UNIVERSITY OF COPENHAGEN
DEPARTMENT OF ECONOMICS

PhD Thesis

Christoffer Jessen Weissert

Essays in Macroeconomics

Inflation Inequality, Consumption Baskets and Impulse Responses

Supervisors: Søren Hove Ravn and Jeppe Druedahl

Submitted on: February 24 2022

Table of Contents

Introduktion (på dansk)

I 1980 havde ca. 35 procent af makroøkonomiske forskningsstudier et empirisk fokus. I 2015 var det tal steget til ca. 65 procent. Makroøkonomisk forskning er med andre ord blevet mere empirisk orienteret over de seneste 35 år. Det viser nobelprismodtageren Joshua Angrist og hans medforfattere bl.a. i deres artikel *Economic Research Evolves: Fields and Styles* fra 2017.

En af grundene til at denne udvikling har fundet sted er, at man har fået adgang til mere og bedre data over tid. Et eksempel herpå er det stregkodedatasæt, som anvendes i dele af denne afhandling. Med datasættet har man mulighed for at følge husholdningers daglige supermarkedsforbrug fra 2004 og frem til i dag. Med den data kan man bl.a. undersøge, hvordan indholdet i rige og fattiges indkøbskurve adskiller sig. Datasættet er et eksempel på den type af information, som er blevet mere og mere almindeligt at anvende i makroøkonomiske forskningsstudier og spiller en vigtig rolle i forbindelse med besvarelsen af makroøkonomiske spørgsmål. Denne afhandling dokumenterer for eksempel at forskelle i vareforbrug giver anledning til en substantiel forskel i inflation mellem rige og fattige. I lyset af den aktuelle debat om den stigende inflation verdens vestlige økonomier har oplevet de seneste par måneder, giver indsigten om inflationsulighed anledning til en diskussion om *hvis* inflation, vi snakker om og om den (makro)økonomiske politik, man eventuel igangsætter for at stabilisere udviklingen er rettidig og gavnlig for alle.

En anden årsag til, at det empiriske fokus er blevet større er, at de metoder, som vi anvender til at analysere data, også er blevet flere og bedre. Værktøjskassen, som forskere og praktikere kan gribe til for at analysere data, er med andre ord blevet udvidet. Et eksempel herpå er de metoder, man anvender til at karakterisere fordelingen af inflation blandt alle individer i et samfund. Et andet eksempel er den seneste udvikling i de metoder, man anvender til empirisk at undersøge hvordan økonomien reagerer på forskellige typer af stød. Selve spørgsmålet er ældgammelt og har mere eller mindre fulgt makroøkonomer siden fagets spæde leveår. Den dag i dag forfines metoderne stadig og forskere udgiver artikler i toptidsskrifter, hvor de deler nye indsigter om metodernes fordele, ulemper og sammenfald. Forfinelsen af de metoder, vi anvender til at analysere data, har både været en vigtig årsag til, at vi kan skrive overbevisende og troværdige empiriske forskningsartikler men er samtidig også vigtige for det arbejde, makroøkonomer udfører udenfor forskingsverdenen. Lad mig eksemplificere det med en personlig anekdote. I året inden jeg blev indskrevet som PhD-studerende, var jeg ansat i den modelgruppe, der har udviklet Danmarks Finansministeriums nye makroøkonomiske model (kaldet MAKRO). Et af de projekter, jeg arbejdede på, var empirisk at undersøge hvordan den danske økonomi reagerer på forskellige typer af stød såsom ændringer i den udenlandske efterspørgsel efter danske varer. I projektet anvendte jeg i vid udstrækning de forskellige metoder, der både historisk og mere nyligt er blevet udviklet. I forbindelse med projektet rendte jeg dog også på nogle udfordringer, som ikke kunne forklares med den nuværende viden om metoderne. Problemstillingen, som den eksisterende viden ikke kunne svare på, forklarer behovet for at vi stadig undersøger og udvikler vores metoder. I denne afhandling har jeg fået mulighed for at kaste lys på problemstillingen og forhåbentlig bidrage til, at der i fremtidigt arbejde kan laves endnu bedre empiriske analyser, end vi i forvejen kan i dag.

På mange måder er jeg selv og denne afhandling blevet rundet af den empiriske tradition, som er begyndt at tegne sig indenfor makroøkonomi. Personligt har jeg gennemgået en udvikling fra at være en selvudnævnt elfenbenstårnsteoretiker i mine unge dage som bachelorstuderende til at være en empirisk orienteret makroøkonom i løbet af min kandidatuddannelse og senere PhD-uddannelse. Min interesse for teoretisk makroøkonomi er dog stadig at finde mange steder i denne afhandling, ikke mindst i udviklingen af metoderne til den empiriske værktøjskasse. I hver enkelt del af denne afhandling er det gennemgående tema dog en orientering mod empiriske analyser indenfor makroøkonomiske områder der beskæftiger sig med inflationsulighed, forbrugsadfærd og impuls-responser. Jeg håber, at du kan få lige så meget glæde ud af denne afhandling, som jeg har fået af at frembringe materialet i den. Jeg ønsker dig en insigtsfuld og stimulerende læsning her fra.

Introduction (in english)

In 1980 around 35 percent of macroeconomic research articles had an empirical focus. In 2015 that share had risen to around 65 percent. Macroeconomic research has in other words become more empirically oriented over the last 35 years. Nobel price winner Joshua Angrist and his co-authors show this in their article *Economic Research Evolves: Fields and Styles* from 2017.

One of the reasons for why this evolution has taken place is undoubtedly because more and better data has become available over time. One example of this is the scannerdata used in parts of this dissertation. The dataset allows researchers to follow households' daily purchases of supermarket products from 2004 until today. With the dataset researchers have thus become able to for example investigate, in detail, how consumption behavior differs across households. The dataset is an example of the type of information that has become increasingly popular to use in the search for answers to macroeconomic questions. This dissertation for example documents that differences in consumption baskets between households give rise to inflation inequality. In light of the recent debate on the rising inflation that western economies have experienced over the last couple of months, this insight gives reasons for discussing *whose* inflation, we concern ourselves with and whether the (macro)economic policy we potentially will see effectuated to tackle inflation is timely and beneficial for everyone.

Another reason for the evolution towards more empirical oriented macroeconomics is that the methods we use in empirical analyses have also become more and better. The toolbox researchers and practitioners reach into is in layman terms more advanced and expanded than it has been previously. One example hereof is the methods used for empirically estimating the distribution of inflation rates across households. Another example is the methods we use to empirically investigate how the economy responds to shocks. This question has been central to macroeconomists since the early days of the field but the methods used to answer it are to this day still being explored and advanced. Researchers continue to publish articles in top journals sharing their recent insights on how these methods work and how they can and cannot be used in practice. The advancement of these methods is one of the reasons for why researchers are able to write convincing and credible research articles but its worth also reached all the way into the work of macroeconomists outside academia. Let me exemplify that with a personal anecdote. The year before I enrolled as a PhD-student I worked in the model group who was assigned by the Danish Ministry of Finance to build a new, large-scale model of the Danish economy. One of the projects I worked on was to empirically investigate how the Danish economy responded to different types of shocks such as for example changes in foreign demand for Danish goods. In this project I extensively used the different types of methods available for this type of question. While my work yielded a lot of interesting answers to the questions I was hired to look at, I also realized that some problems could not be answered with the knowledge we had about the different methods. This was in 2017 and therefore serve as an indication of the fact that while the empirical methods developed in macroeconomics to a large extent has paved the way for answering many important and interesting questions, both in research and outside academia, there is always a demand for improvement even on highly advanced methods. During my doctoral studies I have been fortunate enough to have had time to look at and improve on some of the shortcomings that I found these methods had and the insights I have learned are shared in the last part of this dissertation.

In many ways I myself and this dissertation have been shaped by the empirical tradition that has slowly begun to form in macroeconomics. Personally I have transitioned from being a self-proclaimed ivory-tower theorist in my early years as a bachelor student to becoming an empirically oriented macroeconomist during my masters and PhD education. My love for theory still shines through in parts of this dissertation, not least in the development of the methods, that you will read more about if you dive into the chapters where they are presented. The overarching theme in all chapters is however an orientation towards empirical analyses spread across areas within macroeconomics that concern themselves with inflation inequality, consumption behavior and impulse response estimation. I hope that this dissertation will give you the same joy that I have had while producing the material in it. From hereon, I wish you a stimulating and insightful read.

References

Angrist, J., Azoulay, P., Ellison, G., Hill, R. and Lu, S. F. (2017). "Economic research evolves: Fields and styles." *American Economic Review*, vol. 107(5), pp. 293–97 (cited on pages ii, iv).

Acknowledgements

If you read this dissertation and find that the content entertains you in any way, please bear in mind that I would not have been able to put this together, had it not been for so many phenomenal people. I would like to dedicate a few lines of words to some of those who have played the most important roles.

First and foremost, I owe my sincerest gratitude to Søren Hove Ravn. He is the causal factor that led to my doctoral education. Søren is one of the biggest inspirations for young, Danish macroeconomists and he is a force that, by now, has pulled many, including myself, into macroeconomic research. Thank you for being an outstanding supervisor, for offering your unconditional support, countless hours of listening to my rambles and for never questioning the quality of my thoughts. In the same league of gratitude, I owe so much to Jeppe Druedahl. I have been and still am admiring way of thinking Economics which has served as role model in my own shaping as an economist. Thank you for accepting my proposal for being my co-supervisor, so that I had a proper excuse for constantly tracing you down with thoughts that were too big for me to grasp on my own. To Søren and Jeppe, I am immensely thankful and humble to have been working and developing under the supervision of the two of you.

While many people deserve appreciation, four people have had no more direct impact on this dissertation than my co-authors of the different chapters. To Anders F. Kronborg, thank you for luring me into the wonderful world of impulse responses and for teaching me the important role it plays in real-world policy modelling and making. To Rasmus B. Larsen, thank you for always being rock solid and thought through and a cherished friend who plays the role of the prudent Jutlander better than I could ever do. To Phillip Hochmuth, thank you for improving my coding skills beyond imagination and for always being ready to discuss any detail of any matter. To Markus Pettersson, thank you for never shying away from taking on the Cochrane role and for being ever sharp-minded.

I have been fortunate enough to hold a scholarship as a PhD student at Danmarks Nationalbank. I am thankful to them and especially the Head of Research, Federico Ravenna, for granting me the scholarship and for all the colleagues at the Research Unit and ØP who have helped me along the way. In particular, Filip Rozsypal was kind of enough to set me up with Wouter den Haan when I was visiting the London School of Economics in 2019 and Denis Gorea has been an eternal source of fruitful discussions.

My graduate studies arrived at a key turning point in 2020 when I visited the Institute for International Economic Studies at Stockholm University. Kurt Mitman invited me in and everything that was to come was amazing. Per Krussel and Timo Boppart, two of the most impressive minds I have ever engaged with, helped lift my research and economic mindset to new levels and they taught me that the most effective place for new ideas to form is in the post-innebandy and -soccer sauna. To Timo and Per, thank you. At the IIES I also enjoyed tremendous feedback from Kieran Larkin, Paul Klein and Yimei Zou. The IIES IMDBs, who know who they are, are not to forget either. I am overly happy that I got to spend time with your wonderful group.

My doctoral education has taken place at the Department of Economics, University of Copenhagen. I have enjoyed and been proud to be part of such a highly renowned institution. I would like to say thanks to all the great researchers that I have been fortunate to engage in discussions with. In particular, I would like to name Patrick Moran, Emiliano Santoro, and Niels Johannesen. As a special thanks, I would also like to reach out to all the PhD students.

While drawing conclusions about what factors in my past that in the first place made me capable of writing this dissertation may be an ill-advised endeavour, a fairly large and acknowledged literature on the importance of childhood backs me up in accrediting a fair share to my parents. Mom and Dad, thank you for all the things you put into my bag.

Beyond all, I owe no more gratitude than to Thea Nissen. All my acknowledgements would have suffered from all forms of severe endogeneity biases, had I not mentioned her. She is the sole reason for why I have been able to uphold a decent return on all the things that I have been exposed to during the time of writing this dissertation. To Thea, thank you for staying on board for the entire roller coaster ride, this journey has been. Thank you for sharing every excitement, every frustration, every step forward and all the steps backwards and thank you for never loosing your breath.

And let me now finish this section off with a disclaimer in the most classical Danish spirit by rephrasing a famous quote by Sir Isaac Newton: If I didn't always see far enough, I apologise. It is just because I sometimes got altitude sick from standing on the shoulders of giants.

Christoffer Jessen Weissert
February 2022, Copenhagen

Summary (in english)

The dissertation consists of three self-contained chapters on topics within macroeconomics. All chapters share an orientation towards empirical analysis of macroeconomic questions. New methods for empirical analysis are presented in the first and third chapter. The method presented in the first chapter contributes to analyses of inflation inequality. The method presented in the third chapter contributes to analyses on how the economy responds to shocks. The inflation inequality discussed in the first chapter relates to households' consumption behavior. In the first and second chapter specific empirical analyses on US households' consumption behavior is conducted.

Chapter 1 / A Nonhomothetic Price Index and Inflation Heterogeneity
with Phillip Hochmuth & Markus Pettersson
In this chapter we look at a classic and central topic in macroeconomics: inflation. Specifically, we look at inflation at the household level and derive a microfounded, nonhomothetic cost-of-living index. We document that consumption across the expenditure distribution varies and that this gives rise to differences in inflation rates among households. Inflation inequality has been documented in other empirical studies but the methodological approach taken in these studies has some drawbacks as it rests on group- rather than household-specific price indices. The most apparent drawback of this approach is that it thus does not characterize the full distribution of inflation but only approximates it. The index we derive in this chapter is instead continuous in expenditures and therefore overcomes this issue. We couple the US Consumer Expenditure Survey with goods specific price series from the Bureau of Labor Statistics to conduct an analysis of inflation inequality based on our index. The new result of our empirical analysis is that the volatility of inflation has been 2,5 times higher for poorer households than for richer households over the period from 1996 to 2020. The magnitude of the volatility in inflation for the poor is similar to that for the nation-wide US consumer price index, CPI, measured between 1970 and 1980. Our analysis further shows that this result is driven by poorer households' vulnerability to price changes in food and energy coupled with a limited tendency to substitute between goods. The empirical results suggest that it is important in future work to investigate *why* households make the consumption choices they do (is it because they are forced to buy e.g. food and energy

out of necessity or because they simply prefer to spend more on these?) and whether economic policies can be designed such that the welfare costs of price movements can be minimized.

Chapter 2 / Quality and Consumption Basket Heterogeneity
with Rasmus B. Larsen
In this chapter, which in an earlier version also featured in Rasmus B. Larsen's PhD dissertation, we look at another core topic within macroeconomics: consumption behavior. We investigate what households put into their shopping carts and study in detail the differences across the income distribution. Specifically, we look at two decisions households in general make when they shop: how *much* they buy and what *quality* the products they buy have. We use a large scanner dataset with detailed information on US households' purchase of supermarket products. The dataset is a register that tracks more than 60,000 households' daily consumption from 2004 and onwards. The key information, the product purchases, is based on barcode information for each product and gives information on price, quantity, the store in which the product is bought, where the store is located, whether the product was on sale and so on. This gives us a unique opportunity to characterize households' consumption behavior and compare consumption baskets both across households and over time. We use the dataset to first document a fact: richer households buy goods of higher quality than poorer households. This fact has been shown in other empirical studies but since the literature is still young and under development our result adds to the robustness of this fact. The interesting about this finding is that households make product purchases based on several margins. The obvious margin is the choice of quantity. The less obvious one is in light of our findings the quality margin: households do not only adjust the quantity they buy but also the quality of what they buy. After this has been documented we continue with an original analysis of how households respond on these margins when they receive a one-time transfer of money. The new and novel result that we document is that households adjust their purchases on the quality margin following a one-time increase in income. Moreover, the adjustment on the quality margin is unequal across the income distribution: middle-income households make larger adjustments than low- and high-income households who do not seem adjust extensively. The adjustment response in quality following the one-time transfer of money is in other words hump-shaped across the income distribution. This finding is important as it informs us about how we should think about consumption behavior. In the last part of the chapter we embed this type of consumption behavior into an otherwise standard consumption-saving model and show that it can play a crucial role in wealth accumulation. In particular, as households become richer the marginal utility of consumption continues to stay high due to the increase in quality they can tap into and this induces a want for accumulating wealth. The model we study is in the end better, relative to the standard model, at mimicking the wealth distribution that we observe for the US and thus indicates the role the consumption behavior we empirically document might play in other macroeconomic questions.

Chapter 3 / Local projections or VARs? A data-driven selection rule for finite-sample estimation of impulse responses

with Anders F. Kronborg

In this chapter we look at how one in empirical analyses estimates so-called impulse responses which in other words covers over how the economy responds to shocks. A shock could for instance be a sudden and unforeseen fall in foreign demand. The question that one seeks an answer to is in this regard how the economy responds to the fall in foreign demand, both immediately and over time. We contribute in this chapter with a method that builds bridge between the two most prominent and already existing methods: the VAR- and LP-method, respectively. Our method is important as the LP- and VAR-method often give different - and sometimes even opposing - answers to the same question. This poses a challenge to practitioners as it forces them to take a stance on which answer they should rely on. Our method firstly shows that the difference between the VAR- and LP-method does not arise due to unrelated reasons but is a matter of how data is used. The concrete insight we draw is that the answer from the LP-method can be obtained *from* the VAR-method. This in turn allows us to interpret the LP-answer as a contribution to the VAR's. The benefit from this is that the final answer on average lies closer to the true answer. Adding the LP-contribution however comes at the cost of a higher imprecision and the question we address in the end is whether the cost of that imprecision is made up for by the benefit it comes with. Our final theoretical contribution is thus to suggest a selection rule that makes that assessment based on the same data which is used in the overall analysis. A particular feature of the selection rule is that it does not only give an indication of whether the VAR or LP answer should be chosen but also whether the two answers should be mixed by only allowing for a partial LP-contribution to the VAR. The selection rule is important as it tackles a problem that has been known to the literature for a long time but which we, to the best of our knowledge, are the first to solve with a fully data-driven approach. The chapter is finished off by showing that the selection rule does well in simulation studies.

Resumé (på dansk)

Denne afhandling består af tre selvstændige kapitler indenfor makroøkonomi. Kapitlerne deler alle en orientering mod empirisk analyse af makroøkonomiske spørgsmål. I det første og tredje kapitel præsenteres der nye metoder til anvendelse i empiriske analyser. I det første kapitel bidrager den nye metode til empiriske analyser af inflation på tværs af husholdninger. I det tredje kapitel bidrager den nye metode til empiriske analyser af hvordan økonomien udvikler sig i kølvandet på stød. De husholdningsspecifikke inflationsmål, der diskuteres i det første kapitel, hænger tæt sammen med husholdningers forbrugsadfærd. I det første og andet kapitel foretages der specifikke empiriske analyser af amerikanske husholdningers forbrugsadfærd.

Kapitel 1 / A Nonhomothetic Price Index and Inflation Heterogeneity
med Phillip Hochmuth & Markus Pettersson
Vi ser i dette kapitel på et klassisk og centralt emne indenfor makroøkonomi: inflation. Vi ser specifikt på inflation på husholdningsniveau og udvikler en teori, der både kan forklare, hvad vi ser i data og beskrive inflation på tværs af alle husholdninger. Vi dokumenterer, at der på tværs af husholdninger med forskellige niveauer af forbrugsudgifter er forskel på hvad deres varekurve består af og at dette kan føre til forskelle i inflation. Forskelle i inflation på tværs af husholdninger er blevet belyst i en række empiriske studier, men de bagvedliggende metoder og teorier til at belyse og forstå disse forskelle har nogle begrænsninger. Den empiriske metode, der typisk har været anvendt, har bl.a. den begrænsning, at den beror på gruppe-, og altså ikke individ-, specifik forbrugsadfærd. F.eks. samler man fattigere husholdninger i én gruppe og udleder på baggrund af dette fattiges *gennemsnitlige* varekurv. På samme måde gør man også dette for rigere husholdninger og forskellen på de to gennemsnitlige varekurve bruges til at udlede forskelle mellem rige og fattige. Den mest åbenlyse begrænsing ved denne metode er i denne henseende at man altså ikke ser direkte på den individuelle husholdnings inflation. Den eksisterende metodes begrænsinger er anerkendt i litteraturen, men der har hidtil ikke fandtes bedre alternativer. Metoden som vi bidrager med er vigtig, fordi den netop er et alternativ, der udover at overkomme den tidligere metodes begrænsinger også åbner op for langt flere muligheder. For det første viser vi, at den tidligere metode er en approksimativ udgave af vores. Det demonstrerer vi bl.a. ved at vise, at de tidligere gruppe-specifikke resultater kan genfindes med vores metode. Dernæst

udfolder vi vores metode i en analyse hvor vi først estimerer den fulde fordeling af inflation på tværs af husholdninger i USA fra 1996–2020. Vores analyse frembringer et nyt og overraskende resultat: udsvingene i inflation er mere end 2,5 gange større for de fattigste husholdninger end for de rigeste set over det seneste kvarte århundrede. Størrelsen på udsvingene i inflationen for de fattigste er i omegnen af hvad man så i perioden fra 1970–1980 for det brede, nationale forbrugerprisindeks, CPI, i USA. Resultatet er slående, både på grund af den store ulighed der er på tværs af husholdninger og på grund af den størrelsesorden, som udsvingene har. I den sidste del af analysen viser vi, at resultaterne er robuste og vi afslutter med at vise, at forskellen er drevet af fattigere husholdningers følsomhed overfor udsving i fødevare- og energipriser samt en begrænset tilbøjelighed til at tilpasse forbruget. De empiriske resultater lægger op til at man dykker mere ned i præcis hvorfor fattigere husholdninger er så eksponeret overfor fødevare- og energipriser (er det fordi, de er tvunget til at købe disse goder af nødvendighedsårsager eller fordi de bare foretrækker det?) og om man med politiske redskaber kan indrette et system hvor de negative velfærdseffekter af prisudsving kan minimeres.

Kapitel 2 / Quality and Consumption Basket Heterogeneity

med Rasmus B. Larsen

I dette kapitel, som i en tidligere udgave også indgår i Rasmus B. Larsen's PhD afhandling, ser vi på et andet kerneemne indenfor makroøkonomi: forbrug. Vi dykker i dette kapitel ned i hvad folk lægger i deres indkøbsvogn og studerer i detaljen forskelle på tværs af husholdninger. Helt specifikt ser vi på to beslutninger, som folk generelt træffer, når de handler: hvor *meget* køber de af en given vare og hvor høj *kvalitet* har varen? Vi anvender et stort stregkodedatasæt med detaljeret information om amerikanske husholdningers køb af dagligvarer. Datasættet er et register over mere end 60.000 husholdningers daglige køb af varer fra 2004 til i dag. Nøgleinformationen i datasættet, varekøbet, er baseret på stregkodedata for hver enkelt vare og giver os information om alt fra pris, mængde, butikken hvori varen blev købt, hvor butikken ligger, om varen var på tilbud osv. Det giver os en unik mulighed for at kortlægge husholdningernes forbrug og sammenligne deres varekurve både på tværs af husholdninger og over tid. Ved hjælp af datasættet dokumenterer vi først et faktum: rigere husholdninger køber varer af højere kvalitet end fattigere husholdninger. Dette faktum er blevet dokumenteret tidligere, men litteraturen er ung og stadig under udvikling, hvorfor de resultater vi finder med vores data, er et vigtigt bidrag til at understrege gyldigheden af dette faktum. Det interessante ved resultatet er, at husholdningers køb af varer varierer på flere marginer. Den mest åbenlyse margin er mængde: man køber enten mere eller mindre af nogle varer. Denne margin er typisk hvad man også arbejder med indenfor teoretisk modellering af husholdningers forbrug i makroøkonomi. Den mindre belyste margin er kvalitetsmarginen: man vælger ikke kun hvor meget af en given vare, man vil have, men også varens kvalitet. Efter at dette faktum er dokumenteret for vores data, fortsætter vi med at foretage en original analyse af hvordan husholdninger reagerer på disse marginer, når de modtager en pludselig engangsstigning i deres indkomst. Vi dokumenterer først et efterhånden klassisk resultat: fattigere husholdninger bruger en større andel af deres nyerhvervede penge end rigere husholdninger. Dernæst dokumenterer vi et nyt resultat: den største ændring på kvalitetsmarginen sker blandt husholdninger fra mellemindkomstgruppen. Fattigere og rigere husholdninger reagerer derimod ikke på kvalitetsmarginen i samme grad. Indsigterne fra den empiriske analyse i dette kapitel er vigtige, fordi de fortæller os om, hvordan vi skal tænke på husholdningers forbrugsadfærd. Vi belyser i slutningen af kapitlet vigtigheden af disse indsigter ved at inkorporere den empirisk dokumenterede forbrugsadfærd i en ellers standard forbrugs- og opsparingsmodel. Vi viser at denne adfærd kan give anledning til en langt større velstandsakkummulering blandt rigere husholdninger, hvilket i sidste ende fører til en mere ulige velstandsfordeling, som man også observerer i data.

Kapitel 3 / Local projections or VARs? A data-driven selection rule for finite-sample estimation of impulse responses

med Anders F. Kronborg

Vi fokuserer i dette kapitel på hvordan man i empiriske analyser estimerer såkaldte impuls responser, som med andre ord dækker over hvordan økonomien udvikler sig i kølvandet på et stød. Et sådant stød kunne f.eks. være en pludselig og uforudset nedgang i efterspørgslen efter danske varer. Spørgmålet man søger svar på er således hvordan nedgangen i efterspørgslen sætter sig i økonomien både umiddelbart og over tid. Vi bidrager i dette kapitel med en ny metode, der bygger bro mellem de to mest prominente og allerede eksisterende metoder: VAR- og LP-metoden, henholdsvist. Metoden som vi bidrager med er vigtig, da VAR- og LP-metoden ofte giver forskellige - og nogle gange endda modstridende - svar på det samme spørgsmål. Det skaber et problem for praktikere i og med at de tvinges til at tage stilling til hvilket svar, de skal vælge. Vores metode er en løsning på den problemstilling. Vores metode belyser først og fremmest at forskellen mellem VAR- og LP-metodens svar ikke opstår på baggrund af urelaterede årsager, men skyldes den måde hvorpå data behandles. Den konkrete indsigt gør os i stand til at udlede LP-metodens svar *fra* VAR-metodens. Man kan således forstå LP-metodens svar som et bidrag til VAR-metodens. Gevinsten ved LP-bidraget er at det endelige svar i gennemsnit ligger tættere på det sande. At tilføje LP-bidraget til VAR-metodens svar kommer dog med et tab af præcision og det efterlader os i sidste ende med et behov for at afveje gevinsten ved LP-bidraget ift. tabet. Vores sidste teoretiske bidrag er et forslag til en selektionsregel der foretager den afvejning på baggrund af den data, der i øvrigt er anvendt i analysen. Selektionsreglen fortæller helt konkret om VAR- eller LP-metodens svar skal vælges men foreslår også en blanding af de to svar, når dette er bedst. Selektionsreglen er vigtig fordi den løser et problem som har været kendt i literaturen men som vi, til vores bedste kendskab, er de første der kommer med en data-drevet løsning til. Kapitlet afsluttes med at vi i en række af simulationsstudier viser, at redskabet virker godt.

References

Rasmus B. Larsen. (2020). "Quality and consumption basket heterogeneity." *Essays in Macroeconomics: Consumption Behavior, Price Dynamics, and Fiscal Spending* (pp. 1–76). University of Copenhagen. (Cited on pages ix, xii).

Chapter 1

A Nonhomothetic Price Index and Inflation Heterogeneity

A Nonhomothetic Price Index and Inflation Heterogeneity

Philipp Hochmuth[*] Markus Pettersson[†] Christoffer Weissert[‡]

Abstract

We derive a microfounded, nonhomothetic generalization of all known superlative price indices, including the Fisher, the Törnqvist, and the Sato-Vartia indices. The index largely avoids the need for estimation, aggregates consistently across heterogeneous households, and admits different index weights across the expenditure distribution. The latter property rationalizes the methods used in most previous measurements of inflation inequality. In an empirical application to the United States using CEX-CPI data for the period 1995–2020, we find: (i) poor and rich households experience on average the same inflation rate; but (ii) inflation for the poorest decile is more than 2.5 times as volatile as that of the richest decile; and (iii) this higher volatility primarily stems from a larger exposure to price changes in food, gas and utilities. In these findings, substitution between goods as prices change plays only a second-order role. Instead, almost all differences come from mechanical changes in the cost of different base-period reference baskets.

JEL classification: C43, D11, D12, E31, I30

Keywords: cost of living, inequality, nonhomotheticity, price index

[*] Stockholm University. Email: philipp.hochmuth@iies.su.se.
[†] Stockholm School of Economics. Email: markus.pettersson@phdstudent.hhs.se.
[‡] University of Copenhagen and Danmarks Nationalbank. Email: cjw@econ.ku.dk.

1 Introduction

Does inflation vary with income? Conventional price indices that are used for the measurement of inflation cannot answer this question because these rely fully on the assumption of homothetic preferences. That is, consumers are assumed to make identical consumption allocations, regardless of income level. Yet, one of the oldest empirical economic facts, dating back to at least Engel (1857), is that consumption patterns differ systematically between rich and poor consumers. In other words, preferences are *not* homothetic. Differences in consumption bundles raises the possibility for inflation inequality, with implications for any area where inflation matters, not least monetary policy and the measurement of real incomes. By now there exists an abundance of empirical research investigating this issue by computing standard (homothetic) price indices for separate income groups. The question of how to consistently incorporate nonhomothetic preferences into conventional price index formulas, however, remains unsolved.

The goal of this paper is to tackle this problem head-on. In doing so, we make three main contributions. First, we derive a cost-of-living index that is consistent with nonhomothetic consumer demand theory. In its most general form, this index nests all known superlative price indices as special cases, including the Fisher (1922), the Törnqvist (1936), and the Sato (1976) and Vartia (1976) indices. Second, we outline a feasible strategy to compute these price indices without being at the mercy of estimating entire demand systems. Instead, under a relatively mild assumption, estimation reduces to two parameters which are identified from a single equation. Third, we implement this approach using consumption and CPI data to investigate US inflation heterogeneity over the last quarter century.

Our framework allows for a characterization of the cost of living for the full expenditure distribution as well as at the aggregate level. We achieve this by deriving cost-of-living indices from a specification of Muellbauer's (1975, 1976) "price independent generalized linearity" (PIGL) preferences that has recently gained popularity in the structural change literature.[1] These preferences are nonhomothetic but maintain tractable aggregation properties that allow us to account for consumer heterogeneity. Like many conventional price indices, we show that the PIGL preferences induce cost-of-living indices that are weighted geometric means of individual price changes. Unlike their homothetic counterparts, however, the weights on these price changes vary systematically across the expenditure distribution. Specifically, richer households allocate higher weights to price changes of luxury goods. Changes in the cost of living are consequently allowed to differ with the expenditure level.

Due to the nonhomothetic nature of the underlying PIGL preferences, the cost-of-living index in its most general form is not directly computable without estimating a complete consumer

[1] See for instance Boppart (2014), Alder, Boppart and Müller (forthcoming), and Cravino, Levchenko and Rojas (forthcoming).

demand system. We overcome this hurdle by imposing a key assumption: that preferences are weakly separable into necessities and luxuries. Under the weak separability assumption, the cost-of-living index reduces to observed price changes and expenditure shares and only two unknown parameters, which are readily estimated by linear as well as nonlinear regression methods. As in for instance Wachter and Yogo (2010) and Orchard (2021), classification of individual goods as "necessity" or "luxury" is straightforwardly done by investigating Engel curves. The cost-of-living index still nests homothetic price indices as special cases, so the assumption of weakly separable preferences is not a hard restriction when compared to conventional price indices. In our empirical application, we also show that it is well justified in the data.

We illustrate this implementation method by an empirical analysis of US inflation heterogeneity over the years 1995 to 2020. In this exercise, we consider twenty-one consumption good categories from the Consumer Expenditure Survey (CEX) that we match with corresponding CPI sub-indices. We obtain three main empirical results. First, households in the first expenditure decile ("the poor") experienced similar inflation as households in the tenth expenditure decile ("the rich") between 1995 and 2020, although substantial differences arise in the period around the Great Recession. In particular, poor households faced on average a 0.37 percentage points higher annual inflation rate between 2004 and 2015. Second, while the average inflation *rate* is relatively similar between the poor and the rich, inflation *volatility* is more than 2.5 times higher for the poor. Third, we find that this higher volatility primarily stems from a larger exposure for the poor to price changes in food, gas and utilities.

Decomposing the price index, we find that the overall development is almost entirely driven by mechanical price changes on the base-period consumption basket. Substitution behavior, as relative prices change, plays only a minor role. Yet, both differential base-period reference baskets and differential substitution are significant drivers in explaining the *differences* between groups, as poor households substitute away from expensive goods to a larger degree than the rich.

Furthermore, we exploit the fact that the price index can be used to retrieve inflation for a sequence of varying real expenditures and document the average inflation experienced over the life-cycle. The results from this analysis indicate that the inflation volatility of the young and poor is somewhat dampened by their life-cycle path of expenditures, while this is not the case for the old and poor. Overall, aging effects do not substantially change the fact that households who are initially poor experience a 2.5 times higher expected volatility of inflation compared to households who are initially rich.

This paper falls within an old literature on the economic approach to price index theory following Konüs (1939), Samuelson and Swamy (1974), Diewert (1976, 1978), Feenstra (1994), Redding and Weinstein (2020), and many others, whereby cost-of-living indices are derived from consumer theory via the expenditure function.[2] Central to this line of research is Diewert's notion of a *superlative* price index, which includes indices that are exact for some homothetic expenditure function and can approximate other homothetic indices to the second order. The Fisher, the Törnqvist and the Sato-Vartia indices are all shown to satisfy this property (see Diewert, 1976, and Barnett and Choi, 2008). Our paper provides a nonhomothetic generalization of these and all other currently known indices within this class.

Nonhomothetic preferences have generally received little attention in the price index literature. Feenstra and Reinsdorf (2000) derive an index for Deaton and Muellbauer's (1980) almost ideal demand system (AIDS) and Oulton (2012) proposes a numerical algorithm to calculate indices of Banks, Blundell and Lewbel's (1997) generalization of the AIDS. The AIDS is a special case of the PIGLOG class of preferences, which is itself a limit case of the PIGL preferences used here. Unlike the PIGL class, the AIDS does not provide a straightforward generalization of conventional price indices. Redding and Weinstein (2020) also derive a theoretical price index for the nonhomothetic CES specification of Hanoch (1975) and Sato (1975). In contrast to our index, however, the nonhomothetic CES specification does not consistently aggregate across heterogeneous consumers and (to the best of our knowledge) provides no easy implementation empirically without being forced to estimate all parameters in the utility function.

We also add to an empirical strand of literature concerned with inflation inequality which dates back at least to the 1950s (see for instance Muellbauer, 1974, and references therein), with recent advances surveyed by Jaravel (2021). The bulk of this literature approximates nonhomothetic cost-of-living indices by computing conventional price indices separately for different income groups.[3] This "group-specific" approach posits that differences in inflation are driven by differences in *ex ante* tastes and rests on a theoretical foundation where deep preference parameters jump discontinuously between groups.[4] These discrete jumps set aside straightforward comparisons between groups as well as to an aggregate inflation rate, the latter being completely lost. By contrast, inflation heterogeneity here stems endogenously from differences in expenditures. This allows us to characterize the full inflation distribution as well as aggregate inflation, with clear-cut comparisons between subgroups. Nevertheless, the group-specific approach provides an easy way to obtain index weights that vary with income, a feature that lies at the core of our framework. We therefore do not necessarily see these approaches as direct substitutes. Instead, our framework rationalizes previous empirical methodologies within a consistent theory for nonhomothetic consumer demand.

[2] Diewert (1993) surveys the early stages of this literature, which is far too large for us to do justice to here.

[3] Recent papers employing this approach include Hobijn and Lagakos (2005), McGranahan and Paulson (2005), Broda and Romalis (2009), Kaplan and Schulhofer-Wohl (2017), Jaravel (2019), Orchard (2020, 2021), Argente and Lee (2021), and Lauper and Mangiante (2021).

[4] The implication being, for instance, that poor consumers allocate a larger share of expenditures to, say, rented housing due to some innate preference for rental homes obtained at birth.

The paper proceeds as follows. Section 2 covers the theoretical framework, derives the nonhomothetic price index and shows how all superlative price indices can be generalized to a nonhomothetic setting. Section 3 outlines the strategy of our empirical implementation and discusses the assumption we make to render the demand system estimation feasible. Section 4 explains the data we employ, classifies the twenty-one goods into necessities and luxuries and reports estimates from the tractable demand system estimation. Section 5 reports the main empirical results, while Section 6 decomposes the price index and inflation further. Section 7 compares the main results to traditional demand system estimation and Section 8 concludes.

2 Theoretical Framework

The framework we consider is one where consumers maximize utility over a set of goods J with a corresponding price vector $\boldsymbol{p}$ and where we wish to investigate the change in the cost of living between a period t and some base period s. In what follows, we drop time subscripts whenever possible to simplify notation, as long as this causes no confusion. The minimum expenditure e required to obtain some utility level u when faced by the price vector $\boldsymbol{p}$ is given by the expenditure function $e = c(u, \boldsymbol{p})$. Following Konüs (1939), we define a cost-of-living index in period t relative to base period s to be the ratio of minimum expenditures required to maintain a constant utility level:

$$P(u, \boldsymbol{p}_t, \boldsymbol{p}_s) \equiv \frac{c(u, \boldsymbol{p}_t)}{c(u, \boldsymbol{p}_s)}. \tag{1}$$

Hereinafter we typically leave the arguments of the cost-of-living index implicit and simply write $P_t = c(u, \boldsymbol{p}_t)/c(u, \boldsymbol{p}_s)$.

2.1 The Homothetic Case

In general, the Konüs cost-of-living index (1) depends on the reference standard of living u as well as the prices in the two periods. Samuelson and Swamy (1974) show that independence of u occurs if and only if we consider the special case of homothetic preferences. Suppose for instance that consumer preferences are characterized by an indirect utility function of the standard homothetic CRRA form,

$$V(e, \boldsymbol{p}) = \frac{1}{\varepsilon}\left[\left(\frac{e}{B(\boldsymbol{p})}\right)^{\varepsilon} - 1\right], \tag{2}$$

where $B(\boldsymbol{p})$ is a linearly homogenous function of prices and ε is the coefficient of relative risk aversion. Inverting the utility function to obtain the expenditure function $c(u, \boldsymbol{p}) = (1 + \varepsilon u)^{1/\varepsilon} B(\boldsymbol{p})$ and using Equation (1), we get

$$P_t = \frac{B(\boldsymbol{p}_t)}{B(\boldsymbol{p}_s)}, \tag{3}$$

which is evidently independent of the utility level. All conventional price indices that can be derived from economic theory satisfy this property.

2.2 The Nonhomothetic Case: Preferences

Our framework extends the indirect utility (2) to allow for nonhomothetic behavior. To this end, we characterize preferences by an indirect utility function as in Boppart (2014),

$$V(e, \boldsymbol{p}) = \frac{1}{\varepsilon}\left[\left(\frac{e}{B(\boldsymbol{p})}\right)^{\varepsilon} - 1\right] - \frac{\nu}{\gamma}\left[\left(\frac{D(\boldsymbol{p})}{B(\boldsymbol{p})}\right)^{\gamma} - 1\right], \tag{4}$$

where $B(\boldsymbol{p})$ and $D(\boldsymbol{p})$ are linearly homogeneous functions of prices and the parameters satisfy $\varepsilon, \gamma \in (0,1)$ and $\nu > 0$. This utility function belongs to the class of PIGL preferences defined by Muellbauer (1975, 1976) and more generally to the class of "intertemporally aggregable" preferences defined by Alder, Boppart and Müller (forthcoming). Despite being nonhomothetic, these preferences consistently aggregate across individual-level expenditures. Aggregate expenditure shares in this case correspond to a representative expenditure level which is independent of prices and given by the average expenditure level multiplied by a simple inequality measure.

To gain understanding and intuition of Equation (4), it is convenient to think of $B(\boldsymbol{p})$ and $D(\boldsymbol{p})$ as the expenditure functions of some homothetic sub-utility functions. We refer to these sub-utility functions as "goods" or "baskets". The parameter ε controls the degree of nonhomotheticity between the D and B baskets: the expenditure elasticity of demand for the D basket is $1 - \varepsilon$, which is less than one under the restrictions on ε. The D basket therefore covers necessity needs and B conversely covers luxury needs. In the limit case $\varepsilon \to 0$, the expenditure elasticity is one and we obtain homothetic preferences. Comparing Equations (2) and (4), we also obtain homothetic preferences for $\varepsilon \neq 0$ whenever $B(\boldsymbol{p}) = D(\boldsymbol{p})$ or in the limit case $\nu \to 0$. The parameter ν is a scale parameter that controls the level of demand for the D basket and γ controls the non-constant elasticity of substitution between the B and D baskets.

In general, there is nothing restricting an individual good j from occurring in both the B and the D baskets. If there is overlap between the sets of goods within B and D, the allocations to the B and D goods are not directly observable and we obtain what Blundell and Robin (2000) call "latent separability". Latent separability is equivalent to weak separability, but in the latent goods B and D rather than in purchased goods, with weak separability as the special case when there is no overlap between B and D. Two-stage budgeting is still valid under latent separability, meaning that the consumer's allocation problem can be viewed in two stages where consumers first allocate expenditures between the B and D baskets and then, conditional on this first-stage decision, allocate expenditures across individual goods within B and D. Applying Roy's identity, the expenditure shares w_D and w_B allocated to the D and B baskets in the first stage are therefore given by

$$w_D = \nu \left(\frac{B(\boldsymbol{p})}{e}\right)^{\varepsilon} \left(\frac{D(\boldsymbol{p})}{B(\boldsymbol{p})}\right)^{\gamma} \tag{5}$$

and

$$w_B = 1 - \nu \left(\frac{B(\boldsymbol{p})}{e}\right)^{\varepsilon} \left(\frac{D(\boldsymbol{p})}{B(\boldsymbol{p})}\right)^{\gamma}. \tag{6}$$

Similarly, the shares w_j^D and w_j^B of total D and B expenditures allocated to individual good j are given by

$$w_j^D = p_j \frac{D_j(\boldsymbol{p})}{D(\boldsymbol{p})} \qquad \text{and} \qquad w_j^B = p_j \frac{B_j(\boldsymbol{p})}{B(\boldsymbol{p})}, \tag{7}$$

where D_j and B_j denote, respectively, the partial derivatives of D and B with respect to p_j. Equations (5) to (7) imply an expenditure share w_j of good j in total expenditures of the form

$$w_j = p_j \left[\frac{B_j(\boldsymbol{p})}{B(\boldsymbol{p})} + \left(\frac{D_j(\boldsymbol{p})}{D(\boldsymbol{p})} - \frac{B_j(\boldsymbol{p})}{B(\boldsymbol{p})} \right) \nu \left(\frac{B(\boldsymbol{p})}{e}\right)^{\varepsilon} \left(\frac{D(\boldsymbol{p})}{B(\boldsymbol{p})}\right)^{\gamma} \right]. \tag{8}$$

Therefore, nonhomotheticity between B and D also creates nonhomothetic behavior across individual goods, with a good j being a necessity if $D_j/D > B_j/B$ and a luxury vice versa. Aggregating over any N number of consumers indexed by h, the aggregate expenditure share $\overline{w}_j$ of good j across these consumers is

$$\overline{w}_j = p_j \left[\frac{B_j(\boldsymbol{p})}{B(\boldsymbol{p})} + \left(\frac{D_j(\boldsymbol{p})}{D(\boldsymbol{p})} - \frac{B_j(\boldsymbol{p})}{B(\boldsymbol{p})} \right) \nu \left(\frac{B(\boldsymbol{p})}{\bar{e}}\right)^{\varepsilon} \left(\frac{D(\boldsymbol{p})}{B(\boldsymbol{p})}\right)^{\gamma} \kappa \right], \tag{9}$$

where $\bar{e} \equiv \frac{1}{N} \sum_{h=1}^{N} e_h$ is the average expenditure level and κ is an inequality measure defined by

$$\kappa \equiv \frac{1}{N} \sum_{h=1}^{N} \left(\frac{e_h}{\bar{e}}\right)^{1-\varepsilon}. \tag{10}$$

Aggregate shares $\overline{w}_B$ and $\overline{w}_D$ of the B and D baskets are defined similarly. See Alder, Boppart and Müller (forthcoming, Proposition 2) for a derivation of Equations (9) and (10). The representative agent in Muellbauer's (1975, 1976) sense (henceforth the PIGL RA) is the expenditure level e^{RA} that induces the aggregate expenditure share. By Equation (9), this expenditure level is given by $e^{RA} \equiv \bar{e}\kappa^{-1/\varepsilon}$.

2.3 The Nonhomothetic Case: Price Index

The indirect utility function (4) allows us to extend the homothetic cost-of-living index (3). Unlike the homothetic case, the index now depends on a base-period standard of living, represented by the utility level u in the Konüs definition (1). Because preferences are nonhomothetic over the

B and D baskets, with an expenditure elasticity of demand for the D basket always less than one, this utility level is fully captured by w_{Ds}, the base-period expenditure share of the D good. For the remainder of the paper, let

$$L(x,y) = \begin{cases} \dfrac{x-y}{\ln x - \ln y} & \text{if } x \neq y, \\ x & \text{if } x = y, \end{cases}$$

denote the logarithmic mean (Carlson, 1972). Moreover, recalling from Equation (3) that the cost-of-living index of homothetic preferences between some period t and base period s is the ratio of the corresponding expenditure functions over the same periods, we denote the price indices of the B and D baskets by $P_{Bt} \equiv B(\boldsymbol{p}_t)/B(\boldsymbol{p}_s)$ and $P_{Dt} \equiv D(\boldsymbol{p}_t)/D(\boldsymbol{p}_s)$. The following result then shows that the cost-of-living index corresponding to Equation (4) is a function of w_{Ds}, P_{Bt} and P_{Dt} and the two parameters ε and γ.

Proposition 1 (PIGL cost-of-living index)**.** *If preferences are of the PIGL form (4) and the base-period expenditure share w_{Ds} allocated to the D basket is given, the Konüs cost-of-living index is*

$$P_t^{PIGL} = P_{Dt}^{\frac{\gamma\phi_t}{\varepsilon}} P_{Bt}^{1-\frac{\gamma\phi_t}{\varepsilon}} \qquad \textit{with} \qquad \phi_t \equiv \frac{L(\psi_{Dt},\ \psi_{Ds})}{L(\psi_{Dt},\ \psi_{Ds})\ +\ L(\psi_{Bt},\ \psi_{Bs})}, \tag{11}$$

where

$$\psi_{Bt} \equiv \left(1 - \frac{\varepsilon w_{Ds}}{\gamma}\right)\left(\frac{P_{Bt}}{\widetilde{P}_t}\right)^{\gamma} \qquad \textit{and} \qquad \psi_{Dt} \equiv \frac{\varepsilon w_{Ds}}{\gamma}\left(\frac{P_{Dt}}{\widetilde{P}_t}\right)^{\gamma} \tag{12}$$

are shares of a CES-type aggregator $\widetilde{P}_t$ defined by

$$\widetilde{P}_t \equiv \left[\left(1 - \frac{\varepsilon\, w_{Ds}}{\gamma}\right) P_{Bt}^{\gamma} + \frac{\varepsilon\, w_{Ds}}{\gamma} P_{Dt}^{\gamma}\right]^{\frac{1}{\gamma}}. \tag{13}$$

The aggregate cost-of-living index over any N number of consumers is given identically using their average expenditure share $\overline{w}_{Ds}$ in ϕ_t.

Sketch proof (full proof in Section A.1)*.* Set the reference utility to that of the base period expenditure level, $u \equiv V(e_s, \boldsymbol{p}_s)$. It is then possible to write the period-t expenditure function corresponding to Equation (4) as $c(u, \boldsymbol{p}_t) = c(u, \boldsymbol{p}_s)\widetilde{P}_t^{\frac{\gamma}{\varepsilon}} P_{Bt}^{1-\frac{\gamma}{\varepsilon}}$. Since $\widetilde{P}_t$ is of a CES form, it can be rewritten as a Sato-Vartia index with weights $1-\phi_t$ and ϕ_t on P_{Bt} and P_{Dt}, respectively.

