocation of profits that occurs in HANK models. Additionally, it helps the model to match the highly skewed wealth distribution in the data.

The choice of labor supply is greatly simplified by assuming Greenwood-Hercowitz-Huffman (GHH) preferences. They are represented by subtracting the

23. The model setup, with the exception of the modelling of aggregate shocks and the inclusion of liquid stocks, is the same as in Bayer, Born, and Luetticke (2022). This is a shortened version of their exposition.

disutility of work, $G(h_{it}, n_{it})$, from the consumption good *within* the felicity function, i.e. $u(c_{it} - G(h_{it}, n_{it}))$. In this setting, an increase in working hours directly increases the marginal utility of consumption, which offsets the typical consumption-labor tradeoff that arises with separable disutility of labor, namely that more work is only compatible with a smaller consumption level. As a result, optimal labor supply is a function only of the net labor income, independent of consumption.[24] Let $x_{it} = c_{it} - G(h_{it}, n_{it})$ denote the composite demand for consumption and leisure.

Labor income of households is subject to progressive taxation as in Heathcote, Storesletten, and Violante (2017), i.e. net labor income y_{it} is given by

$$y_{it} \;=\; (1-\tau^L)(w_t h_{it} n_{it})^{1-\tau^P}, \tag{2.6}$$

where w_t is the aggregate wage rate and τ^L and τ^P are the level and the progressivity of the tax schedule. Assuming that $G(h, n)$ has constant elasticity γ with respect to n, the first-order condition for labor supply yields $G(h_{it}, n_{it}) = y_{it}\frac{1-\tau^P}{1+\gamma}$. Choosing $G(h_{it}, n_{it}) = h_{it}^{1-\tau^P}\frac{n_{it}^{1+\gamma}}{1+\gamma}$ simplifies the problem further, as labor supply then is only a function of the aggregate (after tax) wage rate. This implies that every household works the same number of hours, $n_{it} = N(w_t)$.

Households can have unsecured debt (i.e. negative holdings of the liquid asset) up to the borrowing limit $\underline{B}$.[25] In this case, their payment to the lender consists of the nominal liquid rate, R_t^L, plus a wasted intermediation cost, $\overline{R}$. Each period, a household's chance of participating in the market for illiquid assets, and being able to adjust k_{it+1}, is given by the fixed probability λ. This trading friction renders capital illiquid. The capital good's price in period t is q_t. From holding capital, households earn a capital rent r_t. The household's budget constraint sums up to

$$c_{it} + b_{it+1} + q_t k_{it+1} \;=\; y_{it} + \mathbb{1}_{h_{it}\neq 0}(1-\tau)\Pi_t^U + \mathbb{1}_{h_{it}=0}y_t^e + (q_t + r_t)k_{it} + \left(\frac{R_t^L}{\pi_t} + \mathbb{1}_{\{b_{it}<0\}}\frac{\overline{R}}{\pi_t}\right)b_{it}, \tag{2.7}$$

where $\pi_t = \frac{P_t}{P_{t-1}}$ denotes realized gross inflation, τ is the average tax rate (see section 2.3.4) and y_t^e denotes the after-tax income of entrepreneurs (see section 2.3.2).

24. Jaimovich and Rebelo (2009) propose a class of preferences that nests both King-Plosser-Rebelo (KPR) and GHH preferences, which was then adopted by Schmitt-Grohé and Uribe (2012) and others in their structural estimation of the impact of news shocks. The reason is that GHH preferences, that shut down the wealth effect on labor supply, are helpful in generating booms from news shocks. Hence, having a preference class where this wealth effect enters as a parameter, which can be estimated, gives news shocks a higher chance to fit the data. Schmitt-Grohé and Uribe (2012), as well as Born and Pfeifer (2014) and Bayer, Born, and Luetticke (2022) in models without news shocks, find that close to GHH preferences provide the best fit to the data.

25. Since all households hold a share of their liquid wealth in stocks, for negative liquid wealth they symmetrically do some of their borrowing in stocks ("short-selling" stocks).

Households maximize the infinite discounted sum of their utility, choosing (composite) consumption, liquid assets, and, if possible, illiquid capital holdings subject to the budget constraint and the inequalities $k_{it+1} \geq 0$ and $b_{it+1} \geq \underline{B}$.

The individual household's optimization problem can be written recursively as

$$\begin{aligned} V_t^a(b,k,h;\Theta,\mathscr{P},\Omega) &= \max_{k',b_a'}\{u[x(b,b_a',k,k',h)] + \beta\,\mathrm{E}_t\,V_{t+1}(b_a',k',h';\Theta',\mathscr{P}',\Omega')\}, \\ V_t^n(b,k,h;\Theta,\mathscr{P},\Omega) &= \max_{b_n'}\{u[x(b,b_n',k,k,h)] + \beta\,\mathrm{E}_t\,V_{t+1}(b_n',k,h';\Theta',\mathscr{P}',\Omega')\}, \end{aligned} \tag{2.8}$$

$$\begin{aligned} \mathrm{E}_t\,V_{t+1}(b',k',h;\Theta',\mathscr{P}',\Omega') &= \mathrm{E}_t[\lambda V_{t+1}^a(b',k',h;\Theta',\mathscr{P}',\Omega')] \\ &\quad + \mathrm{E}_t[(1-\lambda)V_{t+1}^n(b',k,h;\Theta',\mathscr{P}',\Omega')], \end{aligned}$$

where Θ stands for the distribution over asset holdings and productivity, $\mathscr{P}$ are equilibrium prices, and Ω denotes an exogenous shock.

2.3.2 Tradable profit-stocks

Liquid assets consist of government bonds (see section 2.3.4) and profit-stocks. Profit-stocks are claims to a share of smoothed profits of final goods-firms, Π_t^F (see section 2.3.3). The smoothing works through a fixed investment rule: A fraction ξ^Π of excess profits, defined as the deviation from steady-state profits, $\Pi_t^F - \Pi^F$, becomes available for payment to stock-holders and the entrepreneurs (who are the owners of the firms). The rest of the excess profits is saved in a common account, if positive, or withdrawn from the account, if negative. The account is invested in government bonds. Its wealth is denoted by NW_t^Π at end of period t. At times when firms are net borrowers, they do not pay the borrowing wedge that households pay, and are not subject to a borrowing constraint. On average, the account holds zero wealth, $NW^\Pi = 0$. A fraction ξ^Π of the interest payments on the wealth held in the account becomes available to stock-holders and the entrepreneurs, while the rest is reinvested. The smoothed profits then amount to

$$\tilde{\Pi}_t^F := \xi^\Pi(\Pi_t^F + NW_{t-1}^\Pi R_t^b/\pi_t) + (1-\xi^\Pi)\Pi^F \tag{2.9}$$

A fraction of ω^Π of the smoothed profits is traded with a unit mass of shares every period at price q_t^Π. A fraction of ι^Π of those shares retire every period and lose value, while new shares are emitted by the entrepreneurs. The real, after-tax payout to entrepreneurs then becomes

$$y_t^e := (1-\tau^L)((1-\omega^\Pi)\tilde{\Pi}_t^F + \iota^\Pi q_t^\Pi)^{1-\tau^P} \tag{2.10}$$

Ex-ante, the expected return on bonds, R_{t+1}^B, has to fulfill the no-arbitrage condition

$$\mathrm{E}_t\,\frac{R_{t+1}^B}{\pi_{t+1}} = \mathrm{E}_t\,\frac{q_{t+1}^\Pi(1-\iota^\Pi) + \omega^\Pi\tilde{\Pi}_{t+1}^F}{q_t^\Pi}. \tag{2.11}$$

With B_t denoting the total supply of government bonds at time t, the total supply of liquid assets at time t becomes $L_t = B_t + q^{\Pi}_{t-1}$. The average (ex-post) real return on liquid assets is then given by

$$\frac{R^L_t}{\pi_t} = \frac{B_t}{L_t} \cdot \frac{R^B_t}{\pi_t} + \frac{q^{\Pi}_t(1-\iota^{\Pi}) + \omega^{\Pi}\tilde{\Pi}^F_t}{L_t}. \tag{2.12}$$

2.3.2.1 Accounting of capital gains

To be in line with the data (see below), I count capital gains as part of wealth-gains instead of income. Capital gains can accrue from illiquid capital, $\frac{q_t}{q_{t-1}}$, if households can trade their capital holdings in period t, and liquid stocks, $\frac{q^{\Pi}_t}{q^{\Pi}_{t-1}}$. The budget constraint (2.7) is already formalized in a way that illiquid capital gains count as wealth-gains. For the liquid asset, instead, I introduce the liquid asset *value*

$$q^L_t := 1 + \frac{q^{\Pi}_t - q^{\Pi}_{t-1}}{L_t}. \tag{2.13}$$

Subtracting q^L_t from the ex-post real return on liquid assets, $\frac{R^L_t}{\pi_t}$, yields the net return on liquid assets (net of capital gains from stocks and stock depreciation):

$$r^{L,net}_t := \frac{R^L_t}{\pi_t} - q^L_t = \frac{B_t}{L_t} \cdot \left(\frac{R^B_t}{\pi_t} - 1\right) + \frac{\omega^{\Pi}\tilde{\Pi}^F_t - \iota^{\Pi} q^{\Pi}_t}{L_t} \tag{2.14}$$

The value of liquid assets for a household with liquid saving b_{it} can then be rewritten as

$$\left(\frac{R^L_t}{\pi_t} + \mathbb{1}_{\{b_{it}<0\}}\frac{\overline{R}}{\pi_t}\right) b_{it} = \underbrace{\left(r^{L,net}_t + \mathbb{1}_{\{b_{it}<0\}}\frac{\overline{R}}{\pi_t}\right) b_{it}}_{\textit{net liquid income}} + \underbrace{q^L_t b_{it}}_{\textit{liquid wealth}} \tag{2.15}$$

2.3.3 Production sector

The production sector of the economy is made up of labor unions and labor packers, intermediate goods producers, final goods firms, and capital goods producers. Workers sell their labor at the nominal rate W_t to a continuum of unions (indexed by j), who sell their variety of labor to labor packers (for W_{jt}), which produce and sell the final labor service at the price W^F_t. Since unions have market power, they set a price $W_{jt} > W_t$ subject to the demand curve $n_{jt} = (W_{jt}/W^F_t)^{-\zeta} N_t$, and to a Calvo-type adjustment friction. In a symmetric equilibrium, their optimization yields the wage Phillips curve (linearized around the steady state)

$$\log\left(\frac{\pi^W_t}{\bar{\pi}_W}\right) = \beta\, \mathrm{E}_t \log\left(\frac{\pi^W_{t+1}}{\bar{\pi}_W}\right) + \tilde{\kappa_w}\left(\frac{w_t}{w^F_t} - \frac{1}{\mu^W}\right), \tag{2.16}$$

where $\pi_t^W = \frac{W_t^F}{W_{t-1}^F}$ is the gross wage inflation, w_t and w_t^F are the real wages for households and firms, $\frac{1}{\mu^W} = \frac{\zeta-1}{\zeta}$ is the target markdown of wages, and $\tilde{\kappa_w}$ is determined by the probability of wage-adjustment, κ_w[26]. The return to the unions is then given as $\Pi_t^U = (1 - \frac{1}{\mu^W})N_t w_t^F$ in real terms.

The homogeneous intermediate good Y is produced with the constant returns to scale production function

$$Y_t = A_t N_t^{1-\alpha_t}(u_t K_t)^{\alpha_t}, \tag{2.17}$$

where u_t is capital utilization. As is standard, higher capital utilization implies an increased depreciation of capital, $\delta(u_t) = \delta_0 + \delta_1(u_t - 1) + \frac{\delta_2}{2}(u_t - 1)^2$, where $\delta_1, \delta_2 > 0$. A_t and α_t are the level of Total Factor Productivity (TFP) and the capital share, respectively, and follow the stochastic processes

$$\log(A_t) = \rho_A \log(A_{t-1}) + \epsilon_{t-\ell}^{A,\ell} + \epsilon_t^A, \tag{2.18}$$

$$\alpha_t = (1-\rho_\alpha)\overline{\alpha} + \rho_\alpha \alpha_{t-1} + \epsilon_{t-\ell}^{\alpha,\ell} + \epsilon_t^\alpha, \tag{2.19}$$

$$\epsilon_t^A \sim \mathcal{N}\left(0, \sigma_A^2\right), \ \epsilon_t^\alpha \sim \mathcal{N}\left(0, \sigma_\alpha^2\right).$$

Here, $\epsilon_{t-\ell}^{A,\ell}, \epsilon_{t-\ell}^{\alpha,\ell}$ denote news shocks (technology news, either about TFP or the capital share) that households receive in period $t - \ell$, and which are added to (the logarithm of) the fundamental process ℓ periods later (as indicated by the superscript). ℓ is called the anticipation horizon of the news. In other words, the capital share and log-TFP follow an ARMA process, where the moving average part is known ℓ periods in advance, and hence interpreted as news. This interpretation is standard in the literature (e.g. Barsky and Sims (2012) and Schmitt-Grohé and Uribe (2012)). In particular, I assume the news shock to be iid. from the same distribution as the surprise shocks $\epsilon_t^A, \epsilon_t^\alpha$ (i.e., news are not autocorrelated as in Leeper and Walker (2011)).

Let mc_t denote the relative price (compared to the consumption good) at which the intermediate good is sold to final goods firms (which makes it the marginal cost of Y_t for these firms). The intermediate good producers, who operate in a perfect competition environment, set the real wage and the user costs of capital according to the marginal products of labor and capital:

$$w_t^F = (1-\alpha_t) mc_t A_t (u_t K_t / N_t)^{\alpha_t}, \ r_t + q_t \delta(u_t) = u_t \alpha_t mc_t A_t (N_t / u_t K_t)^{1-\alpha_t}. \tag{2.20}$$

Utilization is decided by the owners of the capital goods, who take the aggregate supply of capital services as given, and therefore follow the optimality condition

$$q_t \delta'(u_t) = \alpha_t mc_t A_t (N_t / u_t K_t)^{1-\alpha_t}. \tag{2.21}$$

26. It holds that $\tilde{\kappa_w} = \zeta \kappa_w \frac{\mu^W - 1}{\mu^W}$.

Final goods firms (that are owned by the entrepreneurs) differentiate the intermediate good into final goods of the variety j, y_j. In this environment of monopolistic competition, they maximize profits subject to the demand curve $y_{jt} = (p_{jt}/P_t)^{-\eta} Y_t$ and price adjustment frictions. It is assumed that they discount the future at the same rate as the households, β. Then, their optimization yields a symmetric equilibrium that up to first order is determined by the Phillips curve

$$\log\left(\frac{\pi_t}{\bar{\pi}}\right) = \beta \, \mathrm{E}_t \log\left(\frac{\pi_{t+1}}{\bar{\pi}}\right) + \tilde{\kappa}\left(mc_t - \frac{1}{\mu^Y}\right), \tag{2.22}$$

where $\mu^Y = \frac{\eta}{\eta-1}$ is the target markup, and $\tilde{\kappa}$ is determined by the probability of price adjustment, κ[27]. The rent of the final goods firms is $\Pi_t^F = Y_t(1-mc_t)$ in real terms.

Capital producers transform the investment of consumption goods into capital goods, taking as given the price of capital goods, q_t, and investment adjustment costs. They maximize

$$\mathrm{E}_0 \sum_{t=0}^{\infty} \beta^t I_t \left\{ q_t \left[1 - \frac{\phi}{2}\left(\log \frac{I_t}{I_{t-1}}\right)^2\right] - 1 \right\}. \tag{2.23}$$

Up to first order, the problem reduces to the equation

$$q_t \left[1 - \phi \log \frac{I_t}{I_{t-1}}\right] = 1 - \beta \, \mathrm{E}_t \left[q_{t+1} \phi \log \frac{I_{t+1}}{I_t}\right], \tag{2.24}$$

which determines q_t from the rates of investment. Since all capital goods producers are symmetric, the law of motion for aggregate capital follows as

$$K_t - (1 - \delta(u_t))K_{t-1} = \left[1 - \frac{\phi}{2}\left(\log \frac{I_t}{I_{t-1}}\right)^2\right] I_t. \tag{2.25}$$

2.3.4 Government sector

In the government sector, a monetary authority (the central bank) controls the nominal interest rate on bonds, while a fiscal authority (the government) issues bonds to finance deficits. The monetary policy follows a Taylor rule with interest rate smoothing:

$$\frac{R_{t+1}^B}{\bar{R}^b} = \left(\frac{R_t^B}{\bar{R}^b}\right)^{\rho_R} \left(\frac{\pi_t}{\bar{\pi}}\right)^{(1-\rho_R)\theta_\pi} \left(\frac{Y_t}{Y_t^*}\right)^{(1-\rho_R)\theta_Y}. \tag{2.26}$$

27. It holds that $\tilde{\kappa} = \eta \kappa \frac{\mu^Y - 1}{\mu^Y}$.

$\theta_\pi, \theta_Y \geq 0$ govern the severity with which the central bank reacts to deviations in inflation and the output gap, where Y_t^* is defined as the output that would be obtained at steady state markups. The government issues bonds according to the fiscal rule

$$\frac{B_{t+1}}{B_t} = \left(\frac{B_t}{\bar{B}}\right)^{-\gamma_B} \left(\frac{\pi_t}{\bar{\pi}}\right)^{-\gamma_\pi} \left(\frac{Y_t}{Y_t^*}\right)^{-\gamma_Y}. \tag{2.27}$$

Let $\mathscr{B}_t := \sum_i (w_t n_{it} h_{it} + \mathbb{1}_{h_{it}=0} \Pi_t^F)$ be the tax base for the progressive tax code. The total tax revenue T_t sums up to $T_t = \tau(\mathscr{B}_t + \sum_i \mathbb{1}_{h_{it}\neq 0} \Pi_t^U)$, where the average tax rate τ satisfies

$$\tau \mathscr{B}_t = \mathscr{B}_t - (1-\tau^L)\mathscr{B}_t^{(1-\tau^P)}. \tag{2.28}$$

After the fiscal rule determines the government debt, and taxes are collected, government expenditure G_t adjusts such that the government budget constraint is fulfilled in every period: $G_t = T_t + B_{t+1} - B_t \frac{R_t^b}{\pi_t}$. As a simplification, it is assumed that G_t does not provide any utility to households. This implies that in steady state, in which government expenditure is calibrated to be strictly positive, a fraction of physical production is wasted.

2.3.5 Market clearing and equilibrium

The labor market clears at the competitive wage in (2.20). The market for liquid assets clears when liquid asset demand, which is given by the households' optimal decisions, $L_t^d = \mathrm{E}[\lambda b_{a,t}^* + (1-\lambda) b_{n,t}^*]$, equals the supply of liquidity $L_{t+1} = B_{t+1} + q_t^\Pi$ (as L_t^d is the aggregate over positive and *negative* private liquid asset holdings, the supply of liquid assets is bigger than L_{t+1} in gross terms). Similarly, the price of capital q_t, which is determined by (2.24), clears the capital market when $K_{t+1} = K_t^d = \mathrm{E}[\lambda k_t^* + (1-\lambda) k_t]$ holds (households that do not adjust capital demand the same amount as last period, k_t). By Walras' law, whenever labor, bonds, and capital markets clear, the goods market also clears.

A *recursive equilibrium* is a set of policy functions $\{x_{a,t}^*, x_{n,t}^*, b_{a,t}^*, b_{n,t}^*, k_t^*\}$, value functions $\{V_t^a, V_t^n\}$, prices $\mathscr{P}_t = \{w_t, w_t^F, \Pi_t^F, \Pi_t^U, r_t, q_t, q_t^\Pi, \pi_t, \pi_t^W, R_t^B, R_t^L, \tau_t, \tau^L\}$, stochastic state A_t and shocks $\Omega_t = \{\epsilon_t, \epsilon_t^l\}$, aggregate capital and labor supply $\{K_t, N_t\}$, distributions Θ_t over individual asset holdings and productivity, and a perceived law of motion Γ, such that

(1) Given the functional $\mathrm{E}_t V_{t+1}$ and $\mathscr{P}_t$, the policy functions $\{x_{a,t}^*, x_{n,t}^*, b_{a,t}^*, b_{n,t}^*, k_t^*\}$ solve the households' planning problem, and given the policy functions, $\mathscr{P}_t$, and $\{V_t^a, V_t^n\}$ solve the Bellman equations (2.8).

(2) The labor, the final goods, the bond, the capital and the intermediate good markets clear, and interest rates on bonds are set according to the central bank's Taylor rule.

(3) The actual and the perceived law of motion Γ coincide, i.e. $\Theta' = \Gamma(\Theta, \Omega')$.

To solve the model, I use the methods developed by Bayer and Luetticke (2020).[28]

2.3.6 Definitions and parameter choice

2.3.6.1 Classification in liquid and illiquid assets

For the classification of assets in the data into the liquid and illiquid categories, I largely follow Kaplan, Violante, and Weidner (2014): Illiquid assets, which are assumed to be productive in the model, consist of positive wealth in housing[29], other real estate, pensions and life insurance assets, certificates of deposit, and saving bonds. To compute the net illiquid asset position in the data, illiquid debt is subtracted, namely housing debt on owner-occupied real estate, and other real estate debt. I abstract from car wealth in the analysis.[30]

Conversely, liquid assets comprise the sum of checking, savings and call/money market accounts, as well as holdings in mutual funds, equity and other managed assets, and bonds other than saving bonds. For cash holdings, I use the estimate by Kaplan, Violante, and Weidner (2014). To arrive at net liquid wealth, I subtract credit card debt. As data source, I use the extension of the Survey of Consumer Finances (SCF), SCF+, by Kuhn, Schularick, and Steins (2020), which yields 20 years of cross-sectional data between 1950 and 2016. I restrict the household head to be in working age, i.e. between 22 and 65 years of age.

2.3.6.2 Parameter choice

The portfolio adjustment probability λ is calibrated at 6.5% so that the mean liquidity in households' portfolios roughly matches the data (see table 2.1). This adjustment probability implies an average waiting time of almost four years until capital holdings can be adjusted. This is also consistent with the interpretation of capital holdings as investments in projects that include R&D, in the following sense: as noted by Li and Hall (2020), the average gestation lag is two years, and the yearly depreciation of R&D in the late 1990s and early 2000s is between 20% and 60% in most sectors.[31] Assuming an initial R&D phase of two years on average, in which

28. For the implementation of the methods, I make use of and extend the Julia package "BASEforHANK" by Bayer, Born, and Luetticke (2022), available on https://github.com.

29. This is in accordance with the definition in NIPA, where "the ownership of the house [...] is treated as a productive business enterprise" (BEA, 2019).