The result then follows from the Konüs definition (1). □

Proposition 1 shows that the PIGL cost-of-living index can be written as something akin to a Sato-Vartia index over the B and D baskets. Unlike the homothetic case, however, the ϕ_t in the weights of the two sub-indices varies across the expenditure distribution. Richer consumers spend a smaller share w_D on the D basket, which reduces the weights ψ_{Dt} and, subsequently, ϕ_t. In other words, because richer consumers allocate a smaller share to the D basket, the corresponding price index P_{Dt} is weighted less heavily when determining the overall change in the cost of living. The weights ψ_{Dt} and ψ_{Bt} are not directly observable but are readily computed given price indices P_{Bt} and P_{Dt}, an expenditure share w_{Ds}, and parameter values for ε and γ. In Section A.1, we show that $w_{Ds}(P_{Dt} \,/\, \widetilde{P}_t)^{\gamma}$ is the expenditure share of the D basket at period-t prices that prevails at the same utility level as w_{Ds}. Therefore, the weights ψ_{Dt} and ψ_{Bt} ensure that the consumer remains on the same indifference curve as in the base period.

Two potential caveats to Proposition 1 are that the underlying preferences are identical across consumers with the same expenditure level and that expenditure shares change monotonically in the level of expenditure. Redding and Weinstein (2020) emphasize accounting for taste heterogeneity in cost-of-living indices while Banks, Blundell and Lewbel (1997) highlight the importance of allowing for hump-shaped expenditure shares to match microeconomic data. In Section A.4 we show that it is straightforward to incorporate time- and household-specific tastes between the B and D baskets into the indirect utility function (4) and that this leaves Proposition 1 unaffected. In Section A.5 we discuss a generalization that allows for hump-shaped expenditure shares. This generalization works well for household-level indices but requires more stringent conditions for aggregate price indices to have the same form.

2.4 Generalized Superlative Indices

With Proposition 1 at hand, it is straightforward to generalize standard homothetic indices to the nonhomothetic PIGL case: just plug in two homothetic indices for P_{Bt} and P_{Dt} in Equation (11). To emphasize the importance of Proposition 1, we present generalizations of two classes of indices: Diewert's (1976) *quadratic-mean-of-order-r* class, which consists of all indices of the form

$$P_t = \sqrt{\left\{ \sum_{j \in J} w_{js} \left(\frac{p_{jt}}{p_{js}} \right)^{\frac{r}{2}} \right\}^{\frac{2}{r}} \left\{ \sum_{j \in J} w_{jt} \left(\frac{p_{jt}}{p_{js}} \right)^{-\frac{r}{2}} \right\}^{-\frac{2}{r}}}, \qquad r > 0, \tag{14}$$

and Barnett and Choi's (2008) *Theil-Sato* class, which is defined as

$$P_t = \prod_{j \in J} \left(\frac{p_{jt}}{p_{js}} \right)^{\delta_{jt}}, \qquad \delta_{jt} \equiv \frac{m(w_{jt}, w_{js})}{\sum_{i \in J} m(w_{it}, w_{is})}, \tag{15}$$

where $m(x, y)$ is a *symmetric mean* of two variables, a function class that includes all linearly homogenous functions satisfying $\min\{x, y\} \leq m(x, y) = m(y, x) \leq \max\{x, y\}$. These index classes include several of the most well-known price index formulas. Equation (14) incorporates Fisher's (1922) ideal index ($r = 2$), the arithmetic Walsh (1901) index ($r = 1$) and, as a limit case, the Törnqvist (1936) index ($r \to 0$). Equation (15) nests the Törnqvist index (arithmetic mean, $m(x, y) = (x + y)/2$), the geometric Walsh (1901) index (geometric mean, $m(x, y) = \sqrt{xy}$), the Sato (1976)-Vartia (1976) index (logarithmic mean, $m(x, y) = (x - y)/(\ln x - \ln y)$), and the Theil (1973) index ($m(x, y) = \sqrt[3]{xy(x + y)/2}$). While we could choose any underlying homothetic price indices P_{Dt} and P_{Bt}, these two classes consists of all currently known *superlative* price indices (Diewert, 1976). That is, they are exact cost-of-living indices for some homothetic expenditure functions and are second-order approximations of any other homothetic price index.[5] Therefore, even if the indices corresponding to the true expenditure functions $B(\boldsymbol{p})$ and $D(\boldsymbol{p})$ have some other forms than those in Equations (14) and (15), we should still be able to reasonably approximate their corresponding price indices under this specific parameterization. The nonhomothetic generalization of these indices under Proposition 1 is presented below.

Corollary 1 (Generalized superlative indices)**.** *If preferences are of the PIGL form (4), the base-period expenditure share w_{Ds} allocated to the D basket is given, and $B(\boldsymbol{p})$ and $D(\boldsymbol{p})$ are expenditure functions with price indices of the form (14) or (15), the Konüs cost-of-living index is*

$$P_t^{G\text{-}S} = \prod_{j \in J} \left(\frac{p_{jt}}{p_{js}} \right)^{\chi_{jt}}, \tag{16}$$

where

$$\chi_{jt} \equiv \frac{\gamma \phi_t}{\varepsilon}\, \delta_{jt}^D + \left(1 - \frac{\gamma \phi_t}{\varepsilon}\right) \delta_{jt}^B, \tag{17}$$

with ϕ_t as in Proposition 1. The weights δ_{jt}^C, $j \in J$, $C \in \{B, D\}$, are given by

$$\delta_{jt}^C = \frac{1}{2} \left[\frac{\widetilde{w}_{Ljt}^C}{\sum_i \widetilde{w}_{Lit}^C} + \frac{\widetilde{w}_{Pjt}^C}{\sum_i \widetilde{w}_{Pit}^C} \right] \qquad \text{or} \qquad \delta_{jt}^C = \frac{m(w_{jt}^C, w_{js}^C)}{\sum_i m(w_{it}^C, w_{is}^C)} \tag{18}$$

if P_{Ct} is as in Equation (14) or (15), respectively. In the former case,

$$\widetilde{w}_{Ljt}^C \equiv w_{js}^C \, L\left(\left(\tfrac{p_{jt}}{p_{js}}\right)^{\frac{r}{2}}, \left(P_{Lt}^C\right)^{\frac{r}{2}} \right) \qquad \text{and} \qquad \widetilde{w}_{Pjt}^C \equiv w_{jt}^C \, L\left(\left(\tfrac{p_{jt}}{p_{js}}\right)^{-\frac{r}{2}}, \left(P_{Pt}^C\right)^{-\frac{r}{2}} \right),$$

where $P_{Lt}^C = \left[\sum_j w_{js}^C \left(\frac{p_{jt}}{p_{js}}\right)^{\frac{r}{2}} \right]^{\frac{2}{r}}$ and $P_{Pt}^C = \left[\sum_j w_{jt}^C \left(\frac{p_{jt}}{p_{js}}\right)^{-\frac{r}{2}} \right]^{-\frac{2}{r}}$. The aggregate cost-

[5] This definition differs from Diewert's original definition but is shown by Barnett and Choi (2008, Theorem 1) to be equivalent.

of-living index over any N number of consumers is given identically using their average expenditure shares in χ_{jt}.

Proof. In Section A.2. □

In Corollary 1, we have rewritten the *quardratic-mean-of-order-r* class on a geometric-mean form following Balk (2004) to highlight the intuitive generalization of the homothetic superlative indices that we obtain. In doing so, we denote the weights by L and P to capture the fact that these reduce to standard Laspeyres and Paasche weights when $r = 2$. The resulting cost-of-living index is a weighted geometric average of individual price changes with index weights (17) of the following structure:

$$\underset{\text{on } j}{\text{Weight}} = \underset{\text{on } D}{\text{Weight}} \times \underset{\substack{\text{on } j \\ \text{within } D}}{\text{Weight}} + \underset{\text{on } B}{\text{Weight}} \times \underset{\substack{\text{on } j \\ \text{within } B}}{\text{Weight}} .$$

The weights on j within B and D, given by Equation (18), are standard homothetic weights and affect all consumers similarly. The weights on D and B are the same as in Proposition 1. Therefore, the overall weights χ_{jt} vary across the base-period expenditure distribution in a similar way as before. If $B(\boldsymbol{p}) = D(\boldsymbol{p})$, we get that $\delta_{jt}^{B} = \delta_{jt}^{D}$ for all j and the generalized superlative indices immediately collapse to the homothetic indices in Equations (14) and (15).

Another feature of Corollary 1 is that it rationalizes the methodology used in much of the literature concerned with inflation inequality, whereby homothetic price indices are computed for different income groups separately. In particular, papers like Broda and Romalis (2009), Jaravel (2019), and Argente and Lee (2021) compute homothetic price indices of the geometric-mean form $\ln P_t = \sum_j \delta_{jt} \ln(p_{jt}/p_{js})$, where δ_{jt} are weights computed separately for each income group considered. This generates heterogeneous weights across the income distribution. The method therefore mimics an overall geometric-mean price index with income specific weights, which is exactly what we also have in Corollary 1. In contrast to the group-specific approach, however, Corollary 1 allows for a full characterization of the inflation distribution rather than a discontinuous, discrete distribution.

3 Empirical Implementation

To compute the generalized superlative indices in practice, we need total expenditure shares between the B and D baskets and the expenditure shares within each basket. Yet, if individual goods occur in both the B and the D baskets, these across and within expenditure shares are unobserved in the data. The only feasible approach then is to parameterize $B(\boldsymbol{p})$ and $D(\boldsymbol{p})$, estimate the demand system associated with the expenditure share equations (8) via GMM, and infer these shares from the estimated model. This methodology, however, suffers from the usual drawbacks of nonlinear demand system estimation. In particular, for standard parameterizations the number of parameters to estimate quickly grows out of proportion as we increase the number of goods considered.[6] The nonlinear nature of the demand system also implies that there is no guarantee that the GMM estimator converges to the actual global minimum of the GMM objective function. The latter could in principle be solved by a grid search, but this only exacerbates the curse of dimensionality further. Estimating more than a few goods is therefore generally infeasible. These issues, however, are fully circumvented within our framework when a simple assumption on the structure of the demand system is met.

Assumption 1. Preferences are weakly separable into the B and D baskets. ◁

Under Assumption 1, an individual good occurs in either the B basket or the D basket, but not in both. Since the D basket captures necessity needs and B basket luxury needs, it follows that preferences are also weakly separable into necessities and luxuries. The assumption is therefore easily implemented empirically by allocating luxuries to B and necessities to D.

The immediate consequence of Assumption 1 is that across and within expenditure shares become observable in the data. Summing the total expenditure shares w_j (which are always observable) over goods in D gives the across share w_D. Within shares are then obtained as $w_j^D = w_j / w_D$. The same applies for the B basket. This knowledge is enough to compute price indices P_{Bt} and P_{Dt} for the B and D baskets using Equation (14) or (15). The only additional information needed to compute the generalized superlative indices are, per Proposition 1, the two parameters ε and γ. Using Equations (3) and (5), we may write the period-t expenditure share on the D good as

$$w_{Dt} = \tilde{\nu} \left(\frac{P_{Bt}}{e_t} \right)^{\varepsilon} \left(\frac{P_{Dt}}{P_{Bt}} \right)^{\gamma}, \tag{19}$$

where $\tilde{\nu} \equiv \nu B(\boldsymbol{p}_s)^{\varepsilon-\gamma} D(\boldsymbol{p}_s)^{\gamma}$ is a scale parameter. Since w_{Dt}, e_t, P_{Bt} and P_{Dt} are all known, estimating ε and γ from (19) is easily carried out using either linear (by taking logs of (19)) or nonlinear estimation methods. We summarize this empirical approach by the following proposition:

Proposition 2 (Tractable demand system estimation)**.** *Under Assumption 1, across and within group expenditure shares are observable in the data and computing the generalized superlative indices (16) only requires estimation of two parameters, ε and γ, from the single expenditure share equation (19).*

At first sight, Assumption 1 may seem to be at odds with the nonhomothetic generalization of

[6] As an illustration, suppose we have n goods and parameterize $B(\boldsymbol{p})$ and $D(\boldsymbol{p})$ by the linearly homogeneous translog expenditure function of Christensen, Jorgenson and Lau (1975), for which the Törnqvist index is an exact cost-of-living index (Diewert, 1976). The PIGL demand system considered here then requires the estimation of $n(n+1)+3$ independent parameters.

the superlative indices: Corollary 1 reduces to the standard homothetic case when $B(\boldsymbol{p}) = D(\boldsymbol{p})$, which Assumption 1 excludes by construction. However, the weak separability assumption still nests homothetic preferences. In particular, as $\varepsilon \to 0$ and $\gamma \to 0$, we obtain Cobb-Douglas preferences with $V(e, \boldsymbol{p}) = \ln \left[\frac{e}{B(\boldsymbol{p})^{1-\nu} D(\boldsymbol{p})^{\nu}}\right]$ and a corresponding price index $P_t = P_{Bt}^{1-\nu} P_{Dt}^{\nu}$, where $\nu = w_D$ is the homothetic and time-invariant expenditure share on D. Thus, if preferences truly are homothetic, we still expect Proposition 2 to yield a homothetic price index. This index is approximately equal to the corresponding superlative index when $B(\boldsymbol{p}) = D(\boldsymbol{p})$, by virtue of superlative indices being second-order approximations of any other homothetic index. This highlights that using the generalized superlative indices under Assumption 1 should at least (approximately) be weakly better than using the standard homothetic indices. Since nonhomothetic preferences is the empirically relevant case, we do not expect the "approximately" part to matter much, and the empirical application below confirms this. For cases where it nevertheless might be of importance, it turns out that a special case exists where Assumption 1 exactly nests the corresponding homothetic index when $B(\boldsymbol{p}) = D(\boldsymbol{p})$: the Törnqvist index.

Proposition 3 (Homothetic Törnqvist index under weak separability)**.** *Suppose that preferences are of the homothetic Cobb-Douglas form,* $V(e, \boldsymbol{p}) = \ln \left[\frac{e}{B(\boldsymbol{p})^{1-\nu} D(\boldsymbol{p})^{\nu}}\right]$*, and that* $B(\boldsymbol{p})$ *and* $D(\boldsymbol{p})$ *are such that their corresponding price indices* P_{Bt} *and* P_{Dt} *are Törnqvist indices. The cost-of-living index under Assumption 1 is then the standard Törnqvist index:*

$$P_t = \prod_{j \in J} \left(\frac{p_{jt}}{p_{js}}\right)^{\delta_{jt}}, \qquad \textit{where} \qquad \delta_{jt} = \frac{w_{js} + w_{jt}}{2}.$$

Proof. In Section A.3. □

4 Data and Estimation

We implement the tractable demand system estimation described in the previous section using consumption and price data from two sources. Household consumption is taken from the interview component of the Consumer Expenditure Survey (CEX) and price data are taken from the product-level Consumer Price Index (CPI) series for all urban consumers. Both are provided by the US Bureau of Labor Statistics (BLS). The CEX interview survey is a quarterly rotating panel of households who are representative of the US population. New households are sampled every month and each household is tracked for up to four consecutive quarters. The survey covers around 95 percent of total household consumption and contains additional information on annual income and other background characteristics. The survey has been continuously conducted since 1980, but to ensure consistency across waves and due to the availability of the CPI sub-indices, we focus on the years 1995 to 2020.

As is standard in the literature, we select a sample of respondents between the ages of 25 and 65 who report strictly positive income. To avoid issues with seasonality, we aggregate expenditures to annual levels and, consequently, drop households that do not respond to all four quarterly interviews. To account for differences in household size, we also divide household income and expenditures by the number of adult equivalents in the household using the equivalence scale of the US Census Bureau (see Fox and Burns, 2021). The final dataset on expenditures consists of approximately 3,000 households per year.

We aggregate consumption expenditures into a rather coarse set of consumption goods categories as this allows us to compare the empirical approach in Proposition 2 with a full demand system estimation. All in all, we consider twenty-one categories of nondurable goods using the hierarchical groupings defined by the BLS. We broadly follow Hobijn and Lagakos (2005) and construct prices for these categories by matching them with individual CPI series. Table B.1 in Section B lists the CEX categories and shows their mapping to the CPI item codes.

4.1 Classification of Goods Into Luxuries and Necessities

In order to utilize the tractable demand system estimation in Proposition 2, we impose Assumption 1 by allocating luxuries to B and necessities to D. The classification into B and D is implemented by investigating slopes of the budget share Engel curves: if the Engel curve of a good decreases as expenditures increase, it is a necessity. Conversely, a good is a luxury if its Engel curve increases with increasing expenditures. We split households into expenditure deciles and, for each good j, run a household-level regression of the expenditure share w_{jh} on the expenditure decile d_h of household h:

$$w_{jh} = \alpha_j + \beta_j \mathsf{d}_h + \epsilon_{jh}.$$

If $\beta_j > 0$, we allocate the good to the B basket, otherwise to the D basket. Figure 1 shows the Engel curves by expenditure decile together with the resulting classification from the regressions and Table B.2 in Section B lists the β_j coefficient estimates. The resulting classification is intuitive and all estimates are significantly different from zero. For comparable product groups, our necessity/luxury split is highly similar to those constructed in similar analyses using CEX data (see for instance Wachter and Yogo, 2010, and Orchard, 2021), thus suggesting that this simple classification regression works well on our coarse set of goods.

4.2 Tractable Demand System Estimation

We estimate the preference parameters ε and γ from Equation (19) using a nonlinear GMM estimator. In doing so, we make explicit corrections for two potential issues.

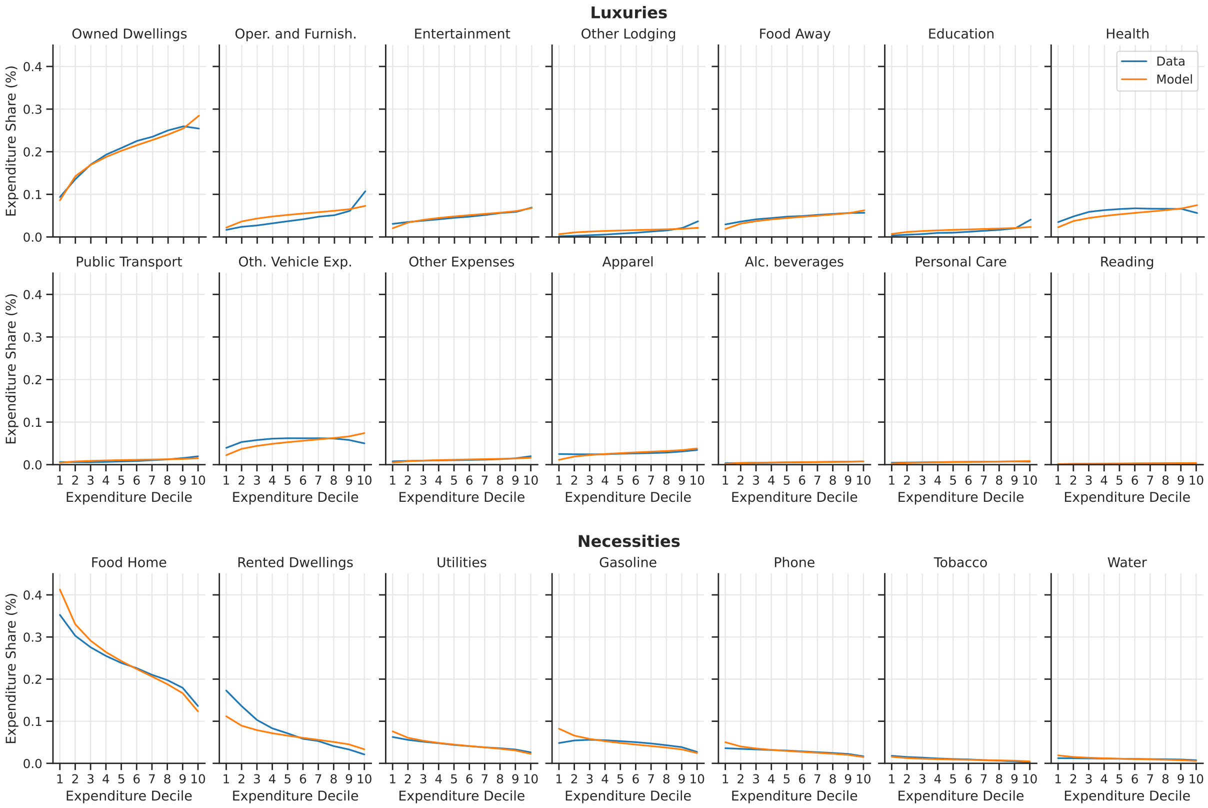


FIGURE 1. Empirical and model implied Engel curves.

Notes. The figure shows the empirical and model implied expenditure shares by expenditure group and expenditure decile averaged over all years. The model implied expenditure shares correspond to those under the assumption of weak separability and the Sato-Vartia specification. They are calculated by first taking the model implied expenditure shares on the B and D goods and then use the empirical expenditures shares of all households to obtain shares within these groups.

First, it is well known that infrequently bought items, like clothing and transportation, create a measurement error in the observed level of expenditures. Although we alleviate much of this concern by excluding durable goods and by aggregating expenditures to an annual level, we follow the literature (Blundell, Pashardes and Weber, 1993; Banks, Blundell and Lewbel, 1997) and control for this endogeneity bias by instrumenting expenditures on household income.

Second, an indirect utility specification like the PIGL requires additional attention with respect to the regularity conditions for utility maximization. Specifically, we need to certify that the parameter estimates yield a symmetric and negative semidefinite Slutsky matrix. Under the indirect utility function (4) and Assumption 1, it follows from Boppart (2014, Lemma 1) that a necessary and sufficient condition for household h to satisfy the Slutsky restrictions in period t is $\tilde{\nu}\left(\frac{P_{Bt}}{e_{ht}}\right)^{\varepsilon}\left(\frac{P_{Dt}}{P_{Bt}}\right)^{\gamma} \leq (1-\gamma)/(1-\varepsilon)$. We enforce this constraint by augmenting the GMM estimation with a classic penalty method. Consequently, the reported parameter estimates below satisfy the Slutsky restrictions for all observations in the sample.

To gauge the sensitivity to different choices of underlying superlative price indices, we estimate ε and γ for six different choices of P_{Bt} and P_{Dt}. These choices correspond to the indices listed in Section 2.4: the Sato-Vartia, the Törnqvist, the Walsh (geometric and arithmetic), the Theil, and the Fisher. This robustness check is instructive since there is generally no guarantee that superlative indices are numerically similar, despite being second-order approximations of each other (see for instance Hill, 2006). For the estimation exercise, we compute these indices on a monthly frequency and, since household expenditures are annualized, construct household-specific annualized price levels by averaging over the months each household is in the sample.

The estimated parameters for the six cases are reported in Table 1. All parameters are statistically different from zero at conventional significance levels and the fact that $\varepsilon > 0$ and $\widetilde{\nu} > 0$ directly rejects homotheticity. Reassuringly, the choice of price indices for the B and D baskets turns out to be completely inconsequential as all specifications yield close to identical estimates. Moreover, Alder, Boppart and Müller (forthcoming, Proposition 3) show that a sufficient condition for expenditure shares to remain globally nonnegative is $0 < \gamma \leq \varepsilon < 1$. This condition is also met in our estimation, though we do not impose the constraint explicitly. Other preference specifications, like the almost ideal demand system, typically violate expenditure share nonnegativity for sufficiently large expenditures levels.

To get an idea of how well the estimated model matches the data, we compute budget-share Engel curves using the parameter estimates for the Sato-Vartia specification and plot these against their empirical counterparts in Figure 1. We construct the model expenditure shares as the product of model-implied across- and within-expenditure shares w_C and w_j^C for goods in $C \in \{B, D\}$. The former is computed from Equation (19) at the representative level of expenditures within each expenditure decile. Since $B(\boldsymbol{p})$ and $D(\boldsymbol{p})$ are homothetic, the latter is given empirically by the average within-shares $\overline{w}_j^B$, $\overline{w}_j^D$ across *all* households. Figure 1 shows that the model fits the data quite well. In particular, the model does a much better job at matching the empirical Engel curves than the constant Engel curves resulting from homothetic preferences would. This underscores that the assumption of weak separability between the B and D baskets is not a strong restriction in our sample.

5 Results

The estimation in the previous section suggests that the results are insensitive to the choice of generalized superlative index.[7] For the remainder of the paper, we therefore focus on the nonhomothetic generalization of the Sato-Vartia index (henceforth G-SV). Since the Sato-Vartia index is the corresponding Konüs index to the canonical CES expenditure function (see Sato, 1976), this choice implies that we are investigating a generalization of homothetic CES preferences.

[7] Figures B.7 and B.8 in Section B show the main results in this section for all choices considered in Section 4.2, which confirms that this is indeed the case.

Table 1. Estimated parameters under weak separability.

	Sato-Vartia	Törnqvist	Geom. Walsh	Theil	Fisher	Arith. Walsh
ε	0.677	0.677	0.677	0.677	0.677	0.677
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
γ	0.211	0.211	0.211	0.211	0.211	0.211
	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
$\widetilde{\nu}$	327.271	327.437	327.173	327.273	324.273	324.600
	(13.358)	(13.365)	(13.354)	(13.358)	(13.217)	(13.233)
N	74,372	74,372	74,372	74,372	74,372	74,372
RMSE	0.1487	0.1486	0.1487	0.1487	0.1486	0.1486

Notes. Robust standard errors in parentheses. "RMSE" refers to the root mean squared error of the expenditure share on the D good: $\sqrt{\sum_h (w_{Dh} - \widehat{w}_{Dh})^2 / N}$.

The prominence of CES preferences within macroeconomics and international trade therefore makes this a case of particular interest. The main reason for selecting the G-SV, however, is that a parameterization of $B(\boldsymbol{p})$ and $D(\boldsymbol{p})$ as CES aggregates contains many fewer parameters than the parameterizations that induce, for instance, the Fisher or the Törnqvist indices. This allows us to compare the G-SV index to a relatively parsimonious estimation of the full PIGL demand system.

Figure 2 shows the evolution of the G-SV price index from 1995 to 2020. We set the base period to 1995 and, in contrast to Section 4.2, use annual indices for P_{Bt} and P_{Dt}.[8] Even though the generalized superlative price indices allow characterizations of the entire distribution of indices, here we focus on expenditure deciles for ease of exposition. Figures B.3 and B.4 in Section B show the full price index and inflation distributions.

Figure 2 corroborates two findings from the literature: inflation rates vary across households and poorer households experienced a larger increase in the cost of living than richer households over the last quarter century. It is noteworthy, though, that the cumulative differences are small. Table 2 makes this point clear: the mean annual inflation rate of the poorest ten percent is only 0.06 percentage points higher than that of the richest ten percent over the full sample period. The changes in the cost of living over the 26 years under study therefore do not diverge dramatically between groups.

The small differences in the cost of living by 2020 is striking given the substantial heterogeneity observed in subperiods of the sample. For instance, if we zoom in on the years 2004 to 2015, the annual change in the cost of living for the poorest ten percent are on average 0.37 percentage

[8] Figure B.1 in Section B shows that the choice of base period does not affect our results.

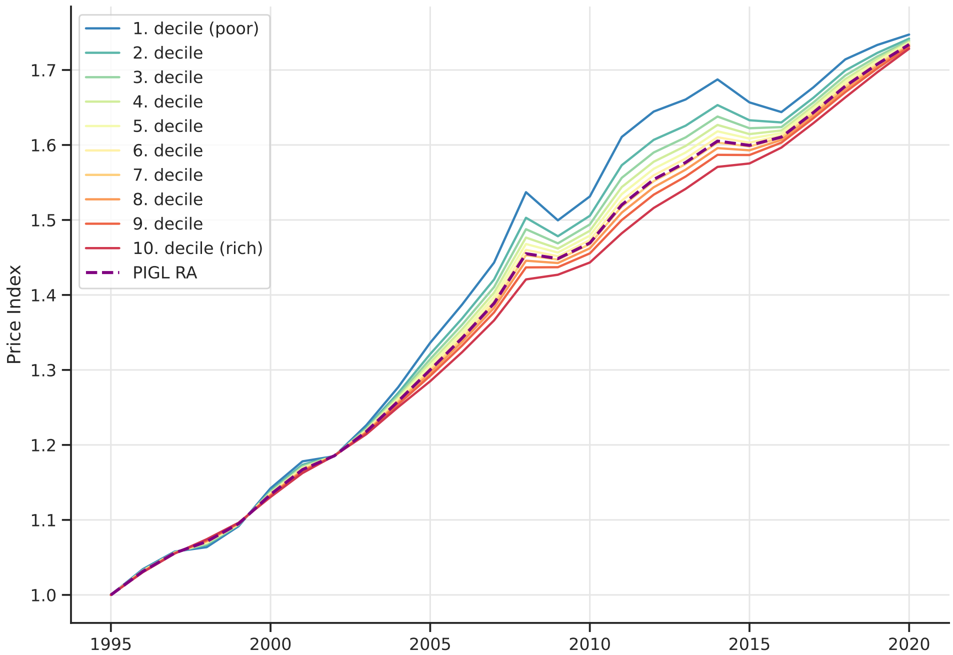


Figure 2. G-SV price index under weak separability by expenditure decile.

Notes. The price index for each expenditure decile is calculated as the price index of the PIGL representative agent over households within each respective decile. "PIGL RA" stands for the PIGL representative agent over *all* households.

Table 2. Inflation heterogeneity in numbers.

	2004 - 2015				1996 - 2020			
Exp. dec.	mean	Δmean	std.	rel. std.	mean	Δmean	std.	rel. std.
1	2.57	0.37	2.71	2.38	2.28	0.06	2.14	2.51
2	2.46	0.26	2.25	1.98	2.26	0.04	1.76	2.06
3	2.41	0.21	2.05	1.79	2.25	0.03	1.59	1.86
4	2.38	0.18	1.89	1.66	2.24	0.03	1.46	1.71
5	2.35	0.15	1.78	1.56	2.24	0.02	1.36	1.60
6	2.33	0.13	1.67	1.46	2.24	0.02	1.28	1.50
7	2.30	0.10	1.57	1.38	2.23	0.02	1.20	1.40
8	2.28	0.08	1.47	1.29	2.23	0.01	1.12	1.31
9	2.25	0.05	1.35	1.18	2.22	0.01	1.02	1.20
10	2.20	0.00	1.14	1.00	2.22	0.00	0.85	1.00

Notes. Arithmetic mean and standard deviation of annual inflation over the respective years. "Δ mean" denotes the difference in the mean annual inflation to the tenth expenditure decile. Rel. std. denotes the relative standard deviation to the standard deviation of the tenth expenditure decile.

points higher than the change for the richest ten percent. Jaravel (2019) and Argente and Lee (2021) focus on the same years, the former using a CEX-CPI dataset similar to ours and the latter using scanner data for the retail sector, and both find results close to ours. To put this difference in perspective, the Boskin Commission Report estimated the total bias in the aggregate US CPI to be 1.1 percentage points (Boskin *et al.*, 1996). Of these, substitution biases alone account for 0.4 percentage points. The difference we find here is therefore substantial when compared to previously estimated biases in aggregate price indices.

That the differences in the change in the cost of living varies across subperiods is also visible from the implied annual inflation rates. Figure 3 plots these rates, which highlights that there are periods where poorer households experience substantially higher *or* lower inflation. In several years, the range of the inflation rate exceeds 2 percentage points. There is another key finding that stands out from Figure 3 however: the inflation rate of the poor is much more volatile than that of the rich. More precisely, Table 2 shows that the standard deviation of inflation has been 2.14 and 0.85 for the poorest and richest, respectively, and the poor have thus experienced a 2.5 times *more volatile* inflation rate than the rich.

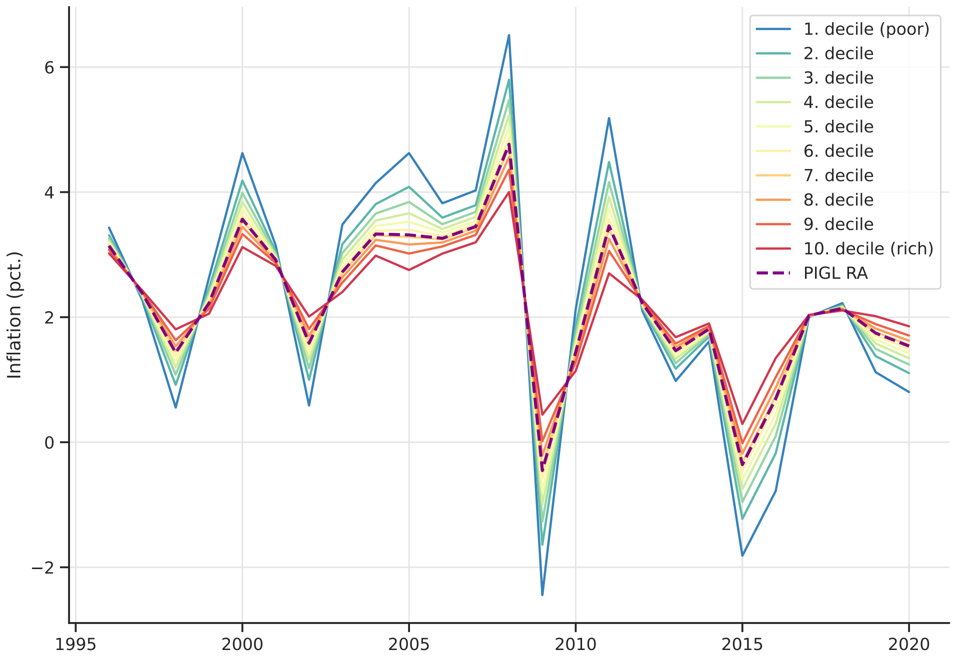


Figure 3. G-SV inflation under weak separability by expenditure decile.

Notes. Inflation for each expenditure decile is calculated as the first difference in the price index of the PIGL representative agent over households within each respective decile. “PIGL RA” stands for the PIGL representative agent over *all* households.

In sum, despite the fact that the overall change in the cost of living has not diverged dramatically between groups, there is a considerable difference in the volatility of inflation rates. This difference in volatility of inflation rates warrants that it is important in future work to understand the economic implications it may have for instance in terms of welfare on individual household levels and economic policies that aim at stabilizing inflation rates.

5.1 Changes in Living Standards

Most empirical evidence on inflation inequality, including the results above, compare the change in inflation rates between two or more groups whose real expenditures stay fixed at some base-period level. In other words, living standards remain constant. Yet, it is a well-known that households experience substantial changes in expenditures over the life-cycle. It is therefore not clear whether the evidence on inflation inequality is of much relevance to the inflation experience of an average household.

Our framework distinguishes itself from previous attempts to study inflation inequality in that it is straightforward to compute inflation rates for households even when real expenditures change over time. The group-specific price index approach poses some challenges in allowing for this due to the discontinuous change underlying preferences across the expenditure distribution. Since our framework expresses the price index as a continuous function of expenditures we do not face that challenge.

Consider an individual whose expenditures, $e_t = c(u_t, \boldsymbol{p}_t)$, change over time. The change in expenditures can arise because prices, $\boldsymbol{p}_t$, change or living standards, u_t, change. The relative change in expenditures between period s and period T may be written as

$$\frac{c(u_T, \boldsymbol{p}_T)}{c(u_s, \boldsymbol{p}_s)} = \left[\prod_{t=s+1}^{T} Q(u_{t-1}, u_t, \boldsymbol{p}_t)\right]\left[\prod_{t=s+1}^{T} \frac{P(u_{t-1}, \boldsymbol{p}_t, \boldsymbol{p}_s)}{P(u_{t-1}, \boldsymbol{p}_{t-1}, \boldsymbol{p}_s)}\right], \tag{20}$$

which is made up of two components: (i) a quantity cost, $Q(u_i, u_k, \boldsymbol{p}_t) \equiv c(u_k, \boldsymbol{p}_t)/c(u_i, \boldsymbol{p}_t)$, that tells how much a household will have to pay to go from living standard u_i to u_k at prices $\boldsymbol{p}_t$, and (ii) a per-period price cost, $P(u_{t-1}, \boldsymbol{p}_t, \boldsymbol{p}_s)/P(u_{t-1}, \boldsymbol{p}_{t-1}, \boldsymbol{p}_s)$, that tells how much the cost of living has changed period by period.[9] This latter part is just a chained price index where the base period is always $t-1$.[10]

[9] The quantity-cost in Equation (20) measures the cost of obtaining living standard u_t relative to maintaining living standard u_{t-1} in present-period prices. That is, it measures the relative cost of obtaining living standard u_t compared to u_{t-1} in prices when the change takes place. A different definition, the quantity-cost in past-period prices defined as $Q_{t-1}(u_{t-1}, u_t)$, gives a similar expression as the one in Equation (20) but is less intuitive.

[10] When living standards do not change throughout period s to period T, we have that

$$\frac{c(u_T, \boldsymbol{p}_T)}{c(u_s, \boldsymbol{p}_s)} = \prod_{t=s+1}^{T} \frac{P(u_s, \boldsymbol{p}_t, \boldsymbol{p}_s)}{P(u_s, \boldsymbol{p}_{t-1}, \boldsymbol{p}_s)} = \frac{P_T(u_s, \boldsymbol{p}_T, \boldsymbol{p}_s)}{P(u_s, \boldsymbol{p}_s, \boldsymbol{p}_s)} = P(u_s, \boldsymbol{p}_T, \boldsymbol{p}_s),$$

which is simply the PIGL price index at time T with base period s.

From Equation (20), we can also compute the inflation cost from the per-period inflation rate $1 + \pi_t(u_{t-1}) \equiv P(u_{t-1}, \boldsymbol{p}_t, \boldsymbol{p}_s)/P(u_{t-1}, \boldsymbol{p}_{t-1}, \boldsymbol{p}_s)$. It forms the basis of the following analysis where we decompose the change in expenditures into quantity and price costs when allowing for changes in expenditures over time.

5.1.1 Life-Cycle Model

Changes in household expenditures over time is a well-documented empirical fact. Fernández-Villaverde and Krueger (2007) for example find that US household expenditures follow a deterministic, hump-shaped pattern over the life-cycle.

Based on our data, we estimate the following deterministic life-cycle model:

$$\ln e_{it} = \alpha_i + \beta_t + \widetilde{\gamma}_{it}, \tag{21}$$

where i denotes age, t time period, α_i are age-fixed effects, β_t year-month-fixed effects, and $\widetilde{\gamma}_{it}$ is the residual. The results of the estimated model in Equation (21) are shown in Figure B.5 of Section B. Our results for the deterministic life-cycle model is in line with what Fernández-Villaverde and Krueger (2007) find.

We use the estimated deterministic life-cycle model to simulate a 26-year nominal expenditure path for 30 different types. Each type has a distinct age-by-initial-expenditure-level combination, where the age is such that in 1995 the type is either young (25 years old), middle-aged (45 years old) or old (65 years old), and the initial expenditure level is equal to the level of one of the expenditure deciles in the CEX data as in the main results. For each type we then compute the price index, inflation series and standard deviation of inflation according to Equation (20).

In panel (a) of Figure 4 we show the simulated path of nominal expenditures for six types: young-poor, young-rich, middle-aged-poor, middle-aged-rich, old-poor and old-rich. For each age type, poor refers to an individual whose 1995 expenditure level was equal to that of expenditure decile 1 and rich refers to an individual whose 1995 expenditure level was equal to that of expenditure decile 10. The figure shows that nominal expenditures varies greatly across age groups and by initial expenditure levels.

Panel (b) of Figure 4 shows the corresponding inflation for each of the six types whose nominal expenditure levels are those in panel (a). The figure shows that different deterministic life-cycle patterns give rise to different inflation levels for individuals who start out with the same initial standard of living. For example, the young-poor experience a somewhat more dampened volatility of inflation over time compared to the old-poor. This is because the young experience changes in expenditures that exceed the change in prices and therefore allows them to increase their living standards. As our main results shows, households with higher living standards have experienced a much lower volatility of inflation and the young-poor transit towards this over time. The broad picture, however, is that the deterministic life-cycle components are far from being able to mitigate the results from the previous section: columns 3–6 in Table 3 show that the relative standard deviation of inflation remains 2.41 times higher for the young-poor compared to the young-rich, 2.46 for the middle-aged-poor and 2.55 times for the old-poor. Although not shown, the mean and standard deviation of inflation are also largely unchanged.

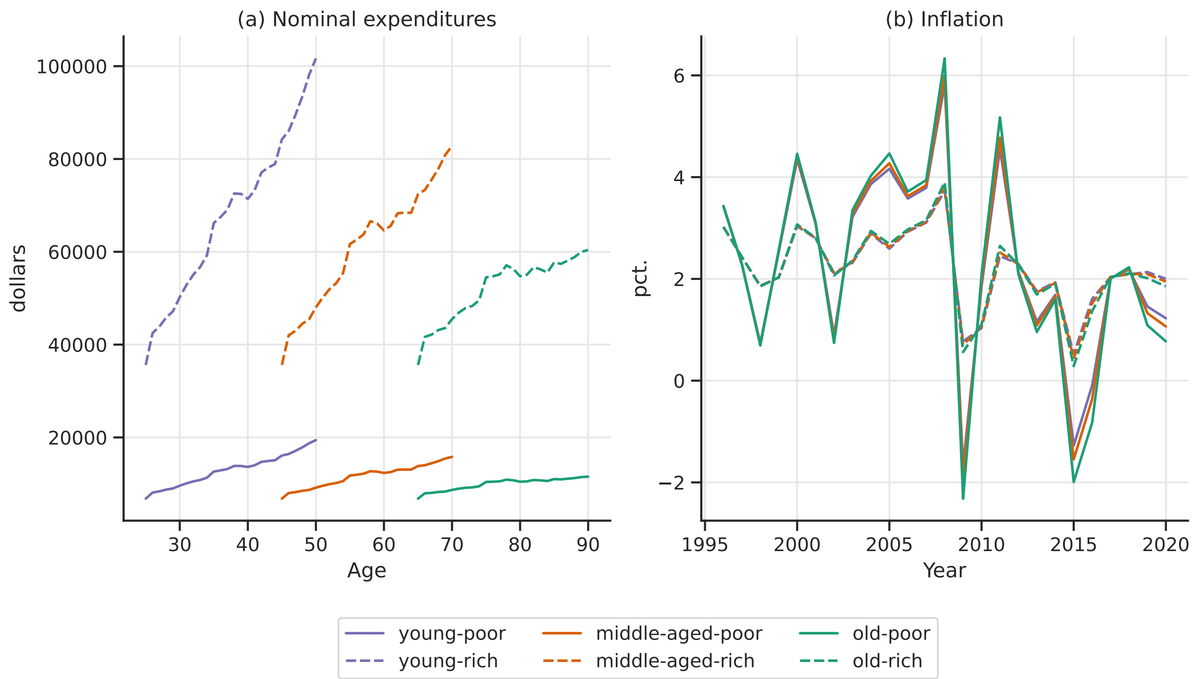


FIGURE 4. Expenditures and inflation in the deterministic life-cycle model.

Notes. Panel (a) shows the nominal expenditure levels for six types simulated from the life-cycle model in Equation (21). Panel (b) shows the corresponding inflation levels for each type based on Equation (20). Young refers to an individual who was 25 years in 1995, middle-aged to an individual who was 45 years and old to an individual who was 65 years. Poor refers to an individual whose 1995 expenditure level was equal to that of expenditure decile 1 in the CEX data. Rich refers to an individual whose 1995 expenditure level was equal to that of expenditure decile 10 in the CEX data.

In addition to the deterministic evolution in the life-cycle expenditures, other empirical studies such as Blundell, Pistaferri and Preston (2008) also find that US households experience significant stochastic shocks to expenditures.

The stochastic component comes from the unexplained change in log expenditures between two ages which is given by

$$\gamma_{it} \equiv \Delta\widetilde{\gamma}_{it} = \ln e_{it} - \ln e_{i-1,t-1} - (\alpha_i - \alpha_{i-1}) - (\beta_t - \beta_{t-1}), \tag{22}$$

Table 3. Relative standard deviation.

Exp. dec.	Constant utility	Life-cycle young	Life-cycle middle-aged	Life-cycle old	Stochastic life-cycle young	Stochastic life-cycle middle-aged	Stochastic life-cycle old
1	2.51	2.41	2.46	2.55	2.50	2.55	2.63
2	2.06	1.97	2.01	2.08	2.05	2.08	2.13
3	1.86	1.78	1.81	1.87	1.84	1.88	1.91
4	1.71	1.64	1.67	1.72	1.70	1.72	1.76
5	1.60	1.54	1.56	1.60	1.59	1.61	1.64
6	1.50	1.44	1.46	1.50	1.47	1.50	1.53
7	1.40	1.36	1.37	1.40	1.39	1.40	1.43
8	1.31	1.27	1.29	1.31	1.30	1.31	1.33
9	1.20	1.17	1.18	1.19	1.19	1.19	1.21
10	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Notes. The table reports the relative arithmetic standard deviations of inflation from 1995 to 2020 between each type whose 1995 expenditure level was equal to expenditure decile $\mathsf{d} \in \{1, \dots, 10\}$ and that of those whose 1995 expenditure level was equal to expenditure decile 10 in the CEX. “Constant utility” refers to the baseline results also presented in Table 2 where real expenditures are fixed. “Life-cycle” refers to the deterministic life-cycle model in Equation (21) where the determinsitic component is shut off (i.e. equal to 0). “Stochastic life-cycle” refers to the life cycle model in Equation (21) with the stochastic component activated. Young refers to an individual who was 25 years in 1995, middle-aged to an individual who was 45 years and old to an individual who was 65 years.

and as in Blundell, Pistaferri and Preston (2008), we assume that γ is independently and normally distributed and we use their estimated variance of γ. In order to preserve the expected life-cycle expenditures from the estimated model, we impose the restriction that $\exp\{\gamma\}$ has a mean of one.

We once again consider the same 30 types as in the deterministic life-cycle model. For each type, we draw 26 γ-shocks and simulate a nominal expenditure path according to the stochastic life-cycle model. We repeat the simulation 10,000 times for each type and for each expenditure path we compute the corresponding price index, inflation series and standard deviation of inflation. We then average over the 10,000 simulations to get the expected inflation for each type.

Columns 6–8 in Table 3 show the relative standard deviation of expected inflation for each type in the stochastic life-cycle model. We see that the stochastic component makes the poorer types worse off in terms of expected inflation volatility. The relative inflation volatility of the young-poor compared to the young-rich, for example, goes from 2.41 in the deterministic life-cycle model to 2.5 in the stochastic model. The broad picture, however, still remains: even when controlling for deterministic and stochastic changes in expenditures, the expected inflation volatility is around 2.5 times larger for the poorest ten percent compared to the richest ten percent.

These findings show that, although households' real expenditures change over time, the average change over a time period of 26 years from ageing and chance is not enough to alter the conclusions drawn in regards to inflation inequality between households whose real expenditures stay fixed.[11] Instead, this suggests that initial expenditure levels are an important determinant for households' expected inflation rates.

6 What Drives the Inflation Differences?

The generalized superlative price index (16) lends itself to a simple decomposition of inflation into its components. Taking first differences of the log of the price index allows us to compute the contribution of inflation coming from each expenditure category j. Figure 5 plots this decomposition for the rich and poor's inflation. The figure shows the primary expenditure categories that have driven inflation since 1996.

Panel (a) in Figure 5 shows that the two primary expenditure categories that drive inflation for the poor are "food at home" and "gas and utilities" (energy). In contrast, panel (b) shows that these categories play a minor role in driving inflation for the rich. Moreover, the key point in regards to what drives inflation for the rich is that no expenditure category is a major driver. In panel (b) we plotted the contributions to the rich's inflation from "food at home" and "gas and utilities" to illustrate that these play a minor role. Additionally, we also show how the most important driver for the rich's inflation, "owned dwellings", and the fourth most important driver, "other vehicle expenditures", contribute. While "owned dwelling" indeed drives a considerable amount of total inflation of the rich, it is still minor and "other vehicle expenditures" is almost invisible in some periods. Figure B.6 in Section B shows the inflation contribution from all expenditure categories to the rich and poor's inflation, respectively.

6.1 Decomposition Into Reference Basket and Product Substitution

In order to shed more light on the inflation heterogeneity between poor and rich households, we decompose the price index into one component that reflects pure price changes of a base-period reference basket and another component that reflects product substitution away from this basket as prices change. The former is simply the Laspeyres price index $P_{Lt} = \sum_j w_{js}(p_{jt}/p_{js})$, which

[11] The results do not reject that other factors such as education could induce changes in expenditures that are strong enough to make a difference.

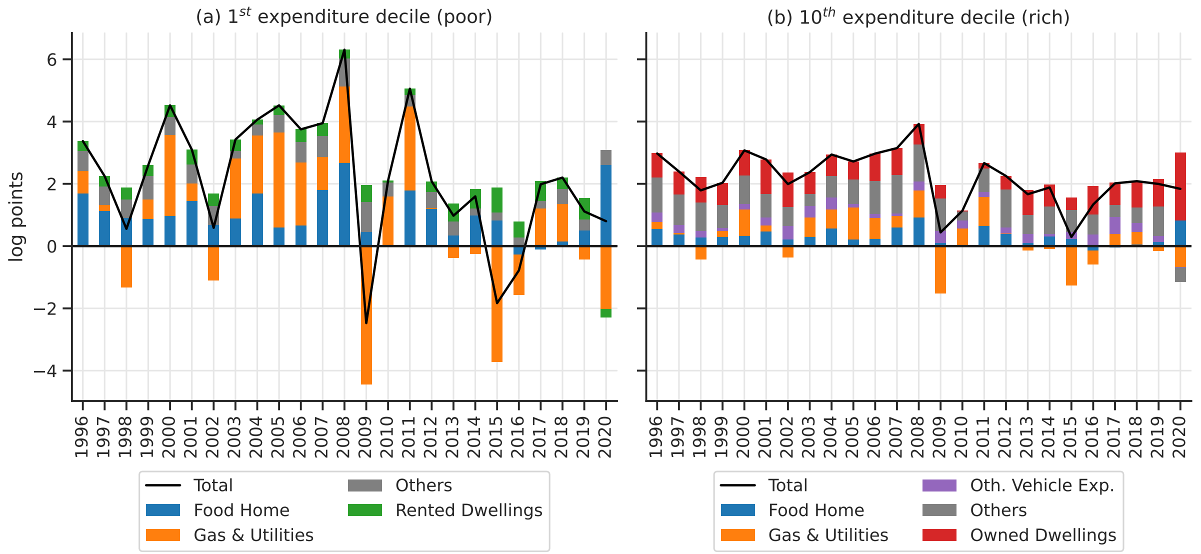


FIGURE 5. Inflation decomposition by expenditure categories.

can be written on a weighted geometric-mean form as[12]

$$P_{Lt} = \prod_{j \in J} \left(\frac{p_{jt}}{p_{js}} \right)^{\delta^L_{jt}}, \qquad \text{where} \qquad \delta^L_{jt} \equiv \frac{w_{js}\, L\left(\frac{p_{jt}}{p_{js}},\ P_{Lt}\right)}{\sum_{i \in J} w_{is}\, L\left(\frac{p_{jt}}{p_{js}},\ P_{Lt}\right)}. \tag{23}$$

The latter component can then be backed out as the residual between the computed price index and the corresponding Laspeyres index. Taking logs of the generalized superlative index (16) and adding and subtracting the log of the Laspeyres index (23), we obtain the decomposition

$$\ln P_t = \underbrace{\sum_{j \in J} \delta^L_{jt} \ln \left(\frac{p_{jt}}{p_{js}} \right)}_{\text{Laspeyres price index}} + \underbrace{\sum_{j \in J} \left(\chi_{jt} - \delta^L_{jt} \right) \ln \left(\frac{p_{jt}}{p_{js}} \right)}_{\text{Product substitution}}. \tag{24}$$

Figure 6 shows the decomposition of the G-SV price index for the first and the tenth expenditure decile. It clearly highlights that the biggest share of the increase in the cost of living is driven by the price changes of the reference baskets and substitution effects only marginally reduce the G-SV price index. Interestingly, however, the substitution effect among the households in the first expenditure decile is considerably bigger than the substitution among the households in the tenth expenditure decile.

Focusing on the difference between the two deciles in Figure 7 shows that the composition of their reference baskets plays an important role but that the difference in product substitution

[12] See Section A.2 or Balk (2004).

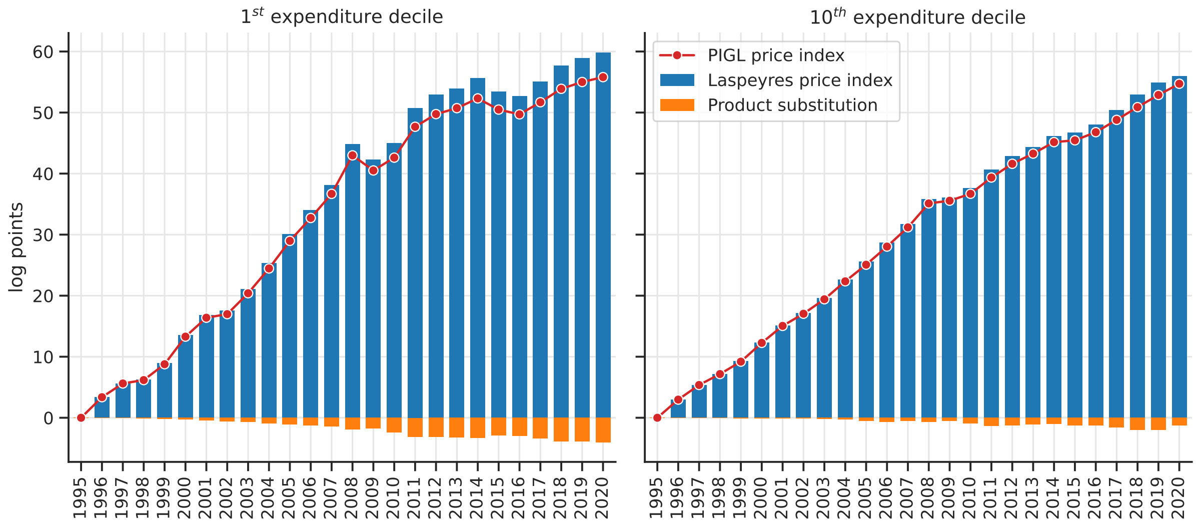


Figure 6. Decomposition of the G-SV price index into the Laspeyres price index and product substitution.

Notes. Decomposition is performed as shown in Equation (24). The G-SV price index for each expenditure decile denotes the price index of the representative agent within each respective decile.

becomes relatively more important.

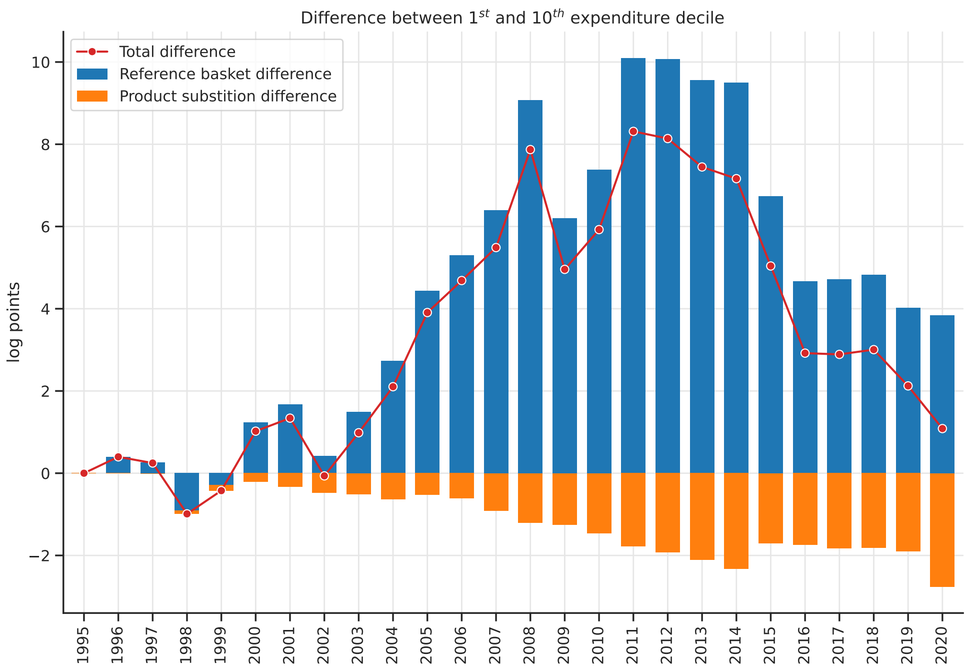


Figure 7. Difference of the decomposed G-SV price index between the first and tenth expenditure decile.

Notes. The figure shows the decomposition of the first expenditure decile minus the corresponding numbers for the tenth expenditure decile in Figure 6.

This section shows that in order to uncover the fundamental drivers of inflation heterogeneity across households it is primarily necessary to understand inflation heterogeneity across products. Therefore, we deem it important for future research to investigate this further. In particular, whether it is a coincidence that the highly volatile product groups are purchased relatively more by poor households, or if there is a deeper and more structural link. In terms of policy, the results show that stabilizing prices of selected product groups, and not just an aggregate price index, are important when inequality is taken into account. Specifically, the results show that the highly volatile inflation of "food at home" and "gas and utilities" affects poorer households disproportionately more.

7 Comparison with Full Demand System Estimation

To assess the validity of the weak separability assumption and the robustness of our results, we perform a full demand system estimation. Full demand system estimation can be carried out via GMM by using the expenditure share equations (8) and the parameterization for $B(\boldsymbol{p})$ and $D(\boldsymbol{p})$. The parameterization of the Sato-Vartia specification used in our empirical analysis is given by

$$B(\boldsymbol{p}) = \left(\sum_{j \in J} \omega_j p_j^{1-\sigma} \right)^{\frac{1}{1-\sigma}} \qquad \text{and} \qquad D(\boldsymbol{p}) = \left(\sum_{j \in J} \theta_j p_j^{1-\varphi} \right)^{\frac{1}{1-\varphi}}, \tag{25}$$

where $\sigma, \varphi > 0$ and the taste parameters satisfy $\sum_{j \in J} \omega_j = \sum_{j \in J} \theta_j = 1$ with $\omega_j, \theta_j \geq 0$ for all $j \in J$.[13]

Since the GMM estimation under the full demand system specification might exhibit several local minima, it generally must not hold that a particular local solution of the GMM estimation is indeed also the global minimum. Therefore, the following results should be interpreted with this in mind. However, by choosing an appropriate guess for the preference parameters in the full demand system that corresponds to the preferences under weak separability, we obtain a local solution for the full demand system which is in the vicinity of the weakly separable case. In particular, we set the values for ε, γ and ν to the values in Table 1 and equally distribute ω_j and θ_j among the goods classified as B and D, respectively.[14]

Table 4 and Figure 8 show the estimates for the full set of preference parameters in the full demand system.[15] We can note that ε is practically unchanged, but γ substantially higher. ν is

[13] This parameterization is also considered in Alder, Boppart and Müller (forthcoming).

[14] In principle we could perform a grid search over the whole set of feasible parameters. However, depending on the initial guess, convergence of the GMM estimator is slow and can sometimes take more than 24 hours. Thus, we deemed a reliable grid search as infeasible.

[15] While point estimates are still consistent when some parameters lie on the boundary of the parameter space,

not important for determining the price index, but still lies fairly close to the estimates under weak separability. Turning to σ and φ, we can note that they both lie below one, which indicates that goods within the B and D basket are complements. Figure 8 shows the point estimates of the complete set of taste parameters ω_j and θ_j for the B and the D basket respectively. Each category has a blue and an orange bar for ω_j and θ_j, but with the minor exception of “tobacco”, every single good only has one strictly positive taste parameter. Thus, the estimated parameters from the full demand system and the implied classification into two baskets is almost perfectly in line with the classification applied in the weakly separable case.

Table 4. Estimates of the preference parameters in the full demand system compared to the estimates under weak separability.

	Full demand system	Weak separability
ε	0.685	0.677
	(0.004)	(0.004)
γ	0.505	0.211
	(0.018)	(0.023)
ν	346.736	327.271
	(11.793)	(13.358)
σ	0.050	
	(0.012)	
φ	0.360	
	(0.006)	
N	74,372	74,372

Notes. Standard errors in parentheses.

Figure 9 shows a comparison of the resulting price index from the full demand system (left panel) with the respective price index under weak separability (right panel). The figure shows that the price indices for the individual expenditure deciles are less spread out under the full demand system, but the general difference is relatively minor.

The differences between the full demand system estimation and estimation under weak separability are even less pronounced when looking at the comparison of inflation rates in Figure 10.

Taking these comparisons at face value indicates that the full demand system estimation does not strongly reject the case of weakly separable preferences. Further, the price index along with inflation from the full demand system and from the weakly separable preferences are remarkably

standard errors obtained from the standard covariance matrix are not (see Andrews, 1999, 2002). This should be kept in mind with the standard errors presented here.

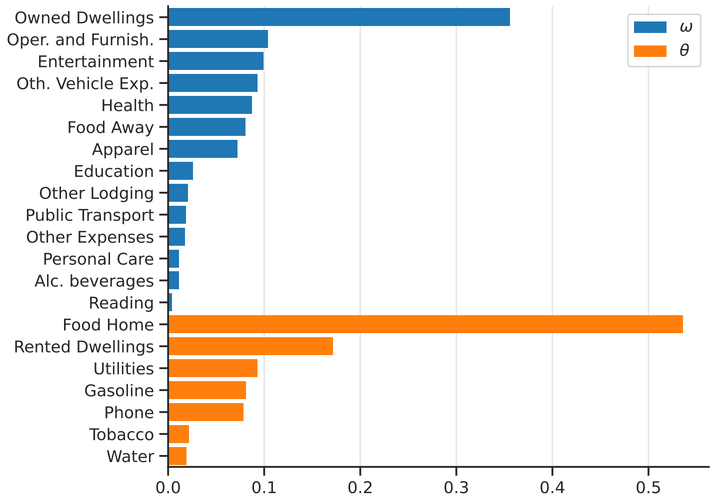


Figure 8. Point estimates for the taste parameters ω_j and θ_j of the full demand system.

Notes. The figure shows the estimated taste parameters from the full demand system estimation. Results are for the closest local minimum to the weakly separable case which has been used as initial guess for the parameters.

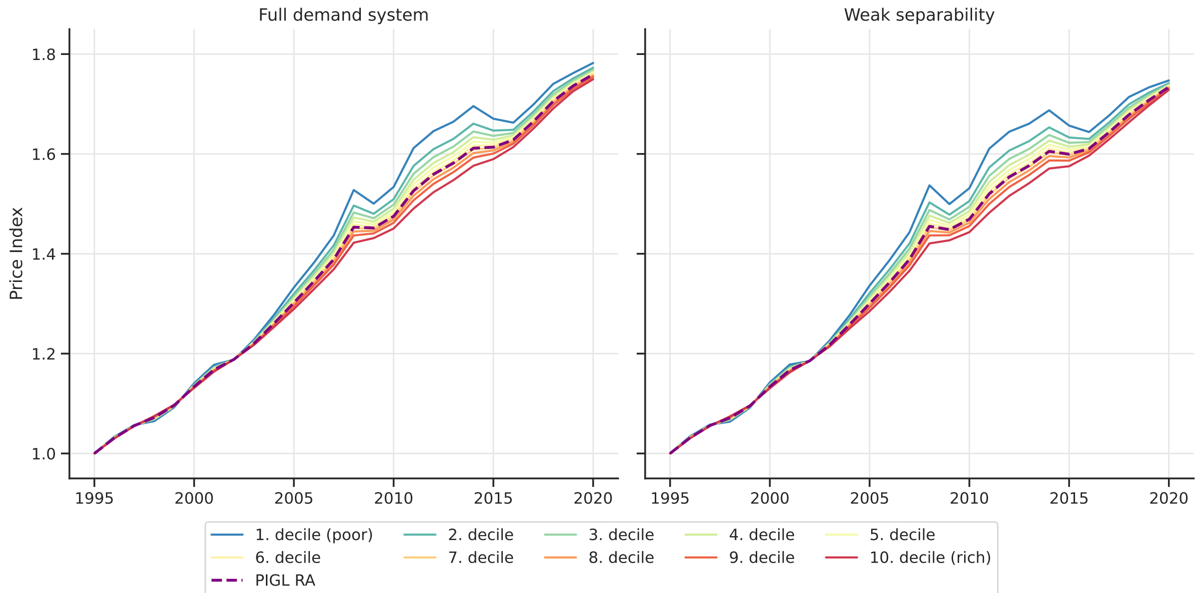


Figure 9. Comparison of the G-SV price index for the full demand system and under weak separability.

close and do not change any of our main conclusions.

8 Conclusion

We derive a nonhomothetic cost-of-living index which allows us to describe inflation heterogeneity along the full expenditure distribution. The cost-of-living index is microfounded with PIGL

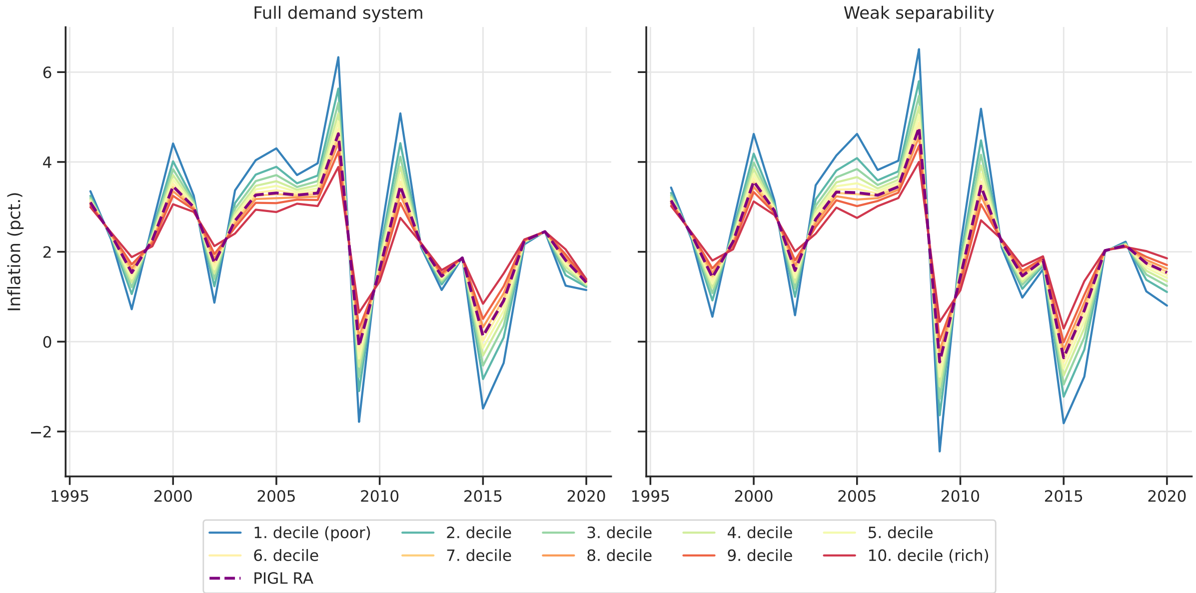


FIGURE 10. Comparison of the G-SV inflation for the full demand system and under weak separability.

preferences and we show that it can be computed without a full demand system estimation if weak separability of consumption goods into necessary and luxury goods is imposed. The theoretical rigor and practical simplicity of our index makes it especially appealing compared to other approaches previously taken in the literature on inflation heterogeneity.

The price index generalizes several classes of homothetic price indices, some of them particularly interesting for their superlative property. We present results for generalized Sato-Vartia, Törnqvist, geometric Walsh, Theil, Fisher and arithmetic Walsh price indices. Homothetic indices such as the Sato-Vartia and Törnqvist are used in previous empirical studies to approximate inflation heterogeneity by computing separate price indices for different income groups. We show that this approach can be rationalized through the lens of our framework but the usual caveats of the group-specific approach still apply.

Our empirical results show that from 1996 to 2020 there was a substantial heterogeneity in inflation between poorer and richer households in the US. The particular striking result we find is that while mean inflation is around 2.25 percent for everyone, the standard deviation of inflation has been 2.14 for the poor compared to 0.85 for the rich. Thus, inflation volatility is 2.5 times higher for the poor.

We find that poorer households are much more exposed to the highly volatile inflation rates of food, gas and utilities compared to the rich. We furthermore show that substitution behavior is only of second-order importance. Our findings hence suggest that in order to uncover the fundamental drivers of inflation inequality it is first and foremost important to understand *why* households make the consumption choices they do and what the explanation for inflation heterogeneity across product groups is.

References

Alder, S., Boppart, T. and Müller, A. (forthcoming). "A theory of structural change that can fit the data." *American Economic Journal: Macroeconomics* (cited on pages 3, 7, 8, 17, 28, 41, 43).

Andrews, D. W. K. (1999). "Estimation when a parameter is on a boundary." *Econometrica*, vol. 67(6), pp. 1341–1383 (cited on page 29).

Andrews, D. W. K. (2002). "Generalized method of moments estimation when a parameter is on a boundary." *Journal of Business and Economic Statistics*, vol. 20(4), pp. 530–544 (cited on page 29).

Argente, D. and Lee, M. (2021). "Cost of living inequality during the Great Recession." *Journal of the European Economic Association*, vol. 19(2), pp. 913–952 (cited on pages 5, 12, 20).

Balk, B. M. (2004). "Decompositions of Fisher indexes." *Economics Letters*, vol. 82(1), pp. 107–113 (cited on pages 12, 26, 38).

Banks, J., Blundell, R. and Lewbel, A. (1997). "Quadratic Engel curves and consumer demand." *Review of Economics and Statistics*, vol. 79(4), pp. 527–539 (cited on pages 5, 10, 16, 41).

Barnett, W. A. and Choi, K.-H. (2008). "Operational identification of the complete class of superlative index numbers: An application of Galois theory." *Journal of Mathematical Economics*, vol. 44(7–8), pp. 603–612 (cited on pages 5, 10, 11).

Blundell, R., Pashardes, P. and Weber, G. (1993). "What do we learn about consumer patterns from micro data?" *American Economic Review*, vol. 83(3), pp. 570–597 (cited on page 16).

Blundell, R., Pistaferri, L. and Preston, I. (2008). "Consumption inequality and partial insurance." *American Economic Review*, vol. 98(5), pp. 1887–1921 (cited on pages 23, 24).

Blundell, R. and Robin, J.-M. (2000). "Latent separability: Grouping goods without weak separability." *Econometrica*, vol. 68(1), pp. 53–84 (cited on page 7).

Boppart, T. (2014). "Structural change and the Kaldor facts in a growth model with relative price effects and non-Gorman preferences." *Econometrica*, vol. 82(6), pp. 2167–2196 (cited on pages 3, 7, 16).

Boskin, M. J., Dulberger, E. R., Gordon, R. J., Griliches, Z. and Jorgenson, D. (1996). *Toward a more accurate measure of the cost of living: Final report to the Senate Finance Committee from the Advisory Commission to Study the Consumer Price Index*. Washington, DC: Advisory Commission to Study the Consumer Price Index. (Cited on page 20).

Broda, C. and Romalis, J. (2009). *The welfare implications of rising price dispersion* [Unpublished manuscript], Booth School of Business, University of Chicago. (Cited on pages 5, 12).

Carlson, B. C. (1972). "The logarithmic mean." *American Mathematical Monthly*, vol. 79(6), pp. 615–618 (cited on page 9).

Christensen, L. R., Jorgenson, D. W. and Lau, L. J. (1975). "Transcendental logarithmic utility functions." *American Economic Review*, vol. 65(3), pp. 367–383 (cited on page 13).

Cravino, J., Levchenko, A. A. and Rojas, M. (forthcoming). "Population aging and structural transformation." *American Economic Journal: Macroeconomics* (cited on pages 3, 39).

Deaton, A. and Muellbauer, J. (1980). "An almost ideal demand system." *American Economic Review*, vol. 70(3), pp. 312–326 (cited on page 5).

Diewert, W. E. (1976). "Exact and superlative index numbers." *Journal of Econometrics*, vol. 4(2), pp. 115–145 (cited on pages 4, 5, 10, 11, 13).

Diewert, W. E. (1978). "Superlative index numbers and consistency in aggregation." *Econometrica*, vol. 46(4), pp. 883–900 (cited on page 4).

Diewert, W. (1993). "The early history of price index research." In W. Diewert and A. O. Nakamura (Eds.), *Essays in index number theory* (pp. 33–71). Elsevier Science Publishers B.V. (Cited on page 5).

Engel, E. (1857). "Die vorherrschenden Gewerbszweige in den Gerichtsämtern mit Beziehung auf die Productions- und Consumtionsverhältnisse des Königreichs Sachsen." *Zeitschrift des Statistischen Büreaus des Königlich Sächsischen Ministeriums des Innern*, vol. 8–9, pp. 153–182 (cited on page 3).

Feenstra, R. C. (1994). "New product varieties and the measurement of international prices." *American Economic Review*, vol. 84(1), pp. 157–177 (cited on page 4).

Feenstra, R. C. and Reinsdorf, M. B. (2000). "An exact price index for the almost ideal demand system." *Economics Letters*, vol. 66(2), pp. 159–162 (cited on page 5).

Fernández-Villaverde, J. and Krueger, D. (2007). "Consumption over the life cycle: Facts from Consumer Expenditure Survey data." *Review of Economics and Statistics*, vol. 89(3), pp. 552–565 (cited on page 22).

Fisher, I. (1922). *The making of index numbers: A study of their varieties, tests, and reliability*. Houghton Mifflin Company. (Cited on pages 3, 11).

Fox, L. E. and Burns, K. (2021). *The supplemental poverty measure: 2020* (Current Population Reports P60–275). Washington, DC: U.S. Census Bureau. (Cited on page 15).

Hanoch, G. (1975). "Production and demand models with direct or indirect implicit additivity." *Econometrica*, vol. 43(3), pp. 395–419 (cited on page 5).

Hill, R. J. (2006). "Superlative index numbers: Not all of them are super." *Journal of Econometrics*, vol. 130(1), pp. 25–43 (cited on page 17).

Hobijn, B. and Lagakos, D. (2005). "Inflation inequality in the United States." *Review of Income and Wealth*, vol. 51(4), pp. 581–606 (cited on pages 5, 15).

Jaravel, X. (2019). "The unequal gains from product innovations: Evidence from the U.S. retail sector." *Quarterly Journal of Economics*, vol. 134(2), pp. 715–783 (cited on pages 5, 12, 20).

Jaravel, X. (2021). "Inflation inequality: Measurement, causes, and policy implications." *Annual Review of Economics*, vol. 13, pp. 599–629 (cited on page 5).

Kaplan, G. and Schulhofer-Wohl, S. (2017). "Inflation at the household level." *Journal of Monetary Economics*, vol. 91, pp. 19–38 (cited on page 5).

Konüs, A. A. (1939). "The problem of the true index of the cost of living." *Econometrica*, vol. 7(1), pp. 10–29 (cited on pages 4, 6).

Lauper, C. and Mangiante, G. (2021). *Monetary policy shocks and inflation heterogeneity* [Unpublished manuscript], University of Lausanne. (Cited on page 5).

McGranahan, L. and Paulson, A. (2005). *Constructing the Chicago Fed income based economic index – consumer price index: Inflation experiences by demographic group: 1983–2005* (Working Paper No. 2005–20). Chicago, IL: Federal Reserve Bank of Chicago. (Cited on page 5).

Muellbauer, J. (1974). "Prices and inequality: The United Kingdom experience." *Economic Journal*, vol. 84(333), pp. 32–55 (cited on page 5).

Muellbauer, J. (1975). "Aggregation, income distribution and consumer demand." *Review of Economic Studies*, vol. 42(4), pp. 525–543 (cited on pages 3, 7, 8, 43).

Muellbauer, J. (1976). "Community preferences and the representative consumer." *Econometrica*, vol. 44(5), pp. 979–999 (cited on pages 3, 7, 8, 43).

Orchard, J. (2020). *Household inflation and aggregate inflation* [Unpublished manuscript], University of California, San Diego. (Cited on page 5).

Orchard, J. (2021). *Cyclical demand shifts and cost of living inequality* [Unpublished manuscript], University of California, San Diego. (Cited on pages 4, 5, 15).

Oulton, N. (2012). "How to measure living standards and productivity." *Review of Income and Wealth*, vol. 58(3), pp. 424–456 (cited on page 5).

Redding, S. J. and Weinstein, D. E. (2020). "Measuring aggregate price indices with taste shocks: Theory and evidence for CES preferences." *Quarterly Journal of Economics*, vol. 135(1), pp. 503–560 (cited on pages 4, 5, 10, 39).

Samuelson, P. A. and Swamy, S. (1974). "Invariant economic index numbers and canonical duality: Survey and synthesis." *American Economic Review*, vol. 64(4), pp. 566–593 (cited on pages 4, 6).

Sato, K. (1976). "The ideal log-change index number." *Review of Economics and Statistics*, vol. 58(2), pp. 223–228 (cited on pages 3, 11, 17).

Sato, R. (1975). "The most general class of CES functions." *Econometrica*, vol. 43(5/6), pp. 999–1003 (cited on page 5).

Theil, H. (1973). "A new index number formula." *Review of Economics and Statistics*, vol. 55(4), pp. 498–502 (cited on page 11).

Törnqvist, L. (1936). "The Bank of Finland's consumption price index." *Bank of Finland Monthly Bulletin*, vol. 16(10), pp. 27–34 (cited on pages 3, 11).

Vartia, Y. O. (1976). "Ideal log-change index numbers." *Scandinavian Journal of Statistics*, vol. 3(3), pp. 121–126 (cited on pages 3, 11).

Wachter, J. A. and Yogo, M. (2010). "Why do household portfolio shares rise with wealth?" *Review of Financial Studies*, vol. 23(11), pp. 3929–3965 (cited on pages 4, 15).

Walsh, C. M. (1901). *The measurement of general exchange-value*. The Macmillan Company. (Cited on page 11).

A Proofs and Extensions

A.1 Proof of Proposition 1

Proof. Inverting the indirect utility function (4) gives the expenditure function

$$c(u,\boldsymbol{p}) = \left[1+\varepsilon\left(u+\frac{\nu}{\gamma}\left\{\left(\frac{D(\boldsymbol{p})}{B(\boldsymbol{p})}\right)^{\gamma}-1\right\}\right)\right]^{\frac{1}{\varepsilon}} B(\boldsymbol{p}).$$

Suppose that the reference utility u corresponds to the expenditure level in some base period s such that $c(u,\boldsymbol{p}_s)=e_s$ and $u \equiv V(e_s,\boldsymbol{p}_s)$. Using the indirect utility function (4) to substitute this into some period-t expenditure function and rearranging terms yields

$$\begin{aligned}
c(u,\boldsymbol{p}_t) &= e_s\left[1+\frac{\varepsilon\,\nu}{\gamma}\left(\frac{B(\boldsymbol{p}_s)}{e_s}\right)^{\varepsilon}\left(\frac{D(\boldsymbol{p}_s)}{B(\boldsymbol{p}_s)}\right)^{\gamma}\left\{\left(\frac{D(\boldsymbol{p}_t)}{D(\boldsymbol{p}_s)}\right)^{\gamma}\left(\frac{B(\boldsymbol{p}_t)}{B(\boldsymbol{p}_s)}\right)^{-\gamma}-1\right\}\right]^{\frac{1}{\varepsilon}}\frac{B(\boldsymbol{p}_t)}{B(\boldsymbol{p}_s)} \\
&= e_s\left[1+\frac{\varepsilon\,w_{Ds}}{\gamma}\left\{\left(\frac{P_{Dt}}{P_{Bt}}\right)^{\gamma}-1\right\}\right]^{\frac{1}{\varepsilon}} P_{Bt} \\
&= e_s\left[\left(1-\frac{\varepsilon\,w_{Ds}}{\gamma}\right)P_{Bt}^{\gamma}+\frac{\varepsilon\,w_{Ds}}{\gamma}P_{Dt}^{\gamma}\right]^{\frac{1}{\gamma}\cdot\frac{\gamma}{\varepsilon}} P_{Bt}^{1-\frac{\gamma}{\varepsilon}},
\end{aligned}$$

where the second equality uses $P_{Bt} \equiv B(\boldsymbol{p}_t)/B(\boldsymbol{p}_s)$, $P_{Dt} \equiv D(\boldsymbol{p}_t)/D(\boldsymbol{p}_s)$ and the expenditure share (5). By the Konüs definition (1), the cost-of-living index is then

$$P_t^{PIGL} = \left[\left(1-\frac{\varepsilon\,w_{Ds}}{\gamma}\right)P_{Bt}^{\gamma}+\frac{\varepsilon\,w_{Ds}}{\gamma}P_{Dt}^{\gamma}\right]^{\frac{1}{\gamma}\cdot\frac{\gamma}{\varepsilon}} P_{Bt}^{1-\frac{\gamma}{\varepsilon}} = \widetilde{P}_t^{\frac{\gamma}{\varepsilon}} P_{Bt}^{1-\frac{\gamma}{\varepsilon}}. \tag{A.1}$$

Since $\widetilde{P}_t$ has a CES form, we may define hypothetical budget shares corresponding to this price function by

$$\psi_{Bt} \equiv \left(1-\frac{\varepsilon w_{Ds}}{\gamma}\right)\left(\frac{P_{Bt}}{\widetilde{P}_t}\right)^{\gamma} \qquad \text{and} \qquad \psi_{Dt} \equiv \frac{\varepsilon w_{Ds}}{\gamma}\left(\frac{P_{Dt}}{\widetilde{P}_t}\right)^{\gamma}, \tag{A.2}$$

with $\psi_{Bs} = 1-\varepsilon w_{Ds}/\gamma$ and $\psi_{Ds} = \varepsilon w_{Ds}/\gamma$. These shares ensure that the price index remains at the same utility level; $w_{Ds}(P_{Dt}\,/\,\widetilde{P}_t)^{\gamma}$ is the expenditure share of the D basket at period-t prices that prevails at the same utility level as w_{Ds}. To see this, use the expenditure share (5) to get

$$\begin{aligned}
w_{Dt} &= \nu \left(\frac{B(\boldsymbol{p}_t)}{e_t}\right)^{\varepsilon} \left(\frac{D(\boldsymbol{p}_t)}{B(\boldsymbol{p}_t)}\right)^{\gamma} && \text{(by (5))} \\
&= \nu \left(\frac{B(\boldsymbol{p}_s)}{e_s}\right)^{\varepsilon} \left(\frac{D(\boldsymbol{p}_s)}{B(\boldsymbol{p}_s)}\right)^{\gamma} \left(\frac{e_s P_{Dt}^{\frac{\gamma}{\varepsilon}} P_{Bt}^{1-\frac{\gamma}{\varepsilon}}}{e_t}\right)^{\varepsilon} && \text{(by (3))} \\
&= w_{Ds} \left(\frac{P_{Dt}^{\frac{\gamma}{\varepsilon}} P_{Bt}^{1-\frac{\gamma}{\varepsilon}}}{P_t Q_t}\right)^{\varepsilon} && \text{(by (5) and } e_t/e_s = P_t Q_t) \\
&= w_{Ds} \left(\frac{P_{Dt}}{\widetilde{P}_t}\right)^{\gamma} Q_t^{-\varepsilon} && \text{(by (A.1))}.
\end{aligned}$$

The third equality uses the decomposition $e_t/e_s = P_t Q_t$, where P_t is the Konüs price index and Q_t the corresponding quantity index. Along the same indifference curve as w_{Ds}, we necessarily have $Q_t = 1$ for all t, and the result then immediately follows. Equation (A.2) allows us to write $\widetilde{P}_t$ as a Sato-Vartia index. The procedure is the same as in the standard case: solve for $\widetilde{P}_t$ from the shares in Equation (A.2), take logs, multiply by the difference in shares over time, sum over both B and D and solve for $\widetilde{P}_t$. For $C \in \{B, D\}$, the first two steps yields

$$\ln \widetilde{P}_t = \ln P_{Ct} - \frac{1}{\gamma} \ln \left(\frac{\psi_{Ct}}{\psi_{Cs}}\right) \qquad \Longleftrightarrow \qquad -\frac{1}{\gamma} = \frac{\ln \widetilde{P}_t - \ln P_{Ct}}{\ln \psi_{Ct} - \ln \psi_{Cs}}.$$

Multiplying both sides by $\psi_{Ct} - \psi_{Cs}$, summing over both $C \in \{B, D\}$, and rearranging terms results in

$$\ln \widetilde{P}_t \sum_{C \in \{B,D\}} \frac{\psi_{Ct} - \psi_{Cs}}{\ln \psi_{Ct} - \ln \psi_{Cs}} = \sum_{C \in \{B,D\}} \frac{\psi_{Ct} - \psi_{Cs}}{\ln \psi_{Ct} - \ln \psi_{Cs}} \ln P_{Ct}.$$

Then solving for $\widetilde{P}_t$ yields

$$\widetilde{P}_t = P_{Dt}^{\phi_t} P_{Bt}^{1-\phi_t}, \qquad \text{where} \qquad \phi_t \equiv \frac{L(\psi_{Dt}, \psi_{Ds})}{L(\psi_{Dt}, \psi_{Ds}) \ + \ L(\psi_{Bt}, \psi_{Bs})}. \tag{A.3}$$

Plugging Equation (A.3) into Equation (A.1) gives the household-level price indices.

Because a representative level of expenditures e^{RA} exists over any group of households, group-level behavior is characterized by the same indirect utility function and expenditure function as household-level behavior.[16] Aggregate-level cost-of-living indices are therefore derived identically

[16] This can be shown by substituting for $e^{RA} = \bar{e}\kappa^{-\frac{1}{\varepsilon}}$ in the aggregate expenditure share (9) and integrating to obtain the group-level indirect utility function.

to above, with the only difference that group-level expenditure shares $\overline{w}_{Ds}$ and representative levels of expenditure e^{RA} are used instead of household-level ones. □

A.2 Proof of Corollary 1

Proof. The result is immediate by setting both P_{Dt} and P_{Dt} to either the *quadratic-mean-of-order-r* index (14) or the *Theil-Sato* index (15) and substituting these into the general PIGL index (11). We only need to rewrite (14) into a geometric-mean form. Balk (2004) does this for the Fisher ideal index ($r = 2$), and the generalization to any $r > 0$ is analogous. As in Corollary 1, define

$$P_{Lt} = \left[\sum_{j \in J} w_{js} \left(\frac{p_{jt}}{p_{js}} \right)^{\frac{r}{2}} \right]^{\frac{2}{r}} \qquad \text{and} \qquad P_{Pt} = \left[\sum_{j \in J} w_{jt} \left(\frac{p_{jt}}{p_{js}} \right)^{-\frac{r}{2}} \right]^{-\frac{2}{r}},$$

such that the *quadratic-mean-of-order-r* index (14) can be written $P_t = \sqrt{P_{Lt} P_{Pt}}$. P_{Lt} weighs price changes by base-period expenditure shares while P_{Pt} uses current-period expenditure shares, and the definition nests the Laspeyres and Paasche indices as the special case where $r = 2$, thus motivating the L and P notation. By the definition of P_{Lt} and the logarithmic mean, it holds that

$$\begin{aligned} 0 &= \sum_{j \in J} w_{js} \left(\frac{p_{jt}}{p_{js}} \right)^{\frac{r}{2}} - P_{Lt}^{\frac{r}{2}} = \sum_{j \in J} w_{js} \left[\left(\frac{p_{jt}}{p_{js}} \right)^{\frac{r}{2}} - P_{Lt}^{\frac{r}{2}} \right] \\ &= \sum_{j \in J} w_{js} \, L \left((\tfrac{p_{jt}}{p_{js}})^{\frac{r}{2}}, \, P_{Lt}^{\frac{r}{2}} \right) \, \ln \left(\frac{p_{jt}/p_{js}}{P_{Lt}} \right)^{\frac{r}{2}} \\ &= \frac{r}{2} \sum_{j \in J} \widetilde{w}_{Ljt} \left[\ln \left(\frac{p_{jt}}{p_{js}} \right) - \ln P_{Lt} \right], \end{aligned}$$

with $\widetilde{w}_{Ljt}$ defined as in Corollary 1. Solving for $\ln P_{Lt}$, we get

$$\ln P_{Lt} = \sum_{j \in J} \frac{\widetilde{w}_{Ljt}}{\sum_{i \in J} \widetilde{w}_{Lit}} \ln \left(\frac{p_{jt}}{p_{js}} \right).$$

Identical steps for P_{Pt} yields

$$\ln P_{Pt} = \sum_{j \in J} \frac{\widetilde{w}_{Pjt}}{\sum_{i \in J} \widetilde{w}_{Pit}} \ln \left(\frac{p_{jt}}{p_{js}} \right),$$

with $\widetilde{w}_{Pjt}$ defined as in Corollary 1. Substituting these into the overall index P_t yields

$$P_t = \prod_{j\in J} \left(\frac{p_{jt}}{p_{js}}\right)^{\delta_{jt}}, \qquad \text{where} \qquad \delta_{jt} \equiv \frac{1}{2}\left[\frac{\widetilde{w}_{Ljt}}{\sum_i \widetilde{w}_{Lit}} + \frac{\widetilde{w}_{Pjt}}{\sum_i \widetilde{w}_{Pit}}\right],$$

and we are done. □

A.3 Proof of Proposition 3

Proof. If preferences are of the Cobb-Douglas form $V(e,\boldsymbol{p}) = \left[\frac{e}{B(\boldsymbol{p})^{1-\nu}D(\boldsymbol{p})^{\nu}}\right]$, then the cost-of-living index is $P_t = P_{Bt}^{1-\nu}P_{Dt}^{\nu}$ by Proposition 1. If P_{Bt} and P_{Dt} are Törnqvist indices, Corollary 1 and Assumption 1 allow us to write this index as

$$P_t = \prod_{j\in J_B} \left(\frac{p_{jt}}{p_{js}}\right)^{(1-\nu)\delta_{jt}^B} \prod_{j\in J_D} \left(\frac{p_{jt}}{p_{js}}\right)^{\nu\delta_{jt}^D}, \qquad \delta_{jt}^C = \frac{w_{js}^C + w_{jt}^C}{2}, \quad C \in \{B, D\}, \tag{A.4}$$

where J_B and J_D denote the sets of goods in B and D, respectively. (Under the weak separability assumption, it holds that $J_B \cup J_D = J$ and $J_B \cap J_D = \varnothing$.) Meanwhile, the standard Törnqvist index reads

$$P_t = \prod_{j\in J} \left(\frac{p_{jt}}{p_{js}}\right)^{\delta_{jt}}, \qquad \delta_{jt} = \frac{w_{js} + w_{jt}}{2}. \tag{A.5}$$

Equations (A.4) and (A.5) are equal if $(1-\nu)\delta_{jt}^B = \delta_{jt}$ for $j \in J_B$ and $\nu\delta_{jt}^D = \delta_{jt}$ for $j \in J_D$. Under Cobb-Douglas preferences, $\nu = w_D$ is the homothetic and time-invariant expenditure share on D. Under the weak separability assumption, the total expenditure share on good $j \in J_C$, $C \in \{B, D\}$, is given by $w_j = w_C\, w_j^C$. Thus,

$$\begin{aligned} (1-\nu)\delta_{jt}^B &= w_B \frac{w_{js}^B + w_{jt}^B}{2} = \frac{w_{js} + w_{jt}}{2} = \delta_{jt}, \qquad && j \in J_B, \\ \nu\delta_{jt}^D &= w_D \frac{w_{js}^D + w_{jt}^D}{2} = \frac{w_{js} + w_{jt}}{2} = \delta_{jt}, \qquad && j \in J_D, \end{aligned}$$

and it follows that the Törnqvist index under weak separability (A.4) is the same as the standard Törnqvist index (A.5). □

A.4 Allowing for Heterogeneity in Tastes

Redding and Weinstein (2020) stress the importance of accounting for heterogeneity in tastes for the cost of living and it is possible to extend the baseline framework to allow for this. Following Cravino, Levchenko and Rojas (forthcoming), let the preferences of household h be characterized by an indirect utility function of the form

$$V_h(e_h, \boldsymbol{p}) = \frac{1}{\varepsilon}\left[\left(\frac{e_h}{B(\boldsymbol{p})}\right)^{\varepsilon} - 1\right] - \frac{\nu_h}{\gamma}\left[\left(\frac{D(\boldsymbol{p})}{B(\boldsymbol{p})}\right)^{\gamma} - 1\right], \tag{A.6}$$

where the only difference to the PIGL specification in Equation (4) is that we allow the taste parameter ν_h to vary across households (and time). As before, the expenditure share of the latent good with price function $D(\cdot)$ is given by Roy's identity as

$$w_{Dh} = \nu_h \left(\frac{B(\boldsymbol{p})}{e_h}\right)^{\varepsilon}\left(\frac{D(\boldsymbol{p})}{B(\boldsymbol{p})}\right)^{\gamma},$$

and the corresponding aggregate expenditure share over any N households is now

$$\overline{w}_D = \left(\frac{B(\boldsymbol{p})}{\bar{e}}\right)^{\varepsilon}\left(\frac{D(\boldsymbol{p})}{B(\boldsymbol{p})}\right)^{\gamma}\kappa, \qquad \text{where} \qquad \kappa \equiv \frac{1}{N}\sum_{h=1}^{N}\nu_h\left(\frac{e_h}{\bar{e}}\right)^{1-\varepsilon}.$$

A representative expenditure level $e^{RA} = \bar{e}\kappa^{-\frac{1}{\varepsilon}}$ therefore exists and incorporates any heterogeneity in tastes. Substituting back into the aggregate expenditure share $\overline{w}_D$ and integrating back yields aggregate-level behavior characterized by the indirect utility function

$$V(e^{RA}, \boldsymbol{p}) = \frac{1}{\varepsilon}\left[\left(\frac{e^{RA}}{B(\boldsymbol{p})}\right)^{\varepsilon} - 1\right] - \frac{1}{\gamma}\left[\left(\frac{D(\boldsymbol{p})}{B(\boldsymbol{p})}\right)^{\gamma} - 1\right],$$

with corresponding expenditure function

$$c(u^{RA}, \boldsymbol{p}) = \left[1 + \varepsilon\left(u^{RA} + \frac{1}{\gamma}\left\{\left(\frac{D(\boldsymbol{p})}{B(\boldsymbol{p})}\right)^{\gamma} - 1\right\}\right)\right]^{\frac{1}{\varepsilon}} B(\boldsymbol{p}).$$

This expenditure function is independent of the taste parameters ν_h. We can therefore follow the same steps as in Section A.1 to derive an identical price index as in Proposition 1. Again, this index is only a function of the base-period expenditure share for the D basket, price indices P_{Dt} and P_{Bt}, and the parameters ε and γ. Heterogeneity in the taste parameters ν_h only affect the price index indirectly to the extent that they affect expenditure shares. Whenever expenditure shares are observed in the data, we therefore do not need to know these individual tastes to compute the price index. We obtain the same result for household-level cost-of-living indices as the special case where $N = 1$.

Taste heterogeneity also poses no challenge with respect to estimating ε and γ. Since PIGL preferences aggregate consistently, it is possible to estimate these parameters from a aggregate data without any aggregation bias. Therefore, taking an aggregate time series and estimating

$$\overline{w}_{Dt} = \left(\frac{B(\boldsymbol{p}_t)}{\bar{e}_t}\right)^{\varepsilon} \left(\frac{D(\boldsymbol{p}_t)}{B(\boldsymbol{p}_t)}\right)^{\gamma} \kappa_t,$$

where κ_t is just a standard time fixed effect, is sufficient and we therefore avoid the need to estimate all the household-level effects ν_h.