30. Consumer durables like cars represent a significant share of poorer households' portfolios (e.g. Guiso and Sodini (2013)); however, they are rather evenly distributed across the wealth distribution, so that leaving them out should not bias the results systematically.

31. Fittingly, Adam and Weber (2023) estimate from product data in the UK the median quarterly turnover rate of consumer products as 13.7%.

Table 2.1. Calibrations

Targets	Calibration	Data	Source
Mean illiquid assets (K/Y)	11.04	11.44	NIPA
Mean gvmt bonds (B/Y)	0.8	1.66 (1.1)	FRED
Government share (G/Y)	0.18	0.21	FRED
Top 10% wealth share	0.68	0.66	WID
Mean portfolio liquidity	0.22	0.25	SCF+
Fraction without capital	0.14	0.22	SCF+
Fraction borrowers	0.125	0.115	SCF+

Notes: In general, data values denote long-run averages from 1950 to 2016. When subtracting federal debt held outside the U.S. from total federal debt held by the public (data availabe since 1970), the debt-to-quarterly-GDP ratio of 1.1 is closer to the model-implied. The wealth share of the top 10% of the wealth distribution is available from the World Inequality Database since 1962. Portfolio liquidity is defined as the ratio of net liquid wealth by total net wealth. To compute it in the data, I delete all observations of households with positive liquid wealth, but non-positive total wealth (0.7% of total observations). Borrowers are defined as households holding a negative net position of liquid wealth.

intangible capital is produced (while physical capital is pledged as collateral), followed by the phase in which goods are produced using the physical capital and the depreciating intangible capital, I arrive at an average holding time of physical capital of four years. In line with the interpretation of the TFP news shock as anticipated spill-over from intangible capital, I likewise set the persistence $\rho_A = 1.0 - 2 \cdot 6.5\%$, i.e. log-TFP depreciates at a quarterly rate of 13%. The steady state capital share in production is set as in Bayer, Born, and Luetticke (2022), $\overline{\alpha} = 0.32$. For the persistence of the shock to the capital share, $\rho_\alpha = 0,9552$, I use the mean probability for firms of losing a low labor-share status within 5 years, as estimated by Kehrig and Vincent (2021).

The size of both of the news shocks will be two times the standard deviations of the surprise shocks (see table 2.2). For TFP, this is the estimated value from Bayer, Born, and Luetticke (2022). For the capital share, I calibrate the size of the news shock to fit to the increase of the capital share from the mid 1990s to 2000. To get an estimate of the capital share, I use the NIPA table 1.12 (National Income by Type of Income) and attribute the components to either profit income ($(1-mc)Y$ in the model), wage income (wN in the model), or capital income (rK in the model). Importantly, corporate profits do not enter into capital income (in the model, profit income and capital income are different), while proprietors' income counts towards capital income. While the concrete estimates differ, this exercise is close in spirit to Karabarbounis and Neiman (2019). I find that, between 1995 and 2000, the capital share increased by about 1 percentage point.

Table 2.2. Estimated parameters (selected)

Parameter	Description	Value
ϕ	Capital adj. costs	0.218
κ	Price stickiness	0.105
μ^Y	Target markup final goods	1.08
κ_w	Wage stickiness	0.133
μ^W	Target markdown wages	1.1
ρ_R	R^B autocorr.	0.803
θ_π	Taylor: inflation	2.614
θ_Y	Taylor: output gap	0.078
γ_B	Fiscal: smoothing	0.157
γ_π	Fiscal: inflation	8.57
γ_Y	Fiscal: output gap	5.73
σ_A	TFP std. dev.	0.00608
σ_α	capital share std dev.	0.005

The degree of profit smoothing is calibrated to match the standard deviation of quarterly dividend growth of the S&P 500 at $\xi^\Pi = 0.05$.[32] The fractions ω^Π and ι^Π are calibrated to yield a share of liquid assets held in stocks of 39%[33] and a quarterly stock price-dividend ratio of 144[34], which implies $\omega^\Pi = 4.7\%$ and $\iota^\Pi = 0.074\%$. I set $\bar{\eta} = 13.5$ and $\bar{\zeta} = 11$, which implies price and wage markups of 8% and 10%, respectively. The real liquid rate is chosen to be 2.5% p.a., while the borrowing penalty $\bar{R}$ is set to 7.5% p.a. in order to roughly match the share of borrowers with the data. The steady state capital rent is $\bar{r} = 3.7\%$ p.a., implying a steady state illiquidity premium of 1.2% p.a. As estimate for the capital rent, I take the series by Gomme, Ravikumar, and Rupert (2011) (including housing, without capital gains, after-tax), which has an average yearly return of 5.6% from 1950 to 2016. Since the model abstracts from long-run technological growth, 2% yearly growth should be substracted from the couterpart of the illiquid rate in the data. The model liquid asset is composed both of government bonds, and more risky equity. Computing real (pre-tax)

32. The standard deviation is calculated from simulating the model subject to random innovations in capital share-news shocks, and (suprise) markup and TFP shocks; see section 2.5.

33. From estimations by Saez and Zucman (2016), when defining bonds as fixed income assets plus net deposits and currency, and stocks as equities (other than S corporations), I get a stockshare of 45% in 1995. From the SCF wave of 1995 (see e.g. Guiso and Sodini (2013)), when defining bonds as cash and fixed income, and stocks as directly held equity, I compute a stockshare of 30%.

34. This is the mean of the S&P 500 stock price divided by dividends amassed over the quarter, from 1948 to 2016. Its inverse, the dividend yield, implies an average return on stocks *without* capital gains of 2.9% annualized. Net of stock depreciation, the return becomes 2.5% p.a., as for all liquid assets in the model economy.

returns on the S&P stock index, 10 year treasury bonds (data source: Robert Shiller) and 3-months treasury bills, I compute average yearly returns of 8.3%, 2.5%, and 0.7%, respectively, over the period from 1950 to 2016. The liquid rate in the model should be considered as a weighted average of these rates.[35]

Tax progressivity $\tau^P = 0.18$ is taken from Heathcote, Storesletten, and Violante (2017), while the tax level $\tau^L = 0.1$ is set to achieve a government share of rougly 18%. With respect to the parameters that Bayer, Born, and Luetticke (2022) estimated, I choose those estimates where inequality data was included in the estimation (the HANK* specification). Importantly, I deviate with respect to the fiscal rule, where I estimate γ_π and γ_Y so that the ratio of the magnitude of the profits- and the magnitude of the bonds-response in the anticipation phase of the news shock matches the respective ratio in the late 1990s.[36] Table 2.2 lists the chosen values for a selection of parameters in the model.

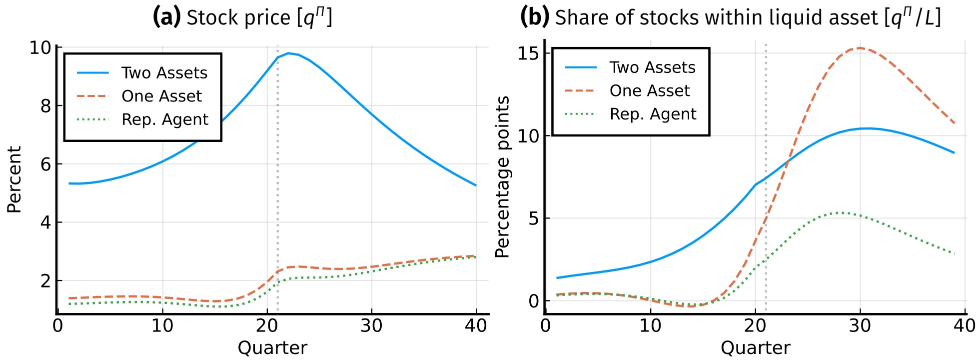
2.4 A news-induced stock price cycle

I consider the following experiment: with an anticipation horizon of 5 years ($\ell = 20$)[37], households become aware that the capital share will increase (by two times its standard deviation). As outlined in the introduction, one can interpret the capital share increase as a temporary change in the production process due to, e.g., more firms employing IT capital. In section 2.4.2, I show that I obtain almost the same impulse responses if the news is instead about a temporary increase in TFP. The reason is that for both news shocks, the expectation of a higher future return on holding capital is identical, which is the decisive impulse to cause the investment-driven boom. The higher expected life-time income that induces households to increase their consumption in the anticipation phase is mainly produced by the higher capital stock,

35. The introduction of aggregate risk, that would allow to differentiate among the classes of liquid assets by their model-implied riskiness, would be an advantage for this part of the calibration. For stock holdings, one should account for the capital gains tax rate of 15-25% over the sample for wealthy households, and discount dividends by 2% long-run growth. Additionally, the financial intermediation wedge of 1.5-2% as calculated by Philippon (2015) reduces the effective rate of financial assets for households.

36. I define the magnitude of the impulse response as the distance between the maximum and the minimum of the percent deviations in the anticipation phase. I constrain both γ_π and γ_Y to the interval $[-10.0, -0.01]$, and search for a global minimum using a Simulated Annealing-algorithm. The estimated bond supply is much more elastic, i.e. the government stabilizes inflation and the output gap more aggressively, than what was estimated by Bayer, Born, and Luetticke (2022) for the whole period since 1960. The reason is that in the late 1990s, the U.S. government strongly reduced their debt.

37. I choose an anticipation horizon of five years to be close to the dotcom-boom example: Karnizova (2012) estimates increased “productivity prospects” around 1995, while in 2000, the NASDAQ peaks.



Notes: Model impulse responses are to news about a temporary capital share-increase in 5 years (quarter 21).

Figure 2.4. Response of stocks across model classes

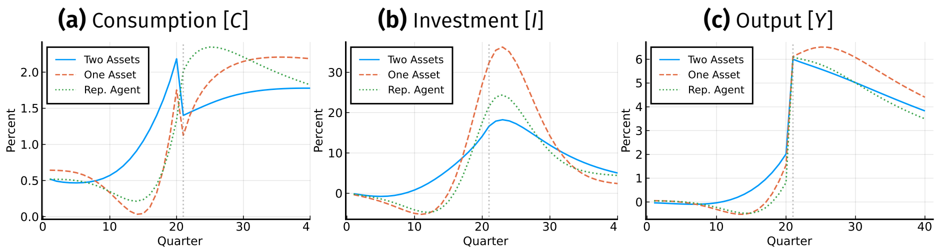
which is accumulated in both scenarios when households rebalance their portfolio towards the productive asset.

Figures 2.4 and 2.5 present the response of the stock price and business cycle variables across three model variants: *Two Assets* denotes the baseline model with heterogeneous agents and portfolio choice between liquid and illiquid assets. *One Asset* retaines the market incompleteness, but takes away the portfolio choice: every household holds a representative portfolio, which is determined by the bond supply rule and the ex-ante illiquidity premium being fixed at a steady state level of zero.[38] This implies that capital becomes liquid in this setting. *Rep. Agent* additionally takes away market incompleteness, and is thus a model of the RANK variety.[39]

Only the HANK model with portfolio choice exhibits a peak in the stock price around the time of the capital share increase (quarter 21), and generates the uniformly accelerating stock price growth that is typical for stock price booms. It is clear that the decisive difference for whether the news drives the business cycle is the portfolio choice. In the full HANK model, richer households start shifting their portfolio towards the illiquid capital after around 2.5 years. This crowds out government bonds (which increases the share of stocks within liquid assets) and thus government expenditures. The higher goods-demand increases wages (since prices are sticky) and lowers the negative labor gap (since wages are sticky), so that households increase their labor supply. Aggregate consumption rises on impact as households expect to have a higher lifetime income, and increases gradually with higher

38. The ex-ante illiquidity premium is defined up to first order as the difference between the expected return on capital and the expected return on liquid assets, $\frac{\mathrm{E}_t(q_{t+1}+r_{t+1})}{q_t} - \frac{R_t^b}{\mathrm{E}_t\,\pi_{t+1}}$.

39. The household's time-preference β is calibrated in the RANK and the One Asset-varieties such that the real rate on the asset in steady state equals that of the baseline Two Asset-model. This implies that also the steady state stock price-dividend ratio is equal across all three varieties.



Notes: Model impulse responses are to news about a temporary capital share-increase in 5 years (quarter 21).

Figure 2.5. Response of the business cycle across model classes

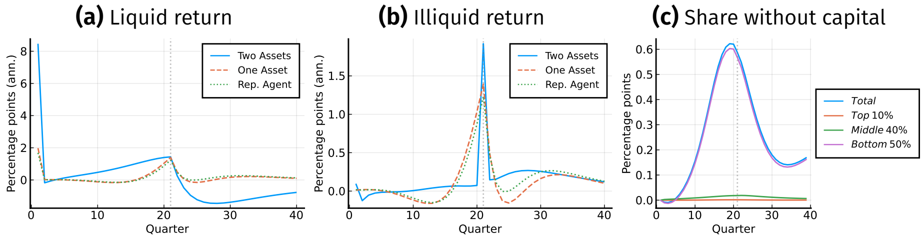
incomes. This gradual consumption increase (by most households) supports a higher real interest rate in equilibrium.

Figure 2.6 shows the response of the (ex-post) returns to the two asset classes, liquid and illiquid assets, across the model varieties. It is clear that without a time-varying illiquidity premium, the expected returns are the same between asset classes (the liquid asset return jumps up at the onset of the news, as the stock appreciates unexpectedly). In contrast, with illiquid capital, the illiquidity premium declines during the anticipation period (the return on liquid asstes increases) and rises after the stock price-peak (the return on liquid assets falls). I also show the change in the share of households without capital. While rich households increase their capital holdings during the boom (intensive margin), poor households are deterred of holding capital by the lower premium (extensive margin). Since the liquiditiy premium rises after the boom, the demand for capital rises, which increases the capital price.

The increasing real interest rate in the anticipation period does not depress the economy; to the contrary, it stabilizes the income of richer households by increasing their return on liquidity (figure 2.7), which enables the middle class (households in between median wealth and the highest wealth decile) to invest in capital, inducing the boom. Is the investment boom driven by the middle class? Households in the top 10% of the wealth distribution own 70% of the capital stock in the economy, so that their incentive to invest in new capital is low. However, if the profit losses of entrepreneurs were higher, or interest income lower, more of the richest household would sell capital to offset their income losses, thereby depressing aggregate investment.

2.4.1 Comparison to the dotcom-boom

Since both the capital share shock as well as several parameters were calibrated to the 1990s in the U.S., I can make a quantitative comparison of the shock responses



Notes: Model impulse responses are to news about a temporary capital share-increase in 5 years (quarter 21). The return on capital includes capital gains. Wealth groups in Panel c) are defined in the *cross-section* each quarter.

Figure 2.6. Response of *ex-post* returns and capital holding across model classes

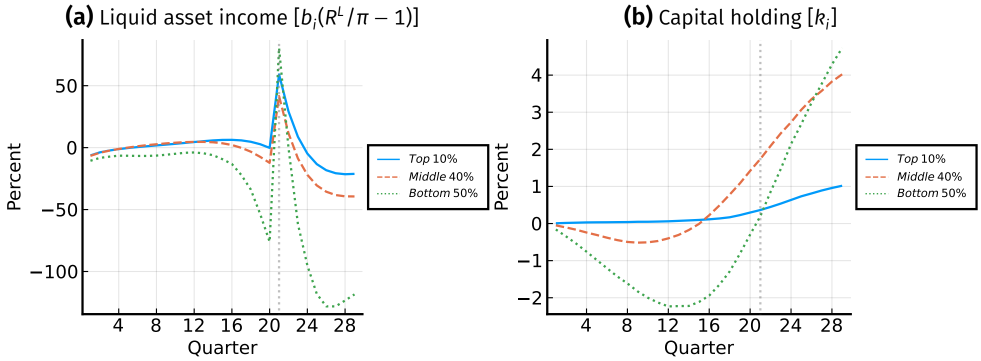
to the aggregate observations from 1995 to 2000.[40] In terms of real business cycle variables, the model exactly replicates the 6% rise in output and the 15% increase in investment, while it only accounts for one third to one half of the observed increase in consumption. As noted above, I calibrate the fiscal rule so that the model responses match the ratio of the decline in U.S. government debt to the decline in corporate profits during the late 1990s. In absolute size, the model explains about 75% of the observed declines in government bonds and profits (notably, federal debt held by the public declined by 20% during that time).

The shortcoming with respect to aggregate consumption may be due to the fixed debt limit in the model, while in reality, financial innovation related to collateral borrowing might have allowed households to consume more. Considering only unsecured borrowing, I find that the model accounts for half of the 30% increase in consumer credit. In the model, the increase in borrowing, mostly by the bottom 50% of households, contributes to the overall increase in wealth inequality during the anticipation period. From the World Inequality Database, the Gini index of wealth increased by 1.25% in that time span; the model explains about half of this increase.[41] Finally, with respect to the share of stocks within the liquid asset class, using the estimates by Saez and Zucman (2016), during the dotcom boom this share increased by 20 percentage points. The model accounts for around a 25% of this increase.[42]

40. I detrend all time series by a constant growth rate of 2%, following McGrattan and Prescott (2010), and deflate nominal series with the GDP deflator [GDPDEF].

41. This is remarkable, since the model does not feature heterogeneous stock shares; in the data, rich households gain disproportionally from stock price booms, see Kuhn, Schularick, and Steins (2020).

42. A more detailed model of stocks and their difference compared to other liquid assets, namely the different aggregate risk they carry, could help explaining this gap. Institutional changes, or agents that learn about the fundamentals over time, receiving observed prices as signals, would be other possible explanations.



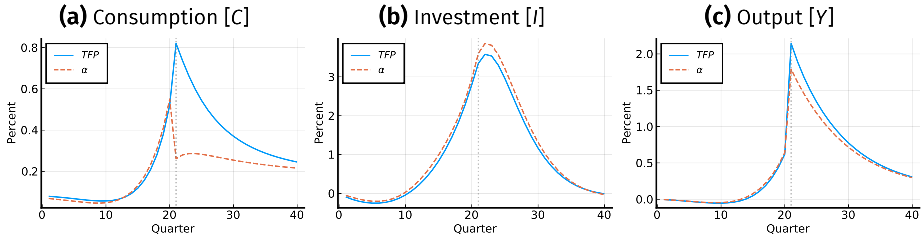
Notes: Model impulse responses are to news about a temporary capital share-increase in 5 years (quarter 21). Wealth-groups are defined from their position *at period 0*.

Figure 2.7. Response of income and investment over the wealth distribution

2.4.2 Alternative news shock

Figure 2.8 compares the response of the business cycle to news about a temporary TFP-increase with the response to the capital share-news (I adjust size and persistence of the shocks to make them comparable). The responses are virtually identical in the anticipation phase. This shows that the portfolio rebalancing towards capital, which is incentivized in both cases by the expectation of higher future returns on holding capital, drives the boom also in consumption and output. Differences only occur once the fundamental shock realizes: a higher capital share redistributes from households with a high marginal propensity to consume to those with a low propensity, so that consumption falls, while higher TFP implies more income for all households. Therefore, output also rises a little less in the case of the capital share increase. Still, in the long run, the levels of consumption and output converge across the two shock responses. The reason is that, when the direct effect of the transitory shocks subsides, the indirect effect of the higher capital stock, built up during the identical anticipation phase, dominates.

In a further clarifying exercise, I also shock the model economy with news about future transitory increases in the markup μ (i.e., market power), and news about future increases in investment-specific technology productivity, which increases the marginal productivity of the transformation from consumption to capital goods. Both variables are prominent candidates in the literature to explain the secular decline (increase) in the labor (capital) share (e.g. in Karabarbounis and Neiman (2014) and Greenwald, Lettau, and Ludvigson (2019)). I find that both news shocks depress the economy in the anticipation phase. The markup shock implies an expected redistribution from capital to profit income, which disincentivizes the holding of capital, so that investment falls. On the other hand, the investment-specific technology shock increases the capital rent, but it lowers the cost of capital; therefore,



Notes: Model impulse responses are to news about a temporary TFP-increase in 5 years, and to news about a temporary capital share-increase in 5 years (both quarter 21). The size of the capital share-impulse is scaled to fit to the TFP news shock (given by two times σ_A). For comparability, the persistence of the capital share-process is adjusted to ρ_A.

Figure 2.8. Response of business cycle to alternative news shock

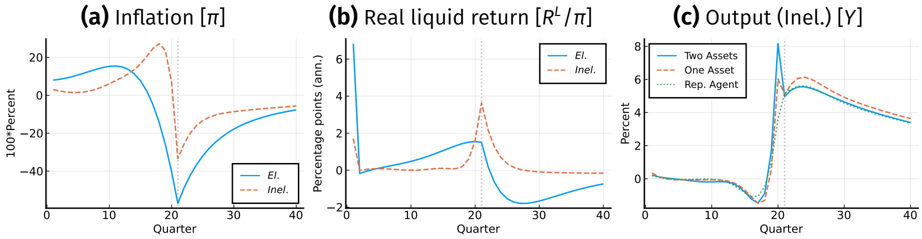
households wait with the investment until capital becomes cheap. This illustrates how only the anticipation of high rents *and* returns for capital causes an investment-driven business cycle and stock price boom in the model.

2.4.3 Importance of the fiscal rule

The investment boom is enabled by an elastic bond supply and a government that is willing to temporarily reduce its expenditure. To illustrate this point, I compare the response of inflation and the real liquid return in the baseline model with the impulse responses in an alternative environment (*Inel.*), where the government does not stabilize the output gap, and stabilizes inflation less strongly (figure 2.9). With the alternative fiscal rule that allows for a prolonged rise of inflation during the anticipation phase, middle class households do not invest enough to start the business cycle (and stock price) boom. The reason is that inflation depresses asset returns and magnifies the increase in the marginal costs of firms (affecting the entrepreneurs) and of unions (affecting the workers) late in the anticipation phase. The expectation of being exposed to these income losses discourages the households' capital investment earlier in the cycle. As a result, even in the model with portfolio choice, government expenditure is crowded out too late to drive the boom, and therefore all three model variants exhibit roughly the same output-response (as well as consumption-response) to the news shock.

2.4.4 Wealthy hand-to-mouth households

Following Kaplan, Violante, and Weidner (2014), wealthy hand-to-mouth households are households that have non-zero wealth in the illiquid asset ($k_i > 0$), while being at a kink in the budget set: either at zero liquid savings ($b_i = 0$), or at the borrowing limit ($b_i = \underline{B}$). Motivated by my numerical findings, I focus on the case when households hold the illiquid asset, while being at the borrowing constraint.



Notes: Model impulse responses are to news about a temporary capital share-increase in 5 years (quarter 21).
The changed fiscal rule parameters of the *Inelastic* specification are $\gamma_Y = 0.007$, $\gamma_\pi = 6.58$.

Figure 2.9. Responses for different bond supply elasticities.