A.5 Allowing for Hump-Shaped Expenditure Shares

Banks, Blundell and Lewbel (1997) stress the importance of allowing for hump-shaped expenditure shares to match the microeconomic data and it is possible to extend the baseline framework to allow for this at the household level. Following Alder, Boppart and Müller (forthcoming), let preferences be characterized by an indirect utility function of the form

$$V(e, \boldsymbol{p}) = \frac{1}{\varepsilon}\left[\left(\frac{e - A(\boldsymbol{p})}{B(\boldsymbol{p})}\right)^{\varepsilon} - 1\right] - \frac{\nu}{\gamma}\left[\left(\frac{D(\boldsymbol{p})}{B(\boldsymbol{p})}\right)^{\gamma} - 1\right], \tag{A.7}$$

where the only difference to the PIGL specification in Equation (4) is the addition of a linearly homogeneous function $A(\boldsymbol{p})$ of prices. The expenditure shares of the three latent goods with price functions $A(\cdot)$, $B(\cdot)$ and $D(\cdot)$ are given by Roy's identity as

$$w_A = \frac{A(\boldsymbol{p})}{e}, \tag{A.8}$$

$$w_B = \left(1 - \frac{A(\boldsymbol{p})}{e}\right)\left[1 - \nu\left(\frac{B(\boldsymbol{p})}{e - A(\boldsymbol{p})}\right)^{\varepsilon}\left(\frac{D(\boldsymbol{p})}{B(\boldsymbol{p})}\right)^{\gamma}\right], \tag{A.9}$$

$$w_D = \left(1 - \frac{A(\boldsymbol{p})}{e}\right)\nu\left(\frac{B(\boldsymbol{p})}{e - A(\boldsymbol{p})}\right)^{\varepsilon}\left(\frac{D(\boldsymbol{p})}{B(\boldsymbol{p})}\right)^{\gamma}. \tag{A.10}$$

The shares w_j^A, w_j^B and w_j^D of total A, B and D expenditures allocated to an individual good j are given as before by $w_j^C = p_j C_j(\boldsymbol{p})/C(\boldsymbol{p})$, $C \in \{A, B, D\}$. Together with Equations (A.8) to (A.10), this implies an expenditure share w_j of good j in total expenditures of the form

$$\begin{aligned} w_j = p_j \Bigg\{ & \frac{A(\boldsymbol{p})}{e}\,\frac{A_j(\boldsymbol{p})}{A(\boldsymbol{p})} \\ & + \left(1 - \frac{A(\boldsymbol{p})}{e}\right)\left[\frac{B_j(\boldsymbol{p})}{B(\boldsymbol{p})} + \left(\frac{D_j(\boldsymbol{p})}{D(\boldsymbol{p})} - \frac{B_j(\boldsymbol{p})}{B(\boldsymbol{p})}\right)\nu\left(\frac{B(\boldsymbol{p})}{e - A(\boldsymbol{p})}\right)^{\varepsilon}\left(\frac{D(\boldsymbol{p})}{B(\boldsymbol{p})}\right)^{\gamma}\right]\Bigg\}. \end{aligned} \tag{A.11}$$

Since the first term on the right-hand side of (A.11) is decreasing in e while the second term can be either increasing or decreasing in e, this allows for expenditure shares that are non-monotonic in expenditures. The derivation of the exact price index of (A.7) is virtually identical to the PIGL case in Section A.1. The corresponding expenditure function of (A.7) is

$$c(u,\boldsymbol{p}) = \left[1+\varepsilon\left(u+\frac{\nu}{\gamma}\left\{\left(\frac{D(\boldsymbol{p})}{B(\boldsymbol{p})}\right)^{\gamma}-1\right\}\right)\right]^{\frac{1}{\varepsilon}} B(\boldsymbol{p})+A(\boldsymbol{p}).$$

Suppose again that the reference utility u is that corresponding to the expenditure level in some base period s such that $c(u,\boldsymbol{p}_s)=e_s$ and $u\equiv V(e_s,\boldsymbol{p}_s)$. Using the indirect utility function (A.7) to substitute this into some period-t expenditure function, rearranging terms, and using $P_{Ct}=C(\boldsymbol{p}_t)/C(\boldsymbol{p}_s)$ together with Equations (A.8) to (A.10) yields

$$c(u,\boldsymbol{p}_t) = e_s\left\{(1-w_{As})\left[\left(1-\frac{\varepsilon}{\gamma}\frac{w_{Ds}}{1-w_{As}}\right)P_{Bt}^{\gamma}+\frac{\varepsilon}{\gamma}\frac{w_{Ds}}{1-w_{As}}P_{Dt}^{\gamma}\right]^{\frac{1}{\gamma}\cdot\frac{\gamma}{\varepsilon}}P_{Bt}^{1-\frac{\gamma}{\varepsilon}}+w_{As}P_{At}\right\},$$

and it follows that the price index is

$$P_t^{IA} = (1-w_{As})\widetilde{P}_t^{\frac{\gamma}{\varepsilon}}P_{Bt}^{1-\frac{\gamma}{\varepsilon}}+w_{As}P_{At}$$

where

$$\widetilde{P}_t \equiv \left[\left(1-\frac{\varepsilon}{\gamma}\frac{w_{Ds}}{1-w_{As}}\right)P_{Bt}^{\gamma}\;+\;\frac{\varepsilon}{\gamma}\frac{w_{Ds}}{1-w_{As}}P_{Dt}^{\gamma}\right]^{\frac{1}{\gamma}}.$$

Writing $\widetilde{P}_t$ as a Sato-Vartia index, we finally obtain the household-level cost-of-living index

$$P_t^{IA} = (1-w_{As})P_{Dt}^{\frac{\gamma\phi_t}{\varepsilon}}P_{Bt}^{1-\frac{\gamma\phi_t}{\varepsilon}}+w_{As}P_{At},$$

where ϕ_t is a Sato-Vartia weight as in Proposition 1 with

$$\psi_{Bt} \equiv \left(1-\frac{\varepsilon}{\gamma}\frac{w_{Ds}}{1-w_{As}}\right)\left(\frac{P_{Bt}}{\widetilde{P}_t}\right)^{\gamma} \qquad \text{and} \qquad \psi_{Dt} \equiv \frac{\varepsilon}{\gamma}\frac{w_{Ds}}{1-w_{As}}\left(\frac{P_{Dt}}{\widetilde{P}_t}\right)^{\gamma}.$$

This index is a direct generalization of Proposition 1 and, as before, is computable given expenditure shares w_{As}, w_{Ds}, price indices P_{At}, P_{Bt}, P_{Dt} and parameter values for ε and γ. Similarly to Proposition 2, under Assumption 1 and appropriate choices for P_{At}, P_{Bt} and P_{Dt}, estimation reduces to only the two parameters ε and γ which are readily obtained from Equations (A.8) and (A.10).

Unlike the baseline framework, however, these preferences do not aggregate as easily. As shown in Alder, Boppart and Müller (forthcoming, Proposition 2), aggregate expenditure shares over N households are now

$$\overline{w}_A = \frac{A(\boldsymbol{p})}{\bar{e}},$$

$$\overline{w}_B = \left(1 - \frac{A(\boldsymbol{p})}{\bar{e}}\right) \left[1 - \nu \left(\frac{B(\boldsymbol{p})}{\bar{e} - A(\boldsymbol{p})}\right)^{\varepsilon} \left(\frac{D(\boldsymbol{p})}{B(\boldsymbol{p})}\right)^{\gamma}\right] \kappa,$$

$$\overline{w}_D = \left(1 - \frac{A(\boldsymbol{p})}{\bar{e}}\right) \nu \left(\frac{B(\boldsymbol{p})}{\bar{e} - A(\boldsymbol{p})}\right)^{\varepsilon} \left(\frac{D(\boldsymbol{p})}{B(\boldsymbol{p})}\right)^{\gamma} \kappa.$$

where

$$\kappa \equiv \frac{1}{N} \sum_{h=1}^{N} \left(\frac{e_h - A(\boldsymbol{p})}{\bar{e} - A(\boldsymbol{p})}\right)^{1-\varepsilon}.$$

Unlike the PIGL case, there is no representative level of expenditure in Muellbauer's (1975, 1976) sense, even though a representative agent exists.[17] Therefore, it is not possible to bake in the parameter κ into some representative level of expenditure and proceed as for an individual household. Instead, the expenditure function of the representative agent is now

$$c(u^{RA}, \boldsymbol{p}) = \left[1 + \varepsilon \left(u^{RA} + \frac{\nu\,\kappa}{\gamma} \left\{ \left(\frac{D(\boldsymbol{p})}{B(\boldsymbol{p})}\right)^{\gamma} - 1 \right\}\right)\right]^{\frac{1}{\varepsilon}} B(\boldsymbol{p}) + A(\boldsymbol{p}).$$

Similar steps as before gives an aggregate price index of the same form as above, $P_t^{IA} = (1 - \overline{w}_{As}) P_{Dt}^{\frac{\gamma \phi_t}{\varepsilon}} P_{Bt}^{1 - \frac{\gamma \phi_t}{\varepsilon}} + \overline{w}_{As} P_{At}$, but with weights given by

$$\psi_{Bt} \equiv \left(1 - \frac{\varepsilon}{\gamma} \frac{\kappa_t}{\kappa_s} \frac{\overline{w}_{Ds}}{1 - \overline{w}_{As}}\right) \left(\frac{P_{Bt}}{\widetilde{P}_t}\right)^{\gamma} \qquad \text{and} \qquad \psi_{Dt} \equiv \frac{\varepsilon}{\gamma} \frac{\kappa_t}{\kappa_s} \frac{\overline{w}_{Ds}}{1 - \overline{w}_{As}} \left(\frac{P_{Dt}}{\widetilde{P}_t}\right)^{\gamma}.$$

where

$$\widetilde{P}_t \equiv \left[\left(1 - \frac{\varepsilon}{\gamma} \frac{\kappa_t}{\kappa_s} \frac{\overline{w}_{Ds}}{1 - \overline{w}_{As}}\right) P_{Bt}^{\gamma} + \frac{\varepsilon}{\gamma} \frac{\kappa_t}{\kappa_s} \frac{\overline{w}_{Ds}}{1 - \overline{w}_{As}} P_{Dt}^{\gamma}\right]^{\frac{1}{\gamma}}.$$

Thus, to compute aggregate price indices, we now either need to know the inequality measures κ in the time periods considered, or we need to impose the rather strong assumption that these measures remain constant over time for *all* groups considered.

[17] The expenditure level e^{RA} that induces the average expenditure shares for A and D are given by $e^{RA} = \bar{e}$ and $\left(1 - \frac{A(\boldsymbol{p})}{e^{RA}}\right) \left(e^{RA} - A(\boldsymbol{p})\right)^{-\varepsilon} = \left(1 - \frac{A(\boldsymbol{p})}{\bar{e}}\right) \left(\bar{e} - A(\boldsymbol{p})\right)^{-\varepsilon} \kappa$, respectively, and these generally differ.

B Additional Figures and Tables

Table B.1. CEX-CPI crosswalk.

	CEX category	**CPI name**	**CPI code**
1	Food at home	Food at home	SAF11
2	Food away from home	Food away from home	SEFV
3	Alcoholic beverages	Alcoholic beverages	SAF116
4	Rented dwellings	Rent of primary residence	SEHA
5	Owned dwellings[a]	Owners' equivalent rent of primary residence	SEHC
6	Other lodging	Lodging while out of town[b]	MUUR0000SE2102
		Lodging away from home[b]	SEHB
7	Utilities	Household energy	SAH21
8	Water	Water and sewerage maintenance	SEHG01
9	Phone	Communication	SAE2
10	Household F&O[c]	Household furnishings and operations	SAH3
11	Apparel	Apparel	SAA
12	Gasoline	Motor fuel	SETB
13	Other vehicle expenses	Motor vehicle maintenance and repair	SETD
		Motor vehicle insurance	SETE
		Motor vehicle fees	SETF
14	Public transportation	Public transportation	SETG
15	Health	Medical care	SAM
16	Entertainment	Recreation	SAR
17	Personal care	Personal care	SAG1
18	Reading	Recreational reading materials	SERG
19	Education	Education and communcation	SAE
20	Tobacco	Tobacco and smoking products	SEGA
21	Other expenses	Miscellaneous personal services	SEGD

Notes. The CEX categories follow the hierachical groupings defined by the BLS. CPI are non-seasonally adjusted nationwide data for urban consumers. *a.* Rental equivalence value of owned dwellings as reported by the households. *b.* "Lodging away from home" from 1995–1997 and "lodging while out of town" afterwards. *c.* Furnishing and operations, includes "household operations", "housekeeping supplies" and "household furnishings and equipment".

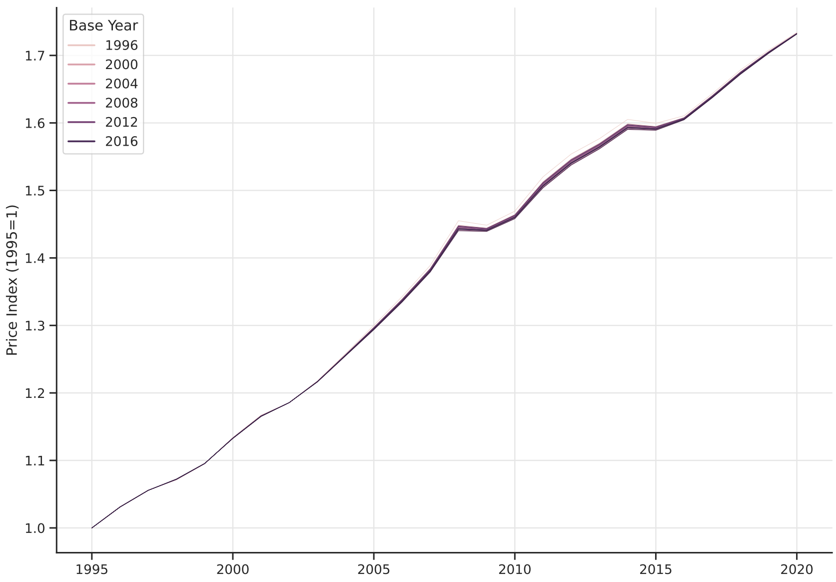


Figure B.1. PIGL representative agent G-SV price index for different base years.

Notes. The price index is calculated under weak separability. Each line represents the representative agent price index for a different base year, but normalized to one in 1995.

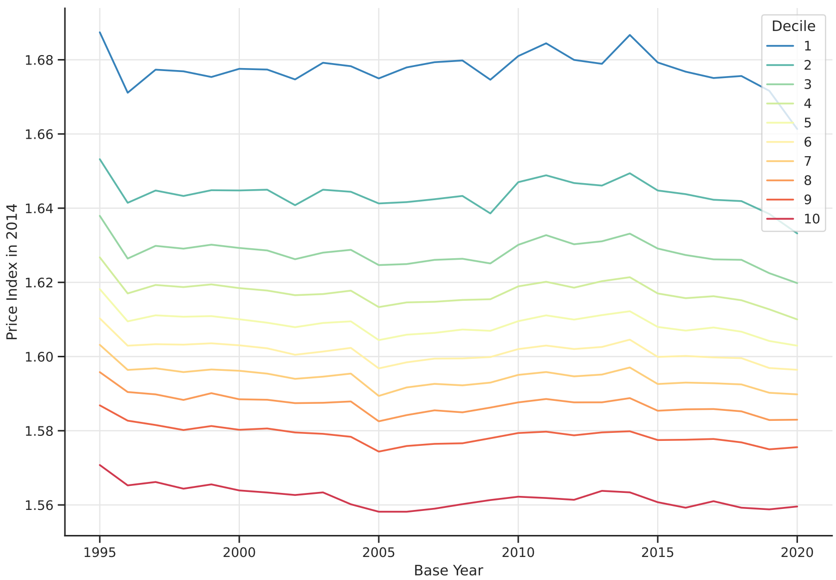


Figure B.2. G-SV price index in 2014 by expenditure decile for different base years.

Notes. The price index is calculated under weak separability. The horizontal axis describes the base year of the price index and the vertical axis the respective value of the price index in 2014. Price indices are all normalized to one in 1995. The price index for each expenditure decile is calculated as the price index of the PIGL representative agent over households within each respective decile.

Table B.2. Marginal effect of a change in expenditure decile on expenditure share.

	Dependent variable: Expenditure share (in %)
Luxuries	
Owned Dwellings	1.757
	(0.019)
Oper. and Furnish.	0.690
	(0.010)
Entertainment	0.386
	(0.006)
Other Lodging	0.303
	(0.005)
Food Away	0.299
	(0.005)
Education	0.280
	(0.006)
Health	0.215
	(0.007)
Public Transport	0.137
	(0.003)
Oth. Vehicle Exp.	0.124
	(0.006)
Other Expenses	0.105
	(0.002)
Apparel	0.098
	(0.004)
Alc. beverages	0.044
	(0.001)
Personal Care	0.031
	(0.001)
Reading	0.028
	(0.001)
Necessities	
Food Home	-2.031
	(0.012)
Rented Dwellings	-1.496
	(0.017)
Utilities	-0.362
	(0.004)
Gasoline	-0.219
	(0.004)
Phone	-0.186
	(0.003)
Tobacco	-0.158
	(0.003)
Water	-0.045
	(0.002)
Expenditure Category Dummies	Yes
Observations	1,562,211
Adjusted R^2	0.579

Standard errors in parentheses.

Notes. The table shows the coefficient estimates from a weighted least square regression of expenditure share on the expenditure decile interacted with individual good dummies using household weights. We include expenditure category dummies and cluster standard errors on household level.

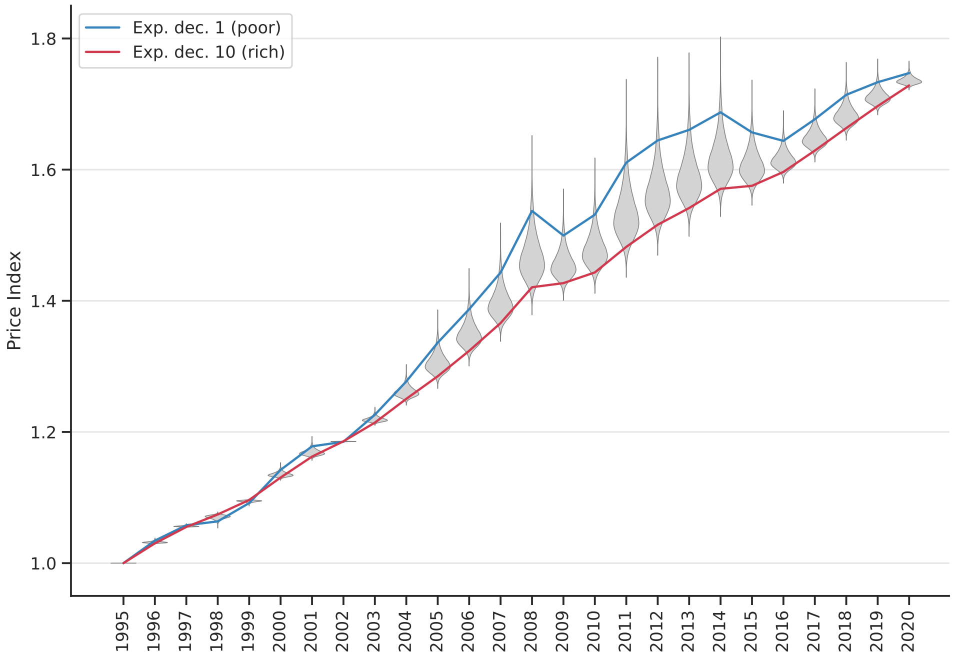


Figure B.3. Distribution of the G-SV price index under weak separability.

Notes. The gray shaded area shows the kernel density estimate of the distribution of G-SV price indices. The index is calculated for each household in the sample of 1995. The blue line shows the PIGL RA price index for the poorest 10 percent. The red line shows the PIGL RA price index for the richest 10 percent.

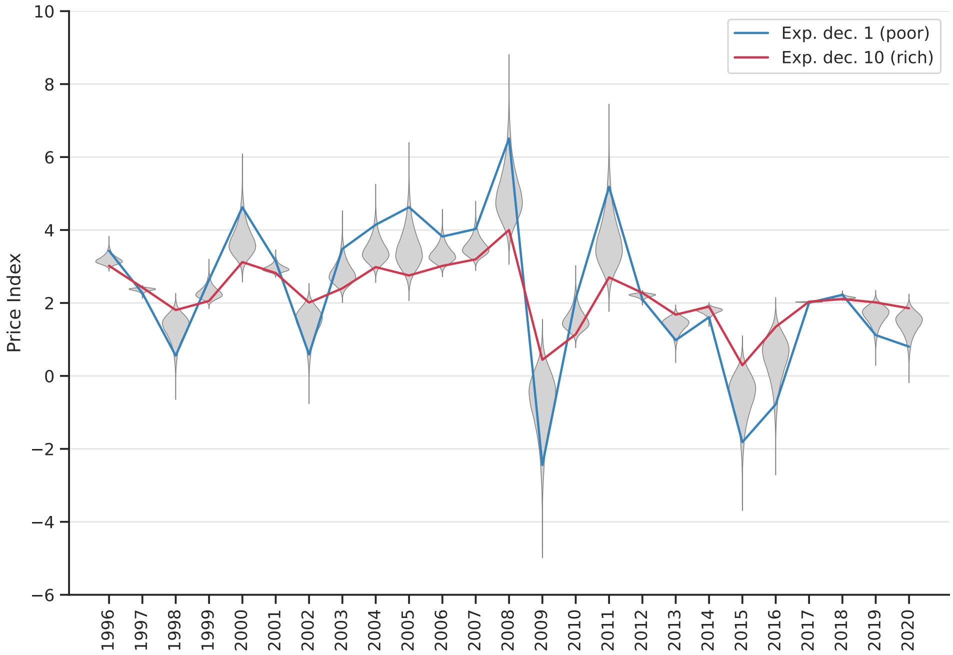


Figure B.4. Distribution of the G-SV inflation under weak separability.

Notes. The gray shaded area shows the kernel density estimate of the distribution of inflation rates. The inflation is calculated for each household in the sample of 1995. The blue line shows the PIGL RA inflation for the poorest 10 percent. The red line shows the PIGL RA inflation for the richest 10 percent.

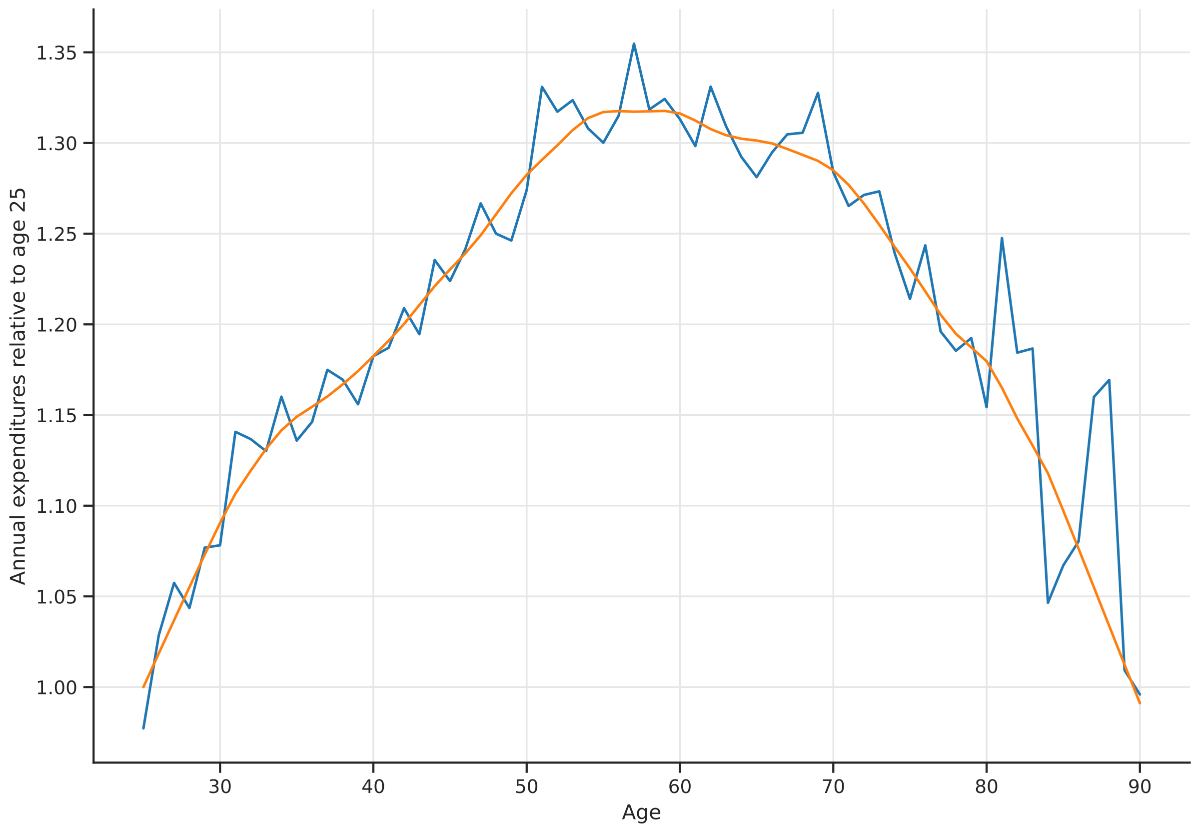


FIGURE B.5. Expenditures over the life cycle.

Notes. The blue line shows the estimated life-cycle expenditures, $\hat{\alpha}_i$'s, from the model in Equation (21). The orange line shows smoothed expenditure levels using a Locally Weighted Scatterplot Smoothing (LOWESS). All measures are relative to the smoothed expenditure level at age 25.

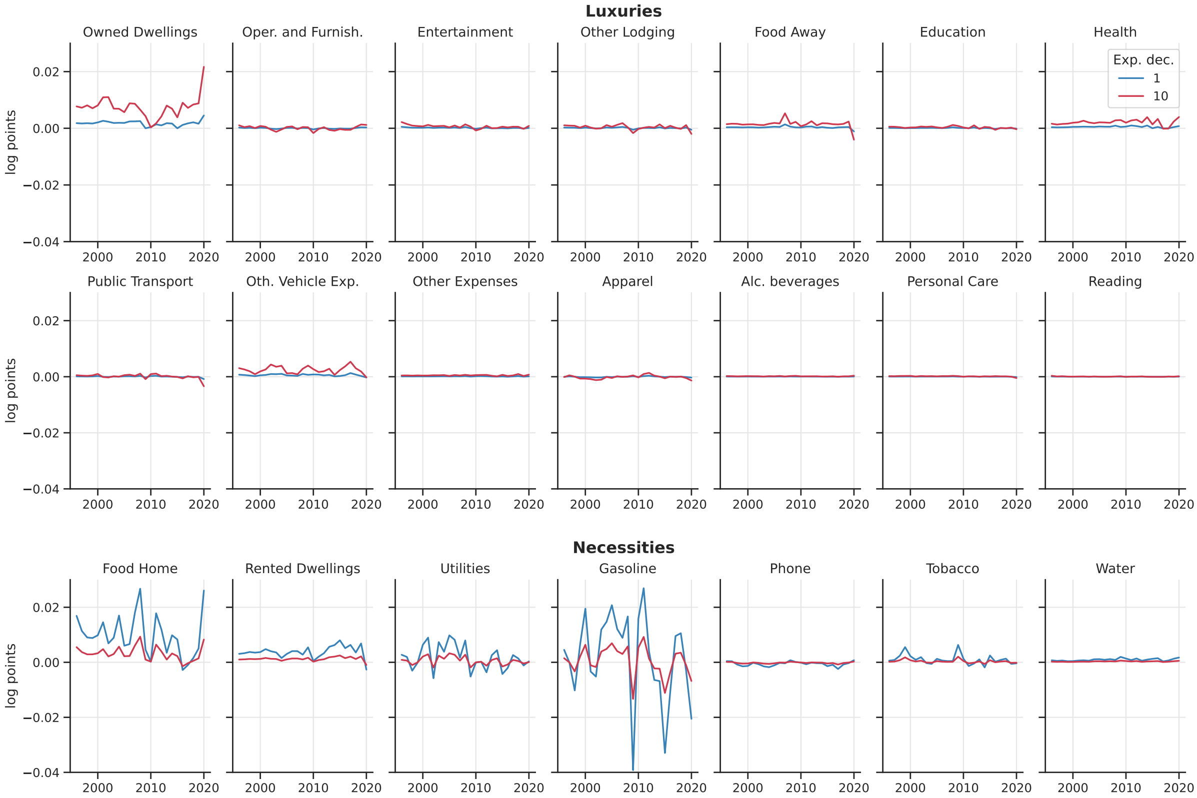


Figure B.6. Inflation decomposition by expenditure categories.

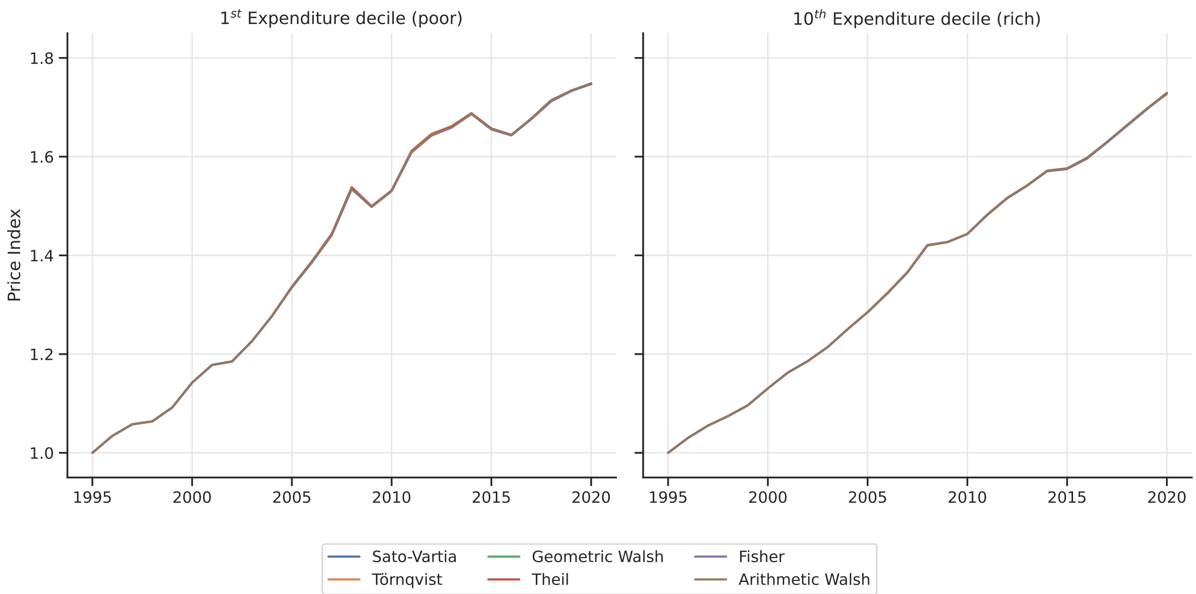


Figure B.7. Comparison of different generalized superlative price indices.

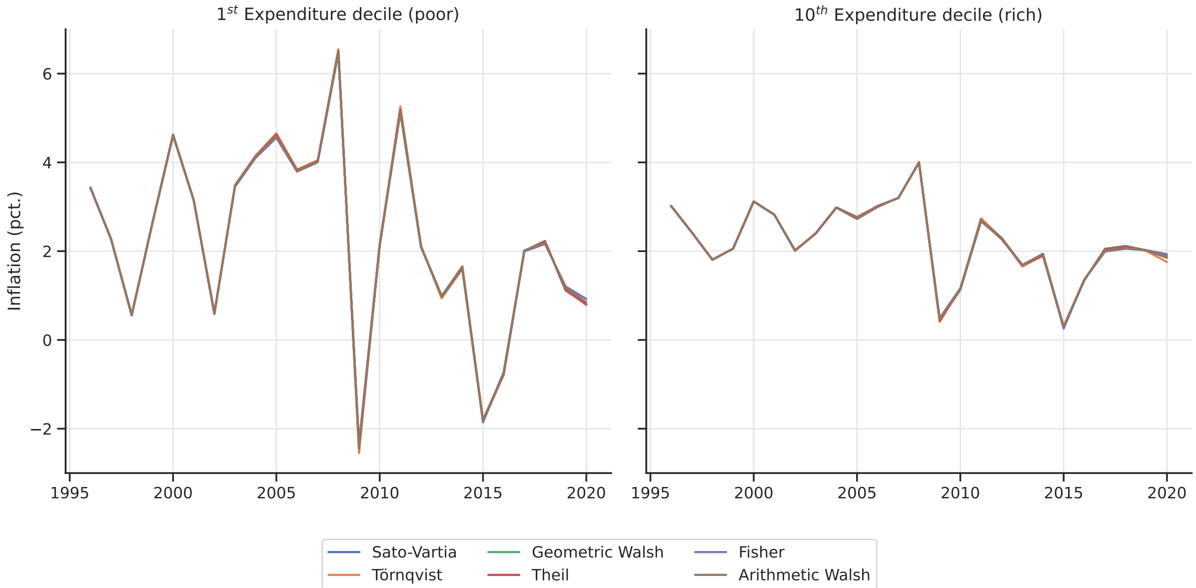


Figure B.8. Comparison of inflation across different generalized superlative indices.

Chapter 2

Quality and Consumption Basket Heterogeneity

Transitory Shocks and Implications for Consumption-Saving Behavior

Quality and Consumption Basket Heterogeneity[*]

Transitory Shocks and Implications for Consumption-Saving Behavior

Rasmus Bisgaard Larsen[†] Christoffer Weissert[‡]

Abstract

We study how the quality of households' consumption baskets varies with income using detailed household-level panel data on purchases. By exploiting the randomized disbursement timing of the Economic Stimulus Payments of 2008, we show that households increased spending when receiving the payment *and* spent more money on goods of higher quality. While the spending effects are concentrated among low-income households, the quality effects are driven by middle-income households. These findings support the theory of nonhomothetic demand. To model this, we embed nonhomothetic preferences over quantity and quality in an otherwise standard buffer-stock model. Contrary to the standard model, the nonhomothetic model can be used to match that the marginal propensity to spend is decreasing in income. Moreover, the calibrated model implies that households trade up in the quality of consumption when receiving a transitory income payment. Compared to the standard model, our nonhomothetic model also generates a more unequal wealth distribution, which is closer to the data.

[*] An earlier version of this chapter also featured in Rasmus B. Larsen's PhD dissertation "*Consumption Behavior, Price Dynamics, and Fiscal Spending*" (p. 1–76).

[†] Danmarks Nationalbank. Email: rbl@nationalbanken.dk.

[‡] University of Copenhagen and Danmarks Nationalbank. Email: cjw@econ.ku.dk.

1 Introduction

We explore one of the key aspects underlying households' consumption-saving decisions: the composition of their consumption baskets. The paper makes two empirical contributions using detailed data on U.S. household purchases. First, we show that households not only increase their spending but also the quality of products purchased when they receive an exogenous and positive transitory shock to income. Second, we show that the quality response is hump-shaped over the income distribution. For several reasons, this is important. In its own right, it deepens our understanding of consumer behavior. When studying aggregate consumption-saving dynamics, it furthermore delivers two key implications. Firstly, household preferences are nonhomothetic. Secondly, the hump-shaped quality response to a transitory income shock delivers a new fact to test model predictions against. To demonstrate the importance of our findings, we develop a model with heterogeneous household demand. The key novelty of the model is that it features nonhomothetic preferences, which stem from a microfounded consumption choice, where quality of the goods consumed enters the decision problem. Consistent with our empirical findings, the model predicts a hump-shaped quality response following a transitory income shock.

To set the stage, we first build a static model which embeds quality of goods in the utility function of the household. This allows us to show how consumption behavior depends on income via the quality channel. The model is similar to that of Handbury (2021) and Faber and Fally (2021) and distinguishes itself from standard models in two ways. First off, goods are grouped into product modules – such as fresh milk, shampoo and beer – and the expenditures allocated to each group depends on income via a Cobb-Douglas aggregator. Second, the quality of each good enters multiplicatively with the quantity of the same good in a constant elasticity of substitution (CES) utility function over all goods within a product module. Specifically, households' tastes for quality depend on income, which makes preferences nonhomothetic.

The static model lends itself in a useful way to an empirical investigation of the postulated channels. Most importantly, we answer the following questions: Does demand for quality depend on income? If yes, does quality demand also respond to transitory income shocks? Do product module expenditure shares depend on income? If yes, do they also depend on transitory income shocks? In chronological order, the answers are yes, yes, yes and no, and this serves as the justification of the exact specifications in the static model. We establish these results using detailed household-level panel data on purchases in 2008 from the Nielsen Consumer Panel Data (CPD) combined with scanner data for retail prices from the Nielsen Retail Scanner Data (RSD). Since the data set does not contain any measures of product quality, we construct a proxy for the quality of an individual product as its price relative to other products in the same product module. This approach to measuring quality is traditionally used in the literature (e.g. Jaimovich, Rebelo and Wong, 2019; Jaimovich *et al.*, 2020; Argente and Lee, 2021; Michelacci, Paciello and Pozzi, 2021) and is based on the assumption that consumers are willing to pay more for a product relative to other similar products because they perceive it to be of higher quality. Using this approach, we construct various quality indices that control for product size and link these to each household's purchases to construct a household-level measure of consumption quality.

Our empirical analysis proceeds in two steps. In the first step, we show that households with high income buy higher quality products than poorer households. Similarly, households that spend more also buy products of higher quality. These findings hold across almost all product modules and is robust to controlling for various demographic factors. We also document heterogeneity in the spending shares of product modules across the income distribution. In step two, we estimate households' spending and quality response to a positive transitory income shock in an event study research design. This is done by following the methodology of Broda and Parker (2014) and exploiting the randomized disbursement timing contained in the Economic Stimulus Act of 2008. Following this act, U.S. households received, on average, $900 in Economic Stimulus Payments (ESPs) during the spring and summer of 2008. Our estimates show that households not only temporarily increase spending when receiving an ESP, but also the quality of their purchases. When splitting our estimates by tertiles of annual income, we estimate that while the nominal spending response is higher for low-income households, the quality response is driven by both low and middle-income households. We find no significant evidence of spending switching across product modules when receiving an ESP.

Lastly, we incorporate the static model into a dynamic consumption-saving setup. This implies that the dynamic model features nonhomotheticicities in consumption. Except from these nonhomotheticities, the model is similar to the classical Deaton-Carroll buffer-stock model. While heterogeneous agent models have emerged as one of the most popular modeling frameworks in contemporary macroeconomics, only very few papers have used this framework to study heterogeneity in consumption baskets. In this paper, we bridge the gap between the recent literature on quality in consumption and the rapidly growing literature on heterogeneous agent models. As a key building block in this, we show how the static model can be expressed in a tractable way and subsequently built into the buffer-stock model.

We use the relative marginal propensity to consume (MPC) out of the ESPs between income groups as moments to calibrate the nonhomothetic model. A feature of the model is that it allows us to match these moments whereas the standard model does not.[1] Unconvincingly, the standard model not only misses the quantitative aspect but also predicts a positive relationship between permanent income and the MPC out of transitory income shocks. We do not target the estimates for the consumption quality response to the ESPs in our calibration, but assuringly

[1] Throughout, we will refer to the model with quality in consumption as the nonhomothetic model and to the classical buffer-stock model as the standard model.

our model predicts an inverse U-shaped relationship between permanent income and the quality response to a transitory income shock, which we also find in the data. We take both of these features of the nonhomothetic model to be evidence of the model successfully accommodating our empirical findings. To demonstrate the implications of taking quality in consumption into account, we show that this, among other things, implies that the wealth inequality increases more than threefold compared to the standard model.

Our work is related to four strands of literature. First, several papers have highlighted the link between business cycles and quality of consumption. Argente and Lee (2021) and Jaimovich, Rebelo and Wong (2019) show that households traded down in their consumption quality during the Great Recession, Jaimovich *et al.* (2020) document that household spending on high-quality products rises with income, and Jørgensen and Shen (2019) find that households' consumption quality is negatively correlated with local unemployment fluctuations. These papers emphasize how households' quality choice creates heterogeneity in inflation rates due to heterogeneous consumption baskets across the income distribution (Argente and Lee, 2021), restricts the ability of low-income households' to smooth consumption (Jørgensen and Shen, 2019), and that the relatively high labor-intensity of high-quality products amplifies output and employment fluctuations in business cycle models as well as affects skill premia in the labor market (Jaimovich, Rebelo and Wong, 2019; Jaimovich *et al.*, 2020). Our work differs from these paper by relating quality choice to a clearly transitory increase in income and exploring theoretically how consumption-saving behavior is affected in a buffer-stock model.

Second, this paper is related to an extensive literature on the estimation of marginal propensities to consume out of transitory income shocks. Most related are the papers by Sahm, Shapiro and Slemrod (2010), Parker *et al.* (2013), Broda and Parker (2014), Parker (2017) and Parker and Souleles (2019). They also exploit the 2008 ESPs to estimate marginal propensities to consume. However, these papers only consider responses in the dollar amount of spending without analyzing what kind of products enter households' consumption basket. To our knowledge, the only other paper that touches upon this is Michelacci, Paciello and Pozzi (2021). However, they focus on the adoption of new products rather than adjustments in the quality of products.

Third, our theoretical exercise is related to the literature studying consumption-saving behavior in buffer-stock models going back to the seminal work of Angus Deaton and Christopher Carroll (Deaton, 1991; Carroll, 1992; Deaton, 1992). Some authors have incorporated nonhomothetic preferences into these types of models through a bequest motive (De Nardi, 2004; Straub, 2019) or by including wealth directly into the utility function (Carroll, 2002). A few of these papers also analyze how households choose between quantity and quality but do so by modeling the choice between quantity and quality as the choice between basic and luxury goods (Wachter and Yogo, 2010; Campanale, 2018).

Lastly, our modeling approach is based on a framework in which nonhomotheticities are modeled as changing tastes for quality as in the work by Handbury (2021) and Faber and Fally (2021). This framework has been used extensively in international trade (e.g. Feenstra (1994) and Verhoogen (2008)) and the literature on estimation of price indices such as Broda and Weinstein (2010) and Redding and Weinstein (2019).

The paper proceeds in the following way. In Section 2, we present a static model in which preferences for different types of goods with different levels of quality depend on income. Section 3 describes the data while Section 4 presents the empirical evidence on the relationship between consumption quality and transitory income shocks that we use to discipline our model. Next, we incorporate the static model into a dynamic setup in Section 5 and explore the implications for and of consumption-saving behavior. Section 6 concludes.

2 Static model

In this section, we present a static model in which households derive utility from consuming goods that vary in terms of quality. It is important for two reasons. Firstly, it is directly related to our data presented in Section 3 and thus constitutes a close link between the empirical analysis and our modeling framework. It further disciplines our empirical analysis and acts as a guiding tool for understanding exactly how our empirical results feed back into the model. Secondly, when we set up the dynamic consumption-saving model, we build it on the microfoundation outlined in this section. Hence, this section provides intuition for the forces acting in the dynamic model.

2.1 Microfoundation with demand for quality

Borrowing directly from Handbury (2021) and Faber and Fally (2021), households receive an instantaneous utility from consuming goods that are characterized by belonging to a product module, m, being of a specific brand/product, i, and being of a given quality, φ_{mi}. For every module m, we denote the set of brands/products G_m. The quality assessment both has an "intrinsic" term, which is brand/product and module-specific and a "perceived quality" term, which is household-specific and depends on the income profile, $\{\xi, P\}$, of the household. We distinguish between transitory income shocks, ξ, and permanent income, P, which is conventional in the consumption-saving literature. The functional form of the instantaneous utility function is given by

$$U = \prod_m \left[\sum_{i \in G_m} \left(c_{mi} \varphi_{mi}(\xi, P) \right)^{\frac{\sigma-1}{\sigma}} \right]^{\alpha_m(P) \frac{\sigma}{\sigma-1}}, \tag{1}$$

where σ is the elasticity of substitution between brands/products, $\alpha_m(P)$ is the product module Cobb-Douglas weight, which depends on permanent income, and c_{mi} denotes quantity of good mi with quality $\varphi_{mi}(\xi, P)$. The way we let the product module weights and the quality term depend on permanent and transitory income is directly motivated by our empirical findings in Section 4. Specifically, we show that both α and φ depend on permanent income, and in Section 4.1 we show that only φ depends on transitory income. The quality assessment of a good is given by

$$\log \varphi_{mi}(\xi, P) = \gamma(\xi, P) \log \phi_{mi}, \tag{2}$$

where ϕ_{mi} denotes the intrinsic quality and $\gamma(\xi, P)$ denotes the income-specific quality term.

Before proceeding, we note that the utility function in Equation (1) may be re-written in a more conventional form as

$$U = \prod_m \left[\sum_{i \in G_m} c_{mi}^{\frac{\sigma-1}{\sigma}} b_{mi}(\xi, P) \right]^{\alpha_m(P)\frac{\sigma}{\sigma-1}},$$

where $b_{mi}(\xi, P) \equiv \varphi_{mi}(\xi, P)^{\frac{\sigma-1}{\sigma}} = \phi_{mi}^{\gamma(\xi,P)\frac{\sigma-1}{\sigma}}$ is the CES weight. From this, two things are worth noting. Firstly, the way households change their consumption baskets may be thought of as stemming from changes in the CES weights in the utility function. Second, the effect of an income shock (irrespective of it being a permanent or transitory income shock) can move these weights up *and* down, depending on the intrinsic value of the good.

To get a first impression of how quality matters in our setup, consider the relative demand of two goods, i and k, within module m, which is given by

$$\log \frac{x_{mi}}{x_{mk}} = (\sigma - 1) \left[\log \frac{\varphi_{mi}(\xi, P)}{\varphi_{mk}(\xi, P)} - \log \frac{\mathcal{P}_{mi}}{\mathcal{P}_{mk}} \right], \tag{3}$$

where $\mathcal{P}_{mi}$ denotes price of good mi and $x_{mi} \equiv \frac{c_{mi}\mathcal{P}_{mi}}{X}$ is the expenditure share out of total expenditures X. From this, it is clear that in the face of an income change, demand is shifted towards the goods that receive higher relative quality ratings. More so, given that the relative price and the elasticity of substitution between the two goods is constant, a change in relative expenditure shares must be due to a change in relative quality assessments. As we shall demonstrate in our empirical analysis, when households receive a positive, transitory income shock, relative expenditure shares are shifted toward more expensive goods, and Equation (3) shows that this may be explained by a relative increase in the quality assessment of the more expensive good.

Now, an important step in making the problem more tractable in the dynamic setup is to reformulate it in terms of indirect utility. In Section A, we show how we can represent Equation (1) as a function of prices, total expenditures, and income. Specifically, the aggregate price index is income-specific and given by $\mathcal{P}(\xi, P) \equiv \prod_m \mathcal{P}_m(\xi, P)^{\alpha_m(P)}$ with the module-specific price index, $\mathcal{P}_m(\xi, P)$, defined as

$$\mathcal{P}_m(\xi, P) = \left(\sum_{i \in G_m} \mathcal{P}_{mi}^{1-\sigma} \varphi_{mi}(\xi, P)^{\sigma - 1} \right)^{\frac{1}{1-\sigma}}, \tag{4}$$

by which we have that

$$U = \frac{X}{\mathcal{P}(\xi, P)} \prod_m \alpha_m(P)^{\alpha_m(P)} = \frac{X}{\mathcal{P}(\xi, P)} \cdot K(P). \tag{5}$$

Hence, utility maximization implies finding the optimal expenditure level given prices and income. This leads us to specify the utility function more generally as

$$U = X \cdot f(\xi, P),$$

where $f \equiv \frac{K(P)}{\mathcal{P}(\xi,P)}$ captures the nonhomotheticities in consumer demand.

Before proceeding, we note that the utility function implies that households will optimally consume a positive amount of each product within a module conditional of purchasing products from that module. It does not imply that households will buy products from all modules. As we show in Section 3.4, the latter is in accordance with our data since households only purchase products from around a fifth of the modules.[2]

2.2 An illustrative, two-period perfect foresight example

To get an idea of how the dynamic model works, we here present a simple two-period perfect foresight model with instantaneous utility as in Equation (5). Let the problem of the household

[2] Households typically only purchase a unique product within a module in a given week, which is at odds with the CES structure. Nonetheless, we use the CES structure for tractability. Product choice could alternatively be modeled using a logit discrete-choice framework with quality shifters and household-level taste shocks. This type of preferences implies that households only consume a unique good within a module but as Faber and Fally (2021) show, the preferences in equation (1) can be derived from aggregation of discrete-choice preferences across many households. Equivalently, the preferences in Equation (1) hold at the household level in expectations in the logit model. These results mirror those of Anderson, Palma and Thisse (1987).

be given by

$$\max_{X_1,X_2} \frac{\left(\frac{X_1 \cdot K(P_1)}{\mathcal{P}(\xi_1,P_1)}\right)^{1-\rho}}{1-\rho} + \beta \frac{\left(\frac{X_2 \cdot K(P_2)}{\mathcal{P}(\xi_2,P_2)}\right)^{1-\rho}}{1-\rho},, \qquad \text{s.t. } X_1 + X_2 = \bar{X},$$

which yields the solution

$$X_1 = \bar{X} \frac{1}{\beta^{\frac{1}{\rho}} \left(\frac{\mathcal{P}(\xi_2,P_2)}{\mathcal{P}(\xi_1,P_1)} \frac{K(P_1)}{K(P_2)}\right)^{\frac{\rho-1}{\rho}} + 1}.$$

For ease of understanding, consider the case under which the household does not discount future consumption and prefers perfect consumption smoothing, i.e. $\beta = 1$ and $\rho \to \infty$, by which the expression reduces to

$$X_1 = \bar{X} \frac{1}{\frac{\mathcal{P}(\xi_2,P_2)}{\mathcal{P}(\xi_1,P_1)} \frac{K(P_1)}{K(P_2)} + 1} = s \cdot \bar{X},$$

where $s \equiv \frac{1}{\frac{\mathcal{P}(\xi_2,P_2)}{\mathcal{P}(\xi_1,P_1)} \frac{K(P_1)}{K(P_2)}+1}$ denotes the share, which is spent in period 1 out of total expenditures.

Now, in the standard case, $\mathcal{P}(\xi_1, P_1) = \mathcal{P}(\xi_2, P_2) = K(P_1) = K(P_2) = 1$, and hence the household divides expenditures evenly across the two periods ($s = \frac{1}{2}$).

In our case, however, the share depends on the income profile of the household in the two periods. Say, for example, that the household has a low income in the first period and high income in the second period. For simplicity, assume that this is purely transitory so that $K(P_1) = K(P_2)$. Clearly, s then depends on whether $\mathcal{P}(\xi_1, P_1) \lesseqgtr \mathcal{P}(\xi_2, P_1)$. A priori, we cannot determine the inequality. The calibrated dynamic model in Section 5, however, implies that $\mathcal{P}(\xi_1, P_1) > \mathcal{P}(\xi_2, P_1)$. This also corresponds to the household valuing quality more when income is high. In this case, $s > \frac{1}{2}$ and the household smooths utility by front-loading expenditures to the first period. The intuition behind this is that the household, rather than distributing utility unequally over the two periods, forgoes some consumption of high-quality goods in the second period in order to increase quantity of the low-quality good in the first period.

3 Data description

We construct a weekly panel of households covering 2008 using the Nielsen Consumer Panel Data, which is combined with a survey among the households on the Economic Stimulus Payments of 2008. This panel is linked to data from the Nielsen Retail Scanner Data to measure the quality of goods purchased at the household level.

3.1 The Retail Scanner Data

The RSD contains weekly pricing and quantity information at the product level from more than 90 retail chains across the contiguous United States. The data set covers approximately 3.2 million products – both food and non-food groceries – sold from over 35,000 different stores making up about half of all sales from food and drug stores and a third of all sales from mass merchandisers. Data is recorded at the point-of-sale, which can be matched with geographic identifiers for each store down to the zip-code level.

Products are identified by their Unique Product Code (UPC) – i.e. barcode – and we treat each of these UPCs as an individual product indexed by i. Additionally, the brand of each product is indicated in the data. UPCs are grouped into an hierarchical structure by Nielsen. At the most granular level, UPCs are grouped into 1,086 product modules, which we index by m.[3] The modules are grouped into around 120 product groups, which are aggregated to 11 product departments. Figure 1 shows an example of the rich detailedness of the data. The product department *dry grocery* has a product group called *snacks*, which has a product module called *snacks – potato chips*. One of the UPCs in this module is a *2-pack of Pringles Sour Cream and Onion tubes*, which belongs to the brand *Pringles*.

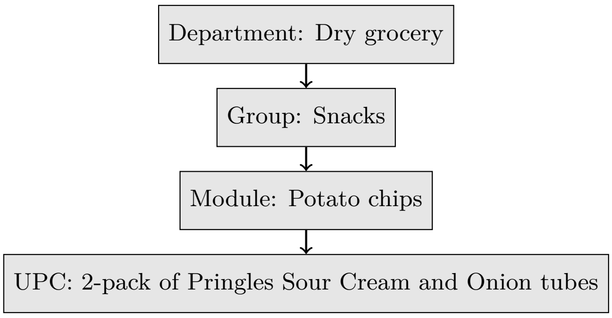


Figure 1. Example of data structure

Nielsen also provides information on the attributes of each UPC such as size in physical units (e.g. 2 liters of milk or 1 pounds of nuts) and multi-pack information on how many of those goods appear in a given pack (e.g. a six-pack of soda or a carton of 8 eggs). We treat all possible combinations of physical units and multi-pack information within a module m as a unique product size indexed by $s \in S_m$.

The UPCs of private-label products are altered by Nielsen who assign the same UPC to private-label products with identical core attributes, while the brand code assigned to all private-label

[3] In addition, there are slightly fewer than 200 modules consisting of so-called magnet products that do not use regular UPCs and are only tracked in the CPD (typically fresh produce). We exclude products from these modules from our data.

products is the same. This is done to preserve the anonymity of the retail stores reporting data to Nielsen. We include all products in our analysis, which effectively means that we might lump some different private-label UPC with identical attributes together into a single UPC within a module. More importantly, all private-label products are lumped into the same brand within each module. For example, two private-label products in the module *ground and whole bean coffee* get the same brand code even though they are two different products sold by two different retail chains.[4]

3.1.1 Measuring quality in the Retail Scanner Data

We construct a number of quality indices for products in the RSD using the relative prices of similar products as a proxy for quality.[5] The assumption underlying this approach is that quality is an intrinsic product attribute that all consumers agree on. As in the static model of quality choice presented in Section 2, consumers agree on the quality ordering of products within a module through the intrinsic quality term, ϕ_{mi}, in Equation (2). However, they do not necessarily agree on how they value quality as a product attribute. This leads to a quality ranking of products within each product module that is equal to the ranking of prices on average.[6]

Our first index – henceforth, the size-based quality index – measures the quality of a product relative to other products of the same size, s, sold within its product module, m, and core-based statistical area (CBSA), c.[7] E.g., we compare the price of a six-pack of 12 oz Coke cans to the price of a six-pack of 12 oz Pepsi cans that are both sold in the Dallas-Forth Worth-Arlington metropolitan area. We construct CBSA-specific quality indices to account for geographic differences in product assortment that limit households' ability to climb up or down the entire national quality ladder. Indeed, Handbury (2021) shows that there are large, systematic differences in product assortment across cities since stores in wealthy cities cater to high-income households by skewing their assortment toward high-quality products. Moreover, we compare the prices of similar-sized products to take into account that large sized items are often cheaper (Nevo and Wong, 2019).

The weekly prices of each product are converted into an annual quantity-weighted average

[4] Dubé, Hitsch and Rossi (2018) show that there has been a rise in the market share of private-label products in the Nielsen data over the last decade, while Nevo and Wong (2019) document that households purchased more private-label products during the Great Recession. Also Stroebel and Vavra (2019) show that homeowners purchase fewer private-label products when local house prices rise, while Dubé, Hitsch and Rossi (2018) estimate a negative effect of income on private-label purchase shares.

[5] As mentioned earlier, Jørgensen and Shen (2019), Jaimovich *et al.* (2020), and Argente and Lee (2021) take a similar approach to measuring quality of products in the Nielsen data.

[6] Other authors have confirmed a positive correlation between prices and other measures of quality. For example, Jaimovich, Rebelo and Wong (2019) find a positive correlation between customer ratings and price of goods and services using data from Yelp!.

[7] CBSAs are geographic areas consisting of one of more counties anchored by an urban center of at least 10,000 people plus adjacent counties that are socioeconomically tied to the urban center by commuting.

price, $p_{i,m,s,c}$, using the prices of all stores selling the product within CBSA c. This removes seasonalities from the quality measure.

Let the size-based index, $q^j_{i,m,s,c}$, be denoted by superscript j. For a product i of size s belonging to module m and sold in CBSA c, the index is constructed as the standardized log-distance from the product's annual price to the median annual price, $\bar{p}_{m,s,c}$, of all products of the same size s in the same module m and sold in the same CBSA c:

$$q^j_{i,m,s,c} = \frac{\ln p_{i,m,s,c} - \ln \bar{p}_{m,s,c}}{\sigma_{m,s,c}} \tag{6}$$

where $\sigma_{m,s,c}$ is the standard deviation of the log-distance of annual prices, $\ln p_{i,m,s,c} - \ln \bar{p}_{m,s,c}$. The index is standardized to allow for comparison of the quality measure across product modules.[8]

Naturally, the index cannot be computed for a product that is the only product sold of a given size in its module within a CBSA. We exclude these products from our analysis.

As alternatives to the first index, we construct two other indices that are based on unit prices (that is, the price per physical unit of the product). For 334 of the 1,086 product modules, the products are measured in multiple physical units – e.g. mass and volume – which complicates the construction of unit prices. However, 181 of these multi-unit modules contain products for which at least 99 percent are measured in the same physical unit. In this case, we remove the fewer than 1 percent of products that are of another physical unit and calculate unit prices for the remaining products. This leaves us with 933 modules for which we calculate unit prices.

Let $p^u_{i,m,c}$ be the annual quantity-weighted average of the unit price for product i sold in CBSA c. The unit price-based quality index, $q^k_{i,m,c}$, denoted by superscript k, is then the standardized log-distance to the median price of products in the same module sold in the CBSA:

$$q^k_{i,m,c} = \frac{\ln p^u_{i,m,c} - \ln \bar{p}^u_{m,c}}{\sigma^u_{m,c}}. \tag{7}$$

To assess if households also switch between brands of different quality, we lastly construct a brand-based quality index, $q^l_{b,m,c}$, denoted by superscript l. The index is constructed in the same manner as for the unit price-based index but using the quantity-weighted average unit price,

[8] A histogram of the quality index is shown in Figure B.1 of Section B. We have also confirmed that the ranking of products is similar in other years as shown in Section C.

$p^u_{b,m,c}$, of products belonging to a given brand b:

$$q^l_{b,m,c} = \frac{\ln p^u_{b,m,c} - \ln \bar{p}^u_{m,c}}{\sigma^b_{m,c}} \tag{8}$$

As with the size-based index, both of these indices are standardized to allow for comparison of the indices between product modules.

3.2 Measuring household-level quality using The Nielsen Consumer Panel Data

The CPD is a household panel that includes 61,440 households in 2008.[9] The households record information about which products they buy as well as where and on which date the products were purchased. In addition, the households provide demographic information such as income, education, employment status and household composition in the fourth quarter prior to the panel year. Most of the demographic information is provided in brackets (e.g. income is reported in 19 brackets). Since panelists are not representative of the U.S. population, Nielsen provides weights to make the sample representative of the population.

A word of caution is warranted regarding the income variable, which we will use when exploring heterogeneity in consumption responses. Income in the Nielsen data is self-reported, likely suffers from non-classical measurement error, and households are asked to report annual income that they earned two calendar years prior to the panel year. To be exact, households in the 2008 CPD are surveyed in the Fall of 2007 about their annual income in 2006. However, Nielsen believes that households are actually reporting their annualized income as of the time of the survey. Thus, the income variable is likely a noisy measure of income in the Fall of 2007.[10] For our analysis, we exclude households with annual income below $5,000 – the lowest income bracket – since we suspect that income reported by these households does not reflect their actual income. These low-income households constitute very little, only 0.8 percent, of the household panel in 2008.

Households record information about shopping behavior by scanning barcodes after each shopping trip using a scanning device. Prices are automatically filled in if the purchase was done at a store partnering with Nielsen. If not, households must enter the prices themselves. Additionally, households must enter the number of units purchased of each product and indicate if each product was on sale or purchased using a coupon. Not all products purchased by the households

[9] Panelists are randomly recruited either via mail or through the Internet. They are not paid but provided incentives to join and stay active. These incentives are designed to be non-biasing in selection of retailers and products. About 80 percent of panelists are retained each year.

[10] Kueng (2018) highlight similar concerns about using self-reported income in the context of estimating the MPC to payments from the Alaska Permanent Fund.

are scanned and registered as individual products.[11] Some products – such as most apparel – are not coded by Nielsen and therefore not tracked as individual products. However, the total expenditures on these not-coded products are still tracked.

We link each product purchase made in week t by household h residing in CBSA c to each of the three quality indices for CBSA c.[12] For each of the three quality indices $o \in \{j, k, l\}$ defined in Section 3.1, we then construct an aggregate quality measure, $Q^o_{h,t}$, for purchases made in week t by household h as the expenditure-weighted averages of the quality of the households' purchases:

$$Q^o_{h,t} = \sum_m \sum_{i \in G_m} w_{i,h,t} q^o_{i,m,c(h)}, \qquad o \in \{j, k, l\}, \tag{9}$$

where $w_{i,h,t}$ is the expenditure share of good i in household h's consumption basket in week t and $q^o_{i,m,c(h)}$ is one of the three quality indices.

Not all purchases can be matched with the quality indices. This is either because 1) the product only occurs in the CPD data but not the RSD, 2) because the product was bought in another CBSA and not sold by any store within the household's CBSA of residence, or 3) the product is a magnet product for which we do not construct the quality index. In addition, a missing match occurs if the product is a unique size for the size-based index, if a product is in one of the modules with multiple physical units for the two indices based on unit prices, or if a product is a unique brand for the brand-based index.

As mentioned above in Section 3.1, the size-based index has the benefit of comparing products of the same size to each other. By contrast, the two other quality indices are based on unit prices and therefore compare products of different sizes. For these two indices, this has the unfortunate by-product of introducing a negative correlation between the indices and product size unrelated to actual product quality because larger products are often cheaper per physical unit. To illustrate this, Figure B.2 in Section B shows binned scatter plots of households' weekly expenditure share of their purchases that are in the top 40 percent of the size distribution of products within product modules against weekly spending in Panel (A) as well as the quality of their weekly purchases according to the three quality indices in Panels (B)-(D).[13] Panel (A) shows that weekly spending and purchases of large products are positively correlated. Hence, when households increase spending, they tend to buy larger products as well. There is no systematic correlation between the quality of purchases according to the size-based index and the purchases of large products as seen in Panel (B). However, there is a clear negative correlation between

[11] Nielsen estimates that around 30 percent of household consumption is covered by the categories tracked in the data.
[12] 3,901 of the households do not live in a CBSA. We exclude these households from the data.
[13] This definition of large-sized products follows Nevo and Wong (2019).

the purchases of large products and quality of purchases according to the unit price-based and brand-based indices as shown in Panels (C) and (D). Hence, we prefer the size-based index over the two other indices since it is not affected by product size.

Another weakness of the brand-based index besides it being influenced by product size, is the presence of private-label products. As mentioned above, all private-label products are lumped into the same brand within a module. Thus, any switching between private-label brands, either within or across stores, will not affect the brand-based quality index.[14] Private-label products make up 16.5 percent of households' annual purchases on average. Moreover, this share is decreasing in annual spending and income as shown in Figure B.3 of Section B. While 12.8 percent of purchases are private-label products on average for households in the top income category (those with an annual income above $200,000), the same share is 20 percent on average for households in the bottom income category (those with an annual income between $5,000 and $8,000). Similarly, the binned scatter plot of annual spending against the share of private-label products shows that the private-label share ranges from 9 to 23 percent.

3.3 Nielsen Consumer Panel Data survey on ESP

We get information on ESPs received by the CPD households using a survey that was originally conducted by Nielsen on behalf of Christian Broda and Jonathan A. Parker. A detailed description of the survey is presented by Broda and Parker (2014) but we provide basic information about it below.

The survey consisted of two parts, which were to be answered by the adult most knowledgeable about the household's income and tax returns. The first part of the survey contained questions about the household's liquid assets and household behavior, while the second part described the ESP program and asked the household if it had received the ESP. If the household responded yes to receiving the payment, it was also asked about the amount, date of arrival and whether it was received by check or direct deposit in addition to some questions about the household's usage of the ESP.

The survey was fielded in multiple waves by either email or regular mail to all households meeting Nielsen's static reporting requirement for January through April 2008. This amounted to 46,620 households receiving the survey by email and 13,243 receiving the survey by regular mail. Households with internet access and in contact with Nielsen by email received the survey in three waves in a web-based version, while other households received the survey in two waves in a paper/barcode scanner version. Households were surveyed repeatedly conditional on their

[14] Coibion, Gorodnichenko and Hong (2015) show that households shift expenditures toward lower-price retailers when local economic conditions deteriorate. If such a shift is made from private-label to private-label product, it will not be picked up by the brand-based quality index.

earlier responses.[15] The response rate after all waves was 80 percent.

Some households reported not receiving any ESP or provided inconsistent survey answers. We handle this by dropping households from the sample following the procedure by Broda and Parker (2014). First, we drop all households that do not report receiving any ESP (around 20 percent of the respondents) or do not report a date for receiving the ESP. This is done because non-recipients of the ESP do not make up an appropriate control group due to selection into receiving an ESP. Additionally, we want to rule out the possibility of households misreporting that they did not receive ESP even though they actually did. Second, we remove households reporting in one survey that they did not receive an ESP and in a later survey report receiving an ESP prior to the response to the earlier survey. Third, we drop households reporting that they received an ESP on a date after they submitted the survey. Fourth, we drop households reporting that they received an ESP outside the period of randomized disbursement. We allow a grace period of two days for misreporting relative to survey submit dates and a grace period of seven days for misreporting relative to the disbursement period. This procedure reduces the sample to 29,205 households. The survey is then linked with the CPD giving us a final sample of 20,174 households for the transitory income shock analysis.[16]

3.4 Summary statistics

Table 1 shows some summary statistics for the full sample in 2008 as well as the sample used for the analysis of the ESP.

Table 1. Summary statistics

	ESP sample			Full sample		
	Mean	S.d.	Median	Mean	S.d.	Median
Annual spending, $	7673.4	4692.3	6599.8	7816.7	4723.1	6767.5
Products bought	1044.5	590.3	931.0	1043.3	583.8	937.0
Unique modules bought	215.6	70.7	217.0	216.3	70.4	217.0
Unique groups bought	76.5	12.9	79.0	76.6	12.8	79.0
Unique products bought	572.9	284.2	530.0	571.1	279.0	531.0
% spending in size-based index	48.1	16.9	48.8	47.7	16.9	48.3
% spending in unit price-based index	45.3	16.0	45.8	44.8	16.0	45.4
Household size	2.4	1.4	2.0	2.4	1.3	2.0
No. of households		20,174			57,049	
% with income below $35,000		37.0			38.3	
% with income $35,000-$70,000		29.6			33.8	
% with income above $70,000		33.4			27.8	

Notes. The table shows summary statistics for the sample used in the ESP analysis and the full CPD panel of 2008.

[15] If households completed part one of the survey, they were not asked part one again but resurveyed with part two only. Households reporting ESP information in part two were not resurveyed, while households reporting that they had not received an ESP in part two were resurveyed using part two only.

[16] Although our ESP sample is based on the same data as Broda and Parker (2014) use, it contains fewer households than their sample of 21,760 households since we exclude households that do not live in a CBSA.

Number of goods purchased and household size in the ESP sample are roughly the same on average as in the full sample. However, there are slightly more high-income households in the ESP sample, while annual spending is on average somewhat higher in the full sample. In both the full sample and the ESP subsample, we can link almost half of annual purchases to the quality indices on average. It is also worth noting that although there are 1,086 product modules available in the data, the typical household only buys product from around a fifth of these modules.

Next, we divide the ESP sample into 3 income groups that are of roughly equal size as by Broda and Parker (2014). Table 2 shows, for each group, the same summary statistics as before along with statistics for the ESP received. The cutoffs for these groups also correspond to the tertiles in the 2007 household income distribution reported in the Current Population Survey released by the Census.

TABLE 2. Summary statistics for ESP sample by income tertile

	Below $35,000			$35,000-$70,000			Above $70,000		
	Mean	S.d.	Med.	Mean	S.d.	Med.	Mean	S.d.	Med.
Annual spending, $	6054.6	3848.3	5167.1	7806.4	4498.2	6883.9	9332.9	5162.5	8295.6
Products bought	935.4	538.7	821.0	1074.2	610.6	956.0	1130.3	601.7	1035.5
Unique modules bought	199.2	67.6	198.0	218.7	72.1	221.0	230.2	68.7	233.0
Unique groups bought	73.6	13.2	76.0	76.9	12.9	79.0	79.1	11.8	81.0
Unique products bought	513.1	258.7	470.0	587.6	294.3	545.0	622.0	287.0	589.0
% spending in size-based index	47.4	17.1	48.0	48.5	16.8	49.2	48.6	16.8	49.3
% spending in unit price-based index	44.4	16.2	44.7	45.6	15.8	46.0	45.9	15.9	46.5
Household size	1.9	1.2	2.0	2.4	1.4	2.0	2.8	1.3	2.0
ESP received	595.5	369.9	600.0	949.0	495.3	900.0	1128.0	502.9	1200.0
No. of households		6,737			7,463			5,974	

Notes. The table shows summary statistics for the ESP sample by income groups.

Annual spending, household size and the number of purchases is increasing in annual income. So is the ESP received, which will be important to keep in mind when analyzing the effects of the ESP across income groups. Households with higher income buy a larger number of unique products as well as a larger number of product categories as measured by either products modules or group. Reassuringly, the share of spending that we can match with the quality indices is almost the same across the three income groups. Thus, our analysis of consumption quality across income groups is not affected by heterogeneity in matching purchases to the quality indices across income.[17]

Our analysis of the effect of the ESP on quality is complicated by quality only being observed in weeks, where households actually purchase goods. Moreover, purchasing patterns across weeks are not random. Panel (A) of Figure 2 shows the distribution of the number of weeks for which

[17] Figure B.4 in Section B shows the average spending coverage for the income bins in our data. Coverage is approximately constant across the income bins.

we observe purchases in 2008 across the households in our ESP sample. The distribution is negatively skewed, and the median household made purchases in 44 weeks of 2008. There is no notably difference in this pattern by the three income groups as shown in Figure B.5 of Section B. Panel (B) of Figure 2 shows that the number of households making at least one purchase in a given week is evenly spaced across the year except for fewer purchases in the first and last weeks of the year.

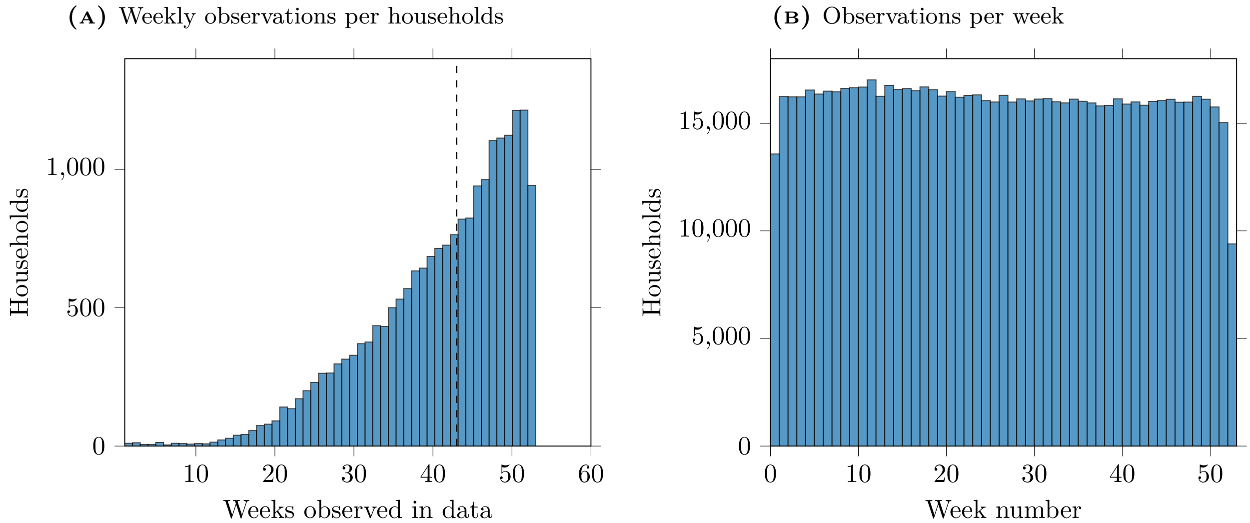


Figure 2. Weekly purchasing patterns in the ESP sample

Notes. Panel (A) shows the distribution of households by the number of weeks in 2008 that we observe purchases for each household in the ESP sample. The vertical, dashed line indicates the number of weeks with observed purchases for the median household. Panel (B) plots the number of households making at least one purchase for each week of 2008.

To get a sense of the variation driving the ESP estimates, we plot the total ESPs per week in our sample in Panel (A) of Figure 3 along with the total amount of ESPs disbursed according to the Daily Treasury Statements. Panel (B) of Figure 3 shows the number of households in our sample receiving an ESP per week.

The ESPs were disbursed in every week from April 14 until July 25 but there is significant variation in the weekly disbursement amounts. Our sample tracks the weekly ESP disbursements reported in statements from the Treasury reasonably well although the survey tends to underreport payments in the later weeks of the ESP program.

4 Empirical results

We begin our analysis by exploring the relationship between quality of purchases, annual spending and income in the cross-section. This is done using the full sample of households that we observe

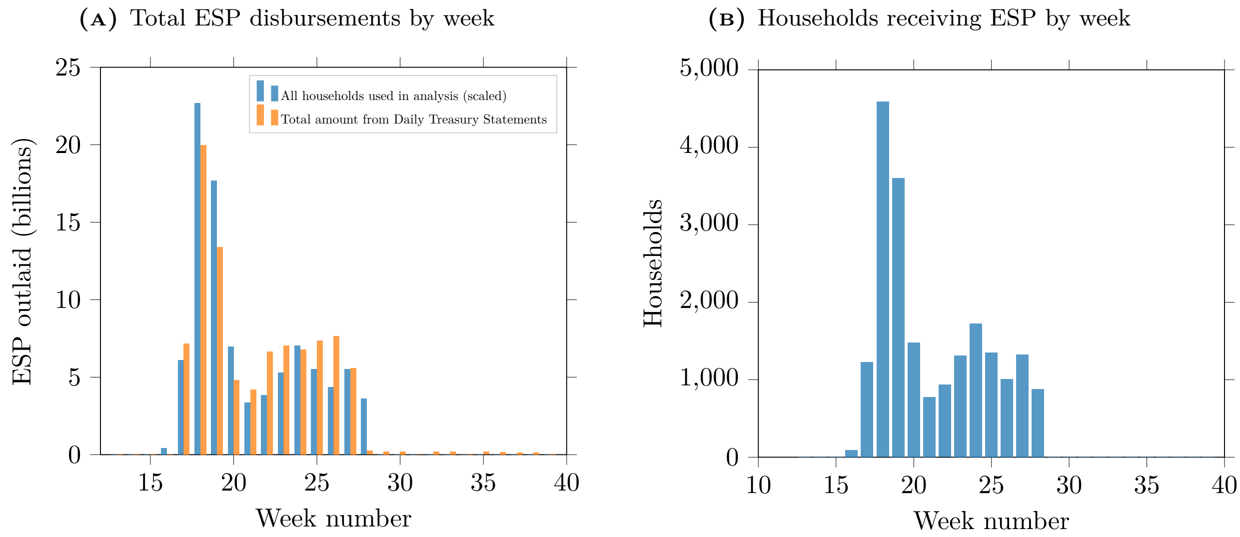


FIGURE 3. ESP disbursements in sample

Notes. Notes: Panel (A) shows the total weekly disbursements of ESPs according to the survey (blue bars) along with the disbursements according to the Daily Treasury Statements (orange bars). The weekly ESP disbursements are scaled such that the sum of ESPs in the survey matches the sum of ESPs in the Treasury data. The Treasury data have been adjusted for the 4th of July holiday. Panel (B) shows the number of households receiving an ESP payment per week in our sample.

in 2008. Although we do not show this, only including households that are in the ESP sample yields virtually identical results.

Figure 4 shows the average quality of the households' annual consumption basket by deciles of the annual spending distribution in the top panels and income brackets in the bottom panels. Using the terminology of Bils and Klenow (2001), the figure plots "quality Engel curves" that trace out the relationship between quality of consumption and income or spending across households. Panels (A) and (D) use the quality index, which compares prices of products of identical size, Panels (B) and (E) in Figure 4 use the quality index that compares unit prices of products, and Panels (C) and (F) use the quality index comparing the average unit prices of different brands. The blue lines indicate the unconditional averages, while the orange lines plot conditional averages that control for household size, race, CBSA of residence, age bracket of both household heads and the education level of both household heads using fixed effects. 95 percent confidence intervals are indicated by error bands.

Households with higher annual spending or higher annual income consume goods of higher quality. This is in line with previous findings from the literature using CEX data (Bils and Klenow, 2001; Jaimovich *et al.*, 2020) as well as CPD data (Jaimovich *et al.*, 2020; Argente and Lee, 2021; Faber and Fally, 2021). The positive relationship holds across the spending and income distribution and is precisely estimated. When we control for other factors that might be

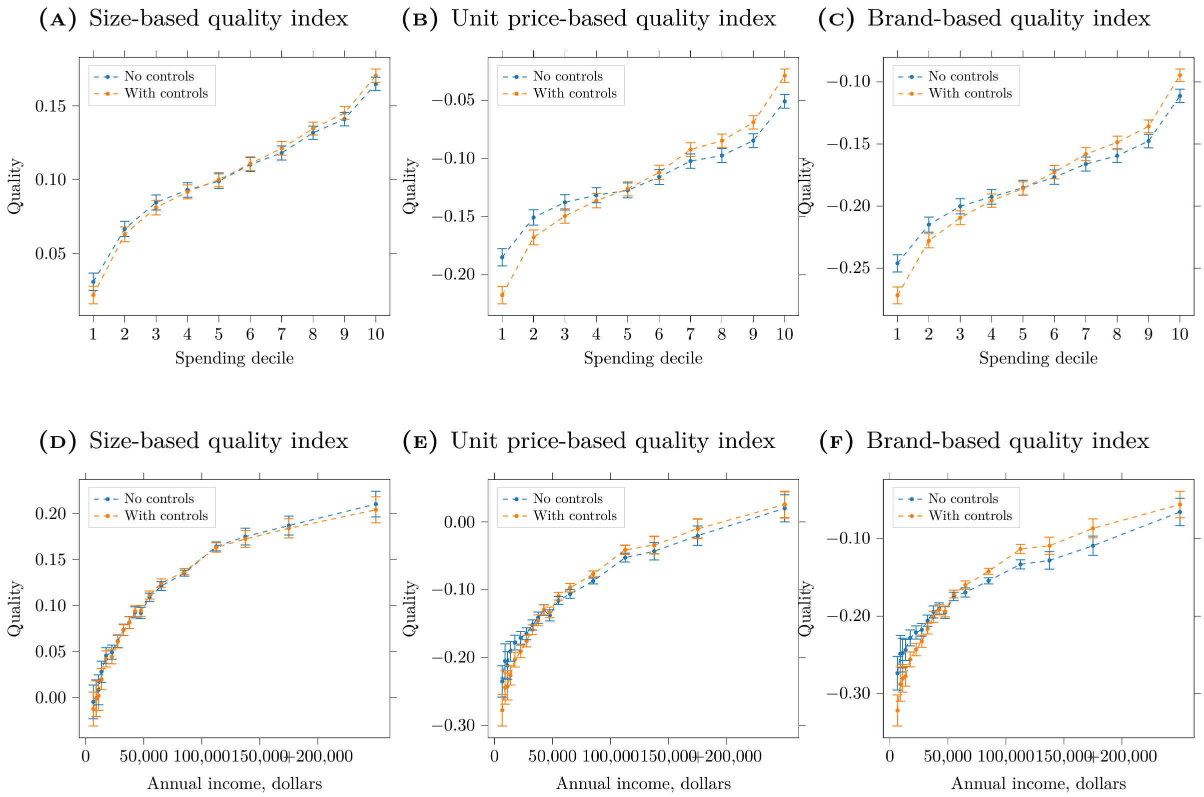


Figure 4. Quality across the spending and income distribution

Notes. The top three panels show the average quality of households' annual purchases for 10 annual spending deciles, while the bottom three panels show average quality of households' annual purchases for the income brackets. We assign income levels to the midpoints in these brackets. Panel (A) and (D) plot the quality using the index that compares products of same sizes, Panel (B) and (E) plot quality using the index based on unit prices, and Panel (C) and (F) plot quality using the index based on brands. The blue lines show the average quality, while orange lines control for household size, race, CBSA of residence, age brackets of both household heads and education levels of both household heads. 95 percent confidence bands based on heteroskedasticity-robust standard errors are indicated by error bars.

correlated with both spending/income and quality – age, household size, race, CBSA of residence, and education – the positive relationship is even more pronounced.

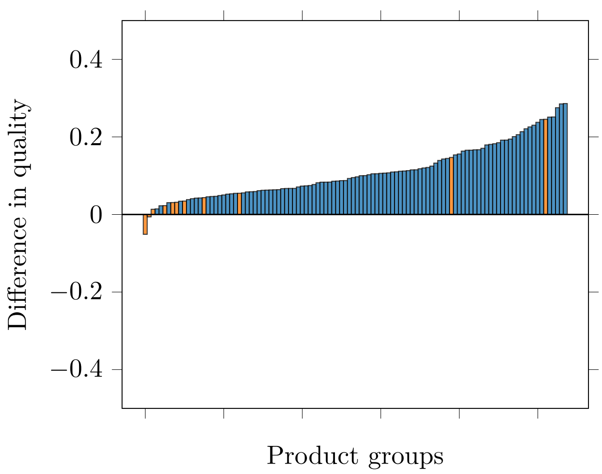
The relationship between quality, spending and income generally holds within the product groups. To show this, we divide the sample into expenditure and income quintiles and estimate the average quality of households' purchases for each quintile within product groups, g, using the following regression equation:[18]

$$Q_{h,g} = \sum_{k=1}^{5} \beta_{g,k} \mathbf{1} \{\text{Quintile}_h = k\} + \Gamma_g X_h + \varepsilon_{h,g}, \tag{10}$$

where X_h is as a vector of controls (household size, race, age brackets of the household heads, education of the household heads, and CBSA of residence).

$\beta_{g,k}$ in Equation (10) is the average quality of purchases by households in expenditure/income quintile k on products from product group g conditional on the controls, X_h. We rank the estimates according to the difference between average quality in the top and bottom expenditure/spending quintiles, $\hat{\beta}_{g,5} - \hat{\beta}_{g,1}$, and plot these differences in Figure 5. Blue bars indicate that the difference is significant at the 5 percent level. The estimates based on spending quintiles are shown in the left panel, while estimates based on income quintiles are shown in the right panel. For all but 2 of the 108 product groups, the top spending or income quintile households consume higher quality goods on average. The difference in average quality is significant for the majority of the groups. For the two groups where $\hat{\beta}_{g,5} - \hat{\beta}_{g,1} < 0$, the difference is insignificant.

(A) Difference in quality by spending quintiles



(B) Difference in quality by income quintiles

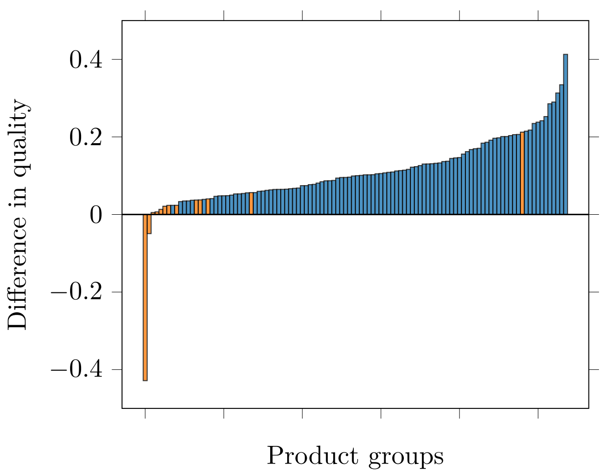


Figure 5. Difference in quality by product group

Notes. The figure shows the difference between average quality of products purchases by households at the top and bottom annual spending (Panel A) and income (Panel B) quintiles for each product group ($\hat{\beta}_{g,5} - \hat{\beta}_{g,1}$ from the regression in Equation (10)). The quality variable is the quality of households' annual consumption basket measured using the size-based quality index. The dependent variable is winsorized at the 0.01 and 99.99 percentile to limit the influence of outliers, and we only include product groups that are purchased by at least 50 households in our sample. Blue bars denote that the difference is significant at the 5 percent level.

[18] We estimate the regression at the group level instead of the more granular module level since the typical household only buys products from 217 modules as shown in Section 3.4.

We also observe heterogeneity in the annual expenditure shares of the product modules across income. This is investigated by estimating Equation (10) with expenditure shares of the modules as the dependent variable and testing that the shares are equal across quintiles, $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5$, using a Wald test. Figure 6 plots a histogram of the resulting F-statistics together with the distribution of the statistic under the null hypothesis.[19]

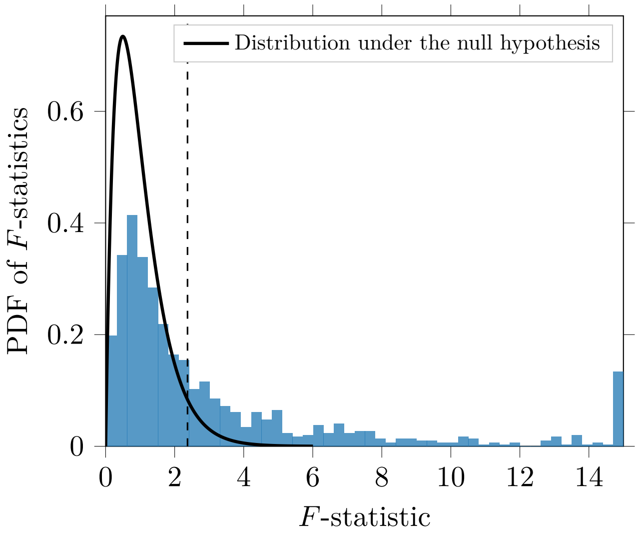


FIGURE 6. Distribution of test statistics for constant product module spending shares across income

Notes. The figure shows a histogram of F-statistics for the Wald test under the null of $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5$ in regression Equation (10), where the dependent variable is the household's annual spending share of a product module. The F-statistics are winsorized at 15 for illustrative purposes. The solid line shows the distribution of the test statistic under the null, which is an F distribution with 4 and $N \to \infty$ degrees of freedom. The vertical, dashed line is the distribution's 95th percentile.

The estimated F-statistics do not fit the distribution under the null hypothesis of no difference in spending shares across the income quintiles. There is a large mass of test statistics above the 95th percentile of the distribution, where 37 percent of the F-statistics lie.[20] Thus, spending on each module is not scaled up proportionally with total spending when comparing households across the income distribution. This implies that the expenditure share for each module, $\alpha_n(P)$, in our structural model introduced in Section 2, should depend on income.

4.1 Transitory income shocks and quality of consumption

The results above show that households with higher spending and income consume products of higher quality compared to households with lower expenditures and income. This was a purely cross-sectional result. Although the correlation is robust to controlling for various demographic

[19] The statistic under the null has an F-distribution with 4 and N degrees of freedom, where N is the number of observations in the regression. We plot the distribution for $N \to \infty$ as most of the regressions include large N.

[20] More formally, a Kolmogorov–Smirnov test rejects that the F-statistics follow an F distribution at any conventional significance level.

factors, it does not necessarily reflect a causal link from spending or income to quality of consumption.

In the following section, we show that a transitory income shock in the form of an ESP results in a temporary increase in the quality of products consumed. That is, the relationship between quality and income not only holds across but also within households. Additionally, these results have a clear causal interpretation due to the randomized timing of the ESP disbursement: a temporary increase in income causes a temporary increase in the quality of consumption.

4.1.1 Empirical framework

The Economic Stimulus Act of 2008 was signed by Congress in January 2008 and enacted on February 13, 2008. The act authorized the distribution of stimulus payments, the ESPs, to tax payers during the Spring and Summer of 2008. A basic payment was distributed as the maximum of $300 ($600 for joint filers) and a taxpayer's tax liability up to $600 ($1,200 for joint filers). Households received this payment as long as they had at least $3,000 of qualifying income. An additional payment of $300 was given per child that qualified for the child tax credit. The total payment was reduced by five percent of the amount by which adjusted gross income exceeded a threshold of $75,000 ($150,000 for joint filers). Hence, payments were made to the bulk of households along the income distribution except those at the very bottom or those at the very top. These payments were disbursed to households by either paper check or direct deposit.[21]

It is clear that whether or not a household received an ESP was not random, nor was the payment size. As emphasized by Broda and Parker (2014), however, the timing of payment was randomized since the week of payment disbursement within method of disbursement depended on the second-to-last digit of the recipient's Social Security number, which is effectively random.[22] Hence, conditional on disbursement method, payment week is random across households. The randomization allows us to identify the effects of an ESP on quality and spending off the differential shopping behavior of households receiving the ESP in different weeks by having timing groups act as controls for each other.

We follow the approach by Broda and Parker (2014) and use the following baseline regression

[21] Recipients that had provided the Internal Revenue Service (IRS) with their personal bank routing received their payments by direct deposit. Each household also received a statement from the IRS a few business days before the electronic transfer of the ESP.

[22] The last four digits of a Social Security number are assigned sequentially to applicants within geographic areas and group numbers (the middle two digits of the number).

equation to estimate the effect of the ESP on shopping behavior for household h in week t:

$$X_{h,t} = \mu_h + \eta_t + \sum_{s=-L}^{L'} \beta_s ESP_{h,t+s} + \varepsilon_{h,t}, \tag{11}$$

where $X_{h,t}$ is one of the measures for quality in Equation (9) or total spending in week t by household h.[23]

$ESP_{h,t+s}$ is a dummy variable taking the value of 1 in the week s periods after household h receives the ESP. Thus, the sequence of coefficients $\tilde{\beta} = (\beta_{-L}, \beta_{-L+1}, \ldots, \beta_0, \beta_1, \ldots, \beta_{L'})$ captures the dynamic effect of the ESP before receiving the payment, at impact and in the weeks following the payment. Since $ESP_{h,t+s}$ is a dummy variable, the estimates for $\tilde{\beta}$ can be interpreted as average treatment effects.[24] Because quality is only observed in weeks in which households actually go shopping, non-responders to the ESP are excluded from the regressions with quality as the dependent variable since weeks without any shopping activity are excluded from the regression. However, this problem is not severe as 80 percent of the households in our data make a purchase in the same week as they receive the ESP, while an additional 18 percent make a purchase within four weeks after receiving the payment. It is nonetheless difficult to estimate the full lead and lag structure for $\tilde{\beta}$ due to the occasionally missing values. Therefore, we constrain the parameters in $\tilde{\beta}$ such that they are constant within four-week periods relative to the week of ESP receipt. Within these four-week periods, 97-98 percent of households are observed at least once. The number of leads is set to 16 ($L = 16$), while the number of lags including the contemporaneous response is set to 24 ($L' = 23$). This ensures that we observe all households for the entire set of leads and lags in Equation (11).

We include two fixed effects in the regression. First, we include week fixed effects, η_t, to absorb any common changes over time in shopping behavior across households. Second, we control for household-level fixed effects, μ_h, to account for household-specific differences in shopping behavior unrelated to receiving the ESP. Although the timing of ESP is random within disbursement method, the ESP was disbursed later by check (from May 16 until June 11) than by electronic transfer (from May 2 until May 16). Hence, selection into method of disbursement – e.g. households receiving the ESP by electronic transfer had a higher income on average – might be an issue if there is a correlation between shopping behavior and household type (for example, through the positive correlation between annual income and quality of purchases documented

[23] Weekly household spending is constructed by aggregating each household's total spending by trip to the weekly level. This implies that the spending variable includes products that we could not match to the quality indices in addition to some products not tracked by Nielsen as mentioned in Section 3.2.

[24] Recent econometric papers study the interpretation of event study estimates and propose alternative estimators of average treatment effects in event study frameworks in the presence of time-varying treatment effects or cross-sectional treatment effect heterogeneity (Borusyak, Jaravel and Spiess, 2021; Goodman-Bacon, 2021; Sun and Abraham, 2021). We stick to the conventional OLS estimator with a flexible set of leads and lags.

in Figure 4 above). Without household fixed effects, this would bias $\tilde{\beta}$ due to the changing composition of the sample (Borusyak, Jaravel and Spiess, 2021). A related issue is that our panel features occasionally missing household quality of consumption in some weeks as discussed in Section 3.4. Including household fixed effects controls for a possible correlation between household type and the tendency to have a missing quality variable.

When studying heterogeneity by income and liquidity groups in the data, we estimate the following model:

$$X_{h,t} = \mu_h + \eta_{j,t} + \sum_j \sum_{s=-L}^{L'} \beta_{j,s} \mathbf{1}\left\{\text{Group}_h = j\right\} ESP_{h,t+s} + \varepsilon_{h,t}, \tag{12}$$

where j denotes groups in the data.

This regression equation is identical to regression Equation (11) except that the coefficients of interest, $\beta_{j,s}$, are allowed to differ by group, and that we include week $\times$ group fixed effects to allow for common changes in shopping behavior within groups. Moreover, we scale the ESP dummies, $ESP_{h,t+s}$, by the ratio between the average ESP received within group j and the average ESP in the full sample. This allows for quantitative comparison of the estimates between groups.

4.1.2 Results

Table 3 presents the estimates of $\tilde{\beta}$ from Equation (11) for the month prior to receiving the ESP as well as the following 3 months. The estimates for total spending are shown in column 1, while columns 2 through 4 show the estimates for spending quality using the three quality indices described in Section 3.1. The quality estimates have been scaled by 100 for illustrative purposes. Standard errors are robust to heteroskedasticity and clustered at the household level to account for intertemporal within-household correlation of the error term. All regressions are carried out using the Stata package REGHDFE for estimation of high-dimensional fixed effect models by Correia (2019).

Column 1 in Table 3 shows that the households increase their spending when receiving the ESP. Weekly spending increases by \$12.6 on average for the 4 weeks after receipt of the ESP, while the three-month cumulative increase in spending is around \$95 (or 10.7 percent of the ESP since the average ESP was \$884), which is broadly in line with what was originally documented by Broda and Parker (2014). Although this estimate seems small relative to the existing evidence on MPCs, we need to remember that spending recorded in the Nielsen data is only a subset of households' total spending and predominantly non-durables. According to Broda and Parker (2014), household-level spending in the CPD is 35 percent of spending on non-durables or 19

Table 3. Response of spending and product quality to the ESP

	Weekly spending	Size-based quality	Unit price-based quality	Brand-based quality
1 months before ESP	0.76	0.0048**	0.0039*	0.0046**
	(0.691)	(0.002)	(0.002)	(0.002)
Contemporaneous month	8.41***	0.0091***	0.0095***	0.0073***
	(0.772)	(0.002)	(0.003)	(0.002)
2 months after ESP	1.70**	0.0049**	0.0056**	0.0020
	(0.750)	(0.002)	(0.003)	(0.002)
3 months after ESP	0.68	0.00047	0.0028	0.0015
	(0.666)	(0.002)	(0.002)	(0.002)
Week × household obs.	1,087,348	849,508	845,186	845,049
Households	20,516	20,506	20,507	20,507

Notes. The table shows the estimates of $\tilde{\beta}$ from Equation (11). Estimates from regressions with a quality measures as the dependent variable have been scaled by 100. Standard errors are clustered at the household level and reported in parentheses. *, ** and *** denote significance at the 0.1, 0.05 and 0.01 level respectively.

percent of total consumption spending. Scaling by these percentages results in a three-month cumulative increase in spending of $123–$227 or 31–56 percent of the ESP, which is broadly in line with existing evidence on the consumption response to tax rebates.[25]

Households not only increase dollars spent but also the quality of their purchases according to columns 2 through 4 although the estimates are not as statistically significant as the spending estimates. The effect is most significant for the size-based and unit price-based indices compared to the brand-based indices. However, the brand-based index will only be affected when households switch the brand of their purchases, while within-brand changes in purchases affects the two other indices. This might be why the effects on the brand-based index are muted.

There are indications of a statistically significant effect on both spending and quality in the 4 weeks leading up to receiving the ESP. Due to the truly randomized timing of the ESP, the presence of effects on spending and quality prior to treatment does not invalidate the research design or yield biased estimates. Rather they reflect anticipation effects that are part of the treatment effect (Borusyak, Jaravel and Spiess, 2021). However, as we discuss in Section 4.1.5 below, formal tests cannot reject no presence of a pre-ESP trend, and the significance of the pre-ESP coefficients are not robust to the number of lags included in the regression. Hence, we are cautious about interpreting these estimates as actually reflecting effects prior to ESP receipt.

[25] Parker *et al.* (2013) use the Consumer Expenditure Survey to estimate the response of consumption to the 2008 ESPs and find that consumers increase non-durables spending by 12–30 percent of their stimulus payment, while the response increases to 50–90 percent when including the response of durable goods. Johnson, Parker and Souleles (2006) analyze the non-durables spending response to the ESPs distributed in 2001 and estimate a slightly higher estimate compared to the 2008 rebates (a response of 20–40 percent of the rebates).