Kaplan, Violante, and Weidner propose a stylized 3-period life-cycle model without uncertainty to highlight the conditions under which it is optimal for households to be wealthy hand-to-mouth: Suppose that in the first period, households allocate their initial endowment between the liquid and the illiquid asset. Next period, they receive income and can sell their liquid asset (or borrow) to increase their consumption, but can not sell the illiquid asset until the third (and last) period, where they consume their income and the return to all asset holdings.

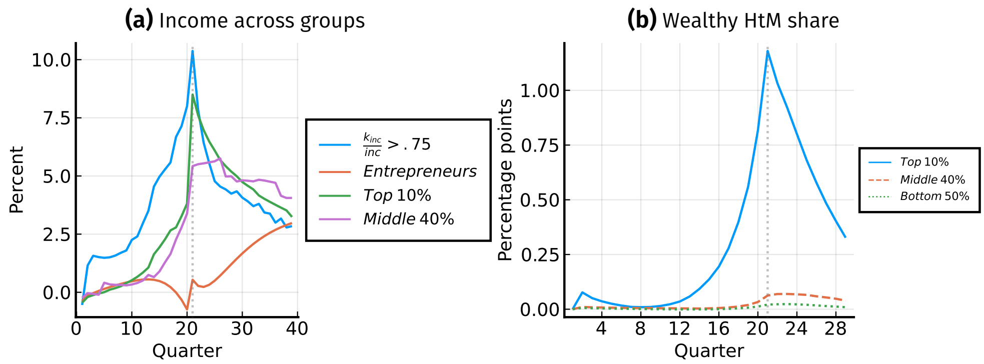
In this setup, households are more likely to be wealthy hand-to-mouth at the end of the second period if:

1. the capital rent and price in the last period are high relative to the borrowing rate,

2. their initial endowment is high, and both capital rent and their income are increasing from the second to the last period.

The news shock raises the expected capital rent and prices in the future. As I argued in section 2.4.3, extreme profit swings towards the end of the cycle depress investment. Part of the reason is that a big output gap late in the cycle requires monetary policy to hike the nominal rate, so that the real rate spikes in the last quarter before the TFP increase. This makes it more expensive to finance illiquid asset holdings with debt accumulated over the anticipation period, so that more households will refrain from doing so (as discussed above, higher real rates *earlier* in the cycle instead are beneficial for investment).

While the income of the average household in the upper half of the wealth distribution rises during the stock price boom, the most income gains are incurred by households whose income is dominated by capital rents (see figure 2.10a). Entrepreneurs, who receive the profit income, experience an income rise at the onset of the capital share increase, but lose in the anticipation period. Therefore, entrepreneurs are less likely to become wealthy hand-to-mouth households in the an-



Notes: Model impulse responses are to news about a temporary capital share-increase in 5 years (quarter 21).
Left panel: $\frac{k_{inc}}{inc} > .75$ denotes households whose main source of income (> 75%) is capital rents ([r] in the model). All groups are defined in the *cross-section* each quarter.
Right panel: Wealth-groups are defined from their position *at period 0*.

Figure 2.10. Response of income and shares of wealthy hand-to-mouth

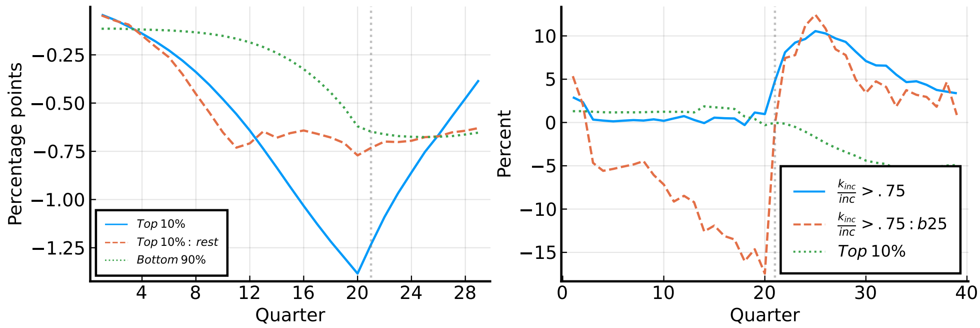
ticipation phase.[43] Hence, by virtue of capital rents, holding (a high amount of) the illiquid asset and experiencing income gains reinforces each other, making point 2) more likely to hold.

For these reasons, it is mostly households at the top of the wealth distribution who become wealthy hand-to-mouth households during the anticipation phase (see figure 2.10b). In steady state, only 0.2% of households are wealthy hand-to-mouth (at the borrowing limit). 73% of those households are in the top 10% of the wealth distribution. I find that during the stock price boom, the share of wealthy hand-to-mouth households among the wealthiest decile grows by 10%. Hence, by far the largest inflow into the group of wealthy hand-to-mouth households comes from capital-wealthy households, who optimally choose to get at or near the borrowing constraint so that they can hold on to the capital a little longer.

2.4.5 Marginal traders

How can it be known whether the mechanism highlighted in section 2.2 is at work in the full HANK model? To show this, I split up households into those that were wealthy hand-to-mouth at some period s after the news shock, and became unconstrained at the subsequent period $s + 1$, and the *rest*. The idea is that it should be the saving behavior of the first group, and not of the *rest* of households, that explains the time-varying returns on liquid assets during the cycle. Figure 2.11a reports the response of the households' saving rate (defined as the fraction that is saved of all

43. What is more, entrepreneurs on average hold much larger liquid asset stocks than workers, as they face the largest idiosyncratic risk (becoming a worker).



(a) Saving rates **(b)** Portfolio liquidity

Notes: Model impulse responses are to news about a temporary capital share-increase in 5 years (quarter 21).
Left panel: The saving rate is defined as $1 - c_{it}/\{\text{cash at hand}_{it}\}$, where

$$\text{cash at hand}_{it} \;=\; y_{it} + b_{it}R^L_t/\pi_t + k_{it}(r_t + \mathbb{1}_{\{k \text{ adjustable}\}}q_t) - \underline{B}.$$

Wealth-groups are defined from their position *at period 0*.
Right panel: Portfolio liquidity is with respect to the *chosen* portfolio, i.e., households' wealth position next period. $\frac{k_{inc}}{inc} > .75$ denotes households whose main source of income (> 75%) is capital rents ([r] in the model). $\frac{k_{inc}}{inc} > .75 : b25$ denotes the mean of the *lowest quartile* of the portfolio liquidity-distribtion for these households. All groups are defined in the *cross-section* each quarter.

Figure 2.11. Response of portfolio choice across groups of households

funds available to the household in a given period) to the news shock across the wealth distribution. It shows the average response of all households in the top 10% and bottom 90% of the wealth distribution, and only that of the *rest* in the top 10%. Clearly, within the top wealth decile, wealthy hand-to-mouth households save less during the anticipation period, and save more after the capital share increase. In particular, it is the only group of households where the saving rate is trending upwards strongly after the 5th year, which indicates that these households drive down the return on liquid assets.[44] Note that, since the aggregate supply of liquid assets is down, also a saving rate below its steady state-level can depress the return on liquid assets in equilibrium.

Figure 2.11b shows the portfolio liquidity of households in the richest decile in the cross section. Among the rich households, it is the households whose income is

44. One may be worried that, since aggregate consumption also decreases after the temporary shock to the capital share, the lower rates are due to a general decline in consumption. However, the results are robust for a news shock about a very persistent TFP increase ($\rho_A = 0.992$). In that scenario, almost all households in the economy *decrease* their savings after the TFP increase, as their incomes continue to rise (and aggregate consumption rises as well). Only the wealthy hand-to-mouth households within the top decile of the wealth distribution increase their savings. The results are available from the author upon request.

dominated by capital income who decrease their portfolio liquidity early on. During the anticipation phase, the distribution of portfolio choices of households with dominating capital income widens. One of the reasons is a *composition effect*: households with less capital wealth enter the group by virtue of higher capital rents during the business cycle boom. This alone drives up the portfolio liquidity of households in this group compared to the steady state.[45] Therefore, I also show the mean response of the lowest quartile in the portfolio liquidity distribution of these households. The marginal traders will be in this region of the distribution during the anticipation phase. I find that households with high capital income in that region of the distribution lower their portfolio liquidity during the anticipation phase. After the boom, the "rentiers" increase their liquid saving - their portfolio liquidity rises - as they are exposed to high consumption risk at that point. This depresses the real rate on liquid assets in equilibrium.

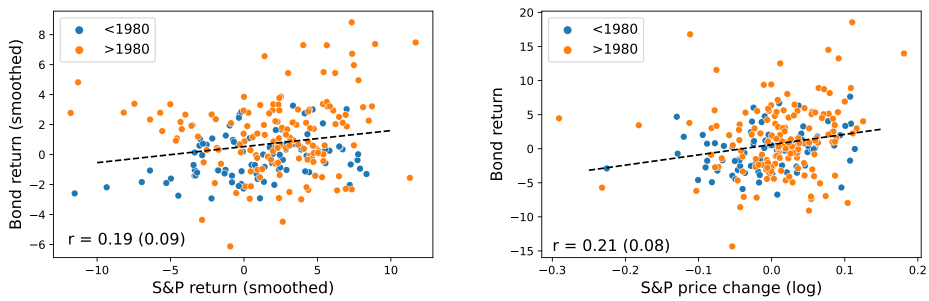
2.5 Asset returns, heterogeneous portfolio choices, and the stock market

In this section, I provide empirical evidence for the relation between the returns on liquid and illiquid assets and stocks, and the relation between portfolio choices of households and stocks, using micro-level data. Then, I simulate the model in order to assess the quantitative success of the model in explaining stock price fluctuations. Additionally, I use the Campbell-Shiller decomposition of the model stock price to highlight the effects of different assumptions about the cyclicality of dividends and the accuracy of the news for the explanatory power of the mechanism.

The theory implies that the expected return on liquid assets, like government bonds, is positively correlated with the expected stock return / the expected stock price growth. Figure 2.12 shows that ex-post returns in the data provide weak evidence for these links. In order to cancel out noise, which is mainly driven by innovations in dividends, I compute a moving average when comparing bond returns and stock returns. The results are robust to different specifications, with longer maturity bonds, or a larger moving average-window, leading to higher correlations with stock returns. Plotting the real 3-months treasury bill rate together with the stock price-dividend ratio along the time-dimension gives an impression of the relevance of this correlation as evidence for my theory (Figure 2.13). It shows that the larger swings in stock prices in the last decades, namely the downturns in the 1970s and 2000s, and the booms in the 1980s and 1990s, all occur in times of lower than average,

45. In the data, this composition effect rather goes in the opposite direction: since empirically, capital rents increase less in stock price booms than real bond rates, there is some evidence that the overall share of households with dominant capital income decreases in stock price booms. However, this does not drive the overall reduction in portfolio liquidity: see section 2.5.

(a) Bond return and S&P return (smoothed) **(b)** Bond return and S&P price change



Notes: Data by Robert Shiller (S&P and 10 year treasury bond). All returns are ex-post (realized) quarterly observations from 1955.Q4 to 2016.Q4. Smoothed series were computed by taking a moving average with a 4-quarter window. Lower maturity-bonds have a similar positive correlation with stock prices: the respective correlation coefficients for the (smoothed) real 3-months treasury bill rate are 0.22 (left panel) and 0.12 (right panel, without outlier 2008.Q3). Newey-West standard errors (1 lag) in parentheses.

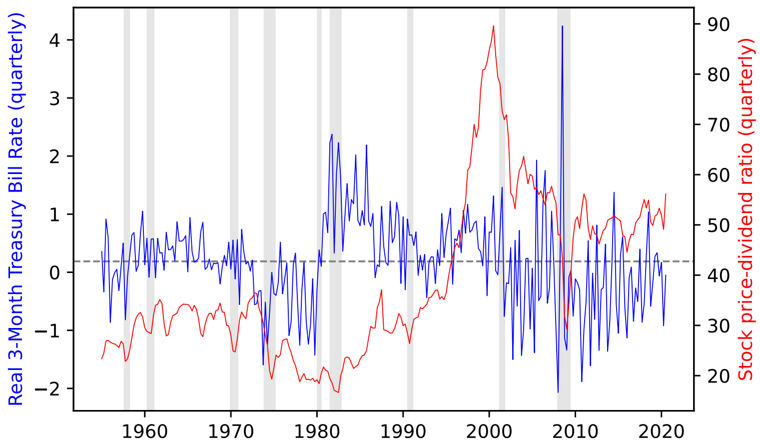
Figure 2.12. Correlations among liquid asset returns

respectively higher than average, real interest rates. The time after the Great Recession seems an abnormality, which may be due to the effect of quantitative easing on asset prices during that time.

Next, I take the capital return series by Gomme, Ravikumar, and Rupert (2011) as a proxy for returns on illiquid assets (no capital gains, after-tax), and look whether the change in capital returns is related to stock returns, as in the theory. Specifically, the proposed mechanism hinges on capital-wealthy households to drive down the return on liquid assets, and thus also stock returns, when capital returns fall. Figure 2.14a shows the correlations. During the boom phase, there is no correlation, but when stock prices are falling, there is a weak correlation. For investment growth, the correlations are more strongly positive. In a regression exercise (see appendix 2.B.1), I check that the positive correlations are unaffected by the inclusion of dividends and other business cycle variables. In sum, the data is consistent with a theory of investment-driven stock price-booms, where a fall in capital rents depresses stock returns after the boom.

2.5.1 Evidence from survey data

Turning to heterogeneous portfolio choice, which is a crucial part of the proposed theory, I use the SCF+ by Kuhn, Schularick, and Steins (2020) to isolate the group of households for whom capital income (excluding capital gains) is the main share



Notes: Stock market data from S&P500 (Robert Shiller), recession years (grey areas) by NBER. The real 3-months treasury bill rate is computed with realized inflation. The dotted line marks the average quarterly real 3-months treasury bill rate over the sample (0.19 pp).

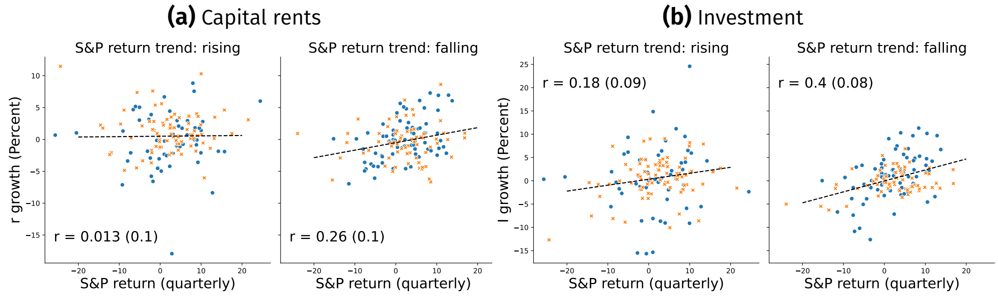
Figure 2.13. Real 3-months treasury bill rate and the stock market

(at least 75% in the baseline) of their overall income.[46] On average over all sampled years, 2.3% of households are in that category (2.7% in the model). The theory implies that their portfolio choice is decisive in affecting the illiquidity premium, and thus stock prices, over the cycle. In order to abstract from secular trends in the portfolio liquidity of the different wealth groups, I take the relative portfolio liquidity of the households with high capital incomes compared to the portfolio liquidity of the top 10% of the wealth distribution as the main measure of comparison between model and data.[47]

While in the model, households with high capital incomes are all in the top decile of the wealth distribution, in the data, only 41% are in that wealth group, while 39% have wealth that lies between the median and the top 10% of the wealth distribution. The likely reason for this discrepancy is that the model abstracts from negative illiquid assets: mortgage debt in particular systematically lowers the net worth of households with high capital income in the data. Due to this overlap of the "rentiers" with lower wealth groups, I also compute the relative portfolio liquidity of the bottom 50% and middle 40% relative to the top 10%. This allows me to see if movements in the relative portfolio liquidity of the households with high capital incomes are spuriously driven by movements across the wealth groups.

46. In the older waves of the Survey of Consumer Finances before 1983, capital income is only available as a measure that lumps together income from illiquid and liquid investments (like dividend income), while only the former counts as capital income in the model. Therefore, I treat separately the time periods before and after 1983. See appendix 2.B.2.

47. I show the time series of the portfolio liquidities of the different groups, as well as other characteristics of their portfolio choices over time, in appendix 2.B.2.

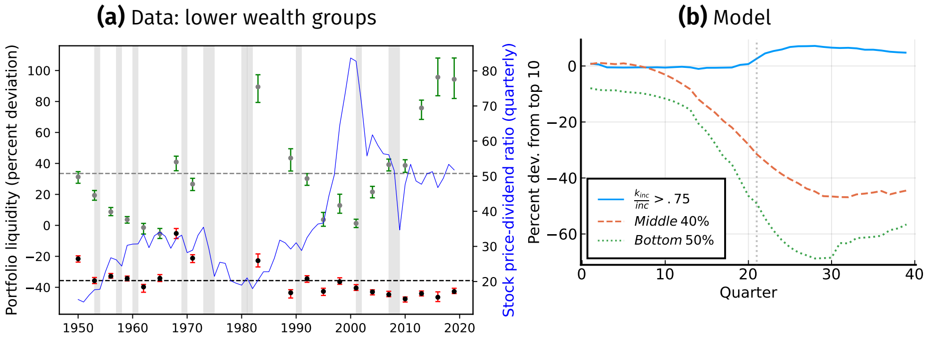


Notes: Data by Robert Shiller (S&P 500) and Gomme, Ravikumar, and Rupert (2011) (capital rents). Quarterly observations (1948.Q2 - 2016.Q4). S&P return trend computed using HP-filter ($\lambda = 1600$). Blue dots: before 1980. Orange crosses: after 1979. Newey-West standard errors (1 lag) in parentheses.

Figure 2.14. Capital rents and investment over the stock price-cycle

Figure 2.1 shows the relative portfolio liquidity of "rentiers" over time, and in comparison to the stock price-dividend ratio of the S&P 500. Figure 2.15 shows the same plot for the relative portfolio liquidities across wealth groups (left panel), as well as the model-implied prediction of the relative portfolio liquidities following a news shock (right panel). The model predicts that in response to the news, households in the bottom 90% of the wealth distribution reduce their portfolio liquidity relative to the top 10% as well. Different from the households with high capital income, however, they do not increase their portfolio liquidity (as much) in the years after the boom, especially so for the middle class.

To put this prediction to the test, I conduct the following exercise: let $\{y_i\}_i$ denote the sequence of two sets of subsequently sampled years, respectively, contained in the SCF+: years between 1950 and 1971, and years between 1983 and 2019. For each sequence of the relative portfolio liquidities of households in group g, computed from the survey data, denoted by $\{\text{pflq}^g(y_i)\}_i$, I compute the difference between subsequent years: $\Delta_i \text{pflq}^g = \text{pflq}^g(y_i) - \text{pflq}^g(y_{i-1})$. I also collect the stock price-dividend ratios for the years where survey data is available, and compute the same differenced sequence, $\Delta_i \frac{q^\Pi}{\Pi^F} = \frac{q^\Pi}{\Pi^F}(y_i) - \frac{q^\Pi}{\Pi^F}(y_{i-1})$. Then, I combine the differenced variables of both sets of years into one pooled sample. Column (I) in table 2.3 shows the results of regressing $\Delta_i \frac{q^\Pi}{\Pi^F}$ on the change in relative portfolio liquidity $\Delta_i \text{pflq}^g$ of the groups $g \in \{\text{high capital income}, \text{middle 40\%}, \text{bottom 50\%}\}$. As predicted by the model, the relative portfolio liquidity of households with high capital income comoves negatively with stock price-dividend growth, with a correlation of -0.37 (standardized), when controlling for the portfolios of the other two wealth groups. Notably, the relative portfolio liquidity of the poor half of the wealth distribution also correlates negatively with the stock market. There is a zig-zag pattern of the portfolio liquidity between the bottom 50% and the top 10% over the sample: it falls in the 1950s, rises thereafter, falls from the 1980s to 2000, and increases since



Notes: Survey evidence from SCF+ (Kuhn, Schularick, and Steins, 2020), stock market data from S&P500 (Robert Shiller), recession years (grey areas) by NBER. Portfolio liquidity is defined as the ratio of liquid assets by total wealth.
Left panel: Left axis shows the relative deviation of portfolio liquidity of households in the bottom 50% (grey dots, green CIs) / middle 40% (black dots, red CIs) from portfolio liquidity of the top 10% of wealth distribution. Whiskers are 68%-confidence intervals.
Right panel: Model responses of relative portfolio liquidity deviations (with respect to top 10%) across groups in the cross section. Responses are net of steady state deviation.

Figure 2.15. Relative portfolio liquidity in model and data

then. This is roughly consistent with the secular trends in the stock price-dividend ratio, with a trough in 1980 and a peak in 2000. I find that the portfolio liquidity of the "rentiers" and the bottom 50% explain mostly different parts of the variation, as leaving the latter out of the regression yields largely the same result for the "rentiers".

One issue with the interpretation of the results is that they could arise mechanically, through a composition effect with respect to stock shares: On average over the sampled years, households in the top 10% of the wealth distribution hold 13.4% of their total wealth in stocks, while households whose income is dominated by capital income hold 10% of their wealth in stocks.[48] Since stocks are liquid, the higher valuation of stock wealth during stock price-booms mechanically increases the liquid wealth and, ceteris paribus, also the portofolio liquidity of the top 10% relative to the households with high capital incomes. To check if this mechanism drives the results, I add the relative stock share of the "rentiers" compared to the top 10% as an additional regressor, where the stock share is defined as the ratio of the wealth held in equity and other managed assets by total wealth of the household. Columns (III) and (V) in table 2.3 show the results. When controlling for the stock share, the evidence for a negative relation between the relative portfolio liquidity of the high capital income-households and the stock market becomes stronger. The reason

48. The share of wealth that the top 50% of the wealth distribution holds in stocks decreases markedly from the first to the second half of the sample, see appendix 2.B.2.