While we observed some heterogeneity in the spending shares of the product modules across the income distribution as shown above in Figure 6, we do not find much evidence of switching across modules when receiving the ESP. We estimate Equation (11) module-by-module using the weekly spending share for the module as the dependent variable. Panel (A) in Figure 7 shows a histogram of the estimates of the coefficient on the ESP indicator in the four weeks following ESP disbursement along with a histogram of their t-statistics in Panel (B). The red line in Panel (B) shows the fitted normal distribution of the t-statistics, while the black line shows their distribution under the null hypothesis of no change in the spending share of the module when receiving the ESP.

(A) Estimates for module-switching

(B) t-statistics for module-switching

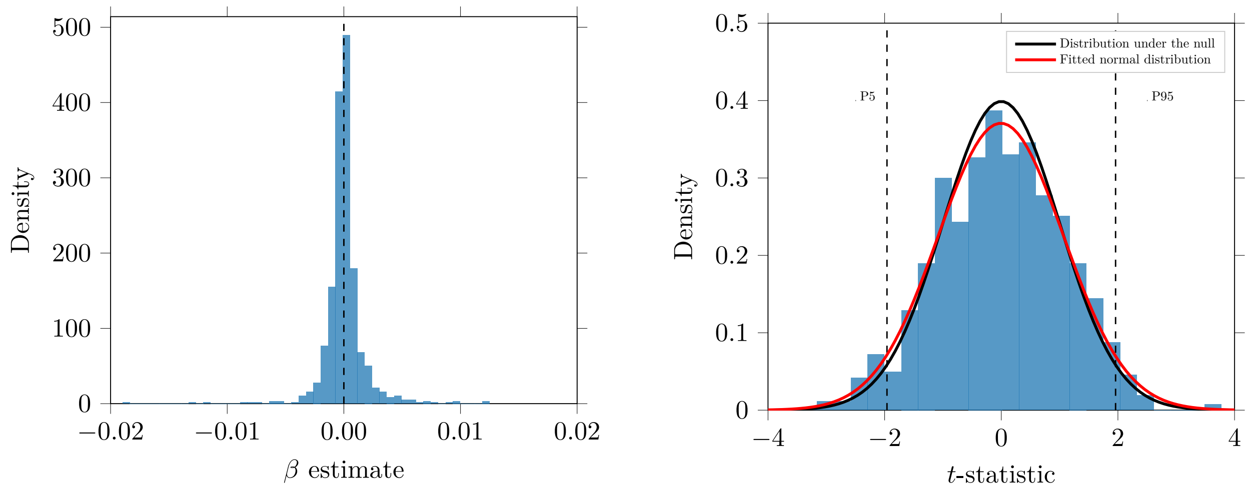


FIGURE 7. Response of module-switching to the ESP

Notes. Panel (A) shows the distribution of estimates of $\tilde{\beta}$ in the 4 weeks after receiving an ESP from Equation (11) at the module level, where the dependent variable is the weekly spending share of the products in the module. Panel (B) shows the distribution of the estimates' t-statistics based on standard errors clustered at the household level. The red line shows a normal distribution fitted to the t-statistics. The black line is the distribution of the t-statistics under the null of $\tilde{\beta} = 0$, while vertical lines indicate the distribution's 5th and 95th percentiles.

The estimates for module switching are tightly centered around zero. Correspondingly, most of their t-statistics are insignificant with 6.9 percent having a p-value below 0.05. Hence, the number of significant estimates for module switching is close to the number of type I errors that we would observe under a true null hypothesis of no module switching.[26]

[26] A Kolmogorov-Smirnov test of the null hypothesis that the estimated t-statistics are distributed according to a standard normal distribution has a p-value of 0.15.

4.1.3 Heterogeneity of response to ESP by income

Next, we look at how the ESP effects differ by income.[27] We split our sample into 3 approximately equally large groups based on annual income using the same groups as Broda and Parker (2014): a low-income group with income less than \$35,000, a middle-income group with income between \$35,000 and \$70,000 and a high-income group with income above \$70,000. We then estimate Equation (12) by these groups. Results are presented in Table 4.

Table 4. Heterogeneity of ESP response by income groups

	Weekly spending	Size-based quality	Unit price-based quality	Brand-based quality
	Income below \$35,000			
1 month before ESP	1.64	−0.0014	0.0075	0.0076
	(1.982)	(0.007)	(0.008)	(0.007)
Contemporaneous month	14.5***	0.014*	0.021***	0.018***
	(2.017)	(0.007)	(0.008)	(0.007)
2 months after ESP	3.54**	0.0016	0.013*	0.0062
	(1.761)	(0.007)	(0.007)	(0.006)
3 months after ESP	2.16	−0.0012	−0.0019	−0.0013
	(1.422)	(0.006)	(0.006)	(0.006)
Week × household obs.	369,145	289,460	287,858	287,792
Households	6,965	6,962	6,963	6,963
	Income between \$35,000 and \$60,000			
1 month before ESP	3.83**	0.0098**	−0.00046	0.0036
	(1.564)	(0.005)	(0.005)	(0.005)
Contemporaneous month	10.5***	0.012***	0.0099*	0.0082*
	(1.529)	(0.005)	(0.005)	(0.005)
2 months after ESP	4.43***	0.011**	0.0081*	0.0030
	(1.357)	(0.005)	(0.005)	(0.004)
3 months after ESP	2.76**	0.0025	0.0062	0.0050
	(1.215)	(0.004)	(0.004)	(0.004)
Week × household obs.	317,788	249,930	248,678	248,635
Households	5,996	5,991	5,991	5,991
	Income above \$60,000			
1 month before ESP	0.85	0.0054	0.0032	0.0038
	(1.375)	(0.004)	(0.004)	(0.003)
Contemporaneous month	5.59***	0.0054	0.0023	0.0017
	(1.402)	(0.004)	(0.004)	(0.003)
2 months after ESP	−0.050	0.0028	−0.00088	-0.00070
	(1.289)	(0.003)	(0.004)	(0.003)
3 months after ESP	−0.78	0.00058	0.0023	0.00070
	(1.061)	(0.003)	(0.003)	(0.003)
Week × household obs.	400,415	310,118	308,650	308,622
Households	7,555	7,553	7,553	7,553

Notes. The table shows estimated $\tilde{\beta}$s from Equation (11). Standard errors are clustered at the household level and reported in parentheses. *, ** and *** denote significance at the 0.1, 0.05 and 0.01 level respectively.

The estimates reveal some heterogeneity in the response to receiving an ESP across the income distribution. Households spent a smaller share of the ESP, the higher annual income they had, which is in line with the findings by Parker (2017). From bottom to the top of the income groups,

[27] We have also analyzed how the ESP effects differ by households' access to liquid wealth. The results are shown in Section D.

the cumulative three-month MPCs are 16.5 percent, 11.0 percent and 6.7 percent. The ratios between these MPCs are shown in table 5 and will be used to calibrate the structural model in section 5.

Table 5. Relative marginal propensities to consume

	Bottom-to-top	Bottom-to-middle	Middle-to-top
Relative MPC	2.463 (1.046)	1.506 (0.451)	1.635 (0.686)

Notes. The table shows the relative 12-week marginal propensities to consumed between income tertiles based on the estimates shown in Table 4. Standard errors are calculated using the delta method and shown in parentheses.

Quality of consumption does not increase for all income groups when they receive the ESP. The quality of consumption for both the low-income and middle-income groups increases in the month after receiving the ESP. The response, however, is most significant and longer-lived for the middle-income group. Lastly, we find no significant effect on the quality of consumption among high-income households even though we find an economically small but statistically significant increase in spending.[28]

We interpret the three income groups as groups for permanent income. Although annual income in one year is a crude measure of permanent income, we show in Section 4.1.5 that the results are robust to other ways of grouping by income that use income reported in years after 2008 in addition to controlling for household size and age.

4.1.4 MPC heterogeneity in relationship to the literature

The empirical literature studying MPC heterogeneity across the income distribution is not conclusive. Our estimates add to those papers finding that MPCs are higher for households with lower income such as Parker *et al.* (2013), Broda and Parker (2014), Parker (2017) and Parker and Souleles (2019) for the case of the 2008 ESPs. In a related study on the spending response to ESPs enacted in The Economic Growth and Tax Relief Reconciliation Act of 2001, Johnson, Parker and Souleles (2006) also find that low-income households have the highest MPCs. Lastly, Jappelli and Pistaferri (2014) use survey data among Italian households on reported MPCs out of a fictitious income shock equal to one month's income and find that the MPC is decreasing in income.

On the contrary, Shapiro and Slemrod (2003), Shapiro and Slemrod (2009) and Sahm, Shapiro and Slemrod (2010) use the University of Michigan's monthly Survey of Consumers to study

[28] The statistically significant effects on spending but not quality for the high-income group could reflect that the spending regressions have more statistical power than the quality regressions because of the larger number of observations.

expected spending responses following the two tax rebates in 2001 and 2008. These papers conclude that MPCs are, if anything, increasing in income. However, whereas our estimates stem from actual behavior, these authors estimate intended behavior. Misra and Surico (2014) add to these findings with revealed-preference estimates by using CEX data to estimate the distribution of MPCs to both the 2001 and 2008 ESPs using quantile regressions. They look at how covariates change between quantiles of the estimated MPC distribution and show that while the low-income households primarily belong to the middle-MPC groups, rich households have either higher or lower MPCs. Lastly, Lewis, Melcangi and Pilossoph (2019) take an agnostic stand on the source of heterogeneity using machine learning methods to group households. They estimate the distribution of MPCs out of the 2008 ESPs across households without imposing any ex ante assumptions on how households are assigned to consumption response groups. Afterwards, they analyze how the estimated MPCs relate to observable variables and document a positive relationship between the MPC and total income, mortgage interest payments and the ratio between annual consumption and annual income. Their best linear regression of the estimated MPCs on observables in their data, however, can only account for 13 percent of the variance in MPCs.

4.1.5 Robustness

Our results regarding the response to income shocks are reasonably robust. In this section, we perform some robustness checks of our findings.

Balancing the sample around ESP receipt Our baseline estimates are estimated using observations that cover the entire year. As discussed in Section 4.1.1, the inclusion of household fixed effects controls for the potential bias between the level of the dependent variable and the timing of ESP receipt. Treatment effect heterogeneity by ESP timing, however, could still affect the estimates because of compositional changes in the households used to identify $\tilde{\beta}$ (Dobkin *et al.*, 2018). To address this, we have estimated Equation (11) with a sample balanced relative to ESP receipt such that households are only included from 16 weeks prior to ESP receipt until 23 weeks after.[29] This ensures that all households are present the same number of weeks in the sample for the estimation of spending effects (and potentially the same number of weeks for quality effects). The estimates from the balanced regression are shown in Table B.1 of Section B. Albeit the spending estimates are slightly smaller and equivalent to a reduction in the average 12-week spending response from \$95 to \$81, the estimates are similar to those reported in Table 3. Additionally, there are no longer any significant effects one month before ESP receipt on quality measured using the unit price-based and brand-based quality indices. Although not

[29] This also requires us to drop two leads as a normalization. We normalize coefficients on the leads 16 and 5 weeks before ESP receipt to zero. If the quality variable is missing in one of these two weeks, we impute it with the mean of consumption quality within its four-week period lead (e.g. if the 16-week lead is missing, we impute it with the average quality of weeks 15, 14 and 13).

shown, there are no statistically significant coefficients prior to the one-month lead for all four regressions.

Weekly estimates We have so far imposed that the coefficients on the ESP indicator variables are identical within four-week periods. Imposing this restriction yields more precise estimates – especially for the estimates concerning consumption quality – but does not allow for the analysis of high-frequency movements. To estimate Equation (11) without restricting coefficients and also formally test for pre-ESP effects, we follow Borusyak, Jaravel and Spiess (2021). First, we exclude the lead furthest away from ESP receipt (16 weeks before) as well as the lead one week before ESP receipt. If the quality variable is missing in one of these two weeks, we impute it as for the balancing robustness check with the mean of consumption quality within its four-week period lead. This is a normalization pinning down a constant and a linear trend between the two leads, which allows us to test for the presence of a non-linear trend prior to receiving the ESP.[30] We then balance the sample relative to the ESP receipt such that we only include households from 16 weeks prior to ESP receipt until 24 weeks after. Table B.2 in Section B shows the estimates on spending and the size-based quality measure from this regression. There is a very significant and positive effect on spending from receiving the ESP, which lasts about 4 weeks, while there is also a positive – albeit not as significant compared to the spending estimates – effect on quality lasting about 2 weeks. Besides a couple of non-adjacent weeks, the pre-ESP coefficients are not statistically different from zero and display no systematic pattern. The bottom of the table also contains the p-values for an F-test of the hypothesis that the coefficients on all leads are equal to zero. For spending and quality, the p-values are 0.100 and 0.484 respectively. Thus, we can clearly not reject that there is no trend in quality before ESP receipt, while no pre-ESP trend in spending is borderline not rejected at the 10 percent confidence level.

Sensitivity to leads and lags The number of leads and lags in our regressions were chosen such that all households are observed for the entire set of leads and lags. To analyze the sensitivity of our estimates to this choice, we have estimated Equation (11) using all combinations of lead and lag lengths up until the 4 four-week leads and 6 four-week lags for a total of 24 different combinations. The results of this exercise for the spending estimates and the size-based quality estimates are shown in Panels (A) and (B) of Figure B.6 in Section B. Point estimates from the same regression are joined by a dashed line in the figure, and a blue dot indicates that the point estimate is significantly different from zero at the 5 percent confidence level. There are three main takeaways from the figure. First, irrespective of the specification, there is a significant and positive effect on both spending and quality in the 4 weeks following the ESP receipt. For the spending regression, there is always a significant effect present in the second month after

[30] As emphasized by Borusyak, Jaravel and Spiess (2021), the event study design can only identify $\tilde{\beta}$ up to a common linear trend. This is because the passing of absolute time cannot be disentangled from time relative to ESP receipt when household fixed effects are included. Hence, $\tilde{\beta}$ can be interpreted as deviations from a common linear trend (if there is any).

receiving the ESP, while the significant effect on quality in the second month is present in all but 2 of the specifications. Second, increasing the number of leads or lags in the regressions tends to increase the point estimates. Finally, the estimates of the pre-ESP receipt coefficients are not significantly different from zero across all specifications.

Income groupings The income split shown in Table 4 is based on income reported by households in the 2008 CPD. We will use these groups as proxies for permanent income groups when calibrating the structural model in Section 5. As mentioned in Section 3.2, however, the income measure in our data is likely a measure of households' annual income in 2007. One might worry that this is a too crude measure of households' permanent income. As an alternative, we base groups on income reported in subsequent years in the following way inspired by Dynan, Skinner and Zeldes (2004). For the first alternative grouping, we use income reported in the years 2008 through 2017 to create year-by-year income groupings using the same income brackets as in Table 4. We then use the modal value of each household's income group through the years and base the income split on this value. For the second alternative grouping, we adjust for household size and age in the following way. We first use the method by Handbury (2021) to adjust income in all years for household size by dividing the midpoint of a household's income category with the square root of the number of family members.[31] We then divide households into income tertiles year-by-year based on the household size-adjusted income. Similar to the method by Dynan, Skinner and Zeldes (2004), we create the tertiles within the 9 age bins in the Nielsen data to account for age differences in income levels.[32] Finally, we use each household's modal value of these tertiles to group households. Table B.3 in Section B shows the ESP results for spending and the size-based quality measure by these alternative splits when only including households that are observed at least in the years 2008, 2009 and 2010. Although this reduces the sample size by about a third due to attrition, the results are similar to our original estimates and robust across grouping methods. If anything, the heterogeneity in the quality response across the income groups is even more pronounced.

Category-level estimates Although Figure 7 does not show any strong indications of households switching spending across product modules when they receive the ESP, we have estimated Equation (11) at the three different product category levels (module, group and department levels). When doing so, we use spending and quality measured for each household at the product category $\times$ week level as the dependent variable and also include product category $\times$ household fixed effects and product category $\times$ week fixed effects. Hence, these estimates reflect the average effects on spending and quality *within* product categories of the ESP. Note

[31] For example, a household consisting of 4 members with annual income in the interval \$25,000-\$30,000 has an adjusted income of $\frac{\$27{,}500}{\sqrt{4}} = \$13,750$. This way of adjusting income for household size has also been used in the OECD Income Distribution Database since 2012 (www.oecd.org/els/soc/IDD-ToR.pdf).

[32] We group by the age of the male household head or the female head if there is no male head. Grouping by the female head yields almost identical results.

that this implies that the spending estimates will mechanically be smaller since they measure the average spending increase within each product category.[33] Table B.4 in Section B shows the estimates from these regressions. Although the magnitude of the estimates are reduced, there are still significant effects on quality from the ESP at all levels of aggregation. The table also illustrates that the additional information gained from estimating regressions at the most disaggregated level is limited since households do not buy products from all modules. Estimating Equation (11) at the department and group level instead of the aggregated level increases the number of observations by a factor of almost 4 and 8, respectively. However, estimating the regression at the product module level only increases the number of observations by around 5 percent compared to the product group level.

Disbursement method Finally, Table B.5 in Section B shows the results from estimating Equation (11) when we include week × disbursement method fixed effects to control for average spending and quality in each week specific to households receiving the ESP by check or direct deposit. This reduces the variation available for estimation since we are now treating the two disbursement methods as two different experiments by only exploiting within-disbursement method variation in ESP disbursement timing. Consequently, the estimates reported in Table B.5 is a weighted average of the disbursement method-specific ESP effects (Gibbons, Serrato and Urbancic, 2018). We still find significant effects on quality although they are less precisely estimated (especially for the unit price-based and brand-based quality indices). Moreover, the spending effects are shorter lived and slightly reduced.[34]

5 Dynamic model

The results presented in the previous section lend themselves to two important characterizations of our consumption-saving model. First, our findings suggest that the microfoundation for the intra-temporal problem of the households are better described by a setup as in Section 2. Second, as will be clear momentarily, the setup allows us to target the relative MPC moments found in Table 5 in the empirical section and to externally validate the model using the quality responses found in Table 4.

The model we present in this section is an extension of the standard buffer-stock model. That is, the economy is populated by N households who all live for T periods. Each household receives an exogenous stream of income and from this income it chooses how much to save and

[33] Additionally, the spending variable for this robustness check is created by summing all purchases recorded by households since these can be matched to product categories. All of our other results use spending constructed as the sum of total spending recorded for all shopping trips. For almost all households, total purchases are below total spending, and the average ratio between total purchases and total spending is 0.58.

[34] On the contrary, Broda and Parker (2014) estimate larger effects on spending when including week × disbursement method fixed effects but they weight households by the weights provided by Nielsen. We are able to replicate this finding.

how much to spend. The optimal choice of expenditures is chosen such that lifetime utility is maximized. We extend the model to allow for quality to affect the optimal expenditure choice. The extended model nests the standard model and thus makes a leveled playground for comparison. Throughout, we use the standard model as a benchmark to our model. We use this comparison to highlight important shortcomings of the standard model and features of the extended model.

5.1 Setup

In the dynamic problem, households live for T periods and seek to maximize lifetime utility by choosing the optimal level of expenditures and savings each period.[35] The per-period utility function has a CRRA form with relative risk aversion parameter ρ and the households discount future utility by a factor β. Each period, the households receive an exogenous stream of income. The income process is made up of a permanent and transitory component, denoted by P and ξ, respectively. The optimal expenditure choice is affected both by the transitory and permanent income state of the household. The optimal expenditure choice is further governed by how much cash-on-hand, M, the household holds. Being a dynamic problem, the level of expected cash-on-hand and transitory and permanent income in the subsequent periods also affect the optimal choice of current expenditures. Formulated recursively, the Bellman equation of the household problem is given by

$$V_t(M_t, P_t, \xi_t) = \max_{X_t} \frac{U_t^{1-\rho}}{1-\rho} + \beta \mathbb{E}_t[V_{t+1}(M_{t+1}, P_{t+1}, \xi_{t+1})],$$

where U_t is the utility function in Equation (1) and V_t is the value function. Using Equation (6), we can write the problem as

$$V_t(M_t, P_t, \xi_t) = \max_{X_t} \frac{(X_t \cdot f(\xi_t, P_t))^{1-\rho}}{1-\rho} + \beta \mathbb{E}_t[V_{t+1}(M_{t+1}, P_{t+1}, \xi_{t+1})]. \tag{13}$$

Except for the $X_t \cdot f(\xi_t, P_t)$ term, the rest of the model is specified exactly as the standard buffer-stock model. Households obey their inter-temporal budget constraint

$$A_t = M_t - X_t, \tag{14}$$

$$M_{t+1} = RA_t + Y_{t+1}, \tag{15}$$

where A_t denotes end-of-period assets, and M_t is thus beginning-of-period cash-on-hand, and R

[35] Note that, building on the intuition from Section 2, "choosing the optimal level of expenditures" is equivalent to choosing the optimal level of consumption in the standard model, since in the standard model, only one price index prevails, and hence consumption and expenditure level coincides.

is an exogenous and constant gross interest rate. Y_{t+1} denotes income and is further explained below. Households can borrow up to a fraction of their permanent income and thus

$$A_t \geq -\lambda P_t. \tag{16}$$

Households cannot leave life in debt and therefore

$$A_T \geq 0. \tag{17}$$

The income process is exogenous and given by

$$Y_{t+1} = \xi_{t+1} P_{t+1}, \qquad \xi_{t+1} = \begin{cases} \mu & \text{with prob. } \pi, \\ \frac{\varepsilon_{t+1} - \mu\pi}{1-\pi} & \text{else,} \end{cases} \tag{18}$$

$$P_{t+1} = G P_t \psi_{t+1}, \tag{19}$$

where

$$\log \psi_{t+1} \sim \mathcal{N}(-0.5\sigma_\psi^2, \sigma_\psi^2), \tag{20}$$

$$\log \varepsilon_{t+1} \sim \mathcal{N}(-0.5\sigma_\xi^2, \sigma_\xi^2). \tag{21}$$

The income process is similar to that found in e.g. Carroll (2021). We refer to this paper for further details. For instance, the possibility of a μ-income event ($\xi_{t+1} = \mu$) is consistent with having a model with unemployment benefits.

Lastly, in the computation, we use a functional form of $f(\xi_t, P_t)$ given by

$$f(\xi_t, P_t) = \kappa \mathrm{e}^{-\iota \mathrm{e}^{-\delta \cdot \xi_t P_t}}, \tag{22}$$

which belongs to the class of sigmoid functions.[36] This function has some particularly nice features. In general, sigmoid functions follow an S-shape and in terms of Equation (22), the specific shape of $f()$ is governed by κ, ι and δ. As a special case, one may notice that for $\delta = 0$ and $\iota = \ln \kappa$, including the natural restriction $\kappa > 0$, our model nests the standard buffer-stock model with homothetic preferences and thus generalizes the framework.[37] When calibrating the model, we thus also allow for the possibility that the standard buffer-stock model is favored.

[36] Specifically, the function in Equation (22) is the so-called Gompertz function.

[37] Note that the "natural restriction" $\kappa > 0$ indeed is obvious since $\kappa < 0$ would imply that utility is declining in consumption. Alternatively, since $f()$ proxies the price index derived in equation (5), $\kappa > 0$ is a restriction that ensures that the price index is positive.

Under the more general specification, the shape of $f()$ is governed by the parameters in the following way: given that ι and δ are positive, **i)** κ is the upper asymptotic level for $f()$, which is approached as income goes to infinity, **ii)** ι determines the minimum value of $f()$, which is obtained in the zero-income event, where $\xi_t P_t = 0$, **iii)** δ determines the rate at which $f()$ goes from its minimum level to its upper bound.[38]

Figure 8 shows how $f()$ is affected by the parameters. The black line shows the calibrated $f()$-function used in the model solution. For every panel, we fix two parameters and vary the last one.

The model is solved using an extension of the Endogenous Grid Method (EGM) first proposed by Carroll (2006). Specifically, we implement the fast multi-linear interpolation algorithm by Druedahl (2021) combined with an upper envelope algorithm as in Druedahl and Jørgensen (2017). In Section F, we provide details on the computational part.

5.2 Calibration

For the parts of the model which are specified as a standard buffer-stock model, we calibrate it in close alignment with the previous literature. Specifically, we use the exact same values for the standard parameters as in Carroll (2021). Households live in the economy for 50 years. Some parameters are conventional in the literature: The coefficient of relative risk aversion is set to 2, the time discount factor is set to 0.96, and the gross interest rate is set to 1.04. The income process is set to match that found for U.S. households in Carroll (1992): the standard deviation of the log of the two income shocks equals 0.1, the permanent income growth rate is 3 percent and with a probability of 0.5 percent the household ends up in a zero-income event.

We once again point out that the standard buffer-stock model is nested for $\iota = \ln \kappa$ and $\delta = 0$, highlighted in the second part of Table 6. In the coming sections, we show how we calibrate κ, ι and δ and how the standard buffer-stock model disagrees with some important empirical moments. We further discuss that, under the common calibration, there is no room for improving on this disagreement in the standard model but that our model does provide such an opportunity.

5.2.1 Matching relative MPCs

In order to pin down κ, ι and δ, we calibrate these to match our findings in Section 2.2. Specifically, we target the relative MPCs between income groups reported in Table 5 and re-iterated below. As seen from Table 5 and as we discuss further in Section 5.2.2, these moments are poorly matched by the standard buffer-stock model but can be targeted using the model

[38] We also allow for the possibility where either ι, δ or both are negative. We do, however, not find any support for this. Section E shows how $f()$ is changed under these scenarios. We also discuss this case when we interpret our results.

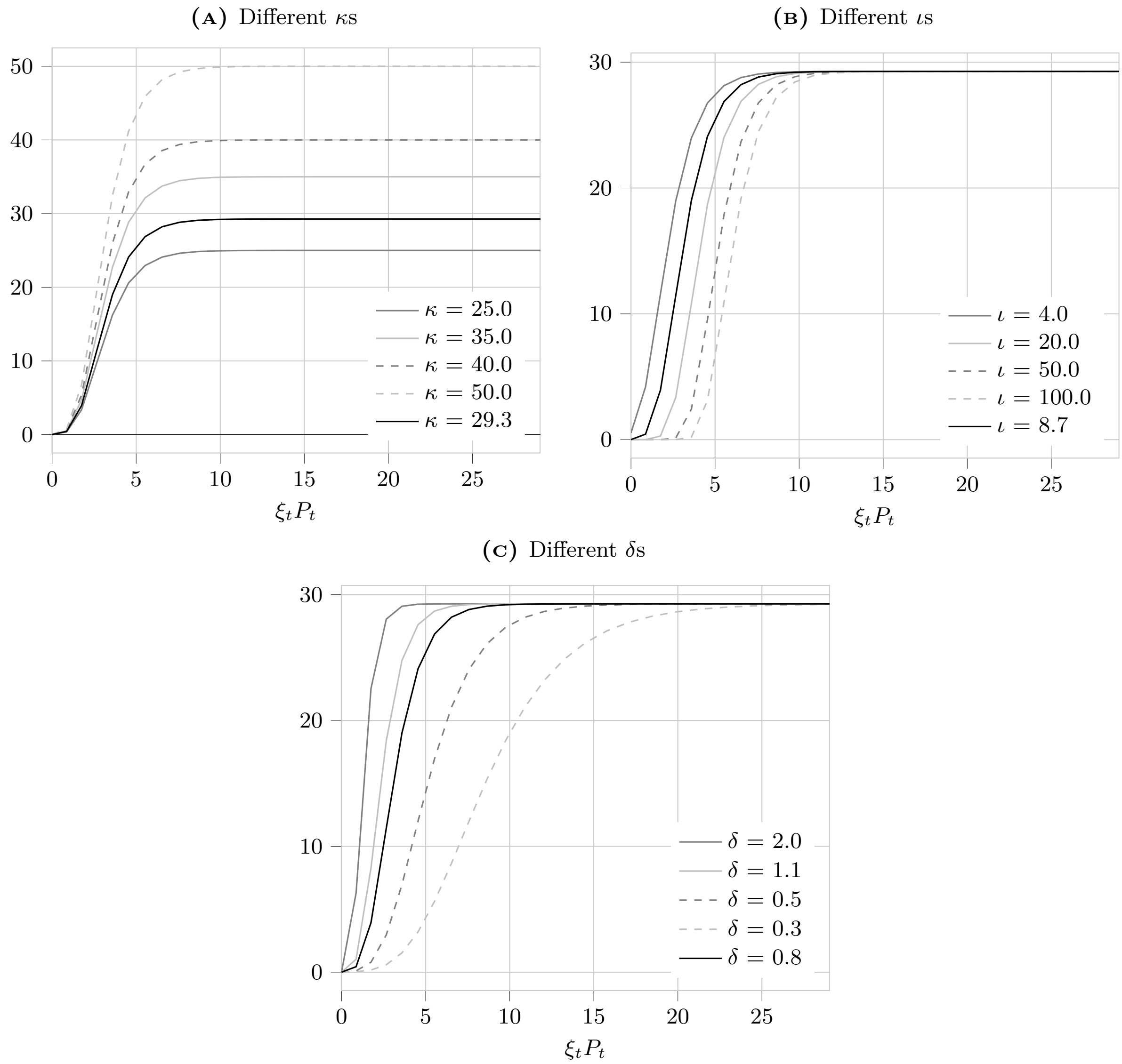


Figure 8. $f()$ function for different parameter values

Notes. The black line shows the calibrated $f()$ function used in the solution of the model. The parameter values found in the solution are $\kappa = 29.3$, $\iota = 8.7$ and $\delta = 0.8$. The maximum point on the income grid is the 99.99th earnings percentile in the simulation.

with nonhomothetic preferences. The reason why the nonhomothetic model is able match these moments is exactly as discussed in Section 2.2: the choice of expenditure allocation over different periods is affected by the quality demand of the household, which, through changes in price indices, affects the real expenditures required to smooth utility. Ultimately, this affects the MPC of the households. The optimal values are found to be $\kappa = 29.3$, $\iota = 8.7$ and $\delta = 0.8$ and as seen in Table 7, it allows us to match the empirical moments fairly well.

Table 6. Model parameters

	Parameter	Description	Value
Demographics	T	Years lived	50
Preferences	β	Time discount factor	0.96
	ρ	Relative risk aversion	2.0
Borrowing/saving	R	Gross interest rate	1.04
	λ	Borrowing limit	0
Income process	G	Growth rate of permanent income	1.03
	σ_ψ	Std. dev. of log permanent shock	0.1
	σ_ξ	Std. dev. of log transitory shock	0.1
	μ	Low-income event	0.0
	π	Probability of low-income shock	0.005
Standard buffer-stock model	κ	No function	$\in \mathbb{R}_{>0}$
	ι	No function	$\ln \kappa$
	δ	No function	0
Buffer-stock model w. non-homothetic preferences	κ	Upper limit	29.3
	ι	Scaling of lower limit	8.7
	δ	Rate of transition	0.8

Table 7. Calibration results for relative MPCs

		Model	
	Empirical	**Standard**	**Nonhomothetic**
Bottom-to-top	2.46	0.86	2.46
Bottom-to-middle	1.51	0.76	1.51
Middle-to-top	1.63	0.86	1.64

Notes. The model-implied relative MPCs are means within income groups.

5.2.2 Discussion on relative MPC moments

It is easiest to understand the properties of the nonhomothetic model if we compare it with the standard model. As documented by e.g. Carroll (1997), households in the standard model have the same buffer-stock target of wealth. Due to this feature, all households save in order to maintain a level of cash-on-hand consistent with this target. This, among other things, implies that rich households dissave and build their asset position down to a level which is consistent with the buffer-stock target. This explains why rich households have high MPCs. In our nonhomothetic model, however, households adjust their buffer-stock target in accordance with their demand for quality. When households become richer, they instead save more, in absolute terms, in order to be able to maintain a high level of quality consumption. This feature lowers their MPCs.

To further understand what is going on, it is helpful to look at the normalized policy functions.[39]

[39] We normalize the policy function by P, as this is in direct analogy to the standard model in ratio form. Due to the homogeneity of the standard model, the normalization implies that the policy function is unaffected by

From these, it is possible to infer the MPC from the gradient on a given point on the curve.

Figure 9 plots the policy functions for both models under various scenarios. In each of the four panels, we show how the policy function is affected by different levels of permanent income, ranging from a low level ($P = 0.86$) to a very high level ($P = 3.57$).[40] *Within* each panel, we show the policy function in the standard model (dotted line) and the policy function in the nonhomothetic model when transitory income is high (black line, $Y = 2.66$) and low (dashed line, $Y = 0.86$), respectively. At this point, it is important to notice the central difference between the two models: In the standard model, the (normalized) policy function is unaffected by changes in income. In the nonhomothetic model, the policy functions vary with income as highlighted in Figure 9. This helps explain the underlying reason why the nonhomothetic model is able to match the relative MPC moments.

Consider a household which has a medium level of permanent income but experiences a bad transitory income shock (low Y). Let this be illustrated by point A in Panel (B) of Figure 9. Imagine now that the permanent income of the household increases but that it still holds the same level of cash-on-hand. This implies that M/P falls. In the standard model, this would be illustrated by a shift from point A to point B. Since the gradient in B is higher than the gradient in point A, this results in an increase in the MPC. This is the only effect in the standard model. In the nonhomothetic model, however, the change in permanent income implies that the household also starts demanding goods of higher quality. For a fixed level of Y, this is illustrated via the change from point B in Panel (B) to point C in Panel (D). At point C, the household has a very high level of permanent income but Y is still assumed to be low. Comparing the gradient in point B to the gradient in point C, we realize that the MPC is even higher given that this is where the household ends up. However, since P increased, Y also increases which, in the example, implies that it ends up in point D. In point D, the gradient is lower than in any of the other points and the final result is that for a given increase in permanent income, the now richer household may exhibit a *lower* MPC.

What is the difference between a household being in point C and a household in point D? Remember that in both points, the level of normalized cash-on-hand (M/P) is the same. Thus, using Equation (15), we see that the difference between the two households is that the household with low Y must have a high level of assets in order to have gained the same level of cash-on-hand as the household with high Y. For the household with a high level of assets (the household in point C), it spends more out of a windfall than the household with lower level of assets. The reason is that the household with low level of assets wishes to save more in order to wear off future bad income shocks. Additionally, the high Y also implies that the household has a demand

varying income. Thus, it is fairly easy to compare the two models under various scenarios.

[40] The four levels roughly correspond to the 10th (low), 50th (medium), 75th (high) and 90th (very high) percentiles in the simulation.

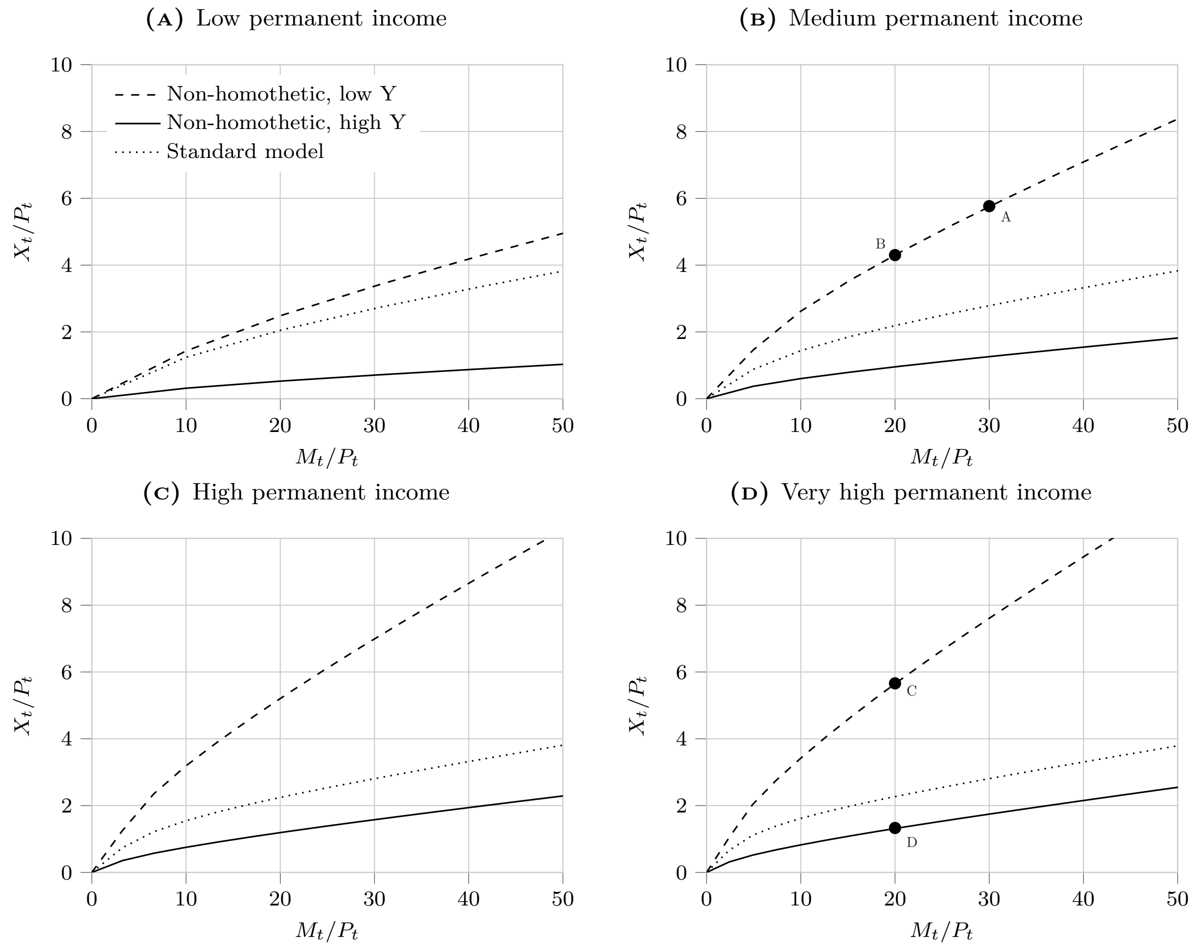


Figure 9. Policy functions in the standard and nonhomothetic model. Different levels of Y and P.

Notes. Policy functions are normalized by P. In Panel (A), $P = 0.86$. In Panel (B), $P = 1.74$. In Panel (C), $P = 2.65$. In Panel (D), $P = 3.57$. In all scenarios, low corresponds to $Y = 0.86$ and high to $Y = 2.66$.

for high quality goods, which it further wishes to maintain consumption of. This adds to its savings demand.

Now, obviously what lacks in this simplified example is that rich households may, arguably, have (much) higher levels of cash-on-hand than poor households which, despite their high level of permanent income, could still imply that M/P is higher for rich households. Thus, to serve justice to the standard model, we should mention that the standard model *could* predict MPC ratios in Table 7 higher than 1 (implying that poor households have higher MPCs than rich households) but for the given calibration, which is the most commonly used calibration in this type of models, this is not the case.

5.3 What does the dynamic model say in terms of quality?

In this section, we put our approximation to how quality enters the consumption choice of the households under scrutiny. Specifically, we show here that our empirical findings in Section 4.1.3 are indeed consistent with what the model predicts. In essence, we show that for the $f()$ function to have the shape as in our calibration, it must stem from an overall increase in quality. Secondly, we show that the quality response following a windfall is hump-shaped over the income distribution, exactly as our empirical results suggest.

Remember that the $f()$ function is an approximation to the function in Equation (4). Specifically, $f(\xi, P) = \frac{K(P)}{\prod_m \mathcal{P}_m(\xi,P)^{\alpha_m(P)}}$ with

$$\mathcal{P}_m(\xi, P) = \left(\sum_{i \in G_m} \mathcal{P}_{mi}^{1-\sigma} \varphi_{mi}(\xi, P)^{\sigma - 1} \right)^{\frac{1}{1-\sigma}}.$$

In our empirical analysis, we study the response in quality following a transitory income shock. It is straightforward to show that our approximation to $f()$ is increasing in ξ. Hence, what remains to be shown is what determines the marginal change in the above expression for $f(\xi, P)$ and specifically what the requirement for $\frac{\partial f(\xi,P)}{\partial \xi} > 0$ is. We show in Section H.1 that the requirement is

$$\sum_m \sum_{i \in G_m} A_m B_{mi} \frac{\partial \varphi_{mi}(\xi, P)}{\partial \xi} > 0, \tag{23}$$

where it holds that

$$A_m \geq 0, \qquad \text{and} \qquad B_{mi} > 0,$$

which leads us to conclude that in order for $f()$ to be increasing in ξ, it must be that *i)* the quality assessment of some goods increases, *ii)* the quality assessment, overall, increases and *iii)* those modules, m, which the household attaches most weight to, is where this quality increase happens.[41] Lastly, coupling this with our study of relative demand of two goods in Section 2, in

[41] A simple way to see that a quality increase must be present is to consider the extreme case where all quality assessments decrease, ie. $\frac{\varphi_{mi}(\xi,P)}{\partial \xi} < 0$ for all m, i. Alternatively, one could also consider the homothetic case where $\frac{\partial \varphi_{mi}}{\partial \xi} = 0$ for all m, i. In this case, $f()$ should be theoretically flat and our approximation would have been proven wrong.

particular Equation (3) which we re-iterate here

$$\log \frac{x_{mi}}{x_{mk}} = (\sigma - 1)\left[\log \frac{\varphi_{mi}(\xi, P)}{\varphi_{mk}(\xi, P)} - \log \frac{\mathcal{P}_{mi}}{\mathcal{P}_{mk}}\right],$$

we see that for those goods where quality increases the most, a substitution towards these goods happens. That is, for those goods where quality increases, relative expenditure shares increase which is exactly in alignment with our empirical findings in Section 3.

To assess how the quality response in the model matches the quality response found using the ESP shock in Section 3, we now study the change in the $f()$ function following a windfall. Based on the definition of the MPC, we define a similar term for the change in $f()$, which we denote the marginal quality change (MQC). For each household, i, we compute the MQC as[42]

$$MQC^i \equiv \lim_{\Delta \downarrow 0} \frac{f(P_t^i, \xi_t^i + \Delta) - f(P_t^i, \xi_t^i)}{\Delta}, \tag{24}$$

and for each income group, we then take the average MQC. To get a sense of how the MQC changes at a more granular level, we divide households into 100 groups based on their permanent income. Figure 10 shows the MQC over the income distribution based on these 100 income groups. The dashed, vertical lines show the cut-off between low-middle income and middle-high income, respectively. That is, all households to the left of the first dashed line belong to the low-income group. All households between the two dashed lines belong to the middle-income group, and all households to the right of the second dashed line are rich households.

Figure 10 reveals that the the hump-shaped pattern in the quality response suggested by our empirical findings is also present in the model. That is, what we found in section 4.1.3, is that the quality response is low for the low-income group, high for the middle-income group and low for the high-income group. [43] In the model, the poorest 1 percent have a MQC of 0.5, the richest 1 percent of the households have a MQC close to zero, while the lowest MQC among the middle-income households is 4.3.[44]

However, as Figure 10 also shows, the dispersion within the three broader groups is high. Among the rich households, the MQC ranges between 0 and 8.8 and the average MQC is 6.4. For the

[42] Note that we scale the shock so that all households receive the same windfall in absolute terms.

[43] Remember that we have not done anything to match this shape of the quality response and the calibration of $f()$ could easily have given us a different MQC distribution. As an example, consider Section H.2 where $\kappa = 50$, $\iota = 4$ and $\delta = 4$ and the average MQC is 15.6, 1.4 and 0.1 for the low income group, middle income group and high income group, respectively.

[44] Note that the MQC of 0.5 is the average MQC for the 1 percent poorest households. If we calculate the average of the 0.1 percent poorest household, it is 0.3.

low-income group, the MQC ranges between 0.5 and 4.3. The average MQC for the low-income group is 2.5. For the middle-income group, the MQC ranges between 4.3 and 8.3 and the average MQC is 6.3. That is, in line with our empirical estimates, the average change in quality is low for the low-income group and high for the middle-income group. However, in the model, the average change in quality for the high-income group is higher than suggested by our empirical estimates.

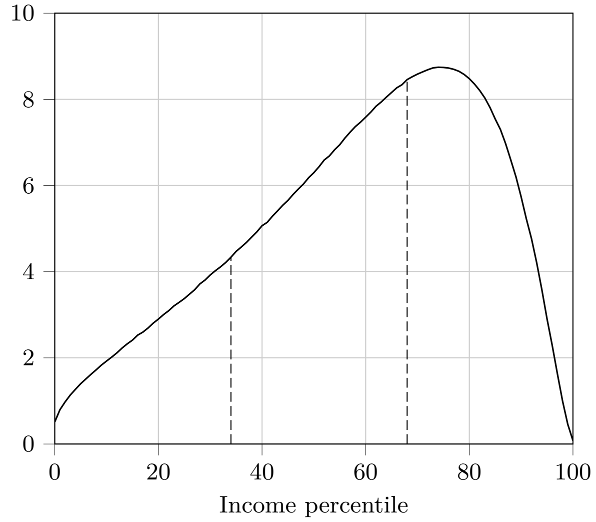


Figure 10. MQC for different income groups

Notes. Households are divided into income percentiles and for each group, we computed the average MQC as defined in Equation (24). That is, group 1 is the 1 percent poorest households and group 100 is the 1 percent richest households. The dashed, vertical lines represent the cut-off between the low-middle income and middle-high income, respectively.

5.4 Inequality in the two models

To finish this section off, we investigate a result which has been central to previous studies looking at the effects of nonhomotheticicities in dynamic consumption-saving models: the wealth distribution. As both highlighted in Carroll (2002) and latest in Straub (2019), nonhomotheticicities make the wealth distribution more unequal and help the model match the empirical distribution better. From the intuition provided for Figure 9, this is also to be expected from our model. In contrast to other studies, the foundation on which our nonhomotheticicities rely are completely different. In Straub (2019), the nonhomotheticicities come from a bequest motive. In Carroll (2002), he also looks at the bequest motive as in Straub (2019) but finds no evidence for such a behavior and instead argues for a direct inclusion of wealth in the utility function. In our model, the nonhomotheticicities come from the microfoundation outlined in Section 2 and rigorously explored in section Section 4. To this end, we feel confident about the source of nonhomotheticity of our model.[45]

Table 8 reports the Gini coefficients of the two models and the corresponding Lorenz curves are shown in Figure 11. We also show the asset distributions in the two models in Section G. Wealth inequality is more than three times higher in the nonhomothetic model compared to the standard model. The standard model provides a fairly equal wealth distribution with a wealth Gini of 0.15. In the nonhomothetic model, the wealth Gini is 0.49. The differences in the wealth distributions in the two models is further highlighted in the last three columns of Table 8. Here we show the wealth holdings of the bottom 50 percent, the top 10 percent and the top 1 percent. In the standard model, the bottom 50 percent hold 39.1 percent of wealth. In the nonhomothetic model, the bottom 50 percent hold 17.7 percent. According to the World Income Inequality Database (WIID),[46] the bottom 50 percent owned around 0 percent of (net) wealth in the U.S. in 2014.[47] The top 10 percent own 14.5 percent of total wealth in the standard model, whereas they own 36.5 percent in the nonhomothetic model. The WIID reports that this figure was 73 percent in the U.S. in 2014. For the top 1 percent, they own 1.6 percent in the standard model and 6.9 percent in the nonhomothetic model. The WIID reports that this was 38.6 percent in the U.S. in 2014.

Why does the model with nonhomothetic preferences generate higher wealth inequality? The intuition for this is similar to that provided in Section 5.2.2. When households become richer in the nonhomothetic model, their increased demand for high quality goods acts as a savings motive because they wish to continue consuming goods of high quality. The increased savings of the richer households exacerbates wealth inequality.

6 Concluding remarks

We use data on households' purchases to show that households trade up in the quality of their consumption when they receive a positive, transitory income payment. Moreover, we show that the response is heterogeneous across the income distribution. In particular, middle-income households exhibited a larger degree of trading up than low-income households, while high-income households did not change the quality of their consumption. We also find that the propensity to spend out of the income payment is decreasing in income.

We incorporate these findings into a canonical buffer-stock model by extending the model with

[45] Obviously, by the nature of being a partial equilibrium model, a rigorous analysis of wealth accumulation and distribution is beyond the scope of our model. However, it does still provide us with some useful insights about what is going on and what may potentially be very interesting to investigate in a full-fledged general equilibrium model.

[46] Latest version: UNU-WIDER, World Income Inequality Database (WIID4). See Piketty, Saez and Zucman (2018) for further details.

[47] Net personal wealth is defined as the total value of non-financial and financial assets (housing, land, deposits, bonds, equities, etc.) held by households, minus their debts in the WIID.