Table 2.3. Regression of price-dividend growth on relative portfolio liquidities

Variables	(I)	(II)	(III)	(IV)	(V)
high cap. inc.	-0.37* (0.2)	-0.3 (0.21)	-0.5** (0.23)	-0.35 (0.2)	-0.51** (0.2)
middle 40%	0.42 (0.24)	-0.07 (0.09)	0.45 (0.27)	0.33 (0.21)	0.31 (0.23)
bottom 50%	-0.75** (0.29)	-	-0.77** (0.32)	-0.66** (0.22)	-0.64** (0.24)
rel. stock share	-	-	0.34** (0.15)	-	0.47** (0.15)
in top 10% share	-	-	-	-0.23 (0.17)	-0.38* (0.18)
Adj. R-squared	0.2	-0.05	0.26	0.2	0.35

Notes: The baseline regression equation is $\Delta_i \frac{q^{\Pi}}{\Pi^F} = \alpha + \sum_g \beta_g \Delta_i \text{pflq}^g + \varepsilon_i$, $i = 1, .., 18$. I divide all variables by their standard deviation. Specifications (III) and (V) include the change in the ratio of the stock share of high capital-households by the stock share of households in the top 10% as a regressor. Specifications (IV) and (V) include the change in the share of high capital-households in the top 10% as a regressor. Newey-West (one lag) standard errors in parentheses. Asterisks indicate that the t-statistic of the coefficient is above the 5% (**) or 10% (*) level.

is that, during stock price booms, the share of stock wealth in total wealth of the "rentiers" *increases* compared to that of the top 10%, even though the top 10% own more stocks on average. This effect — which cannot arise in the model, since it abstracts from aggregate risk — attenuates the negative relation between the relative portfolio liquidity and the stock market in the baseline specification.

To interpret the results as evidence for portfolio choice, one should also account for another composition effect: As shown above, stock price booms coincide with higher returns on liquid assets and business cycle booms. Hence, the overall income of households rises on average in stock price booms. If at the same time, capital rents do not rise (as much), the share of households whose income mainly comes from capital income falls. As a consequence, those households that *remain* above the threshold (>75% of income is capital income) have higher illiquid wealth, and thus a lower portfolio liquidity. The negative correlation of the portfolio liquidity of those households with the stock market would then be a mere restatement of the relation between the stock price cycle and factor incomes.[49] In the columns (IV) and (V), I consider this possibility, by including the change in the share of households with dominant capital income within the top 10% as an additional regressor. I find that, while there is evidence that the share of "rentiers" among the wealthiest households is indeed countercyclical, the negative correlation between relative portfolio liquidity and the stock market remains virtually unchanged. To summarize, the predicted fall in the liquidity of the portfolios of households with high capital incomes and households in the bottom half of the wealth distribution during stock price booms is supported by the evidence from the Survey of Consumer Finances.

49. Note that, in the household survey, capital gains from equity do not count as income. I use the same accounting in the model. Therefore, stock price booms do not mechanically raise liquid incomes.

2.5.1.1 Who are the marginal traders?

The survey data can be used to investigate main characteristics of the high capital income-households, who the model predicts to be the marginal traders of the stock market. For this, I take averages over all sampled years for the “rentiers” and the rest. I find that 40% of high capital income households report no wage income, compared to 20% of the rest. Only 16% of the “rentiers” report positive income from self-employment, while among the rest of households, 21% report such income. At the same time, 42% of the high capital income households are professionals or managers, while this is only the case for 29% of the other households. “Rentiers” hold 26% of their wealth in business wealth, while this share is only 6% on average for the rest of the households. With their high capital income, 32% of these households are in the top 10% of the income distribution.

These characteristics align well with the description of “modern capitalists” by Smith et al. (2019): they find that in the last decades, the top 1% of the income distribution is mostly populated by pass-through business owners. They have a tax incentive to receive compensation through their share of their firm’s profits rather than through wages. Typical pass-through firms are private, single-establishment or regional firms in skill-intensive industries, like law firms, dentists, or auto dealers.

2.5.2 Simulation

In order to evaluate the ability of the quantitative model to explain unconditional moments of the stock market and its correlation with the business cycle, I simulate the model. For the baseline, I pick three shocks: surprise TFP shocks ϵ^A, a surprise shock to the target price markup μ^Y, and the capital share news shock at the 5-year-horizon, $\epsilon^{\alpha,20}$. I set the standard deviation of the price markup shock to 0.01645, taken from the estimation by Bayer, Born, and Luetticke (2022). In one simulation variant, I implement a *noise shock* instead of a news shock. In practice, this is achieved by adding a surprise capital share shock ϵ^α to the system in every period where the capital share change was expected to take place. The surprise capital share shock exactly offsets the effect of the capital share news shock (Chahrour and Jurado, 2018). In other simulation variants, I leave out any anticipatory shocks. I assume all shocks to be normally distributed around zero.

Table 2.4 shows the simulation results for various model variants and shock combinations, and compares them to the unconditional moments of the data. The main result is that the baseline variant, column (I), explains around 75% of the fluctuation in the price-dividend ratio of the S&P 500. The comparison with columns (II) and (III) shows that news shocks and portfolio choice between liquid and illiquid assets are both important for explaining stock price fluctuations. Only the two-asset model allows for a time-varying illiquidity premium, which leads to larger fluctuations in the return to liquid assets and induces comovement between bond returns and stock returns (see the low set of rows). In the one-asset economy, the correlation

Table 2.4. Unconditional moments in data and simulated model

Variables	Data	(I)	(II)	(III)	(IV)	(V)
$mean(P/D)$	152*	151	148	147	146	149
$\sigma(P/D)$	63	48	35	28	28	42
$\rho(P/D)$	0.98	0.986	0.985	0.99	0.996	0.96
$\rho(\Delta P/D)$	0.99	0.11	0.01	0.41	0.41	-0.04
$\sigma(\Delta D)$	1.75%*	1.74%	1.27%	1.81%	1.49%	1.46%
$\rho(I/Y, P/D)$	15.2%	62%	32%	-5%	-24%	41%
$\rho(\Delta I/Y, \Delta P/D)$	17.5%	34%	29%	4.8%	-22%	64%
$\rho(\Delta C/Y, \Delta P/D)$	15.4%	2.1%	-58%	7.9%	-72%	64%
$\rho(R^b/\pi, R^{stocks})$	0.13-0.19	0.24	0.24	0.05	-0.11	0.3
$\sigma(R^{stocks})$	7.28%	5.07%	4.27%	1.63%	1.45%	7.84%
$\sigma(R^{stocks})/\sigma(R^b/\pi)$	1.7-8.9	2.9	5.3	3.7	4.26	12.2

Notes: Unconditional moments in U.S. data, 1950-2016, and in the model. $\sigma(x)$ and $\rho(x)$ denote the standard deviation and the autocorrelation, respectively, of variable x. $\rho(x, y)$ denotes the correlation of x and y. Δx denotes the growth rate of x. Appendix 2.B.3 lists the composition of the aggregate variables. Stock market data by Robert Shiller (S&P 500). The model variants are as follows: (I): Two-Asset HANK with News; (II): Two-Asset HANK without News; (III): One-Asset HANK with News; (IV): One-Asset HANK without News; (V): Two-Asset HANK, only Noise
(*) denotes moments that were targeted during the calibration.

between stock returns and bond returns turns zero or negative, as surprise TFP and markup shocks cause surprise changes in dividend payments that are orthogonal to government bond returns. News shocks cause the illiquidity premium to fluctuate even more, but in a structured way: they add the boom-bust cycle. Thereby, news shocks can explain higher volatility of the price-dividend ratio, while at the same time generating some *momentum* $\rho(\Delta P/D)$, i.e. the autocorrelation in growth rates, which is a salient feature of the data, and causing comovement of the stock price cycle with aggregate consumption. The model predicts that investment and the stock market are more positively correlated than in the data. Adam and Merkel (2019) show that a subset of investment in fixed assets, namely non-residental investment and investment in non-residential structures, correlates more with the stock market. However, since housing is the most important illiquid asset of the majority of households in the data, I cannot abstract from it in my quantitative model. Finally, as presented in column (V), noise shocks are almost equally successful in explaining stock price fluctuations. However, they imply that stock returns fluctuate 12 times more than returns on government bonds, at least a third higher than what is realistic, and fail to generate any momentum.

Next, I use the Campbell and Shiller (1988) decomposition to analyze the degrees to which the model variants explain the salient feature of stock prices (and asset prices in general): return predictability. It is a log-linear approximation of the price-dividend ratio around its (proposed) stationary value, and is given by

Table 2.5. Campbell-Shiller decomposition in data and model

Source / Variant	Dividends	Discount rate	PD-ratio
Cochrane (2011)	0.11	1.01	0.11
Kuvshinov (2022)	0.55	0.45	-
Baseline	0.39	0.52	0.08
One-Asset	0.97	-0.04	0.07
No News	0.29	0.44	0.28
Only Noise	0.25	0.57	0.18

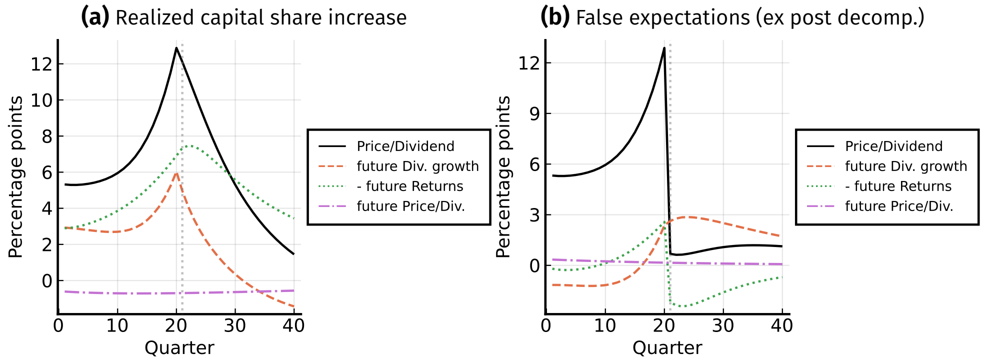
Notes: Variance shares of the Campbell-Shiller decomposition. For the model variants, I use the method by Cochrane (2011) to calculate the variance shares, with a time horizon of 15 years. "No News" and "Only Noise" are two-asset model variants.

$$\log(q_t^{\Pi}/\Pi_t^F) = c + \mathrm{E}_t \sum_{j=0}^{\infty} \rho^j \left[\underbrace{\hat{\Pi}_{t+1+j}^F}_{\text{dividend growth news}} \underbrace{-r_{t+1+j}^L}_{\text{discount rate news}} \right], \tag{2.29}$$

where c and ρ are constants that are computed from long-run averages, and $r_t^L = R_t^L/\pi_t - 1$ is defined as the net real return on the liquid asset (where I assume that the no-arbitrage condition holds up to first order, i.e. r^L is also the expected net return on the stocks). The composition shows that the contemporaneous price-dividend ratio is determined by dividend growth news and negative discount rate news up to first order (in the formula with a finite horizon, a future price-dividend ratio also enters).

Table 2.5 shows the results of decomposing the variance of the log price-dividend ratio into the variances of the two news components and the future price-dividend ratio, in both data and the model variants. In the baseline model, discount rate news explain about half of the variance in the price-dividend ratio. The results for the one-asset variant with news shocks and for the two-asset variant without news shocks show that the main cause of return predictability in the model are not news shocks, but the financial friction on the household side: the existence of wealthy, liquidity-constrained households, whose subjective discount factor varies with asset returns, is the key to generating time-varying discount rates. Naturally, the existence of news increases the predictive power of both dividend growth- and discount rate-news, while noise shocks are able to generate an even higher importance for the discount rate-component, at the expense of the predictive power of the "news".

Why is such a large share of stock price fluctuations in the model explained by expected future dividend growth? In panel a) of figure 2.16, I plot the Campbell-Shiller decomposition as an impulse response to a capital share news shock. One can see that the price-dividend ratio correlates with future dividend growth. The reason is that dividends are countercyclical in the model (although profit smoothing

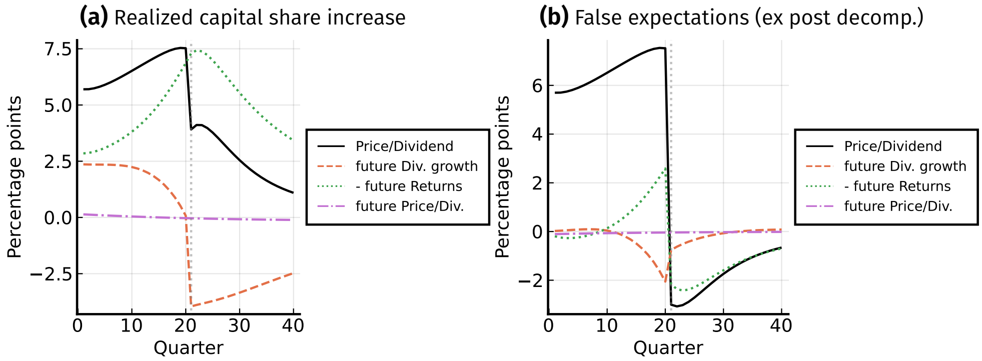


Notes: Model impulse responses are to news about a temporary capital share-increase in 5 years (quarter 21).
In b), the news is offset by a negative capital share surprise shock in quarter 21.

Figure 2.16. Campbell-Shiller decomposition with countercyclical dividends (model)

mitigates this), so that a stock price boom that coincides with a business cycle boom automatically implies positive dividend growth news. In panel b), I plot the impulse response to a noise shock. Specifically, the contemporaneous price-dividend ratio, which up to the 21st quarter is driven by the wrong expectation of a capital share increase, is plotted together with the true future components that are known ex-post. Now, of course, the Campbell-Shiller decomposition does not hold in the anticipation period, as the price is based on a wrong expectation. Indeed, as the real rate also falls after the news-disappointment in the model, the future returns-component can "rationalize" some of the excess price-dividend ratio relative to future dividend growth. Due to profit smoothing, future dividend growth fails to ex-post rationalize the variation of the price-dividend ratio over the cycle.

In order to illustrate the impact of the cyclicality of dividends for the results, I also compute the Campbell-Shiller decomposition for an alternative asset, where the dividend is simply given by a fraction of output (see figure 2.17). For this type of stock, the future returns-component explains the smooth increase of the price-dividend ratio in the anticipation phase, and its smooth decline in the subsequent bust-phase. Future dividends instead explain the jumps in the price-dividend-ratio, one at the onset of the news, and one at the onset of the productivity change. With constant returns (or discount rates), a forward-looking price would already incorporate the future expected decline in dividends at the onset of the news, and thus be mostly declining in the anticipation phase. But since the future dividends will also be discounted less as the demand for liquidity will rise, the price-dividend ratio rises in the anticipation phase. In figure 2.17b, I show that if the capital share-expectation is disappointed, the future returns (which increase quickly after the news-disappointment, as the price level shoots up and then declines slowly) explain



Notes: Model impulse responses are to news about a temporary capital share-increase in 5 years (quarter 21).
In b), the news is offset by a negative capital share surprise shock in quarter 21.
The hypothetical asset yields dividends $\omega^{\Pi} Y$.

Figure 2.17. Campbell-Shiller decomposition with procyclical dividends

most of the subsequent lower stock price, while the future dividend-component converges back to its steady state-level.

2.6 Conclusion

What is the reason for the return predictability on the stock market? I propose a mechanism to explain this pervasive empirical pattern that hinges on incomplete markets and the existence of illiquid assets. I show in a quantitative business cycle model with time-separable preferences that the mechanism can account for a large part of the return predictability, as well as for many other unconditional data moments of stock prices. The main intuition behind the result is that the model accounts for the existence of wealthy marginal traders: wealthy households can be liquidity-constrained when they own mostly illiquid assets. In turn, asset income correlates with productivity shocks and the business cycle, which induces a cyclicality of the stochastic discount factor of the marginal traders. Together with anticipation, these factors generate realistic stock price cycles.

Why are households more risk-loving during a stock price boom? I propose that they anticipate higher future returns on illiquid assets. This induces wealthy household to optimally shift their portfolio towards more illiquid assets, which puts them at a higher idiosyncratic risk. Instead of a time-varying aggregate risk premium, I show that a time-varying illiquidity premium, which reflects the idiosyncratic risk-return calculus of the marginal traders, can account for stock price booms and subsequent busts. This accomodates recent evidence (Kuvshinov, 2022) that the risk factors of assets with different liquidities do not comove in the data.

The empirical evidence is in line with the proposed mechanism: first, returns on liquid and illiquid assets correlate with the stock market as expected. Second, I show using survey data that households who earn mostly capital income shift their wealth towards illiquid assets in stock price booms, and increase the liquidity of their portfolio in the subsequent stock price bust, as predicted by the model. Matching a heterogeneous agent model to micro-level data, I ascribe a large part of stock price fluctuations to the stock-trading of owners of private businesses who are in the top 10% of the income and wealth distribution. I leave the further investigation of this hypothesis — ideally using data on consumption and investment — for future research. On the model side, solving the model nonlinearly, thereby accounting for heterogeneous stock shares, appears to be a promising next step.

Appendix 2.A Challe-Ragot model

Table 2.A.1. Calibration of the model parameters and steady state-levels of variables

Preferences	
β	0.95
σ (risk preference)	1
c^*	10
Environment	
y^l	2
y^h	14
$\mathbb{P}(h \to l)$	10%
$\mathbb{P}(l \to h)$	90%
μ	1
k	50
$\bar{r}$	4%
$\bar{L}$	3
Steady state	
$\mathbb{P}(h)$	90%
$R-1$	3.53%
$\tilde{b}$	3.44

Appendix 2.B Empirical evidence

2.B.1 Stock returns, capital rents, and business cycle variables

This section presents regressions of quaterly S&P 500 stock returns (data by Robert Shiller) on the growth of after-tax capital rents (Gomme, Ravikumar, and Rupert, 2011) and other variables. The sample is split in two, periods where the trend of the S&P 500 return is rising, and periods where it is falling. The trends of the S&P stock return, inflation growth, and GDP are computed using the Hodrick-Prescott filter with a smoothing parameter of 1600. All variables are standardized.

Findings:

- In periods of stock returns trending upwards (panel a)), stock returns are statistically significantly correlated with consumption growth (5% level), and weakly statistically significantly correlated with falling inflation and deviations of GDP from trend (10% level). There is no correlation with capital rents.
- In periods of stock returns trending downwards (panels b) and c)), stock returns are statistically significantly correlated with investment growth and dividend growth (1% level). Capital returns are weakly negatively correlated. However, without investment as regressor, capital returns become positively correlated with stock returns. This shows that investment and capital returns explain similar parts of the variance in downturns.

a) **Subset of observations where S&P return-trend is rising**

Dep. Variable:	Stock return	**R-squared (uncentered):**	0.149
Model:	OLS	**Adj. R-squared (uncentered):**	0.103
Method:	Least Squares	**F-statistic:**	2.999
Date:	-	**Prob (F-statistic):**	0.00599
Time:	03:37:32	**Log-Likelihood:**	-180.15
No. Observations:	135	**AIC:**	374.3
Df Residuals:	128	**BIC:**	394.6
Df Model:	7		
Covariance Type:	HAC		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Cap.rent growth	-0.0945	0.114	-0.829	0.409	-0.320	0.131
Consmpt. growth	0.2247	0.106	2.127	0.035	0.016	0.434
Investm. growth	0.1410	0.104	1.350	0.179	-0.066	0.348
Before 1980	0.0513	0.090	0.573	0.568	-0.126	0.228
Rising Infl.	-0.1436	0.086	-1.673	0.097	-0.313	0.026
GDP deviation	-0.1788	0.103	-1.736	0.085	-0.383	0.025
Dividend growth	-0.0034	0.092	-0.037	0.971	-0.186	0.180

Omnibus:	6.483	**Durbin-Watson:**	1.907
Prob(Omnibus):	0.039	**Jarque-Bera (JB):**	7.238
Skew:	-0.330	**Prob(JB):**	0.0268
Kurtosis:	3.922	**Cond. No.**	1.77

Notes:

[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.

[2] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 1 lags and without small sample correction

b) Subset of observations where S&P return-trend is falling

Dep. Variable:	Stock return	**R-squared (uncentered):**	0.240
Model:	OLS	**Adj. R-squared (uncentered):**	0.200
Method:	Least Squares	**F-statistic:**	9.300
Date:	-	**Prob (F-statistic):**	2.41e-09
Time:	03:37:32	**Log-Likelihood:**	-178.93
No. Observations:	140	**AIC:**	371.9
Df Residuals:	133	**BIC:**	392.4
Df Model:	7		
Covariance Type:	HAC		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Cap.rent growth	-0.0299	0.124	-0.241	0.810	-0.275	0.215
Consmpt. growth	0.1405	0.095	1.484	0.140	-0.047	0.328
Investm. growth	0.3318	0.097	3.422	0.001	0.140	0.524
Before 1980	-0.0507	0.072	-0.700	0.485	-0.194	0.093
Rising Infl.	-0.0892	0.069	-1.285	0.201	-0.227	0.048
GDP deviation	-0.1175	0.091	-1.287	0.200	-0.298	0.063
Dividend growth	0.1760	0.064	2.759	0.007	0.050	0.302

Omnibus:	13.595	**Durbin-Watson:**	2.285
Prob(Omnibus):	0.001	**Jarque-Bera (JB):**	23.130
Skew:	-0.455	**Prob(JB):**	9.49e-06
Kurtosis:	4.771	**Cond. No.**	2.40

Notes:

[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.

[2] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 1 lags and without small sample correction

c) Subset of observations where S&P return-trend is falling; leave out investment

Dep. Variable:	Stock return	**R-squared (uncentered):**	0.171
Model:	OLS	**Adj. R-squared (uncentered):**	0.134
Method:	Least Squares	**F-statistic:**	7.819
Date:	-	**Prob (F-statistic):**	3.16e-07
Time:	03:37:32	**Log-Likelihood:**	-185.04
No. Observations:	140	**AIC:**	382.1
Df Residuals:	134	**BIC:**	399.7
Df Model:	6		
Covariance Type:	HAC		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Cap.rent growth	0.1267	0.124	1.021	0.309	-0.119	0.372
Consmpt. growth	0.1624	0.095	1.714	0.089	-0.025	0.350
Before 1980	-0.0217	0.080	-0.273	0.785	-0.179	0.136
Rising Infl.	-0.0554	0.071	-0.776	0.439	-0.197	0.086
GDP deviation	-0.1578	0.101	-1.565	0.120	-0.357	0.042
Dividend growth	0.2288	0.068	3.375	0.001	0.095	0.363

Omnibus:	24.887	**Durbin-Watson:**	2.230
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	50.473
Skew:	-0.767	**Prob(JB):**	1.10e-11
Kurtosis:	5.510	**Cond. No.**	2.00

Notes:

[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.