Table 8. Inequality in the models

	Wealth Gini	Income Gini	Wealth distribution		
			Bottom 50 %	**Top 10 %**	**Top 1 %**
Homothetic	0.15	0.34	39.1	14.5	1.6
Non-homothetic	0.49	0.34	17.7	36.5	6.9
USA, 2014	0.86	0.60	0.0	73.0	38.6

Notes. Data for the U.S. was collected from the World Income Database, latest version: UNU-WIDER, World Income Inequality Database (WIID4). The wealth Gini is based on net personal wealth defined as the total value of non-financial and financial assets (housing, land, deposits, bonds, equities, etc.) held by households, minus their debts. The income Gini is based on pre-tax national income defined as the sum of post-tax disposable income and public spending.

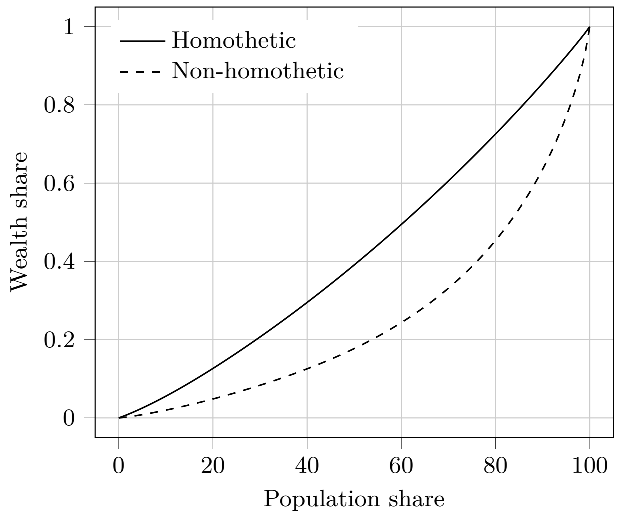


Figure 11. Lorenz curves for wealth in the two models.

Notes. The Lorenz curve shows how much wealth the bottom x percent of the population hold, where x varies along the horizontal axis. The closer the Lorenz curve is to the 45° line, the more equal the distribution is. The more the curve is pushed to the bottom corner, the more unequal is the distribution.

nonhomothetic preferences. In the nonhomothetic model, households not only choose the quantity but also take into account the quality of their consumption. The nonhomothetic model is able to generate a decreasing MPC in permanent income, while the homothetic model predicts the opposite pattern. Moreover, the model predicts that the quality response to a transitory income shock is hump-shaped over the income distribution as we find the data. Lastly, our model echoes the results of recent studies regarding nonhomothetic preferences and savings behavior in that nonhomothetic preferences in our model generate more wealth inequality contrary to a model with homothetic preferences.

There are two avenues for further research, which can build on our work. First, we only analyzed households' quality choice regarding retail spending on products that are predominantly non-durable. If anything, the existing literature suggests that the spending response of durables to transitory income shocks is at least as large as that of non-durables (Parker *et al.*, 2013). Hence, the consumption of durables should also be analyzed to fully understand the extent of quality shifting in consumption. One challenge regarding a high-frequency analysis of households' quality choice of durables, however, is that they are purchased less frequently relative to the products we analyzed.

Second, our structural model of household behavior is intentionally simple but could, for example, be extended with an illiquid asset and return heterogeneity as in the framework of Kaplan and Violante (2014). Embedding the household model into a general equilibrium framework could also be used to provide a more thorough analysis of how the nonhomothetic preferences affect wealth inequality. This poses a computational challenge, however, due to the potential interactions between households' quality choice and firm behavior. As an example of one such interaction, changes in consumption quality can affect firms' price setting through changes in the type of households purchasing products of a given quality as in the static model by Faber and Fally (2021). Such a model with two-sided heterogeneity is computationally demanding to solve since firms' will need to keep track of the distribution of households in order to set prices optimally.

References

Anderson, S. P., Palma, A. D. and Thisse, J.-F. (1987). "The CES is a discrete choice model?" *Economics Letters*, vol. 24(2), pp. 139–140 (cited on page 59).

Argente, D. and Lee, M. (2021). "Cost of Living Inequality During the Great Recession." *Journal of the European Economic Association*, vol. 19(2), pp. 913–952 (cited on pages 54, 56, 62, 70).

Bils, M. and Klenow, P. J. (2001). "Quantifying Quality Growth." *American Economic Review*, vol. 91(4), pp. 1006–1030 (cited on page 70).

Borusyak, K., Jaravel, X. and Spiess, J. (2021). *Revisiting event study designs: Robust and efficient estimation* (tech. rep.). (Cited on pages 75–77, 82).

Broda, C. and Parker, J. A. (2014). "The economic stimulus payments of 2008 and the aggregate demand for consumption." *Journal of Monetary Economics*, vol. 68, pp. S20–S36 (cited on pages 55, 56, 66–68, 74, 76, 79, 80, 84, 113, 114).

Broda, C. and Weinstein, D. E. (2010). "Product Creation and Destruction: Evidence and Price Implications." *American Economic Review*, vol. 100(3), pp. 691–723 (cited on page 57).

Campanale, C. (2018). "Luxury consumption, precautionary savings and wealth inequality." *The B.E. Journal of Macroeconomics*, vol. 18(1) (cited on page 56).

Carroll, C. D. (1992). "The Buffer-Stock Theory of Saving: Some Macroeconomic Evidence." *Brookings Papers on Economic Activity*, vol. 23(2), pp. 61–156 (cited on pages 56, 87).

Carroll, C. D. (1997). "Buffer-stock saving and the life cycle/permanent income hypothesis." *The Quarterly Journal of Economics*, vol. 112(1), pp. 1–55 (cited on page 89).

Carroll, C. D. (2002). "Why Do the Rich Save So Much?" In J. B. Slemrod (Ed.), *Does Atlas Shrug? The Economic Consequences of Taxing the Rich*. Harvard University Press. (Cited on pages 56, 94).

Carroll, C. D. (2006). "The method of endogenous gridpoints for solving dynamic stochastic optimization problems." *Economics Letters*, vol. 91(3), pp. 312–320 (cited on page 87).

Carroll, C. D. (2021). "Theoretical Foundations of Buffer Stock Saving." *Unpublished* (cited on pages 86, 87).

Coibion, O., Gorodnichenko, Y. and Hong, G. H. (2015). "The Cyclicality of Sales, Regular and Effective Prices: Business Cycle and Policy Implications." *American Economic Review*, vol. 105(3), pp. 993–1029 (cited on page 66).

Correia, S. (2019). *REGHDFE: Stata module to perform linear or instrumental-variable regression absorbing any number of high-dimensional fixed effects* (Statistical Software Components). Boston College Department of Economics. (Cited on page 76).

De Nardi, M. (2004). "Wealth Inequality and Intergenerational Links." *The Review of Economic Studies*, vol. 71(3), pp. 743–768 (cited on page 56).

Deaton, A. (1991). "Saving and Liquidity Constraints." *Econometrica*, vol. 59(5), pp. 1221–1248 (cited on page 56).

Deaton, A. (1992). *Understanding Consumption*. (Cited on page 56).

Dobkin, C., Finkelstein, A., Kluender, R. and Notowidigdo, M. J. (2018). "The Economic Consequences of Hospital Admissions." *American Economic Review*, vol. 108(2), pp. 308–352 (cited on page 81).

Druedahl, J. (2021). "A Guide On Solving Non-Convex Consumption-Saving Models." *Computational Economics*, pp. 747–775 (cited on pages 87, 116).

Druedahl, J. and Jørgensen, T. H. (2017). "A general endogenous grid method for multi-dimensional models with non-convexities and constraints." *Journal of Economic Dynamics and Control*, vol. 74, pp. 87–107 (cited on pages 87, 116).

Dubé, J.-P., Hitsch, G. J. and Rossi, P. E. (2018). "Income and Wealth Effects on Private-Label Demand: Evidence from the Great Recession." *Marketing Science*, vol. 37(1), pp. 22–53 (cited on page 62).

Dynan, K. E., Skinner, J. and Zeldes, S. P. (2004). "Do the Rich Save More?" *Journal of Political Economy*, vol. 112(2), pp. 397–444 (cited on page 83).

Faber, B. and Fally, T. (2021). "Firm Heterogeneity in Consumption Baskets: Evidence from Home and Store Scanner Data." *The Review of Economic Studies* (cited on pages 54, 57, 59, 70, 97).

Feenstra, R. (1994). "New Product Varieties and the Measurement of International Prices." *American Economic Review*, vol. 84(1), pp. 157–77 (cited on page 57).

Gibbons, C. E., Serrato, J. C. S. and Urbancic, M. B. (2018). "Broken or Fixed Effects?" *Journal of Econometric Methods*, vol. 8(1) (cited on page 84).

Goodman-Bacon, A. (2021). "Difference-in-differences with variation in treatment timing." *Journal of Econometrics*, vol. 225(2), pp. 254–277 (cited on page 75).

Handbury, J. (2021). "Are poor cities cheap for everyone? non-homotheticity and the cost of living across u.s. cities." *Econometrica*, vol. 89(6), pp. 2679–2715 (cited on pages 54, 57, 62, 83).

Jaimovich, N., Rebelo, S. and Wong, A. (2019). "Trading down and the business cycle." *Journal of Monetary Economics*, vol. 102, pp. 96–121 (cited on pages 54, 56, 62).

Jaimovich, N., Rebelo, S., Wong, A. and Zhang, M. B. (2020). "Trading up and the skill premium." *NBER Macroeconomics Annual*, vol. 34, pp. 285–316 (cited on pages 54, 56, 62, 70).

Jappelli, T. and Pistaferri, L. (2014). "Fiscal Policy and MPC Heterogeneity." *American Economic Journal: Macroeconomics*, vol. 6(4), pp. 107–136 (cited on page 80).

Johnson, D. S., Parker, J. A. and Souleles, N. S. (2006). "Household Expenditure and the Income Tax Rebates of 2001." *The American Economic Review*, vol. 96(5), pp. 1589–1610 (cited on pages 77, 80).

Jørgensen, C. N. and Shen, L. S. (2019). "A New Perspective on the Welfare Implications of Business Cycle Fluctuations: Evidence from Consumption Quality." *Unpublished* (cited on pages 56, 62).

Kaplan, G. and Violante, G. L. (2014). "A Model of the Consumption Response to Fiscal Stimulus Payments." *Econometrica*, vol. 82(4), pp. 1199–1239 (cited on pages 97, 113).

Kueng, L. (2018). "Excess Sensitivity of High-Income Consumers." *The Quarterly Journal of Economics*, vol. 133(4), pp. 1693–1751 (cited on page 64).

Lewis, D. J., Melcangi, D. and Pilossoph, L. (2019). *Latent Heterogeneity in the Marginal Propensity to Consume* (Staff Reports No. 902). Federal Reserve Bank of New York. (Cited on page 81).

Michelacci, C., Paciello, L. and Pozzi, A. (2021). "The Extensive Margin of Aggregate Consumption Demand." *The Review of Economic Studies* (cited on pages 54, 56).

Misra, K. and Surico, P. (2014). "Consumption, Income Changes, and Heterogeneity: Evidence from Two Fiscal Stimulus Programs." *American Economic Journal: Macroeconomics*, vol. 6(4), pp. 84–106 (cited on pages 81, 113).

Nevo, A. and Wong, A. (2019). "The Elasticity of Substitution between Time and Market Goods: Evidence from the Great Recession." *International Economic Review*, vol. 60(1), pp. 25–51 (cited on pages 62, 65).

Parker, J. A. (2017). "Why Don't Households Smooth Consumption? Evidence from a $ 25 Million Experiment." *American Economic Journal: Macroeconomics*, vol. 9(4), pp. 153–183 (cited on pages 56, 79, 80, 113).

Parker, J. A. and Souleles, N. S. (2019). "Reported Effects versus Revealed-Preference Estimates: Evidence from the Propensity to Spend Tax Rebates." *American Economic Review: Insights*, vol. 1(3), pp. 273–290 (cited on pages 56, 80).

Parker, J. A., Souleles, N. S., Johnson, D. S. and McClelland, R. (2013). "Consumer Spending and the Economic Stimulus Payments of 2008." *American Economic Review*, vol. 103(6), pp. 2530–2553 (cited on pages 56, 77, 80, 97).

Piketty, T., Saez, E. and Zucman, G. (2018). "Distributional National Accounts: Methods and Estimates for the United States." *The Quarterly Journal of Economics*, vol. 133(2), pp. 553–609 (cited on page 95).

Redding, S. J. and Weinstein, D. E. (2019). "Measuring Aggregate Price Indices with Taste Shocks: Theory and Evidence for CES Preferences*." *The Quarterly Journal of Economics*, vol. 135(1), pp. 503–560 (cited on page 57).

Sahm, C. R., Shapiro, M. D. and Slemrod, J. (2010). "Household Response to the 2008 Tax Rebate: Survey Evidence and Aggregate Implications." In J. R. Brown (Ed.), *Tax policy and the economy, volume 24* (pp. 69–110). The University of Chicago Press. (Cited on pages 56, 80).

Shapiro, M. and Slemrod, J. (2003). "Consumer Response to Tax Rebates." *American Economic Review*, vol. 93, pp. 381–396 (cited on page 80).

Shapiro, M. and Slemrod, J. (2009). "Did the 2008 Tax Rebates Stimulate Spending?" *The American economic review*, vol. 99, pp. 374–379 (cited on page 80).

Straub, L. (2019). *Consumption, savings, and the distribution of permanent income* (tech. rep.) [(Revise and resubmit at Econometrica)]. (Cited on pages 56, 94).

Stroebel, J. and Vavra, J. (2019). "House prices, local demand, and retail prices." *Journal of Political Economy*, vol. 127(3), pp. 1391–1436 (cited on page 62).

Sun, L. and Abraham, S. (2021). "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." *Journal of Econometrics*, vol. 225(2), pp. 175–199 (cited on page 75).

Verhoogen, E. A. (2008). "Trade, Quality Upgrading, and Wage Inequality in the Mexican Manufacturing Sector." *Quarterly Journal of Economics*, vol. 123(2), pp. 489–530 (cited on page 57).

Wachter, J. A. and Yogo, M. (2010). "Why Do Household Portfolio Shares Rise in Wealth?" *The Review of Financial Studies*, vol. 23(11), pp. 3929–3965 (cited on page 56).

A Additional derivations for the theoretical setup

A.1 Price index and indirect utility function

The household solves the following problem

$$\max \prod_m \left[\sum_{i \in G_m} \left(c_{mi} \varphi_{mi}(\xi, P) \right)^{\frac{\sigma-1}{\sigma}} \right]^{\alpha_m(P) \frac{\sigma}{\sigma-1}},$$
$$\text{s.t.} \sum_m \sum_i \mathcal{P}_{mi} c_{mi} \leq X.$$

The Lagrangian is given by

$$\mathcal{L} = \prod_m \left[\sum_{i \in G_m} \left(c_{mi} \varphi_{mi}(\xi, P) \right)^{\frac{\sigma-1}{\sigma_m}} \right]^{\alpha_m(P) \frac{\sigma}{\sigma-1}} - \Lambda \left(\sum_m \sum_i \mathcal{P}_{mi} c_{mi} - X \right),$$

from which it holds that

$$\frac{\partial \mathcal{L}}{\partial c_{mi}} = \alpha_m(P) \prod_m \left[\sum_{i \in G_m} \left(c_{mi} \varphi_{mi}(\xi, P) \right)^{\frac{\sigma-1}{\sigma}} \right]^{\alpha_m(P) \frac{\sigma}{\sigma-1} - 1} \left(c_{mi} \varphi_{mi}(\xi, P) \right)^{\frac{\sigma-1}{\sigma} - 1} \varphi_{mi}(\xi, P) - \Lambda \mathcal{P}_{mi} = 0,$$

and

$$\frac{\mathcal{P}_{mi}}{\mathcal{P}_{mk}} = \left(\frac{c_{mi}}{c_{mk}} \right)^{-\frac{1}{\sigma}} \left(\frac{\varphi_{mi}(\xi, P)}{\varphi_{mk}(\xi, P)} \right)^{\frac{\sigma-1}{\sigma}}. \tag{A.1}$$

Product module price index:
From Equation (A.1) it follows that

$$\left(\frac{\mathcal{P}_{mi}}{\mathcal{P}_{mk}} \right)^{-\sigma} = \frac{c_{mi}}{c_{mk}} \left(\frac{\varphi_{mi}(\xi, P)}{\varphi_{mk}(\xi, P)} \right)^{1-\sigma},$$
$$\Leftrightarrow c_{mk} = \left(\frac{\mathcal{P}_{mk}}{\mathcal{P}_{mi}} \right)^{-\sigma} \left(\frac{\varphi_{mi}(\xi, P)}{\varphi_{mk}(\xi, P)} \right)^{1-\sigma} c_{mi}.$$

Total module expenditures is then given by

$$X_m \equiv \sum_k \mathcal{P}_{mk} c_{mk} = \sum_k \mathcal{P}_{mk} \left(\frac{\mathcal{P}_{mk}}{\mathcal{P}_{mi}}\right)^{-\sigma} \left(\frac{\varphi_{mi}(\xi,P)}{\varphi_{mk}(\xi,P)}\right)^{1-\sigma} c_{mi},$$

$$\Leftrightarrow c_{mi} = \frac{X_m \mathcal{P}_{mi}^{-\sigma} \varphi_{mi}(\xi,P)^{\sigma-1}}{\sum_k \mathcal{P}_{mk}^{1-\sigma} \varphi_{mk}(\xi,P)^{\sigma-1}}.$$

Next, let $C_m \equiv \left[\sum_{i\in G_m} (c_{mi}\varphi_{mi}(\xi,P))^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}$, from which it follows that

$$\begin{aligned}
C_m &\equiv \left[\sum_{i\in G_m} (c_{mi}\varphi_{mi}(\xi,P))^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} \\
&= \left[\sum_{i\in G_m} \left(\frac{X_m \mathcal{P}_{mi}^{-\sigma} \varphi_{mi}(\xi,P)^{\sigma-1}}{\sum_{k\in G_m} \mathcal{P}_{mk}^{1-\sigma} \varphi_{mk}(\xi,P)^{\sigma-1}} \varphi_{mi}(\xi,P)\right)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} \\
&= \left(\frac{X_m}{\sum_{k\in G_m} \mathcal{P}_{mk}^{1-\sigma} \varphi_{mk}(\xi,P)^{\sigma-1}}\right) \left(\sum_{i\in G_m} \mathcal{P}_{mi}^{1-\sigma} \varphi_{mi}(\xi,P)^{\sigma-1}\right)^{\frac{\sigma}{\sigma-1}} \\
&= X_m \left(\sum_{k\in G_m} \mathcal{P}_{mk}^{1-\sigma} \varphi_{mk}(\xi,P)^{\sigma-1}\right)^{-1} \left(\sum_{i\in G_m} \mathcal{P}_{mi}^{1-\sigma} \varphi_{mi}(\xi,P)^{\sigma-1}\right)^{\frac{\sigma}{\sigma-1}} \\
&= X_m \left(\sum_{i\in G_m} \mathcal{P}_{mi}^{1-\sigma} \varphi_{mi}(\xi,P)^{\sigma-1}\right)^{\frac{1}{\sigma-1}}. \qquad \text{(A.2)}
\end{aligned}$$

Lastly, define the income-specific aggregate price index of module m as $\mathcal{P}_m(\xi,P) \equiv X_m|_{C_m=1}$. Then it holds that

$$\mathcal{P}_m(\xi,P) = \left(\sum_i \mathcal{P}_{mi}^{1-\sigma} \varphi_{mi}(\xi,P)^{\sigma-1}\right)^{\frac{1}{1-\sigma}}.$$

Indirect utility:
From equation (A.2) it follows that

$$\begin{aligned}
U &= \prod_m \left[\sum_{i\in G_m} (c_{mi}\varphi_{mi}(\xi,P))^{\frac{\sigma-1}{\sigma}}\right]^{\alpha_m(P)\frac{\sigma}{\sigma-1}} \\
&= \prod_m \left[\frac{X_m}{\mathcal{P}_m(\xi,P)}\right]^{\alpha_m(P)},
\end{aligned}$$

which defines the household's indirect utility from spending X_m on module m with price $\mathcal{P}_m(\xi, P)$. The budget constraint may likewise be stated as $\sum_m X_m \leq X$. Thus, the household problem now reads

$$\max \prod_m \left[\frac{X_m}{\mathcal{P}_m(\xi, P)}\right]^{\alpha_m(P)} \text{ s.t. } \sum_m X_m \leq X,$$

from which it holds that

$$\alpha_i(P)\left[\prod_m \left(\frac{X_m}{\mathcal{P}_m(\xi, P)}\right)^{\alpha_m(P)}\right] X_i^{-1} = \Lambda,$$

where Λ is the Lagrangian multiplier. Since this holds for all product modules, we have that

$$\frac{\alpha_i(P)}{\alpha_j(P)}\frac{X_j}{X_i} = 1 \Leftrightarrow X_i = \frac{\alpha_i(P)}{\alpha_j(P)} X_j,$$

and imposing the budget constraint yields

$$\begin{aligned} X = \sum_i X_i = \sum_i \frac{\alpha_i(P)}{\alpha_j(P)} X_j = \frac{X_j}{\alpha_j(P)} \\ \Leftrightarrow X_j = \alpha_j(P) X, \end{aligned}$$

where we also use that $\sum_i \alpha_i(P_t) = 1$. This holds for all product modules and thus we may write the utility function of the household as

$$\begin{aligned} U &= \prod_n \left[\frac{X_m}{\mathcal{P}_m(\xi, P)}\right]^{\alpha_m(P)} \\ &= X \prod_m \left(\frac{\alpha_m(P)}{\mathcal{P}_m(\xi, P)}\right)^{\alpha_m(P)}. \end{aligned}$$

Thus, when the household knows its income profile, $\{\xi, P\}$, it maximizes utility by choosing the optimal amount of expenditure, X, given the set of prices $\mathcal{P}_m(\xi, P)$. Lastly, the aggregate price index is given by $\mathcal{P}(\xi, P) \equiv \prod_m P_m(\xi, P)^{\alpha_m(P)}$ and hence

$$U = \frac{X}{\mathcal{P}(\xi, P)} \prod_m \alpha_m(P)^{\alpha_m(P)} = \frac{X}{\mathcal{P}(\xi, P)} \cdot K(P).$$

B Additional figures and tables

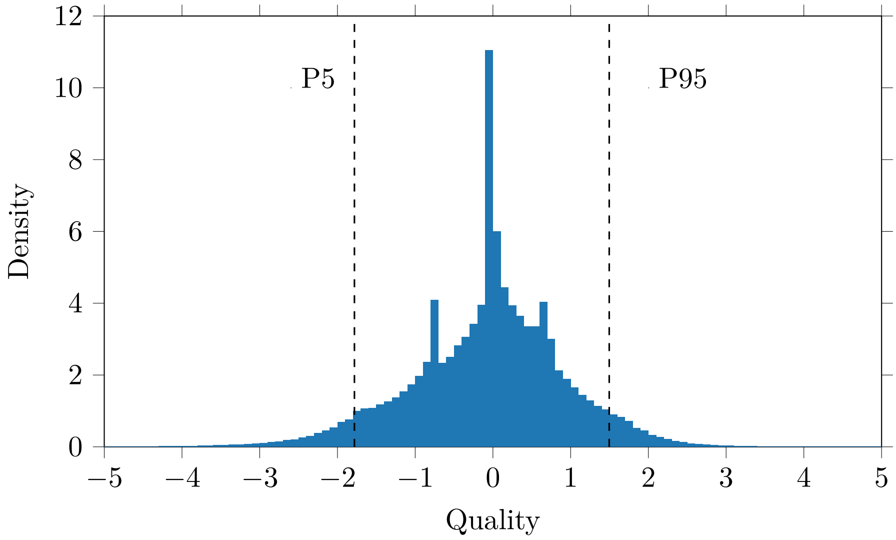


Figure B.1. Distribution of the size-based quality index

Notes. The histogram shows the distribution of the size-based quality index. Dashed lines indicate the 5th and 95th percentiles.

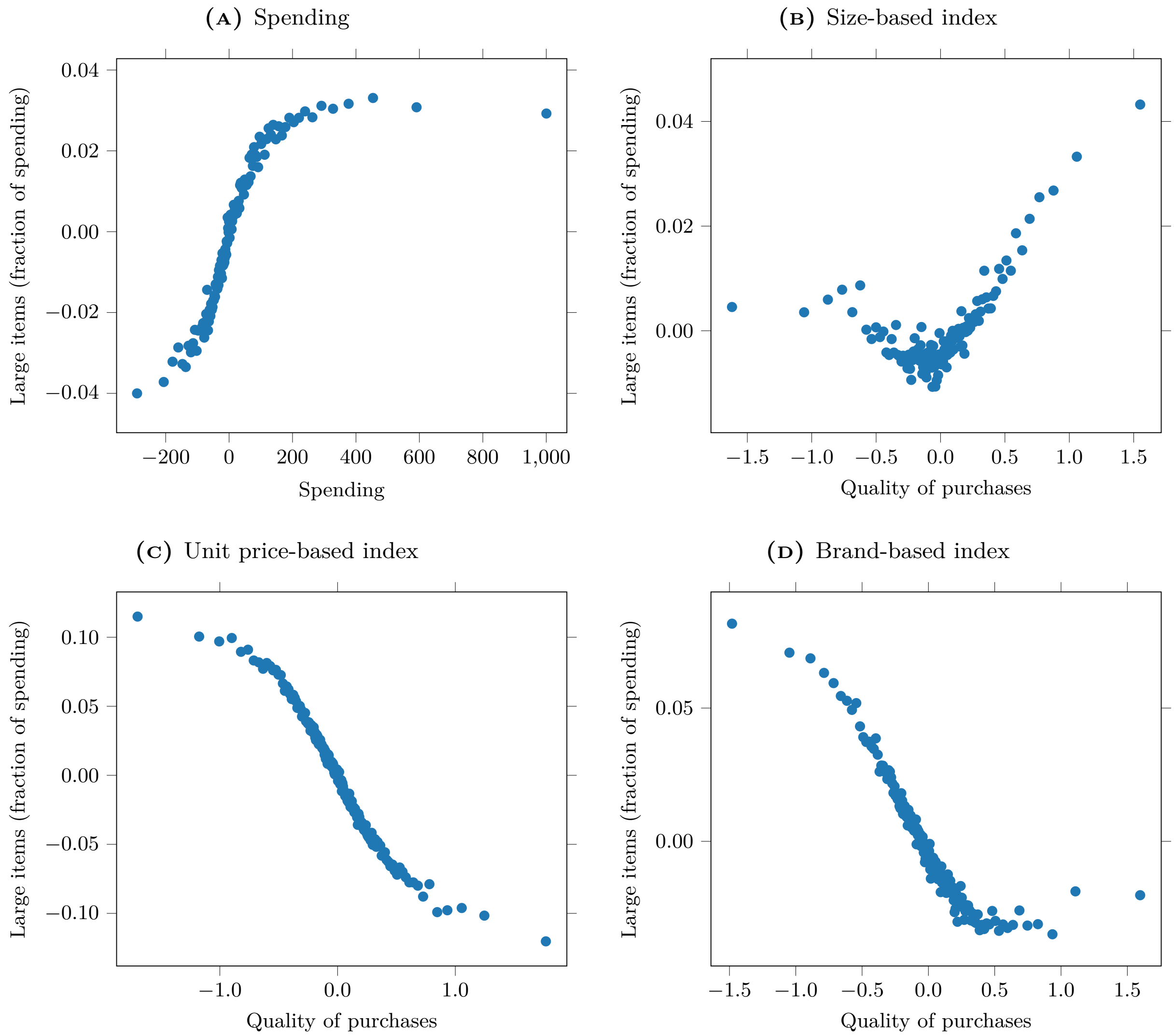


Figure B.2. Households' weekly purchases of large products versus spending and quality of purchases

Notes. Binned scatter plots in which each point is the mean value within bins. y-axes display the fraction of weekly spending on products in the top 40 percent of the product size distribution within product modules. All variables have been residualized with household and week fixed effects.

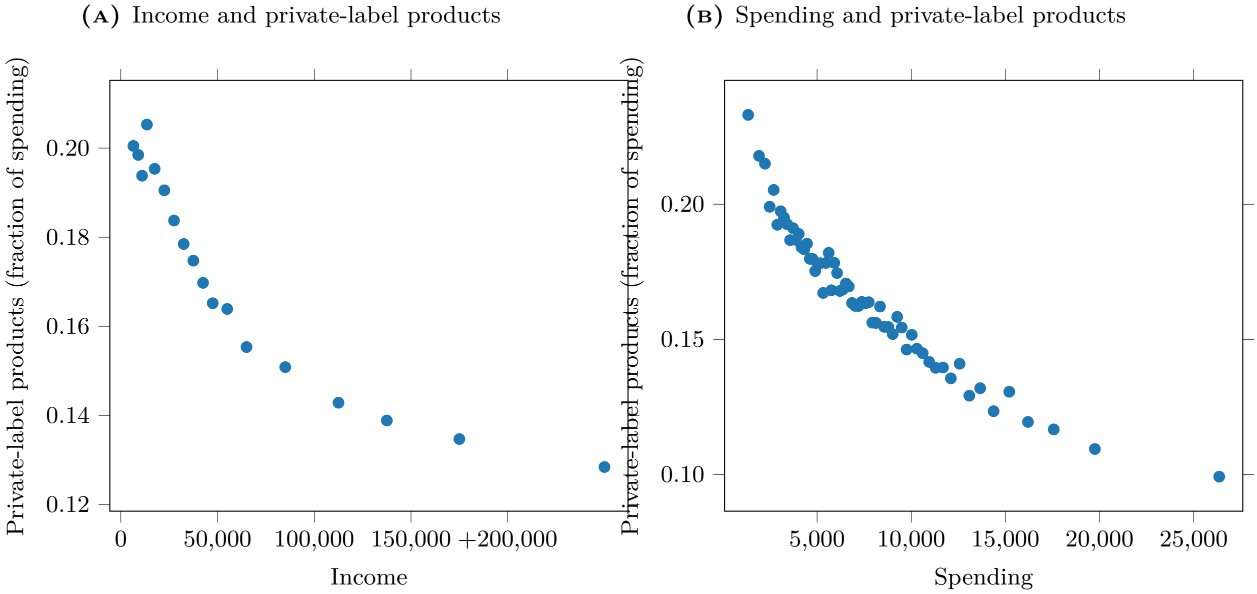


Figure B.3. Households' purchases of private-label products (fraction of spending)

Notes. Panel (a) plots the average expenditure share of private-label products within the midpoint of income bins. Panel (b) is a binned scatter plot of annual spending against the expenditure share of private label products.

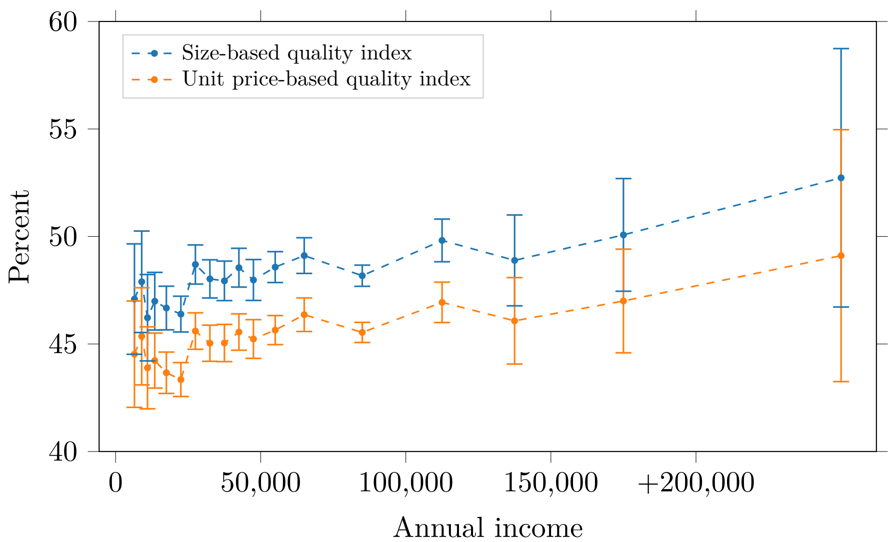


Figure B.4. Spending covered by the quality indices across the income distribution

Notes. The figure shows the average household-level share of annual purchases covered by the quality indices within each income bin. 95 % confidence bands based on heteroskedasticity-robust standard errors are indicated by error bars.

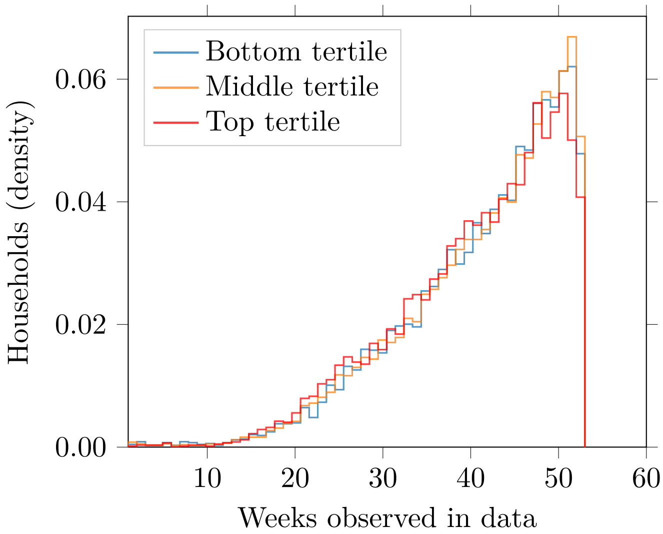


Figure B.5. Weekly purchasing patterns by income tertile

Notes. The figure shows the distribution of households by the number of weeks in 2008 that we observe purchases for each household within three income groups (households with annual income below $35,000, households with annual income between $35,000 and $70,000, and households with annual income above $70,000).

Table B.1. Robustness: Balancing the sample around ESP receipt

	Spending	Size-based quality	Unit price-based quality	Brand-based quality
1 month before ESP	2.81***	0.60**	0.51	0.34
	(0.97)	(0.29)	(0.32)	(0.28)
Contemporaneous month	11.0***	1.15***	1.19***	0.73**
	(1.05)	(0.32)	(0.34)	(0.30)
2 months after ESP	4.87***	0.85**	1.03***	0.34
	(1.11)	(0.34)	(0.37)	(0.33)
3 months after ESP	4.31***	0.42	0.82**	0.36
	(1.14)	(0.35)	(0.38)	(0.34)
Week × household obs.	827,175	662,386	651,663	651,549
Households	20,175	20,160	20,158	20,158

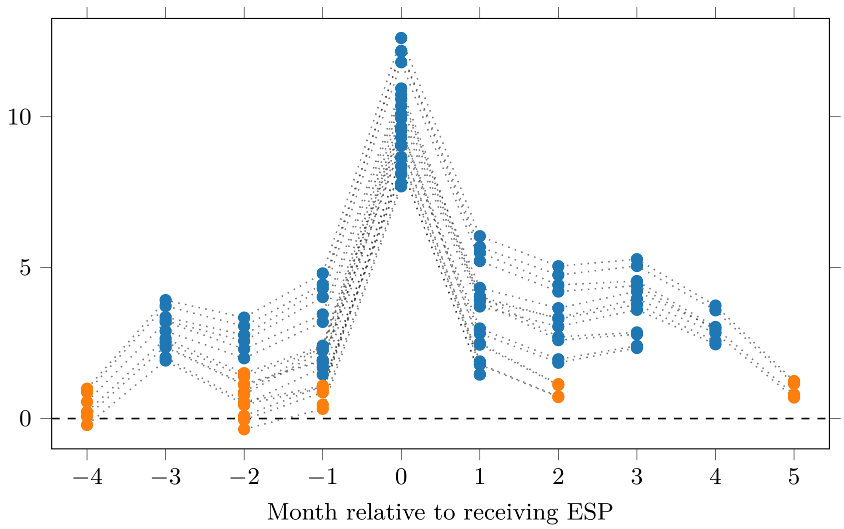
Notes. The table shows the estimates of $\tilde{\beta}$ from equation (11) when balancing the sample around the ESP receipt. Households are included 16 weeks prior to ESP receipt until 23 weeks after. The lead coefficients 16 weeks and 5 weeks prior to ESP receipt are normalized to zero. Estimates from regressions with a quality measure as the dependent variable have been scaled by 100. Standard errors are clustered at the household level and reported in parentheses. *, ** and *** denote significance at the 0.1, 0.05 and 0.01 level respectively.

Table B.2. Robustness: ESP estimates from non-constrained regression

Weeks relative to ESP receipt	Spending	Size-based quality
-15	-1.10	0.06
	(1.49)	(0.41)
-14	-1.82	-0.03
	(1.49)	(0.41)
-13	-2.16	0.10
	(1.51)	(0.42)
-12	0.29	-0.11
	(1.45)	(0.43)
-11	2.48*	-0.21
	(1.48)	(0.41)
-10	0.77	-0.49
	(1.39)	(0.40)
-9	-0.74	-0.09
	(1.39)	(0.39)
-8	-0.58	-0.01
	(1.37)	(0.39)
-7	1.53	-0.06
	(1.35)	(0.39)
-6	-0.76	0.59
	(1.33)	(0.40)
-5	-1.18	-0.16
	(1.34)	(0.41)
-4	-0.65	0.44
	(1.38)	(0.39)
-3	2.64*	0.87**
	(1.43)	(0.39)
-2	0.76	0.05
	(1.48)	(0.39)
0	10.2***	0.63
	(1.62)	(0.44)
+1	10.0***	1.02**
	(1.66)	(0.46)
+2	4.93***	1.01**
	(1.69)	(0.48)
+3	6.31***	0.40
	(1.76)	(0.51)
+4	1.76	-0.09
	(1.85)	(0.53)
+5	1.10	0.76
	(1.93)	(0.56)
+6	0.29	0.49
	(2.06)	(0.60)
+7	-0.05	0.29
	(2.16)	(0.63)
+8	-1.53	0.14
	(2.28)	(0.66)
+9	-1.28	-0.23
	(2.40)	(0.70)
+10	-0.06	-0.47
	(2.56)	(0.72)
+11	0.045	0.02
	(2.65)	(0.76)
+12	-0.60	-0.53
	(2.77)	(0.80)
+13	-0.57	0.36
	(2.90)	(0.83)
+14	-0.17	-0.25
	(3.03)	(0.87)
+15	-1.99	0.28
	(3.14)	(0.90)
+16	-2.68	-0.49
	(3.27)	(0.94)
+17	-1.29	-0.32
	(3.42)	(0.97)
+18	-3.38	-0.53
	(3.54)	(1.01)
+19	-3.76	-0.07
	(3.67)	(1.05)
+20	-6.98*	0.02
	(3.78)	(1.09)
+21	-3.79	-0.29
	(3.93)	(1.12)
+22	-6.37	0.02
	(4.06)	(1.16)
+23	-4.62	-0.11
	(4.24)	(1.20)
+24	-7.81*	-0.01
	(4.40)	(0.01)
p-value for test on leads	0.100	0.484
Week × household obs.	827,175	661,803
Households	20,175	20,162

Notes. The table shows the estimates of $\tilde{\beta}$ from equation (11) when the coefficients are not constrained. The 1 and 16 weeks lead coefficients are normalized to zero. The sample is balanced around ESP receipt. Estimates from regressions with a quality measures as the dependent variable have been scaled by 100. The p-values reported are p-values for an F-test of the hypothesis that all lead coefficients equal zero. Standard errors are clustered at the household level and reported in parentheses. *, ** and *** denote significance at the 0.1, 0.05 and 0.01 level respectively.

(A) Spending estimates



(B) Quality estimates (size-based)

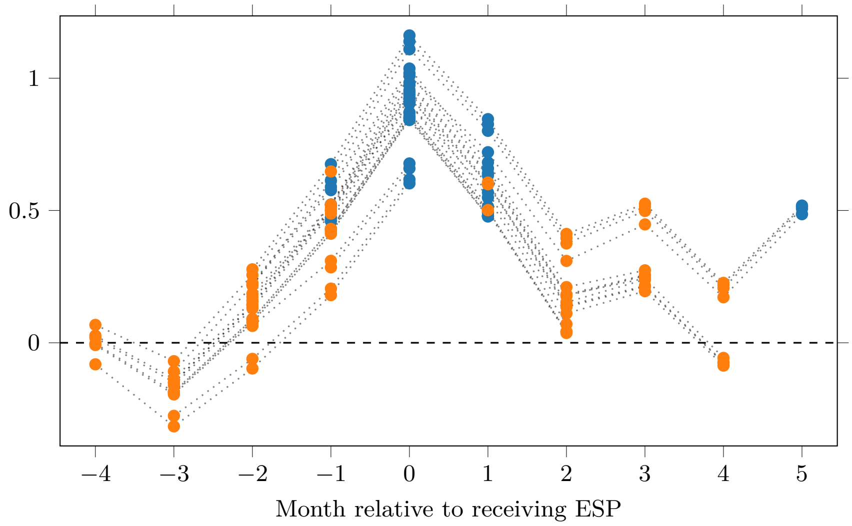


FIGURE B.6. Robustness: Choice of leads and lags in regression

Notes. Figures plot the estimates from equation (11) with different sets of leads and lags. Estimates from the same regression are joined by dashed lines. A blue dot indicates that the point estimate is significantly different from zero at the 5 % level. Panel (a) plots the estimates from the regression with spending as the dependent variable, while panel (b) plots the estimates from the regression with size-based quality variable as the dependent variable. Estimates in panel (b) are scaled by 100. Standard errors are clustered at the household level

Table B.3. Robustness: Heterogeneity of ESP response by alternative income groups

	Baseline grouping		Grouping on future income		Adjust for age and size	
	Spending	Quality	Spending	Quality	Spending	Quality
			Bottom tertile			
1 month before ESP	6.46**	0.26	5.75*	0.19	7.83***	-0.57
	(3.12)	(1.05)	(3.24)	(1.08)	(2.87)	(0.86)
Contemporaneous month	17.1***	1.13	14.4***	1.03	16.9***	0.81
	(3.40)	(1.09)	(3.46)	(1.12)	(3.07)	(0.89)
2 months after ESP	8.86***	0.26	7.40**	0.36	7.39**	0.20
	(3.30)	(1.11)	(3.36)	(1.14)	(2.99)	(0.90)
3 months after ESP	7.21**	0.15	5.94*	-0.23	7.56***	-0.48
	(3.06)	(1.06)	(3.09)	(1.09)	(2.77)	(0.87)
Week × households obs.	241,097	198,841	237,864	196,130	241,574	199,892
Households	4,549	4,547	4,488	4,485	4,558	4,554
			Middle tertile			
1 month before ESP	2.63	1.00*	3.96*	1.37**	0.58	1.72***
	(2.15)	(0.60)	(2.16)	(0.62)	(2.19)	(0.61)
Contemporaneous month	9.98***	1.47**	10.6***	1.61**	7.62***	1.46**
	(2.18)	(0.63)	(2.20)	(0.64)	(2.24)	(0.64)
2 months after ESP	3.17	1.64***	4.28**	1.63**	4.77**	1.75***
	(2.13)	(0.63)	(2.16)	(0.65)	(2.20)	(0.65)
3 months after ESP	1.48	1.04*	2.46	0.86	2.13	1.16*
	(2.12)	(0.62)	(2.17)	(0.63)	(2.14)	(0.63)
Week × households obs.	265,477	219,822	261,926	217,712	243,270	202,184
Households	5,009	5,007	4,942	4,941	4,590	4,590
			Top tertile			
1 month before ESP	0.73	0.56	0.10	0.32	0.88	0.61
	(2.33)	(0.57)	(2.29)	(0.54)	(2.41)	(0.64)
Contemporaneous month	4.47*	0.28	5.83**	0.32	5.87**	0.59
	(2.42)	(0.59)	(2.39)	(0.57)	(2.49)	(0.66)
2 months after ESP	1.72	-0.044	1.81	-0.0092	0.67	0.075
	(2.39)	(0.59)	(2.34)	(0.57)	(2.44)	(0.67)
3 months after ESP	1.08	-0.61	1.16	-0.085	-0.66	-0.11
	(2.27)	(0.55)	(2.21)	(0.53)	(2.37)	(0.63)
Week × households obs.	206,859	170,012	213,643	174,833	228,589	186,599
Households	3,903	3,903	4,031	4,031	4,313	4,313

Notes. The table shows the estimates of $\tilde{\beta}$ from equation (12). Regressions only include households that enter the data in at least 2008, 2009, and 2010. Columns (1) and (2) use the same income groups as in table 4, columns (3) and (4) base income groups on the modal income group in subsequent years, and columns (5) and (6) base income groups on the modal age and size-adjusted income tertile in subsequent years. Standard errors are clustered at the household level and reported in parentheses. *, ** and *** denote significance at the 0.1, 0.05 and 0.01 level respectively.

C Ranking of products in different years

To assess if products are ranked similarly in other years by the size-based index, we have constructed the same quality index but using the prices of 2007 and 2009. The correlation coefficients between the original index and the indices for 2007 and 2009 are 0.74 in both years at the product-CBSA pair level. Note that there is substantial entry and exit of products in

Table B.4. Robustness: ESP estimates at department, group and module level

	Spending	Size-based quality	Unit price-based quality	Brand-based quality
		Product category: Department level		
1 month before ESP	2.44***	0.51*	0.43	0.32
	(0.73)	(0.26)	(0.28)	(0.24)
Contemporaneous month	6.76***	0.96***	0.95***	0.65**
	(0.76)	(0.27)	(0.30)	(0.25)
2 months after ESP	2.78***	0.68**	0.72**	0.32
	(0.73)	(0.27)	(0.30)	(0.26)
3 months after ESP	2.70***	0.32	0.44	0.30
	(0.71)	(0.26)	(0.29)	(0.25)
Household × week × department obs.	9,812,791	3,169,578	2,959,971	2,958,773
Households	20,175	20,159	20,160	20,160
		Product category: Group level		
1 month before ESP	1.03***	0.14	0.033	0.19
	(0.33)	(0.19)	(0.17)	(0.20)
Contemporaneous month	2.69***	0.40**	0.34*	0.48**
	(0.34)	(0.20)	(0.18)	(0.21)
2 months after ESP	1.01***	0.22	0.14	0.33
	(0.33)	(0.20)	(0.18)	(0.21)
3 months after ESP	0.95***	0.06	0.01	0.09
	(0.32)	(0.20)	(0.18)	(0.21)
Household × week × group obs.	67,165,893	6,836,181	6,562,980	6,568,743
Households	20,175	20,146	20,145	20,145
		Product category: Module level		
1 month before ESP	0.55***	0.43**	0.32*	0.16
	(0.19)	(0.18)	(0.18)	(0.14)
Contemporaneous month	1.44***	0.56***	0.55***	0.36**
	(0.20)	(0.18)	(0.18)	(0.14)
2 months after ESP	0.48**	0.24	0.37**	0.28*
	(0.19)	(0.19)	(0.18)	(0.14)
3 months after ESP	0.44**	0.25	0.18	0.12
	(0.19)	(0.18)	(0.18)	(0.14)
Household × week × module obs.	156,318,253	7,179,775	6,908,330	6,903,046
Households	20,175	20,139	20,136	20,136

Notes. The table shows the estimates of $\tilde{\beta}$ from equation (11) at the product department (upper panel), product group (middle panel), and product module (lower panel) level. Estimates from regressions with a quality measures as the dependent variable have been scaled by 100. Regressions include household × product category and week × product category fixed effects. Standard errors are clustered at the household level and reported in parentheses. *, ** and *** denote significance at the 0.1, 0.05 and 0.01 level respectively.

the Nielsen data. Hence, we should not expect to find the exact same quality ranking between products in different years.