[2] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 1 lags and without small sample correction

2.B.2 Survey evidence

2.B.2.1 Data selection and definitions

This section presents more evidence about heterogeneous portfolio choice in the U.S. over time. I use the 20 years available in the SCF+ (Kuhn, Schularick, and Steins, 2020) between 1950 and 2019. I split the sample into two subgroups of years: from 1950 to 1971, and from 1983 to 2019. Year 1977 is left out in the analysis in the main text for two reasons: first, the gap to sampled years before and after 1977 is 6 years, which is double the gap between most of the sampled years in the survey (3 years). Hence, computing differences between sampled years is less consistent when including the year 1977. Second, I find that 1977 is an outlier in terms of the main object of analysis in this paper, the group of households with high capital income: while the share of households with high capital income who are in the top 10% of the wealth distribution is 42% in the median year, it is only 13% in 1977. Conversely, the share of these households who are in the bottom half of the distribution is 19% in the median year, and 61% in 1977. The likely reasons for this discrepancy are issues with the imputations of total and capital income. Over all remaining years, N=84430 households are in the survey.

The first subgroup from 1950 to 1971 is from the older waves of the SCF, where capital income lumps together asset incomes from the following sources:

(1) non-taxable investments (e.g. municipal bonds)

(2) other interest

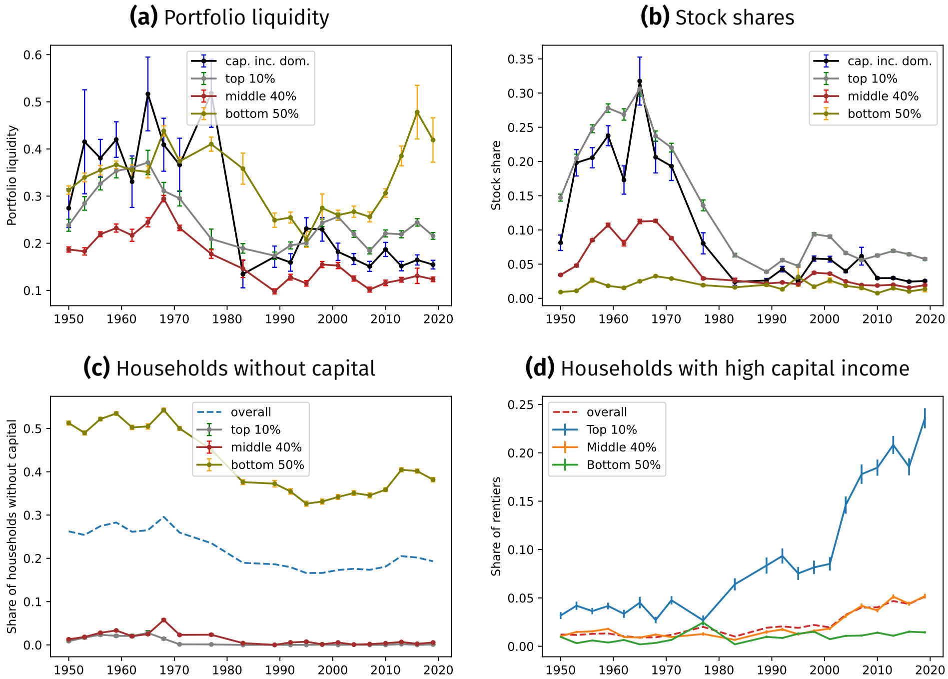
(3) dividends

(4) other business or investments, net rent, trusts, or royalties

Since asset incomes number 2 and 3 likely stem from more liquid sources, namely treasury bonds and stocks, this definition of capital income does not fit to the dichotomy between liquid and illiquid assets suggested by the analysis in the main text. Therefore, starting from year 1983 (the modern waves of the SCF), I sum up as a measure of capital income only income from the sources number 1 and 4.

In line with the quantitative model, I define as high capital income those households where capital income is at least 75% of their total income. In order to make this definition comparable across the old and modern waves of the SCF, where only the modern waves allow to compute the model-consistent definition of capital income, I proceed in the following way: From the modern waves, I calculate the average share of asset income from sources number 1 and 4 among asset income from all sources, which equals 0.19. Then, I categorize households into the high capital income-group in the *older* waves if at least 75% of their total income stems from the original cpaital income measure (with all sources), while for the *modern* waves, households' income must stem from sources number 1 and 4 at least at the rate of $75\% \cdot 0.19 = 15\%$ to be classified as high capital income. The similarity of the average share of households with high capital income in the data with their share in the model economy justifies this procedure.

2.B.2.2 Portfolio choice over time



Notes: Survey evidence from SCF+ (Kuhn, Schularick, and Steins, 2020). *Portfolio liquidity* is defined as the ratio of liquid assets by total wealth. *Stock shares* are defined as the ratio of stock wealth by total wealth. Households *without capital* are defined as households with zero illiquid wealth. Households with *high capital income* are households who earn a large share of capital income (> 75%) compared to their overall income. Whiskers are 68%-confidence intervals.

Figure 2.B.1. Heterogeneous portfolio choice over time

Several secular trends are noticeable:

- For households in the top half of the wealth distribution, portfolio liquidity peaks in the 1960s, and declines since then. Some of this development is due to a larger share of wealth held in stocks in the first half of the sample.
- For the bottom half of the wealth distribution, stocks are mostly irrelevant, and up to half of the households in that wealth category do not hold illiquid assets. The share of households without capital decreases from the 1970s on, and increases again since the Great Recession. Also, the portfolio liquidity of the poorer households increases markedly since 2008.
- While the overall share of households with high capital income stays mostly constant over time, their share within the richest decile increases steadily since the

1970s. Since 2000, more households in the U.S. are becoming high capital income households overall, a trend that is driven by the middle class.

2.B.3 Business cycle data

All series are available at quarterly frequency from the St.Louis FED - FRED database:

Output, Y: Sum of gross private domestic investment (GPDI), personal consumption expenditures for nondurable goods (PCND), durable goods (PCDG), and services (PCESV), and government consumption expenditures and gross investment (GCE) divided by the GDP deflator (GDPDEF) and the civilian noninstitutional population (CNP16OV).

Consumption, C: Sum of personal consumption expenditures for nondurable goods (PCND), durable goods (PCDG), and services (PCESV) divided by the GDP deflator (GDPDEF) and the civilian noninstitutional population (CNP16OV).

Investment, I: Gross private domestic investment (GPDI) divided by the GDP deflator (GDPDEF) and the civilian noninstitutional population (CNP16OV).

References

Adam, Klaus, Albert Marcet, and Johannes Beutel. 2017. "Stock Price Booms and Expected Capital Gains." *American Economic Review* 107 (8): 2352–408. [127]

Adam, Klaus, and Sebastian Merkel. 2019. "Stock Price Cycles and Business Cycles." ECB Working Paper 2316. [127, 162]

Adam, Klaus, and Henning Weber. 2023. "Estimating the Optimal Inflation Target from Trends in Relative Prices." Forthcoming at American Economic Journal: Macroeconomics. [143]

Alves, Felipe, Greg Kaplan, Benjamin Moll, and Giovanni L. Violante. 2020. "A Further Look at the Propagation of Monetary Policy Shocks in HANK." *Journal of Money, Credit and Banking* 52 (S2): 521–59. [120]

Angeletos, George-Marios. 2007. "Uninsured idiosyncratic investment risk and aggregate saving." *Review of Economic Dynamics* 10 (1): 1–30. [120]

Asker, John, Joan Farre-Mensa, and Alexander Ljungqvist. 2014. "Corporate Investment and Stock Market Listing: A Puzzle?" *Review of Financial Studies* 28 (2): 342–90. [123]

Auclert, Adrien, Matthew Rognlie, and Ludwig Straub. 2020. "Micro Jumps, Macro Humps: Monetary Policy and Business Cycles in an Estimated HANK Model." Working paper. Revise and resubmit at American Economic Review. [125]

Bansal, Ravi, and Amir Yaron. 2004. "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles." *Journal of Finance* 59 (4): 1481–509. [121, 126]

Barsky, Robert B., and Eric R. Sims. 2012. "Information, Animal Spirits, and the Meaning of Innovations in Consumer Confidence." *American Economic Review* 102 (4): 1343–77. [140]

Bayer, Christian, Benjamin Born, and Ralph Luetticke. 2022. "Shocks, Frictions, and Inequality in US Business Cycles." CEPR Discussion Papers 14364. [129, 136, 137, 143, 144, 146, 161]

Bayer, Christian, and Ralph Luetticke. 2020. "Solving discrete time heterogeneous agent models with aggregate risk and many idiosyncratic states by perturbation." *Quantitative Economics* 11 (4): 1253–88. [143]

Bayer, Christian, Ralph Luetticke, Lien Pham-Dao, and Volker Tjaden. 2019. "Precautionary Savings, Illiquid Assets, and the Aggregate Consequences of Shocks to Household Income Risk." *Econometrica* 87 (1): 255–90. [122]

Beaudry, Paul, and Franck Portier. 2004. "An exploration into Pigou's theory of cycles." *Journal of Monetary Economics* 51 (6): 1183–216. [128]

Beaudry, Paul, and Franck Portier. 2006. "News, stock prices, and economic fluctuations." *American Economic Review* 96 (4): 1293–307. [128]

Beaudry, Paul, and Franck Portier. 2014. "News Driven Business Cycles: Insights and Challenges." *Journal of Economic Literature* 52: 993–1074. [128]

Ben Zeev, Nadav. 2018. "What can we learn about news shocks from the late 1990s and early 2000s boom-bust period?" *Journal of Economic Dynamics and Control* 87: 94–105. [125]

Beutel, Johannes, and Michael Weber. 2022. "Beliefs and Portfolios: Causal Evidence." Chicago Booth Research Paper No. 22-08. [127]

Bianchi, Francesco, Howard Kung, and Gonzalo Morales. 2019. "Growth, slowdowns, and recoveries." *Journal of Monetary Economics* 101: 47–63. [125]

Bilbiie, Florin O. 2019. "Monetary Policy and Heterogeneity: An Analytical Framework." CEPR Discussion Paper No. DP12601. [127]

Bilbiie, Florin O. 2020. "The New Keynesian cross." *Journal of Monetary Economics* 114: 90–108. [127]

Bilbiie, Florin O., Diego R. Känzig, and Paolo Surico. 2022. “Capital and income inequality: An aggregate-demand complementarity.” *Journal of Monetary Economics* 126: 154–69. [128]

Bordalo, Pedro, Nicola Gennaioli, Rafael La Porta, and Andrei Shleifer. 2020. “Belief Overreaction and Stock Market Puzzles.” Working Paper, Working Paper Series 27283. National Bureau of Economic Research. [127]

Born, Benjamin, and J. Pfeifer. 2014. “Policy risk and the business cycle.” *Journal of Monetary Economics* 68: 68–85. [137]

Broer, Tobias, Niels-Jakob Harbo Hansen, Per Krusell, and Erik Öberg. 2019. “The New Keynesian Transmission Mechanism: A Heterogeneous-Agent Perspective.” *Review of Economic Studies* 87 (1): 77–101. [127]

Brown, James R., Steven M. Fazzari, and Bruce C. Petersen. 2009. “Financing Innovation and Growth: Cash Flow, External Equity, and the 1990s R&D Boom.” *Journal of Finance* 64 (1). [125]

Caballero, Ricardo J, and Alp Simsek. 2020. “Monetary Policy and Asset Price Overshooting: A Rationale for the Wall/Main Street Disconnect.” Working Paper, Working Paper Series 27712. National Bureau of Economic Research. [122]

Campbell, John, and Robert Shiller. 1988. “The dividend-price ratio and expectations of future dividends and discount factors.” *Review of Financial Studies* 1 (3): 195–228. [119, 162]

Campbell, John Y., and John H. Cochrane. 1999. “By Force of Habit: A consumption-based Explanation of Aggregate Stock Market Behavior.” *Journal of Political Economy* 107 (2): 205–51. [121, 126]

Chahrour, Ryan, and Kyle Jurado. 2018. “News or Noise? The Missing Link.” *American Economic Review* 108 (7): 1702–36. [161]

Challe, Edouard, and Xavier Ragot. 2016. “Precautionary Saving Over the Business Cycle.” *Economic Journal* 126 (590): 135–64. [129, 130, 132]

Chien, YiLi, Harold Cole, and Hanno Lustig. 2012. “Is the Volatility of the Market Price of Risk Due to Intermittent Portfolio Rebalancing?” *American Economic Review* 102 (6): 2859–96. [122]

Christiano, Lawrence, Cosmin Ilut, Roberto Motto, and Massimo Rostagno. 2010. “Monetary policy and stock market booms.” *Proceedings - Economic Policy Symposium - Jackson Hole,* 85–145. [128]

Cochrane, John H. 2011. “Presidential Address: Discount Rates.” *Journal of Finance* 66 (4): 1047–108. [119, 163]

Constantinides, George M., and Darrell Duffie. 1996. “Asset Pricing with Heterogeneous Consumers.” *Journal of Political Economy* 104 (2): 219–40. [120, 133]

Domínguez-Díaz, Rubén. 2021. “Precautionary Savings and Financial Frictions.” RTG-2281 Discussion Paper 2021-14. University Bonn. [125]

Favilukis, Jack, Sydney C. Ludvigson, and Stijn Van Nieuwerburgh. 2017. “The Macroeconomic Effects of Housing Wealth, Housing Finance, and Limited Risk Sharing in General Equilibrium.” *Journal of Political Economy* 125 (1): 140–223. [122]

Fernández-Villaverde, Jesús, Samuel Hurtado, and Galo Nuño. 2022. “Financial Frictions and the Wealth Distribution.” Conditionally accepted at Econometrica. [126]

Gomme, Paul, B. Ravikumar, and Peter Rupert. 2011. “The Return to Capital and the Business Cycle.” *Review of Economic Dynamics* 14 (2): 262–78. [145, 156, 158, 166]

Gormsen, Niels Joachim. 2021. “Time Variation of the Equity Term Structure.” *Journal of Finance* 76 (4): 1959–99. [126]

Görtz, Christoph, John D. Tsoukalas, and Francesco Zanetti. 2022. “News Shocks under Financial Frictions.” *forthcoming in American Economic Journal: Macroeconomics.* [129]

Greenwald, Daniel L, Martin Lettau, and Sydney C Ludvigson. 2019. “How the Wealth Was Won: Factors Shares as Market Fundamentals.” Working Paper, Working Paper Series 25769. National Bureau of Economic Research. [150]

Guiso, Luigi, and Paolo Sodini. 2013. “Chapter 21 - Household Finance: An Emerging Field,” edited by George M. Constantinides, Milton Harris, and Rene M. Stulz, 2: 1397–532. Handbook of the Economics of Finance. Elsevier. [143, 145]

Haddad, Valentin, and Tyler Muir. 2021. “Do Intermediaries Matter for Aggregate Asset Prices?” *Journal of Finance* 76 (6): 2719–61. [123]

Heathcote, J., K. Storesletten, and G.L. Violante. 2017. “Optimal tax progressivity: An analytical framework.” *Quarterly Journal of Economics* 132 (4): 1693–754. [137, 146]

Jaimovich, Nir, and Sergio Rebelo. 2009. “Can News about the Future Drive the Business Cycle?” *American Economic Review* 99 (4): 1097–118. [137]

Jermann, U.J., and V. Quadrini. 2007. “Stock market boom and the productivity gains of the 1990s.” *Journal of Monetary Economics* 54: 413–32. [125]

Kaplan, Greg, Benjamin Moll, and Giovanni L. Violante. 2018. “Monetary Policy According to HANK.” *American Economic Review* 108 (3): 697–743. [120]

Kaplan, Greg, Giovanni L. Violante, and Justin Weidner. 2014. “The Wealthy Hand-to-Mouth.” *Brookings Papers on Economic Activity* 45 (1): 77–153. [143, 151, 152]

Karabarbounis, Loukas, and Brent Neiman. 2014. “The Global Decline of the Labor Share.” *Quarterly Journal of Economics* 129 (1): 61–103. [125, 150]

Karabarbounis, Loukas, and Brent Neiman. 2019. “Accounting for Factorless Income.” *NBER Macroeconomics Annual* 33: 167–228. [125, 126, 144]

Karnizova, Lilia. 2012. “News Shocks, Productivity and the U.S. Investment Boom-Bust Cycle.” *B.E. Journal of Macroeconomics* 12 (1). [146]

Kehrig, Matthias, and Nicolas Vincent. 2021. “The Micro-Level Anatomy of the Labor Share Decline.” *Quarterly Journal of Economics* 136 (2): 1031–87. [144]

Kekre, Rohan, and Moritz Lenel. 2022. “Monetary Policy, Redistribution, and Risk Premia.” *Econometrica* 90 (5): 2249–82. [126]

Krusell, Per, Toshihiko Mukoyama, and Anthony A. Smith. 2011. “Asset prices in a Huggett economy.” *Journal of Economic Theory* 146 (3): 812–44. [127]

Kuhn, Moritz, Moritz Schularick, and Ulrike Steins. 2020. “Income and Wealth Inequality in America, 1949-2016.” *Journal of Political Economy* 128 (9): 3469–519. [123, 124, 143, 149, 156, 159, 171, 172]

Kuvshinov, Dmitry. 2022. “The Co-Movement Puzzle.” Proceedings of Paris December 2021 Finance Meeting EUROFIDAI - ESSEC. [119, 122, 163, 165]

Leeper, Eric M., and Todd B. Walker. 2011. “Information flows and news driven business cycles.” *Review of Economic Dynamics* 14: 55–71. [140]

Li, Wendy C. Y., and Bronwyn H. Hall. 2020. “Depreciation of Business R&D Capital.” *Review of Income and Wealth* 66 (1): 161–80. [125, 143]

Luetticke, Ralph. 2021. “Transmission of Monetary Policy with Heterogeneity in Household Portfolios.” *American Economic Journal: Macroeconomics* 13 (2): 1–25. [125]

McGrattan, Ellen R., and Edward C. Prescott. 2010. “Unmeasured Investment and the Puzzling US Boom in the 1990s.” *American Economic Journal: Macroeconomics* 2 (4): 88–123. [129, 149]

Philippon, Thomas. 2015. “Has the US Finance Industry Become Less Efficient? On the Theory and Measurement of Financial Intermediation.” *American Economic Review* 105 (4): 1408–38. [146]

Ravn, Morten O., and Vincent Sterk. 2017. “Job uncertainty and deep recessions.” *Journal of Monetary Economics* 90: 125–41. [127, 128]

Saez, Emmanuel, and Gabriel Zucman. 2016. “Wealth inequality in the united states since 1913: Evidence from capitalized income tax data.” *Quarterly Journal of Economics* 131 (2): 519–78. [145, 149]

Sagi, Jacob S. 2020. “Asset-Level Risk and Return in Real Estate Investments.” *Review of Financial Studies* 34 (8): 3647–94. [120]

Schmitt-Grohé, Stephanie, and Martín Uribe. 2012. “What’s News in Business Cycles.” *Econometrica* 80 (6): 2733–64. [137, 140]

Smith, Matthew, Danny Yagan, Owen Zidar, and Eric Zwick. 2019. “Capitalists in the Twenty-First Century.” *Quarterly Journal of Economics* 134 (4): 1675–745. [126, 161]

Sorensen, Morten, Neng Wang, and Jinqiang Yang. 2014. “Valuing Private Equity.” *Review of Financial Studies* 27 (7): 1977–2021. [120]

U.S. Bureau of Economic Analysis. 2019. “Chapter 6: Private Fixed Investment.” NIPA Handbook. Washington,DC:BEA. [143]

Chapter 3

A Krusell-Smith Type Approximation of the Second Order Solution to a Heterogeneous Agent Model*

3.1 Introduction

The advancement of business cycle models with heterogeneous agents has provided macroeconomists with a powerful toolbox to analyze the effects of monetary and fiscal policy or exogenous shocks on the business cycle and inequality. However, most solution techniques so far rely on the assumption of certainty equivalence[1]: while the agents take their idiosyncratic risks into account, they do not react to aggregate risk. The perturbation approach to the modeling of incomplete markets, which was started by Reiter (2009) and subsequently developed by Bayer and Luetticke (2020) to allow for heterogeneity in several dimensions, in principle allows for solving the model up to higher orders, where the agents' policies react to aggregate risk.[2] However, a large number of system-variables (around 10^3) complicates the computation of the second order solution, as the Hessian has to be computed and stored sparsely, and a large-scale generalized Sylvester equation has to be solved. Even with further

* I thank Christian Bayer, Ralph Luetticke, and Sebastian Hildebrand. I gratefully acknowledge financial support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the RTG 2281 - The Macroeconomics of Inequality and from the ERC-CoG project with agreement ID 724204.

1. The exception are global solution methods, building on Krusell and Smith (1998), that use machine learning to allow for nonlinearities, see for example Fernández-Villaverde, Hurtado, and Nuño (2019) and Azinovic, Gaegauf, and Scheidegger (2022) To be practicable, these methods usually rely on good initial guesses for their iterations. Hence, advancements in solving models by perturbation can be complementary to that literature.

2. This contrasts with the sequence-space method proposed by Auclert et al. (2021), which can account for nonlinearities in aggregate model equations, but necessitates certainty equivalence on the agents' part.

dimensionality-reduction efforts[3], the computational burden of calculating the second order solution may become prohibitively large for some use cases, in particular for bringing the model to the data.

In this paper, we propose several ways to make the computation of the second order solution more efficient. For this, we build on the results of Levintal (2017), who introduces a Kronecker-notation for perturbation methods, and Bayer and Luetticke (2020), who show that the knowledge of the typical structural of a heterogeneous agent-model helps to identify non-zero parts of the Hessian. We make three contributions to the literature: First, we show how to efficiently compute Kronecker-products, that may appear to become prohibitively large with a growing variable-space, in the context of a second order perturbation. Second, we show analytically that the main computational burden of solving a heterogeneous agent model up to second order relates to the nonlinear impact of the distribution (e.g. the distribution of wealth in the economy). Third, we propose a method to approximate the second order solution to the model that hinges on the idea that the average level of a (marginal) distribution should be a sufficient state to determine the optimal policy of agents. We show that this is only the case when agents estimate the overall distribution from its first moment (as in Krusell and Smith (1998)). In an application, we demonstrate that the approximation considerably speeds up the computation of the solution, while capturing well the full second-order model dynamics in the short run.

We provide a thorough analysis of the reason for the approximation's ability to replicate some of the nonlinear dynamics, and describe the nonlinearities that it misses when compared to the correct second order solution. We find that the distribution has direct and indirect nonlinear effects on the aggregate economy. Consider the evolution of the distribution over time. The share of agents who hold a certain amount of wealth — say, one bar of gold — in one period is determined by how many agents held a similar amount of wealth last period, and by their propensity to save or consume their wealth in that period. Now, giving some agents more gold changes both factors: the share of agents with one bar of gold may rise, but agents may also be less willing to save the same amount of their wealth. The interaction of the two factors induces a nonlinearity, which we call the *direct* nonlinear effect of the distribution.

Giving some agents more gold will reduce the price of gold in the economy. For agents who own many bars of gold, this implies a significant wealth-loss. They may optimally try to counteract this by buying more gold. However, they might be less inclined to react this way when the issuance of gold comes at a time when the risk of getting unemployed is higher than in normal times. Then, they might prefer to save in a more liquid asset, like money. This means that the optimal policy reaction

3. Bayer, Born, and Luetticke (2022) propose a two-step approach, where extracting the most important factors from the full linearized model significantly reduces the variable-space and speeds up the estimation of the model.

to a change in the distribution may well be *state-dependent*, and hence induce a nonlinearity. We call this the *indirect* nonlinear effect of the distribution, since it works through the distribution's impact on another (aggregate) variable that itself has nonlinear effects.

We show that the presence of the indirect nonlinear effects of the distribution makes up a large share of the computational burden of solving the model up to the second order. The approximation to the solution that we propose allows us to abstract from the indirect nonlinear effects, and can thus be computed at a significant speed gain. However, note that the direct nonlinear effects only describe the nonlinear effect on the distribution itself, not on any aggregate variables in the system. Therefore, we feed the direct nonlinear effects into the correct first order solution of the model, where aggregate states and controls are influenced by the distribution. The quality of the approximation hinges on the importance of the direct nonlinear effects, that are fed through the first order dynamics, versus the indirect nonlinear effects of the distribution. In our application, we find that the direct nonlinear effects drive the nonlinearity in the short run. Apart from the concrete application, the approximation in general provides a middle ground between the first order solution and the "full" second order solution.

The paper is structured as follows: in section 3.2, we describe the general model setting and equilibrium concept that our theory applies to. In section 3.3, we introduce the perturbation-notation of Levintal (2017) for the first- and second order. In section 3.4, we propose efficient ways for computing the Hessian and the generalized Sylvester equation. In section 3.5, we thoroughly analyze the nonlinear impact of the distribution and propose an approximation to the second-order solution that significantly reduces computation times by disentangling its direct and indirect nonlinear effects. Section 3.6 concludes.

3.2 Setting

We consider the problem of solving a forward-looking state-space system (as usually encountered in macroeconomics) with heterogeneity up to the second order. The states of the system encompass variables that are predetermined and endogenous (called D_t), and variables that follow a stable first-order autoregressive process with exogenous innovations (called X_t). They occur both on the level of the individual agent, denoted by $s_{it} = \begin{bmatrix} x_{it} \\ d_{it} \end{bmatrix}$, and on the aggregate level: $S_t = \begin{bmatrix} X_t \\ D_t \end{bmatrix}$. Additionally, the distribution of agents over the individual states s_{it} is an aggregate state, denoted by μ_t. We call H the autoregressive process that connects X_t and X_{t+1}, i.e. $X_{t+1} = H(X_t) + \sigma_t \epsilon^A_{t+1}$, where σ_t is the perturbation parameter, and define H^D as a function that determines D_{t+1} from the aggregates states in t, i.e. $D_{t+1} = H^D(S_t, \mu_t)$.

The system is forward-looking in the sense that individual agents solve a forward-looking planning problem over the infinite horizon. We present it in the form of a Bellman equation:

$$\begin{aligned} v(x_{it}, d_{it}, S_t, \mu_t) &= \max_{d_{it+1}} \{u(x_{it}, d_{it}, d_{it+1}; P_t) + \beta\, \mathrm{E}_t\, v(x_{it+1}, d_{it+1}, S_{t+1}, \mu_{t+1})\} \\ &\quad \text{s.t. } d_{it+1} \in \Gamma(s_{it}, P_t), \end{aligned} \tag{3.1}$$

where Γ defines a budget set that is determined by the individual state and aggregate prices $P_t = P(S_t, \mu_t)$.[4] We define a time-varying value function $v_t(s_{it}, S_t, P_t) := v(s_{it}, S_t, \mu_t)$ that does not directly depend on the distribution, which is captured by the time dimension. We assume that v_t is separable into a time-constant function, and a time-varying function of the idiosyncratic state: $v_t(s_{it}, S_t, P_t) = v_t(s_{it})\, v_0(s_{it}, S_t, P_t)$.[5] We call h_t^d the (time-varying) optimal policy of the individual planning problem, i.e.

$$h_t^d(x, d) = \underset{d' \in \Gamma(x,d;P_t)}{\operatorname{argmax}} \left\{ u(x, d, d'; P_t) + \beta\, \mathrm{E}_t \left[v_{t+1}(x', d')\, v_0(x', d', S_{t+1}, P_{t+1}) \right] \right\}. \tag{3.2}$$

We call h^x the process that describes the evolution of the idiosyncratic exogenous states, i.e. $x_{it} = h^x(s_{it-1}) + \epsilon_t^I$.

Since we want to solve for a sequence of equilibria, we need to impose that markets clear. Typically, this creates m^d additional constraints, where m^d is the dimension of d_{it}, i.e. the endogenous states the agent chooses (e.g. capital or money). We write $\Phi_t(h_t^d, \mu_t)$ for the excess demand function that takes as arguments the individual optimal policies and the density over the agents. Via the time index, it also depends on the aggregate states and prices.

For general (continuous) state spaces, μ_t and $v_t(s_{it})$ are infinite dimensional objects. In order to solve the system numerically, we approximate them by finite objects. First, we restrict the values of the individual states s_{it} to lie on a fixed grid, $\bar{s} = (\bar{s}_1, ..\bar{s}_n)$. Since the optimal policy may potentially be a non-grid point, we approximate it by the two nearest grid points that the agent reaches with such probabilities that in expectation, she acts optimally (i.e., we force the agent to only use mixed strategies over two neighboring grid points).[6] We denote the discrete distribution (a histogram) over the grid by $d\mu$. Second, we approximate the optimal value function $v_t(s_{it})$ by a linear interpolant between grid points, so that only the on-grid values have to be saved. From now on, we define v_t as the vector of optimal values over the grid.

4. We assume that the set of prices is rich enough to cover all moments of the distribution relevant to the individual optimization problem. Otherwise, μ_t would directly enter the budget set.

5. v_0 can be thought of as governing the marginal return to assets, which may be level-dependent (e.g. because of a borrowing fee).

6. When the dimension of endogenous states is $m^d > 1$, the optimal policy in one dimension depends on those in the other dimensions, and so one may alternatively approximate the optimal solution by a span over 2^{m^d} gridpoints. In the following, we always "flatten" the grid, or functions on the grid, so that they can be represented as vectors.

3.2.1 Fokker-Planck and Bellman equation

The functions h^d_t, h^x, together with the approximation of optimal policies by mixed strategies and the probability distribution of ϵ^I, induce a transition matrix Π_{h_t} for the histograms, $d\mu_{t+1} = d\mu_t \Pi_{h_t}$, which is a discrete time Fokker-Planck equation. At the same time, the approximations imply that the Bellman equation imposes the conditions $v_t = u_{h^d_t} + \beta \Pi_{h_t} v_{t+1} \odot v_0(\bar{s}, S_{t+1}, P_{t+1})$ on the vectors v_t, v_{t+1}, where $u_{h^d_t} := u(\bar{s}, h^d_t(\bar{s}); P_t)$ is the vector of the (optimal) utility values at all grid nodes.[7]

3.2.2 Steady state

We solve the system around the non-stochastic steady state, or stationary equilibrium, which is defined by aggregate determinancy, i.e. $\sigma \equiv 0$. It has the following attributes:

(1) The steady state-optimal policies $\bar{h}^d(\bar{s})$ solve $\bar{h}^d(\bar{s}) = \text{argmax}_{d' \in \Gamma(\bar{s}, \bar{P})} \left\{ u(\bar{s}, d'; \bar{P}) + \beta \Pi_{\bar{h}(d')} \bar{v} \right\}$, where $\bar{P} = P(\bar{S}, d\bar{\mu})$.
(2) The steady state optimal values $\bar{v}$ are consistent with $\bar{h}^d$: $\bar{v} = u_{\bar{h}^d} + \beta \Pi_{\bar{h}} \bar{v}$.
(3) The transition matrix of idiosyncratic states, $\Pi_{\bar{h}}$, induced by $\bar{h}^d, h^x$ and the probability distribution of ϵ^I, is irreducible and has the ergodic distribution $d\bar{\mu}$.
(4) Markets clear: $\Phi(\bar{h}^d, d\bar{\mu}) = 0$.

3.2.3 Sequential equilibrium

We collect the variables and equilibrium conditions to write the sequential equilibrium in the form

$$\mathrm{E}_t F(\hat{C}_{t+1}, \hat{C}_t, \hat{S}_{t+1}, \hat{S}_t) \ = \ 0. \tag{3.3}$$

For this, we define $\hat{S}_t := [d\mu_t, X_t, D_t, \sigma_t]^T$ as the states, and $\hat{C}_t := [v_t, P_t, C_t]^T$ as the controls of the system F.[8] C_t are static variables, which we include here for generality (they are inessential for the system dynamics). In the following, we will also call $d\mu_t$ the "idiosyncratic" and X_t, D_t, σ_t the "aggregate" states of the system, and similarly we call v_t the "idiosyncratic" and P_t, C_t the "aggregate" controls.[9]

7. $\odot$ denotes the Hadamard, i.e. element-wise, product.
8. We denote transpositions as $(\cdot)^T$, while $(\cdot)'$ stands for future variables.
9. Therefore, "idiosyncratic" and "aggregate" now have a different meaning than above: they both relate to aggregate variables, since we write and perturb the system only with respect to aggregate variables, but "idiosyncratic" variables differ by being only relevant in a heterogeneous agent model where differences across agents matter for aggregate dynamics.

The corresponding static equilibrium conditions are subsumed in Φ_t, together with the market clearing conditions. F calculates all equilibrium conditions:

$$F(\hat{C}_{t+1}, \hat{C}_t, \hat{S}_{t+1}, \hat{S}_t) \;=\; \begin{bmatrix} d\mu_{t+1} - d\mu_t \Pi_{h_t} \\ X_{t+1} - H^X(S_t, C_t) + \sigma_t \epsilon^A_{t+1} \\ D_{t+1} - H^D(S_t, C_t, d\mu_t) \\ v_t - (u_{h^d_t} + \beta \Pi_{h_t} v_{t+1} \odot v_0(\bar{s}, S_{t+1}, P_{t+1})) \\ \Phi_t(h^d_t, d\mu_t) \\ \sigma_{t+1} - \sigma_t \end{bmatrix}, \tag{3.4}$$

where h^d_t is the optimal policy function given v_{t+1}, P_t, S_{t+1}, and P_{t+1}, and ϵ^A_{t+1} is given exogenously.

3.3 Perturbation: Derivations

Define as $ns = n^\mu + n^X + n^D + 1$ the number of states, divided in the size of the distribution, n^μ [10], and the aggregate states. Similiarly, let $nc = n^v + nc^A$ denote the number of controls, where n^v is the size of the optimal values on the grid[11], and nc^A the number of aggregate controls.

We follow the approach of Levintal (2017) to solve for the system dynamics caused by perturbations around the steady state. We assume that there exist functions g, h such that

$$\mathrm{E}_t F(g(h(\hat{S}_t) + \sigma_t \eta \epsilon^A_{t+1}), g(\hat{S}_t), h(\hat{S}_t) + \sigma_t \eta \epsilon^A_{t+1}, \hat{S}_t) \;=\; 0 \tag{3.5}$$

on an open set around the steady state. They provide a solution to the system in the form of a state transition h, and a state-to-control mapping g. $\eta \in \mathbb{R}^{ns \times n^X}$ maps the exogenous innovations to the position of X in the state vector.

3.3.1 First order approximation

Let $\hat{S}^{SS}, \hat{C}^{SS}$ denote the steady state values of the states and the controls. Clearly, $\mathrm{E}_t F(\hat{C}^{SS}, \hat{C}^{SS}, \hat{S}^{SS}, \hat{S}^{SS}) = 0$, and hence $h(\hat{S}^{SS}) = \hat{S}^{SS}$, $g(\hat{S}^{SS}) = \hat{C}^{SS}$. The linear approximation to the solution around the steady state is then given by

$$h(\hat{S}_t) \;\approx\; \hat{S}^{SS} + h_{\hat{S}}(\hat{S}_t - \hat{S}^{SS}), \;\; g(\hat{S}_t) \;\approx\; \hat{C}^{SS} + g_{\hat{S}}(\hat{S}_t - \hat{S}^{SS}), \tag{3.6}$$

where $h_{\hat{S}}, g_{\hat{S}}$ are defined as the Jacobian matrices of h, g at the point $\hat{S}^{SS}$. Define $F_1, .., F_4$ as the partial derivatives of F of the first to last argument, at the steady state. Defining the matrices $A := [F_3 \; F_1]$ and $B := -[F_4 \; F_2]$, we can use the method

10. n^μ is the number of variables needed to represent $d\mu$ in the system. With dimensionality reduction techniques, this may be substantially smaller than the ("flattened") grid size n.

11. Analogously to n^μ, n^v can be reduced by dimensionality reduction techniques, e.g. projection.

by Klein (2000) that solves the generalized eigenvalue problem $AZ\Lambda = BZ$, where Z is a unitary matrix and Λ is a diagonal matrix with the generalized eigenvalues of the system on its diagonal, with the Schur decomposition. It yields the solution $h_{\hat{S}}$ and $g_{\hat{S}}$, provided the Blanchard-Kahn conditions are fulfilled.

3.3.2 Second order

To derive a solution to the second order approximation, i.e. $h_{\hat{S}\hat{S}}$ and $g_{\hat{S}\hat{S}}$, we introduce more notation from Levintal (2017). First, we define the vector of all endogenous variables $v_t = [\hat{C}_{t+1}, \hat{C}_t, \hat{S}_{t+1}, \hat{S}_t]^T$. Next, we define the matrix $\zeta_t := [0_{ns\times(ns-1)} \;\; \eta\epsilon_{t+1}]$, which allows us to write the state transition as $\hat{S}_{t+1} = h(\hat{S}_t) + \zeta_t\hat{S}_t$. Dropping the time subscripts, we can write the system as

$$\mathrm{E}F(v(\hat{S},\zeta)) = 0, \text{ with } v(\hat{S},\zeta) = \begin{bmatrix} g(h(\hat{S}) + \zeta\hat{S}) \\ g(\hat{S}) \\ h(\hat{S}) + \zeta\hat{S} \\ \hat{S} \end{bmatrix}. \tag{3.7}$$

In what follows, we reshape the second derivates f_{xx} of multi-dimensional functions $f : \mathbb{R}^{nx} \to \mathbb{R}^{ny}$, which are three dimensional $ny \times nx \times nx$ tensors, into two dimensional $ny \times nx^2$ matrices. The repeated application of the chain rule yields $F_{\hat{S}\hat{S}} = F_{vv}v_{\hat{S}}^{\otimes 2} + F_v v_{\hat{S}\hat{S}}$, where the superscript $^{\otimes 2}$ denotes a Kronecker power, i.e. the Kronecker product of a matrix with itself. All derivatives are taken at the steady state. Since $\tilde{F}(\hat{S}) := \mathrm{E}F(v(\hat{S},\zeta)) = 0$ on an open set around $\hat{S}^{SS}$, it follows that $0 = \tilde{F}_{\hat{S}\hat{S}} = \mathrm{E}F_{\hat{S}\hat{S}}$. We derive

$$v_{\hat{S}} = \underbrace{\begin{bmatrix} g_{\hat{S}}h_{\hat{S}} \\ g_{\hat{S}} \\ h_{\hat{S}} \\ I_{ns} \end{bmatrix}}_{=:V_{\hat{S}}^0} + \underbrace{\begin{bmatrix} g_{\hat{S}} \\ 0_{nc\times ns} \\ I_{ns} \\ 0_{ns\times ns} \end{bmatrix}}_{=:V_{\hat{S}}^1} \zeta \text{ and } \mathrm{E}v_{\hat{S}\hat{S}} = \begin{bmatrix} g_{\hat{S}\hat{S}}h_{\hat{S}}^{\otimes 2} + g_{\hat{S}}h_{\hat{S}\hat{S}} + g_{\hat{S}\hat{S}}\,\mathrm{E}\,\zeta^{\otimes 2} \\ g_{\hat{S}\hat{S}} \\ h_{\hat{S}\hat{S}} \\ 0_{ns\times ns^2} \end{bmatrix}. \tag{3.8}$$

Defining the auxiliary matrices $A_2 := F_{vv}\,\mathrm{E}\,v_{\hat{S}}^{\otimes 2}$ and $B_2 := h_{\hat{S}}^{\otimes 2} + \mathrm{E}\,\zeta^{\otimes 2}$, we finally obtain the equation

$$A_2 + F_1 g_{\hat{S}\hat{S}} B_2 + F_2 g_{\hat{S}\hat{S}} + (F_3 + F_1 g_{\hat{S}})h_{\hat{S}\hat{S}} = 0, \tag{3.9}$$

in the unknowns $h_{\hat{S}\hat{S}}$ and $g_{\hat{S}\hat{S}}$.

We note that the transition of σ (in the last row of F) is linear and independent of all other variables, which implies empty last rows in A_2, F_1, F_2, and $h_{\hat{S}\hat{S}}$. Define $\hat{\mathscr{S}}$ as the state vector without σ, $\mathscr{F}$ as F without its last row and with σ stripped from its arguments, i.e. $\mathscr{F}(\hat{C}', \hat{C}, \hat{\mathscr{S}}', \hat{\mathscr{S}})$, and $\mathscr{H}$ as the state transition from $\hat{S}$ to $\hat{\mathscr{S}}'$. Then, we can split equation (3.9) into

$$\mathscr{A}_{21} + \mathscr{F}_1 g_{\hat{\mathscr{S}}\hat{\mathscr{S}}} \mathscr{H}_{\hat{\mathscr{S}}}^{\otimes 2} + \mathscr{F}_2 g_{\hat{\mathscr{S}}\hat{\mathscr{S}}} + (\mathscr{F}_3 + \mathscr{F}_1 g_{\hat{\mathscr{S}}}) \mathscr{H}_{\hat{\mathscr{S}}\hat{\mathscr{S}}} = 0, \tag{3.10}$$

$$\mathscr{A}_{22} + \mathscr{F}_1 g_{\hat{\mathscr{S}}\hat{\mathscr{S}}} \operatorname{E}(\eta_{-1,:}\epsilon)^{\otimes 2} + (\mathscr{F}_1 + \mathscr{F}_2) g_{\sigma\sigma} + (\mathscr{F}_3 + \mathscr{F}_1 g_{\hat{\mathscr{S}}}) \mathscr{H}_{\sigma\sigma} = 0, \tag{3.11}$$

where $\mathscr{A}_{21}$ and $\mathscr{A}_{22}$ copy A_2 without the last row, and without the columns pertaining to σ ($\mathscr{A}_{21}$), or the columns pertaining to $\hat{\mathscr{S}}$ ($\mathscr{A}_{22}$).

3.4 Computation of the second order perturbation

First, we note that a closed-form solution to equation (3.11) is attainable if we already computed $g_{\hat{\mathscr{S}}\hat{\mathscr{S}}}$, as we then get

$$\begin{bmatrix} \mathscr{H}_{\sigma\sigma} \\ g_{\sigma\sigma} \end{bmatrix} = -\underbrace{\begin{bmatrix} \mathscr{F}_3 + \mathscr{F}_1 g_{\hat{\mathscr{S}}} & \mathscr{F}_1 + \mathscr{F}_2 \end{bmatrix}}_{(ns-1+nc)\times(ns-1+nc)}^{-1} (\mathscr{A}_{22} + \mathscr{F}_1 g_{\hat{\mathscr{S}}\hat{\mathscr{S}}} \operatorname{E}(\eta_{-1,:}\epsilon)^{\otimes 2}). \tag{3.12}$$

Second, we can rewrite equation (3.10) in the form of a generalized Sylvester equation:

$$\mathbb{A}\mathscr{X} + \mathbb{B}\mathscr{X}\mathscr{H}_{\hat{\mathscr{S}}}^{\otimes 2} + \mathscr{A}_{21} = 0, \tag{3.13}$$

where $\mathscr{X} = \begin{bmatrix} \mathscr{H}_{\hat{\mathscr{S}}\hat{\mathscr{S}}} \\ g_{\hat{\mathscr{S}}\hat{\mathscr{S}}} \end{bmatrix}$, $\mathbb{A} = \begin{bmatrix} \mathscr{F}_3 + \mathscr{F}_1 g_{\hat{\mathscr{S}}} & \mathscr{F}_2 \end{bmatrix}$, and $\mathbb{B} = \begin{bmatrix} 0_{ns-1+nc,ns-1} & \mathscr{F}_1 \end{bmatrix}$.

To summarize, the computational challenge is twofold:

(1) Compute and store the Hessian of the system F, F_{vv}. For large systems (with $ns + nc \approx 10^3$ or higher), F_{vv} has to be stored sparsely, which will allow to efficiently compute A_2 (even though $v_{\hat{\mathscr{S}}}^{\otimes 2}$ may itself be too large to store).

(2) Solve the generalized Sylvester equation (3.13).

3.4.1 Computation of F_{vv}

In order to sparsely fill the Hessian F_{vv}, we identify its 0-entries as implied by the model[12], and propose a manner to efficiently compute the cross derivatives that are (possibly) non-zero. It is useful[13] to split up F, as defined in equation (3.4), into the parts F^h and F^A, where

$$F^h(\cdot) := \begin{bmatrix} d\mu_{t+1} - d\mu_t \Pi_{h_t} \\ \nu_t - (u_{h_t^d} + \beta \Pi_{h_t} \nu_{t+1} \odot \nu_0(\bar{s}, S_{t+1}, P_{t+1})) \\ \Phi_t(h_t^d, d\mu_t) \end{bmatrix}, F^A(\cdot) := \begin{bmatrix} X_{t+1} - H^X(S_t, C_t) + \sigma_t \epsilon_{t+1}^A \\ D_{t+1} - H^D(S_t, C_t, d\mu_t) \\ \sigma_{t+1} - \sigma_t \end{bmatrix}. \tag{3.14}$$

12. Here, we follow Bayer and Luetticke (2020).

13. To be more precise, the split up is instructive for identifying zero-columns of F_{vv}, and it shows a way to fill some parts of F_{vv} without having to compute the whole function.

F^h encompasses the equilibrium conditions that require the computation of the optimal policy $h_t^d(\bar{s})$, while F^A contains the rest.

First, we consider the partial derivatives of F^h, and point out the ones that are known to be constant.[14] We know that $F^h_{d\mu_{t+1}} = [I, 0, 0]^T$, $F^h_{\nu_t} = [0, I, 0]^T$, and $F^h_{d\mu_t} = \left[-\Pi_{h_t}, 0, \frac{\partial \Phi}{\partial d\mu_t}\right]^T$. We note that, while the Fokker-Planck equation is linear in the histogram, $F^h_{d\mu_t}$ is not constant with respect to the variables in the system that affect the transition matrix Π_{h_t}. Without much loss of generality, we can assume the following:

Assumption 3.4.1. $\frac{\partial \Phi}{\partial d\mu_t}$ *is constant, i.e. the distribution enters only linearly into market clearing conditions and static variable definitions.*

Derivatives of F^h with respect to variables that influence the individual state transition h_t are possibly non-constant: this includes ν_{t+1}, P_t (directly through h_t^d), S_{t+1}, P_{t+1} (through ν_0), and S_t (through h^x). Identifying the zeros in $\frac{\partial \nu_0}{\partial S_{t+1}}, \frac{\partial \nu_0}{\partial P_{t+1}}$ and $\frac{\partial h^x}{\partial S_t}$ uncovers some of those partial derivatives as zeros. For the remaining variables, $F^h_{C_t} = [0, 0, \frac{\partial \Phi}{\partial C_t}]^T$ and $0 = F^h_{C_{t+1}} = F^h_{\sigma_t} = F^h_{\sigma_{t+1}}$.

Turning to F^A, we first note its independence from ν_t, ν_{t+1} and $d\mu_{t+1}$, and hence $0 = F^A_{\nu_t} = F^A_{\nu_{t+1}} = F^A_{d\mu_{t+1}}$. In contrast, the evolution of the endogenous state variables may depend on the distribution, i.e. $F^A_{d\mu_t} = \left[0, -\frac{\partial H^D}{d\mu_t}, 0\right]^T$.

Assumption 3.4.2. $\frac{\partial H^D}{d\mu_t}$ *is constant, i.e. the distribution affects the evolution of endogenous state variables only linearly.*

The partial derivatives with respect to S_t, P_t, C_t and S_{t+1}, P_{t+1} depend on the specific implementation of H^X and H^D. In particular, if the variables enter the conditions linearly, we have constant $F^A_{S_{t+1}}, F^A_{P_{t+1}}$. Finally, we find the constant partial derivatives $F^A_{C_{t+1}} = 0, F^A_{\sigma_t} = [\epsilon_{t+1}, 0, -1]^T$[15], and $F^A_{\sigma_{t+1}} = [0, 0, 1]^T$.

Let $\mathbb{V}$ denote the set of variables for which all cross-derivatives, as well as their second-order derivatives, are zeros (translating to zero-columns of F_{vv}). We have seen that $\mathbb{V}$ contains $\nu_t, d\mu_{t+1}, C_{t+1}, \sigma_t$, and σ_{t+1}. Even with the assumptions 3.4.1 and 3.4.2, $d\mu_t$ is *not* in $\mathbb{V}$: the Fokker-Planck equation implies non-zero cross-derivatives between the histogram $d\mu_t$ and variables that determine the optimal policy h_t, namely ν_{t+1} and some aggregate variables.[16] In addition, we define the

14. Of course, this knowledge is also useful to make the computation of the first order approximation more efficient.

15. The exogenous innovations ϵ are independent of all variables in the system, and hence count as constant.

16. Note that, aside from this *direct* second order effect on the system, the distribution also affects the optimal policy h_t^d, e.g. through their effects on prices, $\frac{\partial P_t}{\partial d\mu_t}$, and thereby has additional *indirect* second order effects (since h_t^d affects the system non-linearly). See section 3.5.

sets $\mathbb{V}^h$ and $\mathbb{V}^A$ of variables not in $\mathbb{V}$, that yield zero cross- and second-order derivatives of either F^h or F^A. Using knowledge about v_0 and h^x places a share of aggregate states and future prices in $\mathbb{V}^h$, while $\mathbb{V}^A$ contains at least v_{t+1}, and $d\mu_t$ if assumption 3.4.2 holds.

We propose the following procedure to calculate the non-zero cross- and second-order derivatives:

(1) For each F^h and F^A, split up the variables not in $\mathbb{V} \cup \mathbb{V}^{h/A}$ into groups of the same size.[17] For all possible pairings among these groups, define auxiliary functions, wrapping F^h or F^A, that only take variables of the pair as arguments, while the remaining variables are constant at their steady state values.

(2) Calculate the Hessian of each auxiliary function using automatic differentiation.[18] This step is suitable for parallelization.

(3) Fill F_{vv} sparsely with the computed cross- and second-order derivatives.[19]

3.4.2 Solving the generalized Sylvester equation

Let $n^v = 2 * (ns + nc)$ denote the length of the vector v_t. Dividing the columns of F_{vv} into n^v blocks of length n^v, let $F_{vv}(j)$ denote the j-th such block.[20] In order to compute $A2 = F_{vv}\,\mathrm{E}\,v_{\hat{\mathscr{S}}}^{\otimes 2}$, we first use equation (3.8) to see that

$$A2 = \underbrace{F_{vv} {V^0_{\hat{\mathscr{S}}}}^{\otimes 2}}_{=:I} + \underbrace{F_{vv} {V^1_{\hat{\mathscr{S}}}}^{\otimes 2}}_{=:J} \mathrm{E}\,\zeta^{\otimes 2}. \tag{3.15}$$

To calculate matrix I (and analogously J) without generating the (prohibitively large) Kronecker product, we first compute the $ns(ns+nc) \times n^v$-matrix M, defined as[21]

$$M = \left[vec(F_{vv}(1) V^0_{\hat{\mathscr{S}}}) \cdots vec(F_{vv}(n^v) V^0_{\hat{\mathscr{S}}}) \right], \tag{3.16}$$

which is sparse (many zero-columns) due to the sparsity of F_{vv}. Reshaping $M \cdot V^0_{\hat{\mathscr{S}}}$ gives I.

17. Specifically, it may be advisable to break up a large variable block like v_{t+1} into smaller parts.

18. So far, only forward mode automatic differentiation, where dual numbers are seeded into the function, seems to be feasible. In practice, we nest the jacobian-command from the Julia-package ForwardDiff.

19. The drawback of this procedure is that all second-order derivatives will be calculated multiple times.

20. For interpretation, $F_{vv}(j)[:, j]$ contains the second-order derivatives of the j-th variable in v_t, and the remaining columns of block $F_{vv}(j)$ contain the cross-derivatives of this variable with all the other variables.

21. $vec()$ denotes the shaping of a matrix into a column vector.

To solve equation (3.13), we first multiply by the left-inverse of $\mathbb{A}$ to get

$$\mathscr{X} + \mathbb{A}^{-1}\mathbb{B}\mathscr{X}\mathscr{H}_{\hat{\mathscr{S}}}^{\otimes 2} + \mathbb{A}^{-1}\mathscr{A}_{21} = 0. \tag{3.17}$$

The first $ns-1$ columns of $\mathbb{B}$, and hence $\mathbb{A}^{-1}\mathbb{B}$, contain only zeros (additional zero columns in $\mathbb{B}$ may come from zero columns in $\mathscr{F}_1$). Hence, the lower nc rows of equation (3.17) are a condition on $g_{\hat{\mathscr{S}}\hat{\mathscr{S}}}$ only, independently of $\mathscr{H}_{\hat{\mathscr{S}}\hat{\mathscr{S}}}$. What is more, once we have solved for $g_{\hat{\mathscr{S}}\hat{\mathscr{S}}}$, the closed-form solution for $\mathscr{H}_{\hat{\mathscr{S}}\hat{\mathscr{S}}}$ is given by

$$\mathscr{H}_{\hat{\mathscr{S}}\hat{\mathscr{S}}} = -((\mathbb{A}^{-1}\mathbb{B})_{(i^{\mathscr{H}},j^{g})}g_{\hat{\mathscr{S}}\hat{\mathscr{S}}}\mathscr{H}_{\hat{\mathscr{S}}}^{\otimes 2} + (\mathbb{A}^{-1}\mathscr{A}_{21})_{(i^{\mathscr{H}},\cdot)}), \tag{3.18}$$

where $i^{\mathscr{H}}, j^{g}$ denote the upper row and lower column indices of the respective matrices.

We propose a simple iterative algorithm to solve the lower part of (3.17) for $g_{\hat{\mathscr{S}}\hat{\mathscr{S}}}$:

(1) Start with an initial guess $g_{\hat{\mathscr{S}}\hat{\mathscr{S}}}^{(0)}$.

(2) Update with

$$g_{\hat{\mathscr{S}}\hat{\mathscr{S}}}^{(i+1)} = -((\mathbb{A}^{-1}\mathbb{B})_{(i^{g},j^{g})}g_{\hat{\mathscr{S}}\hat{\mathscr{S}}}^{(i)}\mathscr{H}_{\hat{\mathscr{S}}}^{\otimes 2} + (\mathbb{A}^{-1}\mathscr{A}_{21})_{(i^{g},\cdot)}). \tag{3.19}$$

(3) Stop if $|g_{\hat{\mathscr{S}}\hat{\mathscr{S}}}^{(i+1)} - g_{\hat{\mathscr{S}}\hat{\mathscr{S}}}^{(i)}|$ is smaller than the tolerance $\epsilon > 0$.[22]

3.5 Krusell-Smith-type approximation to second order solution

Solving the generalized Sylvester equation is computationally expensive. This can render the algorithm prohibitively slow for some use-cases, for example the estimation of the model. In this section, we investigate more closely the impact of the distribution on the second order changes in the system. Specifically, we discuss the case in which $g_{d\mu d\mu} = 0$, i.e. where the second derivative of the histogram has no impact on the control variables in the system. In that case, the iteration described in section 3.4.2 would be much faster, since most of the columns in $g_{\hat{\mathscr{S}}\hat{\mathscr{S}}}$ do not have to be updated.

First, we impose that the system contains one aggregate state variable for each dimension of individual states s_i, that tracks the mean of the marginal histogram in the respective dimension. For example, if the individual states encompass idiosyncratic productivity h_i, and holding of an asset k_i, we require the existence of state variables H and K s.t. $K_t = \sum_l \overline{k}_l d\mu_{lt}^k$ and $H_t = \sum_l \overline{h}_l d\mu_{lt}^h$ (as above, $\overline{k}$, $\overline{h}$ denote the

22. In our application, we solve equation (3.17) with precision 10^{-8} with respect to the maximum norm.

nodes of the fixed grids).[23] In the following discussion, we focus on the "asset-type", k_i, and the corresponding aggregate state K, that enters the equilibrium conditions of the system e.g. as the aggregate capital stock of the economy.

A suitable approximation of the histogram that only depends on the aggregate state (as in Krusell and Smith (1998)) then allows us to decouple the *direct* second-order effects of the histogram from its *indirect* second-order effects. The former work through the Fokker-Planck equation, where the cross-derivatives between the histogram and variables that affect the optimal policy (like the future marginal utility) are non-zero. The latter describe the effect of the histogram on controls, e.g. prices, that affect the system in a nonlinear way by changing the optimal policy. Those indirect effects constitute a larger part of the Hessian of the state-to-control mapping, $g_{\hat{\mathscr{S}}\hat{\mathscr{S}}}$, so that a Krusell-Smith-type approximation, where these effects vanish, can be computed at a significant speed gain.

3.5.1 The role of $d\mu^k$ in the first order system dynamics

The equilibrium condition for K can be written as

$$f^K(\hat{C}', \hat{C}, \hat{\mathscr{S}}', \hat{\mathscr{S}}) := K' - \sum_{l=1}^{n^k} (d\mu^k \Pi_h^k)_l \bar{k}_l = 0, \tag{3.20}$$

which is dependent on both today's histogram, $d\mu^k$, and today's optimal policy choices h^d that induce the transition matrix, Π_h^k, in the Fokker-Planck equation. Assuming the existence of the implicit functions h and g as above implies the following first order conditions for this equation:

$$f^K_{\hat{\mathscr{S}}} + f^K_{\hat{\mathscr{S}}'} h_{\hat{\mathscr{S}}} + f^K_{\hat{C}} g_{\hat{\mathscr{S}}} + f^K_{\hat{C}'} g_{\hat{\mathscr{S}}} h_{\hat{\mathscr{S}}} = 0. \tag{3.21}$$

With that, we can show the following lemma:

Lemma 3.5.1. *The marginal histogram $d\mu^k$ cannot at the same time be irrelevant for the controls of the system, and be irrelevant for the next period's aggregate states (with respect to first-order dynamics), i.e.*

$$h_{d\mu^k} \text{ has } 0\text{-rows in all aggregate states, esp. } K' \Rightarrow g_{d\mu^k} \neq 0$$

23. To be more precise, the average over the distribution cannot be a state variable if the *whole* marginal histogram is in the state space as well, since then the equation that defines the average over the distribution would not be necessary for the system (the system would be indeterminate). In practice, we therefore reduce the dimension of the marginal histogram by 1 when we introduce the average as another state variable. Then, the complete marginal histogram is backed out from the average within the system. We abstract from this technicality as it is inconsequential for the following results.

Proof: We first note that the derivate of the equilibrium condition of K by the marginal histogram $d\mu^k$ is not zero: $f^K_{d\mu^k} = -\left(\Pi^k_h \cdot \bar{k}\right)^T$ is a non-zero row-vector that stores the expected amounts of asset k that agents will hold next period conditional on their holdings of k this period. Therefore, at least one of the other summands in equation (3.21) must not be zero in all (row-) indices that belong to $d\mu^k$, namely one of $f^K_{\hat{\mathscr{S}}'} h_{d\mu^k}$, $f^K_{\hat{C}} g_{d\mu^k}$, or $f^K_{\hat{C}'} g_{\hat{\mathscr{S}}} h_{d\mu^k}$ must not be zero. Under the condition that all prices relevant to future decisions of agents are included in the set of aggregate variables of the system, the future marginal histogram $d\mu^{k'}$ does not directly affect the agent's optimal policy h^d. Therefore, $f^K_{\hat{\mathscr{S}}'}$ has non-zero entries only at the aggregate states X' and D', which implies that $f^K_{\hat{\mathscr{S}}'} h_{d\mu^k} = 0$ if $h_{d\mu^k}$ has 0-rows in all aggregate states. The lemma follows by noting that, if additionally $g_{d\mu^k} = 0$, both $f^K_{\hat{C}} g_{d\mu^k}$ and $f^K_{\hat{C}'} g_{\hat{\mathscr{S}}} h_{d\mu^k}$ are 0 as well, which violates equation (3.21). □

Lemma 3.5.1 illustrates the challenge that rational agents in the economy face: they need to keep track of the information stored in the whole marginal histogram, since it impacts the *aggregate* system, either by affecting aggregate states *directly* ($h_{d\mu^k} \neq 0$), or by affecting the aggregate system dynamics *indirectly* through controls ($g_{d\mu^k} \neq 0$), or (possibly) both.

Proposition 3.5.1. *The marginal histogram is never irrelevant for the system's controls up to first-order, i.e.* $g_{d\mu^k} \neq 0$.

Proof: With lemma 3.5.1, we only have to show the claim for the case where the entries in the first order approximation of the state transition function h that map the marginal histogram to the average over the distribution are not zero, i.e. $h_{K',d\mu^k} \neq 0$. Suppose that $g_{d\mu^k} = 0$. If it was also the case that the controls were independent of the average over the distribution, $g_K = 0$, the agents' actions would not depend on the state of the economy, which cannot be optimal. Hence, $g_K \neq 0$. The proof proceeds by analysing the first-order conditions that pertain to the Bellman equation. We restate its equilibrium condition here:

$$f^\nu(\hat{C}', \hat{C}, \hat{\mathscr{S}}', \hat{\mathscr{S}}) \;=\; \nu_t - (u_{h^d_t} + \beta \Pi_{h_t} \nu_{t+1} \odot \nu_0(\bar{s}, S_{t+1}, P_{t+1})). \tag{3.22}$$

Clearly, the condition is independent of present or future state variables, i.e. $f^\nu_{\hat{\mathscr{S}}} = f^\nu_{\hat{\mathscr{S}}'} = 0$, while it depends on present and future controls: $f^\nu_{\hat{C}} \neq 0, f^\nu_{\hat{C}'} \neq 0$. The implicit function theorem for the functions h and g yields the first order condition

$$f^\nu_{\hat{\mathscr{S}}} + f^\nu_{\hat{\mathscr{S}}'} h_{\hat{\mathscr{S}}} + f^\nu_{\hat{C}} g_{\hat{\mathscr{S}}} + f^\nu_{\hat{C}'} g_{\hat{\mathscr{S}}} h_{\hat{\mathscr{S}}} \;=\; 0.$$

In the column-indices that belong to $d\mu^k$, this condition reduces to $f^\nu_{\hat{C}'} g_{\hat{\mathscr{S}}} h_{d\mu^k} = 0$. Since $h_{K',d\mu^k} \neq 0$, it must hold that $f^\nu_{\hat{C}'} g_K = 0$, i.e. g_K is in the kernel of $f^\nu_{\hat{C}'}$. But since $f^\nu_{\hat{C}} = I$ (identity matrix), it must also hold that $g_K = f^\nu_{\hat{C}'} g_{\hat{\mathscr{S}}} h_K = f^\nu_{\hat{C}'} g_K h_{K',K} = 0$, which is a contradiction. □

Proposition 3.5.1 is an important result. It shows that, even in a model economy where prices and value functions are affected by distributions only through their average levels, the policy function must map directly from the distribution to the controls. The reason is that, when a surprise shock changes today's marginal histogram, it affects the forward-looking part of the system through two channels: for one, it changes the expected future average over the histogram, which changes future controls, like prices and the value function. At the same time, it affects today's utility levels, as a change in today's average level (caused by the change in the histogram) affects production, factor incomes, and thus consumption. To account for both channels in a way that is consistent with the intertemporal optimization by the agents, the system needs to have enough degrees of freedom, and hence $g_{d\mu^k} \neq 0$ must hold. One way to interpret this result is to note that the first-order solution, being a linear solution, necessarily is a *reduced form*-solution: although the underlying structure of the economic model implies that agents only care about the average level instead of the whole distribution when making choices, this "chain logic" cannot be represented in a linear system. Therefore, the linear system lumps together two distinct effects of the distribution: its effect on the average level, and the average level's effect on the agents' optimal choices.

Next, we use these results about the first-order impact of the distribution to analyze the *nonlinear* effect of the distribution on the system.

3.5.2 Second-order effects of $d\mu^k$

Analogously to the analysis of the first-order effects of the distribution $d\mu^k$, we use the equilibrium conditions on the second order derivatives of the functions h and g to analyze the nonlinear effects of the distribution on the system. We know from section (3.4.2) that $g_{\hat{\mathcal{S}}\hat{\mathcal{S}}}$ solves the fixed point problem

$$g_{\hat{\mathcal{S}}\hat{\mathcal{S}}} = -((\mathbb{A}^{-1}\mathbb{B})_{(i^g,j^g)} g_{\hat{\mathcal{S}}\hat{\mathcal{S}}} \mathcal{H}_{\hat{\mathcal{S}}}^{\otimes 2} + (\mathbb{A}^{-1}\mathcal{A}_{21})_{(i^g,\cdot)}). \tag{3.23}$$

It is easier to analyze the implications of this equation by rewriting the multiplication of $g_{\hat{\mathcal{S}}\hat{\mathcal{S}}}$ with the Kronecker power $\mathcal{H}_{\hat{\mathcal{S}}}^{\otimes 2}$ as the collocation of the following set of matrices:

$$\left\{ \sum_{j=1}^{ns} g_{\hat{\mathcal{S}}\hat{\mathcal{S}}}(j) \cdot \mathcal{H}_{\hat{\mathcal{S}}} \cdot \mathcal{H}_{\hat{\mathcal{S}}}[j,k] \;\middle|\; k = 1,..,ns \right\}, \tag{3.24}$$

where $g_{\hat{\mathcal{S}}\hat{\mathcal{S}}}(j)$ denotes the j-th block of columns of $g_{\hat{\mathcal{S}}\hat{\mathcal{S}}}$, where each block has ns columns.

Proposition 3.5.2. *The marginal histogram has nonlinear effects on the controls of the system.*

Proof: Suppose to the contrary that $g_{\hat{\mathcal{S}}\hat{\mathcal{S}}}(j) = 0$ for all $j \in J^{d\mu^k}$, i.e. all indices that belong to $d\mu^k$. Without loss of generality, we assume that the average level K is the only other state apart from the marginal histogram $d\mu^k$. In order to fulfill equation (3.23) in the column-indices $J^{d\mu^k}$, we see using the set-notation (3.24) that $g_{K,K}$ should fulfill the equations

$$g_{K,K}\mathscr{H}_{K',d\mu^k}\mathscr{H}_{K,l} \;=\; (\mathbb{A}^{-1}\mathscr{A}_{21})_{(i^g,(d\mu^k,l))} \tag{3.25}$$

for all $l = 1,..,ns$. The matrix $\mathbb{A}^{-1}\mathscr{A}_{21}$ is determined by the Hessian of the system F and the first-order approximations of the functions g and h. $\mathscr{A}_{21}$ (essentially) consists of $F_{vv}\,\mathrm{E}\,v_{\hat{\mathscr{S}}}^{\otimes 2}$, which can be split into matrices I and J as described in (3.15). Matrix I (and J analogously), which contains a Kronecker power, can again be rewritten as a collocation of the matrix-set

$$\left\{ \sum_{j=1}^{n^v} F_{vv}(j)\cdot V_{\hat{\mathscr{S}}}^0 \cdot V_{\hat{\mathscr{S}}}^0[j,i] \;\middle|\; i \in \{1,..,ns\} \right\}. \tag{3.26}$$

As shown in proposition 3.5.1, it always holds that $g_{d\mu^k} \neq 0$, while $\mathscr{H}_{d\mu^k}$ may or may not have non-zero rows for some aggregate states. Therefore, $V_{\hat{\mathscr{S}}}^0[j, J^{d\mu^k}] \neq 0$ for some indices j, shown in equation (3.8), that belong to today's or future control variables. We have argued in section (3.4.1) that columns in the Hessian with respect to future value functions, and today's prices, will be non-zero, i.e. $F_{vv}(j) \neq 0$ for some $j \in J^{v',P}$. Taken together, we see that the columns in $\mathscr{A}_{21}$ belonging to interaction terms with $d\mu^k$ will not be zero. If $\mathscr{H}_{K',d\mu^k} = 0$, this leads to a direct contradiction. If $\mathscr{H}_{K',d\mu^k} \neq 0$, the element $g_{K,K}$ would be overdetermined by fulfilling equation (3.25) for all indices $l = 1,..,ns$, which is also contradictory. □

3.5.3 Krusell-Smith type approximation

From the last two subsections, we saw that the first-order dynamics in a system where the distribution has effects on the aggregate variables in the system imply that the distribution has (indirect) second-order effects. From now on, we make the following assumption:

Assumption 3.5.1. *Apart from the state-transition of the average level K, the marginal histogram does not enter any other equilibrium condition that contains aggregate state variables.*

Note that the Fokker-Planck equation is a condition on the marginal histogram and aggregate *control* variables. Hence, in most economic models, the distribution affects the aggregate system only through its average level, so that this assumption is without much loss of generality.[24] Note also that assumption 3.5.1 implies the assumptions 3.4.1 and 3.4.2. Now, we consider an auxiliary system, $\hat{F}$, where we only

24. One can also readily extend the approach by adding state variables for each additional feature of the distribution that matters for the aggregate dynamics in a specific model, for example a measure of inequality if the agents' preferences are influenced by it.

change equilibrium condition (3.20): we impose that today's distribution $d\mu^k$ does not enter the condition, and therefore has no effect on K' (or any other aggregate variable in the system). Instead, we substitute it for an estimator of $d\mu^k$, $d\tilde{\mu}^k(K)$, that only depends on the aggregate state[25]:

$$\hat{f}^K(\hat{C}', \hat{C}, \mathscr{S}', \mathscr{S}) := K' - \sum_{l=1}^{n^k} (d\tilde{\mu}^k(K)\Pi_h^k)_l \bar{k}_l = 0. \tag{3.27}$$

In the auxiliary system, the marginal histogram does not have direct (through $\hat{h}_{\mathscr{S}}$) nor indirect (through $\hat{g}_{\mathscr{S}}$) first-order effects on aggregate state-variables. Still, we show that the marginal histogram has a nonlinear effect on the state-transition of the system, which works through the interaction of controls and the distribution in the Fokker-Planck equation, and which is effective since the controls react to the average level K. We call this the *direct* nonlinear effect of the distribution on the aggregate system.

Lemma 3.5.2. *In the first-order approximation of the auxiliary system $\hat{F}$, the marginal histogram has no influence on the aggregate system, i.e. $\hat{h}_{d\mu^k}$ has zeros in rows with aggregate states, and $\hat{g}_{d\mu^k} = 0$.*

Proof: We show that, if $\hat{h}_{d\mu^k}$ has zeros in rows with aggregate states, and $\hat{g}_{d\mu^k} = 0$, the necessary first-order conditions of the system are fulfilled. The claim then follows with the uniqueness of the first-order solution of the system (Klein, 2000). First, all summands on the left-hand side of the first-order condition for the average level K,

$$\hat{f}^K_{\mathscr{S}} + \hat{f}^K_{\mathscr{S}'}\hat{h}_{\mathscr{S}} + \hat{f}^K_{\hat{C}}\hat{g}_{\mathscr{S}} + \hat{f}^K_{\hat{C}'}\hat{g}_{\mathscr{S}}\hat{h}_{\mathscr{S}} = 0 \tag{3.28}$$

are zero in the columns $d\mu^k$, so that the conditions are trivially fulfilled. For the column K, the condition determines the state-transition from K to K': $\hat{h}_{K',K} = -(\hat{f}^K_K)/(1 + \hat{f}^K_{\hat{C}'}\hat{g}_K)$. Next, the equilibrium condition for the Fokker-Planck equation is linear in the future marginal histogram, $\hat{f}^{d\mu}_{d\mu^{k'}} = I$. Hence, in the columns $d\mu^k$, the first-order condition is a condition on the state-transition function: $\hat{h}_{d\mu^{k'},d\mu^k} = -\hat{f}^{d\mu}_{d\mu^k}$. In the column K, it determines the state-transition from K to $d\mu^{k'}$: $\hat{h}_{d\mu^{k'},K} = -(\hat{f}^{d\mu}_{\hat{C}'}\hat{g}_K\hat{h}_{K',K})$. The Bellman equation is independent of state variables, while the remaining derivatives by control variables in the first-order condition multiplies with zero-entries in $\hat{g}_{d\mu^k}$ and $\hat{h}_{d\mu^k}$ in the columns $d\mu^k$, so that the condition there is trivially fulfilled. For the column K, the Bellman equation restricts the K-to-controls mapping, as column vector $\hat{g}_K$ has to fulfill the equation $(\hat{f}^v_{\hat{C}} + \hat{f}^v_{\hat{C}'}\hat{h}_{K',K})\hat{g}_K = 0$. Note

25. As an example, one can modify the distribution in steady state by shifting probability weights on some grid points (while maintaining its overall shape) to adapt to distribution-means K that differ from steady state-K.

that, in general, the matrix $\hat{f}^{v}_{\hat{C}} + \hat{f}^{v}_{\hat{C}'}\hat{h}_{K',K}$ has more non-zero columns than non-zero rows, as the Bellman equation depends on the value function *and* other controls (like prices) that determine the return on saving and the agents' optimal policy. Therefore, the columns of the matrix will be linearly dependent. Finally, the market clearing condition for asset k depends on the average level K and today's controls. It further restricts the column $\hat{g}_K$: $\hat{f}^{\Phi}_{\hat{C}}\hat{g}_K = -\hat{f}^{\Phi}_{K}$. □

Note that the market clearing condition rules out $\hat{g}_K = 0$, i.e. in general, states influence the controls also in the auxiliary system. This is necessary to obtain the direct nonlinear effects of the marginal histogram, as we show in the following proposition.

Proposition 3.5.3. *Assume $\hat{g}_K \neq 0$. Then, for the auxiliary system, it holds that $\hat{g}_{\hat{\mathscr{S}}\hat{\mathscr{S}}}(j) = 0$ for all $j \in J^{d\mu^k}$, while $\hat{h}_{\hat{\mathscr{S}}\hat{\mathscr{S}}}(j) \neq 0$ for some $j \in J^{d\mu^k}$.*

Proof: To show the claim, we revisit equation (3.9), which is the matrix equation that determines both $\hat{g}_{\hat{\mathscr{S}}\hat{\mathscr{S}}}$ and $\hat{h}_{\hat{\mathscr{S}}\hat{\mathscr{S}}}$. We guess $\hat{h}_{\hat{\mathscr{S}}\hat{\mathscr{S}}} = -A_{2\,(i^{d\mu,K},:)}$ as a solution to the equation, where $i^{d\mu,K}$ denotes the rows in the system $\hat{F}$ that belong to the Fokker-Planck equation and the equilibrium condition $\hat{f}^K$. Then, we verify that it is consistent with $\hat{g}_{\hat{\mathscr{S}}\hat{\mathscr{S}}}(j) = 0$ for all $j \in J^{d\mu^k}$. By uniqueness of the solution, the claim follows. $\hat{F}_3$ is the derivative of the system by future states. As discussed in section 3.4.1, $\hat{F}_3$ has the identity matrix in the rows that belong to the Fokker-Planck equation (as well as to the transition of the average, $\hat{f}^K$), and zeros in the other rows. First, we show that the matrices $\hat{F}_1\hat{g}_{\hat{\mathscr{S}}\hat{\mathscr{S}}}B_2$, $F_2\hat{g}_{\hat{\mathscr{S}}\hat{\mathscr{S}}}$ and $F_1\hat{g}_{\hat{\mathscr{S}}}\hat{h}_{\hat{\mathscr{S}}\hat{\mathscr{S}}}$ have zero columns at the indices $j \in J^{d\mu^k}$. Noting that B_2 essentially consists of the Kronecker product $\hat{h}^{\otimes 2}_{\hat{\mathscr{S}}}$, we can use the set-notation (3.24) to see that $\hat{g}_{\hat{\mathscr{S}}\hat{\mathscr{S}}}\hat{h}^{\otimes 2}_{\hat{\mathscr{S}}}(j) = 0$ for all $j \in J^{d\mu^k}$. The claim follows as in the proof to proposition 3.5.2, due to the zeros in $\hat{g}_{d\mu^k}$ and $\hat{h}_{d\mu^k}$ by lemma 3.5.2. For $F_2\hat{g}_{\hat{\mathscr{S}}\hat{\mathscr{S}}}$, the claim follows immediately. With our guess for $\hat{h}_{\hat{\mathscr{S}}\hat{\mathscr{S}}}$, it remains to show that $\hat{g}_{\hat{\mathscr{S}}}A_{2\,(i^{d\mu,K},:)}(j) = 0$ for all $j \in J^{d\mu^k}$. Set-notation (3.26) for A_2 helps to see that this is the case: while $\hat{g}_K \neq 0$, the state transition (3.28) is only nonlinear in the interaction of the marginal histogram with controls, or possibly with K, depending on the estimator $d\tilde{\mu}^k(K)$. However, the rows in $V^0_{\hat{\mathscr{S}}}$ that belong to controls and K are zero in the columns $j \in J^{d\mu^k}$, by lemma 3.5.2. Finally, for the interaction of the state K with itself, the column $\hat{g}_{KK}$ is determined by being the solution to $A_{2\,(:,KK)} + \hat{F}_2\hat{g}_{KK} - (\hat{F}_3 + \hat{F}_1\hat{g}_{\hat{\mathscr{S}}})A_{2\,(i^{d\mu,K},KK)} = 0$. As seen in section 3.4.1, $\hat{F}_2$ has the identity matrix in the columns that belong to the value function and the rows that belong to the Bellman equation, non-zero entries at columns that belong to prices, and zeros everywhere else.

Finally, we show that $\hat{h}_{\hat{\mathscr{S}}\hat{\mathscr{S}}}(j)$ is non-zero for some $j \in J^{d\mu^k}$, i.e. that the marginal histogram has nonlinear effects in the auxiliary system. We use again set-notation (3.26) to analyze A_2. Setting the column-index $i = i^K$, $\hat{g}_K \neq 0$ implies that $V^0_{\hat{\mathscr{S}}}[j, i^K] \neq 0$ for some control-indices j. At the same time, $V^0_{d\mu^k}$ is the identity matrix in the rows that map to today's marginal histogram. Noting that the Fokker-Planck equation has a nonlinear interaction between the marginal histogram and controls yields the result.[26] □

Our analysis reveals the precise nonlinearity caused by the distribution in the auxiliary system. Since the proofs were constructive, they can also be used to compute the second-order solution to the auxiliary system quickly (thereby reducing the overhead of the method proposed below). In the proofs, we considered a simplified setting where the only state variables were the marginal histogram and the average level K. This has to be accounted for when attempting to compute parts of the auxiliary system with a richer set of variables in closed form.

What are the *indirect* nonlinear effects of the marginal histogram that we miss when computing the auxiliary system? Suppose that $F_{vv}(j) = 0$ for all $j \in J^{d\mu^k}$, i.e. the direct second-order effects of the distribution are muted. As before, the columns belonging to the marginal histogram in matrix I (which is part of A_2) are in the set

$$\left\{ \sum_{j=1}^{n^v} F_{vv}(j) \cdot V^0_{d\mu^k} \cdot V^0_{\hat{\mathscr{S}}}[j,i] \;\middle|\; i \in \{1,..,n^\mu\} \right\}.$$

Let us consider only the part of the sum pertaining to $j \in j^P$, i.e. the indices in v belonging to the prices P. We can expect that $F_{vv}(j) \neq 0$ for $j \in j^P$, namely, that prices have a nonlinear impact on the system. In the correct system, we have also shown that prices react to changes in the distribution, i.e. $g_{d\mu} \neq 0$ (in the respective rows). Hence, the columns in I describing the interaction effects between the distribution and other states will not be zero. We see that in the correct system, A_2 captures not only the direct second-order effects of variables through the Hessian F_{vv}, but also their impact on *other* variables who might have a non-constant effect on the system. That is, part of the nonlinear effect of the marginal histogram goes through its *linear* effect on the average level K, which impacts controls like prices or value functions, which in turn affect the system nonlinearly.

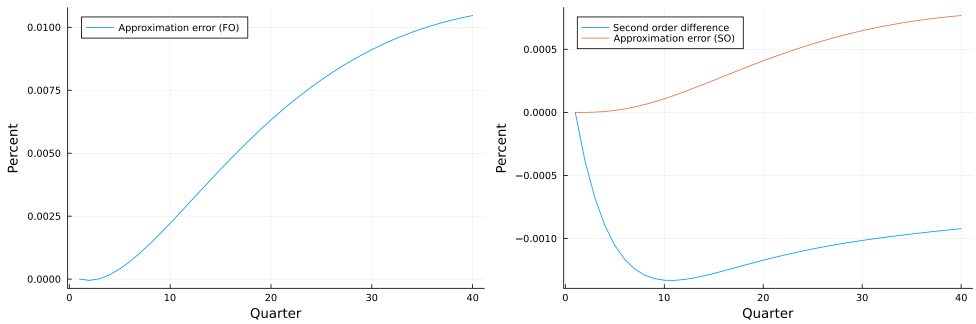
Since the grid-size of the marginal histogram typically outnumbers aggregate states, a majority of columns that have to be found through the iterative algorithm that we propose in section 3.4.2 vanish in the auxiliary system, where $\hat{g}_{d\mu^k d\mu^k} = 0$. Therefore, the Krusell-Smith type approximation should have a significant speed-advantage over the correct solution. Next, we test the accuracy of the approximation as well as the speed-gain in an application of the method.

3.5.4 Application

We apply the approximation technique to a neoclassical heterogeneous agent economy with one asset. The agents are heterogeneous in wealth and idiosyncratic in-

26. The crucial difference to the above finding that $\hat{g}_{\hat{\mathscr{S}}} A_{2\,(i^{d\mu,K},:)}(j) = 0$ for $j \in J^{d\mu^k}$ is that in the state-transition for the average level K, the marginal histogram does not enter (in the auxiliary system), while $V^0_{d\mu^k}$ is zero in rows that pertain to K.

(a) First order difference **(b)** Second order differences



Notes: Relative differences (correct solution as basis) of impulse responses of capital stock K to a TFP-shock (size: 1 std. dev.).
First-/ second order approximations computed using the estimator of the histogram, $d\tilde{\mu}^k(K)$.

Figure 3.1. Differences of impulse response functions

come. We use the dimension-reduction techniques of Bayer and Luetticke (2020) (see section 3.4.1 therein for a short description of the model setup). For the estimator of the histogram over capital holdings from the aggregate capital stock, $d\tilde{\mu}^k(K)$, we impose the steady state-histogram each period t, corrected by a shift of probability weights between the mode and the grid point closest to the true average, K_t. The system has 91 state variables, of which 78 belong to the marginal histogram, and 89 control variables. Therefore, the Hessian of the state-to-control mapping $g_{\hat{\mathcal{S}}\hat{\mathcal{S}}}$ has 89 rows and 8281 columns.

First, we solve both the correct system and the auxiliary system up to first order. Figure 3.1a plots the percent difference in the impulse responses of the aggregate capital stock to a TFP shock between the first order dynamics with and without using the approximated histogram $d\tilde{\mu}^k(K)$. Over time, the approximation of the true (first-order) histogram by the shape of the steady state histogram becomes less accurate. In the long run, the true (first-order) capital stock is overestimated by 10^{-2} percent.

This shows that abstracting from direct and indirect effects of the marginal histogram on the aggregate states in the first-order dynamics comes at a non-negligible approximation error. In particular, the approximation worsens over time, as the distribution deviates from the steady state-distribution, and the estimation errors of K accumulate. Instead, we propose to use the first-order solution of the correct system and add the Hessians of the state transition and state-to-control mappings of the auxiliary system, $\hat{h}$ and $\hat{g}$, to approximate second-order effects. Note that the direct nonlinear effects of the distribution only affect the distribution itself, through the Fokker-Planck equation. Only when feeding these nonlinear effects through the *correct* first order dynamics, where the distribution has an effect on aggregate variables, the method is able to capture a part of the overall nonlinear dynamics. It is important to note that this procedure implies that the method has a significant *over-*

head: We have to solve for the first-order dynamics of *both* the correct system and the auxiliary system. The computations for the second-order solution, namely computing the Hessian (section 3.4.1) and solving the generalized Sylvester equation (section 3.4.2), are only carried out once for the auxiliary system. Thus, the method only yields a reduction in calculation times if the time to additionally compute a first-order solution[27] is outweighed by the quicker solving of the generalized Sylvester equation.

We know from proposition 3.5.3 that in the auxiliary system, all columns in $g_{\hat{\mathscr{S}}\hat{\mathscr{S}}}$ that belong to interactions of state variables with the marginal histogram are zero. In our application, this amounts to 85% of the columns. In order to get a clear comparison of computation times, we run the simple algorithm proposed in section 3.4.2 to solve the generalized Sylvester equation, once for the correct and once for the auxiliary system.[28] In each case, the starting guess of the iteration is a matrix that contains only zeros. Knowing zero-columns of $g_{\hat{\mathscr{S}}\hat{\mathscr{S}}}$ in advance can be implemented into update-step (3.19) in the following way: with $J^{\neq 0}$ denoting the column-indices where the columns are *not known* to be zero, we can compute the iteration step as

$$\hat{g}^{red,(i+1)}_{\hat{\mathscr{S}}\hat{\mathscr{S}}} = -\Big((\mathbb{A}^{-1}\mathbb{B})_{(i^g,j^g)} \hat{g}^{red,(i)}_{\hat{\mathscr{S}}\hat{\mathscr{S}}} \mathscr{H}^{\otimes 2}_{\hat{\mathscr{S}}\ (J^{\neq 0},J^{\neq 0})} + (\mathbb{A}^{-1}\mathscr{A}_{21})_{(i^g,J^{\neq 0})} \Big), \tag{3.29}$$

where $\hat{g}^{red}_{\hat{\mathscr{S}}\hat{\mathscr{S}}}$ denotes the Hessian of the state-to-control mapping *reduced* to the columns that are not known to be zero.

3.5.4.1 Results

We find that it takes about 114 seconds to solve the generalized Sylvester equation by running the simple iterative algorithm for the correct system, while it takes 4 seconds to solve it in the auxiliary system, which is a reduction of computation time by a factor of 28. We expect the difference in computation times to grow in the size of the variable space. Hence, even with the overhead of computing an additional first-order solution, our approximation-method is likely to yield a speed-gain over computing the correct second-order solution for most use-cases.

In terms of the accuracy of the approximation, the graph “approximation error (SO)” in figure 3.1b shows that the approximation error in the impulse response of

27. Note that, since the auxiliary system differs from the original system only by one equilibrium condition (with one row), the Jacobian of both systems is almost identical. Therefore, the time to compute the first order solution to the auxiliary system will be that of doing a Schur-decomposition. Alternatively, one may attempt to use the constructive proofs in section 3.5.3 to compute (part of) the solution to the auxiliary system in closed form.

28. In particular, we pre-calculate the Kronecker product $\mathscr{H}^{\otimes 2}_{\hat{\mathscr{S}}}$ ahead of each iteration, instead of computing the product $g_{\hat{\mathscr{S}}\hat{\mathscr{S}}}\mathscr{H}^{\otimes 2}_{\hat{\mathscr{S}}}$ in each iteration step, which we have found to be faster when using parallelization. Another technical trick that we do not use in the comparison is to reduce the column space by filtering out non-unique columns, thereby making use of the symmetry of the Hessian.

aggregate capital increases over time and is at about 10^{-4} percent over fourty periods. To put this magnitude into perspective, we also show the difference between the true second-order, and the true first-order impulse response ("second order difference" in figure 3.1b). We find that already after a few quarters, solving the system up to first order underestimates the true (second-order) capital stock by about 10^{-3} percent.

We conclude that abstracting from the indirect second-order effects of the histogram is quite unimportant in the short run, as most of the second-order difference can still be captured. However, the approximation worsens with a longer horizon. Not surprisingly, then, the approximation error when calculating the unconditional expectation of the capital stock (taking into account the stochastic influence of the exogenous variables, here TFP shocks) is larger: In the auxiliary system where the histogram is approximated by $d\tilde{\mu}^k(K)$, the capital stock is expected to be 0.02 percent lower than in the correct system.

3.6 Conclusion

In this paper, we give a detailed account of the computational burden that arises when using the perturbation method to solve a heterogeneous agent model up to the second order. The high computational cost stems from the nonlinear impact of the distribution, which makes up the largest part of the state space. We employ the knowledge of the structure inherent to any heterogeneous agent model with forward-looking agents to establish that the computational cost is high as long as the distribution affects aggregate variables in the system directly. Building on this result, we propose an approximation to the second order solution by setting up an auxiliary system where agents estimate the full (marginal) distribution only from its first moment, similar to the approach in Krusell and Smith (1998), when calculating its effect on aggregate variables of the system.

We find that the approximation can be computed at a notable speed gain, while its accuracy is good in the short run, but deteriorates over time. This feature of the approximation can make it attractive for the purpose of model-estimation, where the long-run effect of a particular shock overlaps with, and is typically dominated by, short run effects of other shocks. As a possible extension of the method for this specific use-case, one could attempt to update the estimator $d\hat{\mu}^k$ after a few periods during the estimation in order to partly account for changes in the distribution, which may reduce the accumulation of approximation errors over time. We leave the further validation of the method and its possible extensions in this and other applications for future research.

References

Auclert, Adrien, Bence Bardóczy, Matthew Rognlie, and Ludwig Straub. 2021. "Using the Sequence-Space Jacobian to Solve and Estimate Heterogeneous-Agent Models." *Econometrica* 89 (5): 2375–408. [179]

Azinovic, Marlon, Luca Gaegauf, and Simon Scheidegger. 2022. "Deep Equilibrium Nets." *International Economic Review* 63 (4): 1471–525. [179]

Bayer, Christian, Benjamin Born, and Ralph Luetticke. 2022. "Shocks, Frictions, and Inequality in US Business Cycles." CEPR Discussion Papers 14364. [180]

Bayer, Christian, and Ralph Luetticke. 2020. "Solving discrete time heterogeneous agent models with aggregate risk and many idiosyncratic states by perturbation." *Quantitative Economics* 11 (4): 1253–88. [179, 180, 186, 197]

Fernández-Villaverde, Jesús, Samuel Hurtado, and Galo Nuño. 2019. "Financial Frictions and the Wealth Distribution." Working Paper, Working Paper Series 26302. National Bureau of Economic Research. [179]

Klein, Paul. 2000. "Using the generalized Schur form to solve a multivariate linear rational expectations model." *Journal of Economic Dynamics and Control* 24 (10): 1405–23. [185, 194]

Krusell, Per, and Anthony A. Smith Jr. 1998. "Income and Wealth Heterogeneity in the Macroeconomy." *Journal of Political Economy* 106 (5): 867–96. [179, 180, 190, 199]

Levintal, Oren. 2017. "Fifth-order perturbation solution to DSGE models." *Journal of Economic Dynamics and Control* 80: 1–16. [180, 181, 184, 185]

Reiter, Michael. 2009. "Solving heterogeneous-agent models by projection and perturbation." *Journal of Economic Dynamics and Control* 33 (3): 649–65. [179]