We have also calculated the normalized rank of the quality index for each product within the

Table B.5. Robustness: ESP estimates with disbursement method fixed effects

	Spending	Size-based quality	Unit price-based quality	Brand-based quality
1 month before ESP	1.55	0.87*	0.85	0.81*
	(1.79)	(0.50)	(0.54)	(0.48)
Contemporaneous month	8.63***	1.26**	1.07*	0.98*
	(1.85)	(0.52)	(0.56)	(0.51)
2 months after ESP	3.28*	0.82	1.08*	0.68
	(1.81)	(0.52)	(0.56)	(0.51)
3 months after ESP	1.35	0.58	1.20**	0.67
	(1.74)	(0.50)	(0.55)	(0.50)
Week × household obs.	1,069,275	835,470	831,244	831,107
Households	20,175	20,165	20,166	20,166

Notes. The table shows the estimates of $\tilde{\beta}$ from equation (11) when week fixed effects are replaced with disbursement method × week fixed effects. Estimates from regressions with a quality measures as the dependent variable have been scaled by 100. Standard errors are clustered at the household level and reported in parentheses. *, ** and *** denote significance at the 0.1, 0.05 and 0.01 level respectively.

group of products of the same size in the same module sold in the same CBSA:

$$\frac{\text{rank}_{i,m,s,c} - 1}{N_{m,s,c} - 1} \tag{C.1}$$

where $\text{rank}_{i,m,s,c}$ is the rank of product-CBSA pair (i, c)'s quality index relative to the other products of the same size s in the same module m sold in CBSA c. $N_{m,s,c}$ is the number of product-CBSA pairs in the CBSA-size-module group.

This normalized rank is calculated separately for the years 2007, 2008 and 2009. We then construct binned scatter plots for the normalized rank in 2008 against the two other years in the following way. First, we sort all product-CBSA pairs by their normalized rank in 2008 and divide them into 100 equal-sized bins. Second, we calculate the median normalized rank within these bins for 2007 through 2009 and use these medians to create two scatter plots.

Figure C.1 shows these two binned scatter plots. The median normalized ranks in 2008 is plotted against the corresponding medians in 2007 in panel (a), while panel (b) plots the median normalized ranks in 2008 against the medians for 2009. All points are very close to the 45 degree line. Hence, along with the high cross-year correlation coefficients for the quality indices, these scatter plots show that the quality rank is approximately the same in different years across the entire quality index distribution.

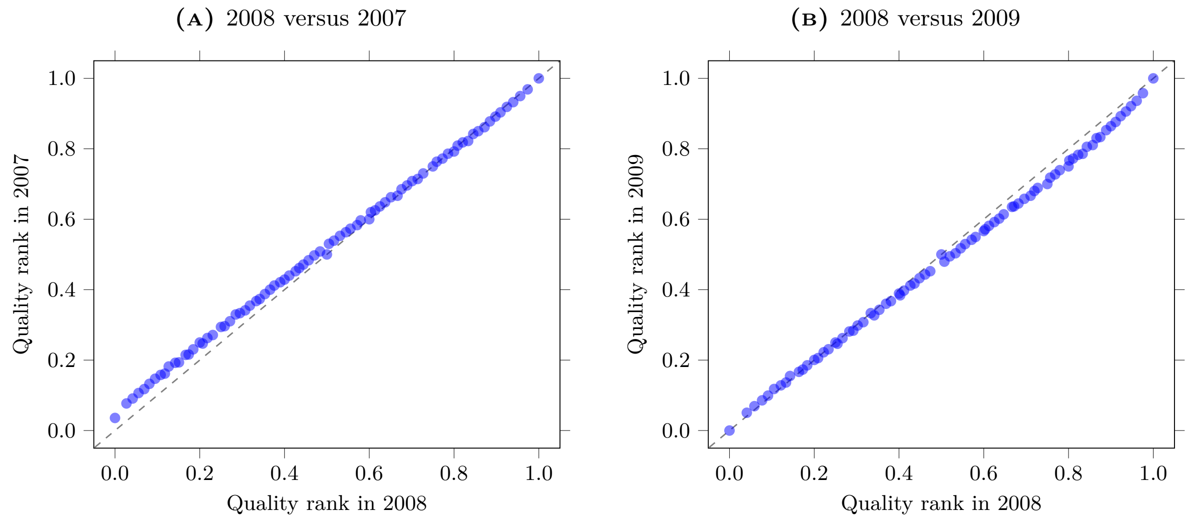


FIGURE C.1. Product-CBSA pairs' quality rank in 2008 versus 2007 and 2009

Notes. Binned scatter plots with 100 equal-sized bins based on the normalized rank of product-CBSA pairs' size-based quality index in 2008. Each point is the median value of the normalized rank within bins. Dashed lines are 45 degree lines.

D Heterogeneity of response to ESP by liquidity

Liquidity constraints are important for shaping households' consumption behavior since liquidity constrained households display larger propensities to consume out of transitory income shocks. Many studies on ESPs also find that liquidity constrained households have a larger propensity to consume out of the ESP relative to non-constrained households (Broda and Parker, 2014; Misra and Surico, 2014; Parker, 2017). Similarly, Kaplan and Violante (2014) estimate that between 17.5 percent and 35 percent of US households are hand-to-mouth consumers due to liquidity constraints and that many of these households are wealthy but still hand-to-mouth consumers since they hold the lion's share of their wealth in illiquid assets. Since liquidity constraints have received considerable attention in the literature on MPCs, we also report estimates of how the quality differ by households' access to liquid wealth.

The ESP survey contains a question asking households if they had access to liquid wealth to buffer against unexpected declines in income or increases in expenses.[48] 35 percent of the households in our sample report that they do not have access to liquid wealth. We now look at how households' quality response differ by their access to liquidity. Regression (12) is estimated according to this split, and the results are presented in table D.1.

[48] The survey question was "In case of an unexpected decline in income or increase in expenses, do you have at least two months of income available in cash, bank accounts, or easily accessible funds?" to which the households could answer "Yes" or "No".

Table D.1. Heterogeneity of ESP response by access to liquidity

	Weekly spending	Size-based quality	Unit price-based quality	Brand-based quality
	Liquidity constrained			
1 month before ESP	6.30***	-0.12	0.082	0.44
	(2.31)	(0.58)	(0.61)	(0.55)
Contemporaneous month	19.0***	1.22**	1.79***	1.37**
	(2.40)	(0.59)	(0.64)	(0.57)
2 months after ESP	7.05***	0.51	0.92	0.27
	(2.31)	(0.60)	(0.64)	(0.58)
3 months after ESP	6.31***	0.24	0.66	0.034
	(2.19)	(0.58)	(0.63)	(0.57)
Week × household obs.	375,770	285,342	284,195	284,147
Households	7,090	7,085	7,085	7,085
	Not constrained			
1 month before ESP	2.34*	0.89**	0.81*	0.88**
	(1.41)	(0.42)	(0.46)	(0.41)
Contemporaneous month	7.41***	0.96**	0.78*	0.81*
	(1.46)	(0.44)	(0.47)	(0.42)
2 months after ESP	3.78***	0.84*	0.87*	0.67
	(1.45)	(0.44)	(0.48)	(0.43)
3 months after ESP	2.58*	0.34	0.56	0.72*
Week × household obs.	693,505	550,128	547,049	546,960
Households	13,085	13,080	13,081	13,081

Notes. The table shows the estimates of $\tilde{\beta}$ from equation (12) with the sample split by being liquidity constrained or not. Estimates from regressions with a quality measures as the dependent variable have been scaled by 100. Standard errors are clustered at the household level and reported in parentheses. *, ** and *** denote significance at the 0.1, 0.05 and 0.01 level respectively.

Over three months, the propensity to consume out of the ESP for the liquidity constrained is almost two and a half times as large as for the non-constrained (14.6 percent versus 6.2 percent). These estimates are in line with those of Broda and Parker (2014). Both groups increase the quality of spending although the effect is most significant for the constrained households.

E The f() function with alternative parameter values

The Gompertz function has the general functional form

$$f(\xi_t, P_t) = \kappa \mathrm{e}^{-\iota \mathrm{e}^{-\delta \cdot \xi_t P_t}}. \tag{E.1}$$

For $\iota > 0$ and $\delta > 0$, $f()$ will be S-shaped. Letting either ι or δ be negative, the shape becomes hyperbolic on its support $\mathbb{R}_+$. We showcase the two scenarios below. Note that in the case where both ι and δ are negative, $f()$ is exploding. We do not show that here.

Case 1: $\iota < 0$ and $\delta > 0$

In this scenario, the asymptotic level is a lower bound with value κ. In the zero-income event, the maximum value of $f()$ is reached at $\kappa e^{-\iota} > \kappa$.

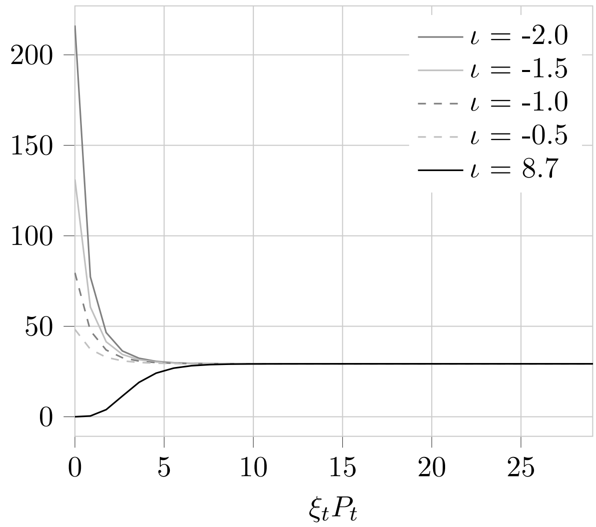


FIGURE E.1. $f()$ function with $\iota < 0$, $\delta > 0$. Varying ι.

Notes. The black line shows the calibrated $f()$ function used in the solution of the model.

Case 2: $\iota > 0$ and $\delta < 0$

In this scenario, the asymptotic level is a lower bound with value 0. In the zero-income event, the maximum value of $f()$ is reached at $\kappa e^{-\iota}$.

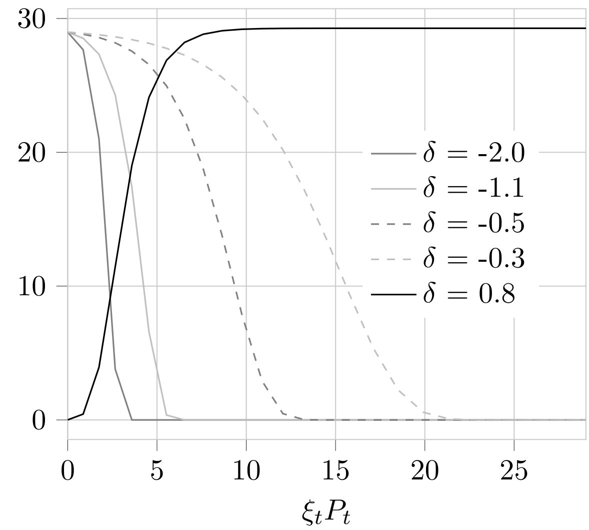


FIGURE E.2. $f()$ function with $\iota > 0$, $\delta < 0$. Varying δ.

Notes. The black line shows the calibrated $f()$ function. For the purpose of exposition, we changed the values of ι to 0.1 in the $\delta < 0$ scenarios.

F Computational appendix

In this section we provide an explanation of how the fast multi-linear interpolation algorithm from Druedahl, 2021 is implemented in the solution of the dynamic programming problem in section 5.1.

F.1 EGM with a fast, multi-linear interpolation algorithm

The Bellman equation is given by

$$V_t(M_t, P_t, \xi_t) = \max_{X_t} \frac{(X_t \cdot f(\xi_t, P_t))^{1-\rho}}{1-\rho} + \beta \mathbb{E}_t[V_{t+1}(M_{t+1}, P_{t+1}, \xi_{t+1})], \tag{F.1}$$

To solve the problem, we employ the Endogenous Grid Method (EGM) combined with an upper envelope as in Druedahl and Jørgensen, 2017.[49] However, implementing the EGM in a multi-dimensional setting like ours is costly due to the need for multi-linear interpolation. In the following, we describe how to alleviate this issue by exploiting some structure of our problem. As we show, we end up only doing a two-dimensional interpolation.

Following Druedahl, 2021, we define two auxiliary variables, w_t and q_t, which we refer to as *post-decision variables*. Common to these is that they can be computed when the so-called *post-decision states* are known. In particular, in the present problem, only two of the post-decision states, namely end-of-period assets, A_t, and permanent income, P_t, are needed to compute w_t and q_t.[50]

Post-decision value function. w_t is defined as

$$\begin{aligned} \beta \mathbb{E}_t[V_{t+1}(M_{t+1}, P_{t+1}, \xi_{t+1})] &= \beta \mathbb{E}_t[V_{t+1}(RA_t + Y_{t+1}, P_{t+1}, \xi_{t+1})] \\ &= \beta \mathbb{E}_t[V_{t+1}(RA_t + \xi_{t+1}\psi_{t+1}GP_t, \psi_{t+1}GP_t, \xi_{t+1})] \\ &\equiv w(A_t, P_t), \end{aligned} \tag{F.2}$$

and we refer to w_t as the *post-decision value function*. From equation (F.2), we see that after knowing A_t and P_t, we can compute w_t for given values of ξ_{t+1} and ψ_{t+1}. To compute the expectation, we can use an appropriate weighting for each of the shocks.[51]

Post-decision marginal value of cash. From the Euler equation of the problem, we have

[49] The upper envelope is needed to rule out scenarios, where the Euler equation is not sufficient for generating points on the consumption curve.

[50] In this terminology, also transitory income shocks, ξ_t, is a post-decision state.

[51] In particular, we use the Gauss-Hermite quadrature to compute the expectation.

that

$$X_t^{-\rho} f(\xi_t, P_t)^{1-\rho} = \beta R \mathbb{E}_t[X_{t+1}^{-\rho} f(\xi_{t+1}, P_{t+1})^{1-\rho}]. \tag{F.3}$$

Defining q_t as the right-hand side of this expression, we have that

$$\begin{aligned}
\beta R \mathbb{E}_t[X_{t+1}^{-\rho} f(\xi_{t+1}, P_{t+1})^{1-\rho}] &= \mathbb{E}_t\left[\beta R \left(X_{t+1}^{\star}(M_{t+1}, \xi_{t+1}, P_{t+1})\right)^{-\rho} \left(f(\xi_{t+1}, P_{t+1})\right)^{1-\rho}\right] \\
&= \mathbb{E}_t\left[\beta R \left(X_{t+1}^{\star}(RA_t + Y_{t+1}, \xi_{t+1}, P_{t+1})\right)^{-\rho} \left(f(\xi_{t+1}, P_{t+1})\right)^{1-\rho}\right] \\
&= \mathbb{E}_t\left[\beta R \left(X_{t+1}^{\star}(RA_t + \xi_{t+1}\psi_{t+1}GP_t, \xi_{t+1}, \psi_{t+1}GP_t)\right)^{-\rho} \left(f(\xi_{t+1}, \psi_{t+1}GP_t)\right)^{1-\rho}\right] \\
&\equiv q_t(A_t, P_t), \qquad \text{(F.4)}
\end{aligned}$$

and we refer to q_t as the *post-decision marginal value of cash*. As for w_t, we also see that after knowing A_t and P_t along with some optimal expenditure choice $X_{t+1}^{\star}$, we can compute q_t for given values of ξ_{t+1} and ψ_{t+1}. Computing the expectation can also be done in the same way as for w_t, using an appropriate weighting for the shocks.

Endogenous grid method. After having solved for q_t, we see that knowing the last post-decision state, ξ_t, we can fully determine the time t expenditure choice, X_t. Specifically, we have that

$$\begin{aligned}
X_t^{-\rho} f(\xi_t, P_t)^{1-\rho} = q_t(A_t, P_t) &\Leftrightarrow \\
X_t &= \left(\frac{q_t(A_t, P_t)}{f(\xi_t, P_t)^{1-\rho}}\right)^{-\frac{1}{\rho}} \\
&= F(A_t, \xi_t, P_t; \xi_{t+1}, \psi_{t+1}). \qquad \text{(F.5)}
\end{aligned}$$

We thus see that, given that the Euler equation is a necessary condition for utility maximization, we can calculate all the points on the expenditure function and beginning-of-period cash-on-hand from

$$X_t = F(A_t, \xi_t, P_t; \xi_{t+1}, \psi_{t+1}), \tag{F.6}$$

$$M_t = A_t + X_t. \tag{F.7}$$

Additionally, we also see that after having calculated X_t, the value function in equation (F.1) readily available.

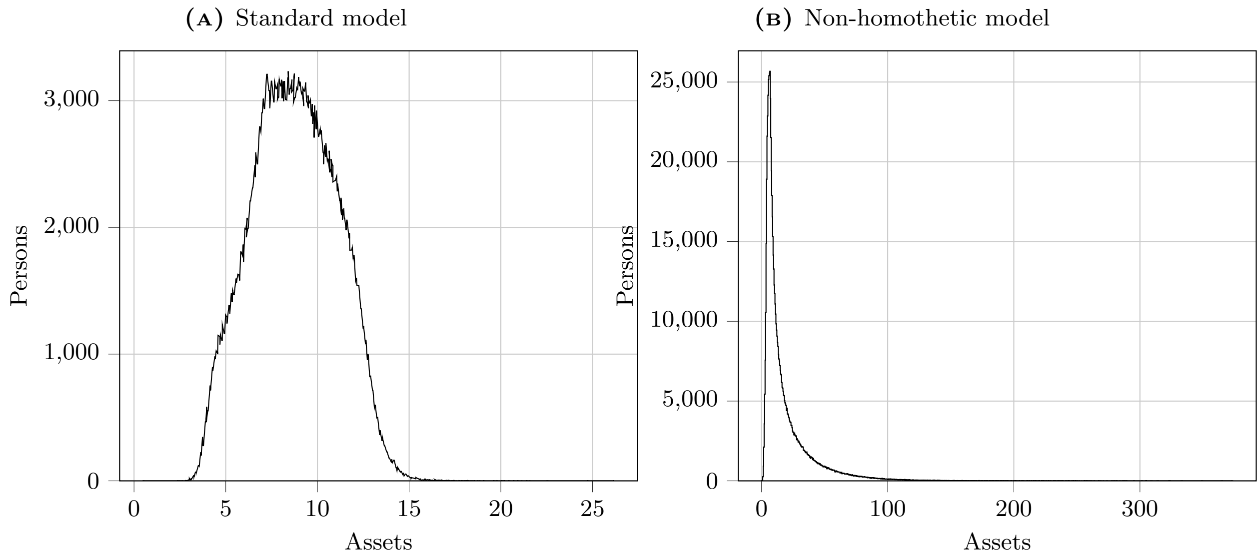


FIGURE G.1. Asset distributions in the standard and non-homothetic model

G Asset distributions in the two models

H Quality in the theoretical model

H.1 Requirement for f() to be increasing in quality

From Section 2, we have that that $f(\xi, P) \equiv K(P)/\mathcal{P}(\xi, P) = K(P)/\prod_m \mathcal{P}_m(\xi, P)^{\alpha_m(P)}$ where

$$\mathcal{P}_m(\xi, P) = \left(\sum_{i \in G_m} \mathcal{P}_{mi}^{1-\sigma} \varphi_{mi}(\xi, P)^{\sigma - 1} \right)^{\frac{1}{1-\sigma}}.$$

Take the partial derivative of $\ln f(\xi, P)$ w.r.t. ξ:

$$\begin{aligned} \frac{\partial \ln \left(\frac{K(P)}{\prod_m \mathcal{P}_m(\xi,P)^{\alpha_m(P)}} \right)}{\partial \xi} &= \frac{\partial \ln K(P)}{\partial \xi} - \frac{\partial \sum_m \alpha_m(P) \ln \mathcal{P}_m(\xi, P)}{\partial \xi} \\ &= - \frac{\partial \sum_m \alpha_m(P) \ln \mathcal{P}_m(\xi, P)}{\partial \xi} \end{aligned} \tag{I}$$

Consider some module, m, and look at the partial derivative:

$$\begin{aligned}
\frac{\partial \alpha_m(P) \ln \mathcal{P}(\xi, P)}{\partial \xi} =& \frac{\partial \alpha_m(P) \frac{1}{1-\sigma} \ln \left[\sum_{i \in G_m} \mathcal{P}_{mi}^{1-\sigma} \varphi_{mi}(\xi, P)^{\sigma-1} \right]}{\partial \xi} \\
=& - \alpha_m(P) \frac{1}{\sum_{i \in G_m} \mathcal{P}_{mi}^{1-\sigma} \varphi_{mi}(\xi, P)^{\sigma-1}} \cdot \sum_{i \in G_m} \mathcal{P}_{mi}(\xi, P)^{1-\sigma} \varphi_{mi}(\xi, P)^{\sigma-2} \frac{\partial \varphi_{mi}(\xi, P)}{\partial \xi} \\
=& - \alpha_m(P) \frac{1}{\mathcal{P}_m(\xi, P)^{1-\sigma}} \cdot \sum_{i \in G_m} \mathcal{P}_{mi}(\xi, P)^{1-\sigma} \varphi_{mi}(\xi, P)^{\sigma-2} \frac{\partial \varphi_{mi}(\xi, P)}{\partial \xi} \qquad \text{(II)}
\end{aligned}$$

Now, the requirement is that $\frac{\partial f(\xi, P)}{\partial P} > 0$. Using this, equation (I) gives us

$$\begin{aligned}
& \sum_m \alpha_m(P) \frac{1}{\mathcal{P}_m(\xi, P)^{1-\sigma}} \cdot \sum_{i \in G_m} \mathcal{P}_{mi}(\xi, P)^{1-\sigma} \varphi_{mi}(\xi, P)^{\sigma-2} \frac{\partial \varphi_{mi}(\xi, P)}{\partial \xi} > 0 \\
&= \sum_m \sum_{i \in G_m} A_m B_{mi} \frac{\partial \varphi_{mi}(\xi, P)}{\partial \xi} > 0,
\end{aligned}$$

where

$$A_m \equiv \alpha_m(P) \frac{1}{\mathcal{P}_m(\xi, P)^{1-\sigma}} \geq 0, \qquad \text{and} \qquad B_{mi} \equiv \mathcal{P}_{mi}(\xi, P)^{1-\sigma} \varphi_{mi}(\xi, P)^{\sigma-2} > 0.$$

H.2 MQC with alternative parameter values

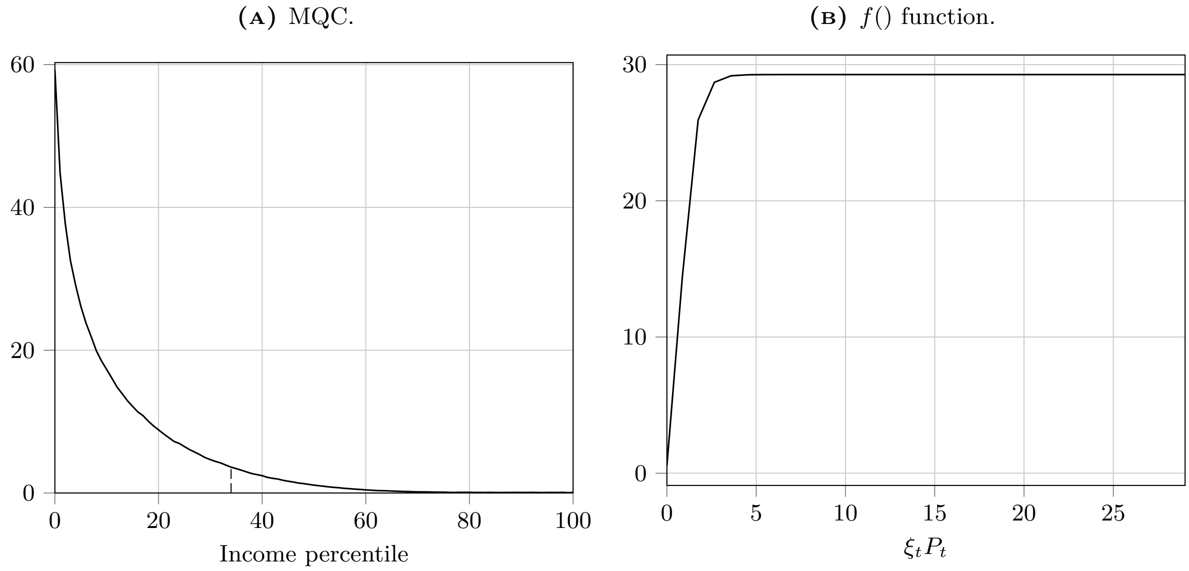


Figure H.1. MQC and $f()$ function with $\kappa = 50$, $\iota = 4$ and $\delta = 4$.

Notes. The dashed, vertical lines in panel H.1a represent the cut-off between the low-middle income and middle-high income, respectively. Average MQC is 15.6, 1.4 and 0.1 for the low-income group, middle-income group and high-income group, respectively.

Chapter 3

Local Projections or VARs? A Data-Driven Selection Rule for Finite-Sample Estimation of Impulse Responses

Local Projections or VARs? A data-driven selection rule for finite-sample estimation of impulse responses

Christoffer J. Weissert[*] Anders F. Kronborg[†]

Abstract

We derive a data-driven rule for when to choose local projection, vector autoregression or a mix of these two methods' estimates of impulse responses in finite samples. We show that local projections and VARs are linked in finite samples: local projections can be derived from a first-step estimate of the VAR impulse responses and a second-step linear correction of the forecast errors of the VAR. The sum of these two estimates yields the local projection estimate and the second-step estimate therefore captures the difference between local projections and VARs. We coin the difference between local projections and VARs the local projections contribution to VARs and since this difference is estimated we can perform inference on it. The selection rule is based on this inference and depends on a simple key statistic for the local projection contribution: the coefficient of variation. When the coefficient of variation is large, the local projection contribution is imprecisely estimated and more weight is put on the VAR estimate. When the coefficient of variation is small more weight is put on the local projection estimate. We show that the selection rule performs well in a host of Monte Carlo studies.

JEL classification: C32, C52, C53.

Keywords: impulse response function, local projection, vector autoregression.

[*] University of Copenhagen and Central Bank of Denmark. Email: cjw@econ.ku.dk.
[†] Danish Research Institute for Economic Analysis and Modelling. Email: ank@dreamgruppen.dk.

1 Introduction

Vector autoregressions and local projections are the two primary methods used for estimating impulse responses in empirical macroeconomics. Plagborg-Møller and Wolf (2021) show that the two methods are asymptotically equivalent but often, however, different - and sometimes even opposing - impulse responses are estimated in finite sample settings. Examples of pronounced differences reach all the way into studies of core macroeconomic questions (Ramey, 2016) and Li, Plagborg-Møller and Wolf (2021) furthermore show that a general feature of U.S. macroeconomic time series is that local projections and vector autoregressions disagree on the impulse responses they estimate. As the finite sample setting is the ballgame for most practitioners, knowing how to choose between vector autoregression and local projection estimates is crucial when these differ.

The difference between vector autoregression and local projection estimates arise because the two methods have different finite sample properties: they trade off bias and variance differently. It is, however, hard in practise to choose between vector autoregressions and local projections because it requires knowledge on how bias and variance influence the two methods in the specific analysis at hand. A key take-away from ample Monte Carlo studies is that no method is superior in general and supple changes in the data generating process alters the preferred estimator at different horizons (Meier, 2005; Marcellino, Stock and Watson, 2006; Kilian and Kim, 2011; Brugnolini, 2018; Li, Plagborg-Møller and Wolf, 2021).

We make two novel contributions. First, we show that local projection estimates can be decomposed into a term that is equal to the vector autoregression estimate and a term which we coin the local projection contribution. An important property of this decomposition is that both terms are identified and can be estimated by standard procedures. Since we identify both these terms, this enables us to make our second contribution: We provide a data-driven selection rule that determines when to choose vector autoregressions, local projections or a mix.

The decomposition of the local projection estimates arise from a two-step estimation procedure. In step 1 we estimate a vector autoregression. In step 2 we estimate the local projection contribution as the parameter estimate from a projection of the vector autoregression forecast error onto the same set of explanatory variables used in step 1. We then show that adding the estimates from step 1 and 2 exactly gives the local projection estimate. This shows that the local projection estimate can be decomposed into a vector autoregression estimate and our local projection contribution estimate. This decomposition, while simple, has a number of implication of interest to practitioners.

First off, it provides a link between local projections and vector autoregressions. In the way we formulate it local projection estimates can be thought of as a correction of the predictable forecast errors the vector autoregression makes at longer horizons. Second, while we focus on the local projection contribution, the two-step estimation procedure opens up for a much broader perspective on impulse response estimation: local projections are only one way we can think of correcting the vector autoregression estimates. As the local projection correction is linear we can think of it as a 'simple' correction but the opportunities to gain from exploring more sophisticated corrections are interesting to pursue. We focus on the the local projection corrections as a showcase and because it allows us to highlight a third implication of the two-step procedure: We can derive a data-driven selection rule for how to choose between vector autoregressions and local projections.

We derive the data-driven selection rule by first multiplying the local projection contribution with a parameter, α. We coin the impulse response estimate obtained from the sum of the vector autoregression and α times the local projection contribution the *local projection corrected vector autoregression* estimate. The local projection corrected vector autoregression estimate then nests three important cases: *i*) there is no local projection contribution (i.e. the vector autoregression estimate is obtained) when $\alpha = 0$, *ii*) the full local projection contribution is added (i.e. the local projection estimate is obtained) when $\alpha = 1$ and *iii*) a mix between the vector autoregression estimate and the local projection estimate is obtained when $\alpha \in (0, 1)$. We then show that minimizing the mean squared error of the local projection corrected vector autoregression estimate w.r.t. α depends on a key statistic: the coefficient of variation of the local projection contribution. The selection rule states: when the coefficient of variation of the local projection contribution is large, more weight should be put on the vector autoregression estimate. When the coefficient of variation is low, more weight should be put on the local projection estimate.

The performance of the selection rule is analyzed in the last part of the paper. We use several Monte Carlo studies intended to represent the main challenges practitioners face: the sample size of each dataset is finite and the true data generating process is unknown. The finite sample size implies that variance plays a crucial role. The unknown data generating process implies that also bias is important. The importance of bias and variance in the mean squared error of the different impulse response estimates differ across the Monte Carlo studies and over the forecast horizons considered in each study. Common to all Monte Carlo studies is that neither local projections nor vector autoregressions estimate impulse responses that are preferred over all forecast horizons. The Monte Carlo studies demonstrate one main result: the mean squared error of the local projection corrected vector autoregression estimate always mix the vector autoregression and local projection estimates such that when the mean squared error of one is large the other is picked. Thus the local projection corrected vector autoregression estimates are always the safest choice.

The contributions made in this paper are related to several strands of literature. The seminal paper by Jordà (2005) showed how to estimate local projections and *compared* them with vector autoregression impulse responses proposed in Sims (1980). We show that local projections can be estimated *from* vector autoregressions. Relative to Jordà (2005), we use this to cast new light on what local projections are and to bridge two seemingly different approaches that seek to estimate the same object. Plagborg-Møller and Wolf (2021) prove that local projections and vector autoregressions indeed do estimate the same impulse responses in population and provide a lot of insights on how, why and when these two estimators differ. Plagborg-Møller and Wolf (2021) briefly consider the difference between the two estimators in finite samples and the link between the two estimators that we derive bears resemblance to what they find. But whereas their focus is on the asymptotic properties of the difference between the two estimators, we use the link we derive to propose a selection rule that combats the bias-variance trade-off in finite samples.

The two-step estimation procedure we propose in this paper relates to similar procedures used in other professions than economics. Judd and Small (2000) proposed the use of a procedure to improve forecast errors of iterative models in physics and related papers are referenced in that paper. Taieb and Hyndman (2012) also propose the use of a similar procedure to generally improve forecasts of iterative models. Relative to both these papers we focus on impulse responses and discuss the link between VARs and LPs using this approach.

Our selection rule emerges from a minimization of the mean squared error of the impulse response estimates. As mentioned, it is meant to address the bias-variance trade-off in impulse response estimation with vector autoregressions and local projections. The same issue has been addressed by notably two other papers. Barnichon and Brownlees (2019) propose to shrink local projection estimates using regularization tools to minimize the variance loss. Relative to Barnichon and Brownlees (2019) we "shrink" the local projection estimate towards the vector autoregression estimate and our selection rule does not depend on tuning parameters set by the practitioner. Our selection rule is instead entirely data-driven. Miranda-Agrippino and Ricco (2021) similarly propose to shrink the local projection estimates but, as in our case, towards the vector autoregression estimates. Our paper contrasts with Miranda-Agrippino and Ricco (2021) in two important ways. First, we formally show that shrinking the local projection estimates towards the vector autoregression estimates is indeed sensible. Second, Miranda-Agrippino and Ricco (2021) propose a Bayesian approach where local projections are estimated with conjugate priors centred around vector autoregression estimates from a pre-sample. Their Bayesian approach is fundamentally different from our classical approach.

In our Monte Carlo studies we find support for the notion that the best impulse response estimates are found by using a mix of vector autoregressions and local projections - not choosing one over the other. This conclusion is reminiscent of the literature on forecast combinations (for a review see for example Timmermann, 2006). We show that the insights carry through to impulse response estimation.

The Monte Carlo studies we use to analyse the performance of vector autoregressions, local projections and local projection corrected vector autoregressions are related to a large literature. A vast amount of papers have looked at the performance of vector autoregressions versus local projections in Monte Carlo studies. An extensive but not exhausting list includes Jordà (2005), Meier (2005), Kilian and Kim (2011), Brugnolini (2018), Choi and Chudik (2019), Jordà, Singh and Taylor (2020), and Li, Plagborg-Møller and Wolf (2021). Relative to these papers we add new Monte Carlo studies and the analyses of the performance of our local projection corrected vector autoregression. Our Monte Carlo studies furthermore demonstrates the key take-away from this paper: We do not only cast light on specific scenarios where practitioners could expect one method to be better than the other, but we also give a tool that ensures the best performing method is followed.

2 Local Projection Corrected Vector Autoregressions

Consider the time series $\boldsymbol{y}_1, ..., \boldsymbol{y}_N$, where $\boldsymbol{y}_t$, $t \in \{1, \ldots, N\}$, is a $(1 \times k)$ matrix of variables. Let $\boldsymbol{P}_t \equiv [\boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \ldots, \boldsymbol{y}_{t-p}]$ be a $(1 \times kp)$ matrix of past observations.

Assume that the economy is hit by a shock at time t such that the variables move by $\boldsymbol{s}$ on impact of the shock. Let $\tilde{\boldsymbol{y}}_t$ denote the time t outcome after the shock such that $\tilde{\boldsymbol{y}}_t = \boldsymbol{y}_t + \boldsymbol{s}$. The shock is specified in this generic way to keep things simple. It may be thought of as any identified immediate response of the variables in $\boldsymbol{y}_t$ to a shock.[1]

The usual LP definition of the time $t + h$ impulse response to the shock is

$$\hat{IR}_{LP}(h, \boldsymbol{s}; \boldsymbol{P}_t) \equiv \boldsymbol{s}\hat{\boldsymbol{C}}^{LP}_{h,1}, \tag{1}$$

where $\boldsymbol{C}^{LP}_{h+1,1}$ is the upper $(k \times k)$ block of the coefficient matrix $\boldsymbol{C}^{LP}_{h+1}$ from the projection

$$\boldsymbol{y}_{t+h} = \boldsymbol{P}_t\boldsymbol{C}^{LP}_{h+1} + \boldsymbol{v}_t, \tag{2}$$

and $\boldsymbol{v}_t$ are projection errors. We use a hat to indicate that an object is estimated.

Equation (1) highlights the well-known nature of LP impulse responses: for each h-step ahead impulse response, the LP impulse responses are based on new coefficient matrices, $\boldsymbol{C}^{LP}_h$. In

[1] Think e.g. of a row of a Cholesky decomposition of the covariance matrix familiar from the identification from short-run restrictions.

contrast, the h-step ahead VAR impulse responses are given by

$$\hat{IR}_{VAR}(h, \boldsymbol{s}; \mathbf{P}_t) = \boldsymbol{s}\hat{\boldsymbol{B}}^h_{VAR,1}, \tag{3}$$

where $\boldsymbol{B}^h_{VAR,i}$ denotes the i'th upper $(k \times k)$ block of $\boldsymbol{B}^h_{VAR}$, with

$$\boldsymbol{B}_{VAR} \equiv \begin{bmatrix} \boldsymbol{B}_1 & \boldsymbol{I}_k & \mathbf{0}_k & \cdots & \mathbf{0}_k \\ \boldsymbol{B}_2 & \mathbf{0}_k & \boldsymbol{I}_k & \cdots & \mathbf{0}_k \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{B}_{p-1} & \mathbf{0}_k & \mathbf{0}_k & \cdots & \boldsymbol{I}_k \\ \boldsymbol{B}_p & \mathbf{0}_k & \mathbf{0}_k & \cdots & \mathbf{0}_k \end{bmatrix},$$

$\mathbf{0}_k$ and $\boldsymbol{I}_k$ denoting the $(k \times k)$ zero and identity matrices, respectively, $\boldsymbol{B}^0_{VAR,1} \equiv \boldsymbol{I}_k$ and $[\boldsymbol{B}_1, \ldots, \boldsymbol{B}_p] = \boldsymbol{C}^{LP}_1$ since the LP projection in Equation (2) coincides with the VAR projection for $h = 0$.

Equation (3) shows how the VAR, in contrast to LP, relies completely on $\boldsymbol{C}^{LP}_1$ to update the coefficient matrix $\boldsymbol{B}^h_{VAR,1}$.

Equations (1) and (3) show that VARs and LPs are different approaches that can be used to obtain the same object: the impulse response to the time t shock. The reuse of $\boldsymbol{C}^{LP}_1$ vis-á-vis the new $\boldsymbol{C}^{LP}_h$ is the fundamental difference between the two approaches and it is the explanation for why practitioners sometimes find that the two methods yield different results for their impulse response estimates.

The reuse of $\boldsymbol{C}^{LP}_1$ implies that the VAR does not require any new information to form h-step ahead predictions. This makes the VAR estimator highly efficient and this feature is typically praised by proponents of VARs. To visualize this efficiency, consult e.g. panel (b) of Figure 7 in the simulation studies of Section 4. The VAR justifies the reuse of $\boldsymbol{C}^{LP}_1$ under the assumption that the data generating process (DGP) follows Equation (2) for $h = 0$ and instead of making new projections for each new h, the VAR instead iterates forward on the equation to obtain the h-step ahead predictions. If, however, the DGP does not follow Equation (2), the VAR model is misspecified and the reuse of $\boldsymbol{C}^{LP}_1$ may result in biased predictions. To visualize such bias, consult e.g. panel (a) of Figure 7. The LP approach tries to alleviate on this bias cost by forming the new coefficient matrices $\boldsymbol{C}^{LP}_h$. Proponents of LP typically highlight the (perceived) smaller bias in justification of the method vis-á-vis VAR. Such improvement on bias is for example also illustrated in panel (a) of Figure 7. LP users, however, have to give up some of the efficiency of the VAR. Panel (b) of Figure 7 e.g. illustrates this efficiency loss. Ultimately, all practitioners face a classical statistical challenge when they have to decide between using VARs or LPs: how should bias and variance be traded off?

The bias-variance trade-off is inherently difficult to solve and practitioners need to have specific knowledge about how severe the bias and variance costs are in the specific analysis they are conducting. Practitioners, however, do not typically hold such valuable knowledge and furthermore, formal guidelines to aid them are scarce. As a result, practitioners are most often left to make their own, subjective judgements.

The reason why there are no formal guidelines available to practitioners to help judge whether VAR or LP impulse responses should be used, is because the two methods build on different approaches that are difficult to compare. In other words, the way the two methods work opens a gap between them and ultimately the impulse responses they yield.

The first theoretical contribution this paper makes is to bridge the gap between the two methods and show that VARs and LPs are linked through an additive decomposition of the LP. Proposition 4 shows this results.

Proposition 4 (Additive LP decomposition)**.** *The h-step ahead $LP(p)$ impulse response can be decomposed into the sum of an h-step ahead $VAR(p)$ impulse response and a local projection correction (LPC) to the VAR:*

$$\hat{IR}_{LP}(h, \boldsymbol{s}; \boldsymbol{P}_t) = \boldsymbol{s}\hat{\boldsymbol{B}}^{h}_{VAR,1} + \boldsymbol{s}\hat{\boldsymbol{C}}^{LPC}_{h,1}, \tag{4}$$

where $\boldsymbol{C}^{LPC}_{h+1,1}$ is the upper $(k \times k)$ block of the coefficient matrix $\boldsymbol{C}^{LPC}_{h+1}$ from the projection

$$\boldsymbol{r}_{t+h} = \boldsymbol{P}_t\boldsymbol{C}^{LPC}_{h+1} + \boldsymbol{u}_{ht}, \tag{5}$$

and $\boldsymbol{r}_{ht}$ is the h-step ahead prediction error of $\boldsymbol{y}_{t+h}$ of the $VAR(p)$ model.

Sketch proof (full proof in Section A). Fit a $VAR(p)$ model using the projection in Equation (2) with $h = 0$ and set $\left[\hat{\boldsymbol{B}}_1, \ldots, \hat{\boldsymbol{B}}_p\right] = \hat{\boldsymbol{C}}^{LP}_1$. Iterate forward on the $VAR(p)$ model to make h-step ahead *predictions* for $\boldsymbol{y}_{t+h}$. The h-step ahead in-sample prediction error the $VAR(p)$ model makes of $\boldsymbol{y}_{t+h}$ is given by $\boldsymbol{r}_{t+h} = \boldsymbol{y}_{t+h} - \hat{\boldsymbol{y}}^{VAR}_{t+h}$. Fit the local projection of the VAR prediction errors in Equation (5) to obtain the local projection corrected VAR (LPCVAR) predictions $\hat{\boldsymbol{y}}^{LPCVAR}_{t+h} = \hat{\boldsymbol{y}}^{VAR}_{t+h} + \boldsymbol{P}_t\hat{\boldsymbol{C}}^{LPC}_{h+1}$ and impulse responses $\hat{IR}_{LPCVAR}(h, \boldsymbol{s}; \boldsymbol{P}_t) = \boldsymbol{s}\hat{\boldsymbol{B}}^{h}_{VAR,1} + \boldsymbol{s}\hat{\boldsymbol{C}}^{LPC}_{h,1}$. To show that the h-step ahead LPCVAR impulse response is equal to the Jordà, 2005 local projections, write the LPCVAR on "projection form", i.e. projection matrix, $P_{\boldsymbol{P}_t}$, times object to be projected, $\boldsymbol{r}_{t+h}$, and use that projection matrices are idempotent and that a once projected object is unchanged by another projection with the same projection matrix. □

Proposition 4 provides an important link between the VAR and LP h-step ahead impulse responses as it shows that we can obtain LP *from* VAR. Moreover, Proposition 4 shows that LP can be thought of as a correction of the VAR model and in particular, LP corrects the prediction errors the VAR model makes. As we shall study in the next section, predictions may be efficient in using the data for the 1-step ahead projection to extrapolate into the future. However, exactly due to this efficient use of data where one prediction is used to form the next, it may introduce a bias cost. This bias cost can be reduced by correcting the prediction error that the VAR makes by a projection back onto the set of explanatory variables. The correction, however, may be poorly estimated (the variances of the coefficients in $\hat{\boldsymbol{C}}_h^{LPC}$ are large) and if this is the case, we might not be interested in the bias reduction after all. Due to the additive decomposition of the LP estimator in Proposition 4 and the fact that we estimate the LP correction, we can assess whether we want to correct the VAR estimate or not and in the coming section we show that we can also allow only for partial correction of the VAR estimate.

2.1 Bias vs. variance

To assess the performance of the LP and VAR estimators we use the mean squared error (henceforth MSE) as the metric. The MSE is a popular choice of metric which is typically used in the literature to compare different methods and an appealing property of this metric is that it can be decomposed into two terms: bias and variance. Our second theoretical contribution, outlined in Corollary 2, utilizes the insights from Proposition 4 to show how the MSE of the LP estimator compares to the VAR.[2]

Corollary 2 (Local Projection Mean Squared Error Decompostion)**.** *Let $\beta_{X,h}$ denote the h-step ahead response of any given variable k obtained by method $X \in \{VAR, LP, LPC\}$. The Mean Squared Error of the local projection impulse response estimate is given by*

$$\begin{aligned} MSE_{LP,h} =& MSE_{VAR,h} + \mathbb{V}[\beta_{LPC,h}] - (bias^2_{VAR,h} - bias^2_{LP,h}) \\ &+ 2\text{Cov}[\beta_{VAR,h}, \beta_{LPC,h}], \end{aligned} \tag{6}$$

where $\mathbb{V}$ is the variance operator and Cov *is the covariance operator.*

Proof. Using the decomposition of the LP impulse response estimates in Proposition 4, we have

[2] To ease notation we suppress the denotation of the response and shock variables such that $\beta_{X,h}$ denotes the h-step ahead response of any given variable k to the time t shock obtained by method $X \in \{VAR, LP, LPC\}$. Thus, $\beta_{VAR,h}$ refers to the k'th element of $\boldsymbol{s}\hat{\boldsymbol{B}}^h_{VAR,1}$, $\beta_{LP,h}$ refers to the k'th element of $\boldsymbol{s}\hat{\boldsymbol{C}}^{LP}_{h,1}$ and $\beta_{LPC,h}$ refers to the k'th element of $\boldsymbol{s}\hat{\boldsymbol{C}}^{LPC}_{h,1}$.

that the MSE is given by

$$MSE_{LP,h} = \mathbb{E}\left[(\beta_{LP,h} - \beta_{TRUE,h})^2\right]$$
$$= \mathbb{E}\left[(\beta_{VAR,h} + \beta_{LPC,h} - \beta_{TRUE,h})^2\right]. \tag{7}$$

Using the standard MSE bias-variance decomposition (provided for convenience in Section B.1), we can also write the h-step ahead LP MSE as

$$MSE_{LP,h} = \mathbb{V}[\beta_{VAR,h}] + \mathbb{V}[\beta_{LPC,h}] + \text{bias}^2_{LP,h} + 2\text{Cov}[\beta_{VAR,h}, \beta_{LPC,h}], \tag{8}$$

where we realize that the MSE of the LP impulse response consists of the usual contribution from the variance of our estimates and the squared bias *plus* an additional contribution from the covariance between the VAR and LPC estimates.

In the VAR case, we have that

$$MSE_{VAR,h} = \mathbb{V}[\beta_{VAR,h}] + \text{bias}^2_{VAR,h}, \tag{9}$$

from which it easily follows that we can write the MSE of the local projection impulse response estimate as in Equation (6) of Corollary 2. □

Equation (6) yields important insights and formalizes some of the typical claims made in the literature regarding the comparison of VARs and LPs. First, we see that the LPC adds variance to the MSE on top of that in the VAR model, something that is often emphasized by proponents of VARs (for example Kilian and Kim, 2011). Second, we see that *given* the LP has a smaller bias than the VAR (typically this is the case, and at least in the simulation studies shown below), this contributes to a reduction in the MSE of the LP relative to VARs. These benefits of accuracy from lower bias is typically highlighted by researchers who prefer LPs (for example Jordà, Singh and Taylor, 2020). Finally, the MSE of the LP model is affected by a covariance term between the VAR estimate and the LP contribution. This term is a novelty of our setup, in the sense that this is not discussed explicitly in the literature comparing VARs and LPs. Since the LPC is the linear mapping of past information to the forecast errors of the VAR model we should, *on average*, expect to see a smaller $\beta_{LPC,h}$ the closer the value of $\beta_{VAR,h}$ is to $\beta_{TRUE,h}$. On the other hand, the further away the estimate of the VAR is from the true value, the more need for a correction from the LPC there is. Thus, we expect the covariance term to be negative and hence lower the MSE of the LP relative to the VAR. In conclusion, Equation (6) formally states that whether the LP or the VAR is the preferred model (i.e. has the lowest MSE) depends on the magnitude of the additional variance introduced by the LP balanced off against the gains made from bias reduction. Naturally, this will depend on the particular dataset at hand.

Now, let us use the framework in the above to consider an alternative to choosing between *either* the VAR or the LP. We propose a simple approach which lets the impulse responses be given by what we call the local projection corrected vector autoregression (LPCVAR):

$$\beta_{LPCVAR,h} = \beta_{VAR,h} + \alpha_h \beta_{LPC,h}. \tag{10}$$

The LPCVAR attempts to balance off the forces that are at play in the VAR and LP estimators: We acknowledge that LP may be correcting the bias of the VAR but we want to dampen the correction if this bias correction is associated with a non-negligible statistical uncertainty. We use the parameter α_h to control the degree of correction that the LP is allowed to add to the VAR. LPCVAR coincides with the VAR if $\alpha_h = 0$. LPCVAR coincides with LP if $\alpha_h = 1$. Whenever $0 < \alpha_h < 1$ LPCVAR reflects a weighting of VAR and LP.

With Equation (10) at hand we are ready to present our last theoretical contribution, presented in Corollary 3, which shows the optimal weighting of VAR and LP and the terms which it depends on.

Corollary 3 (Optimal weighting of VAR and LP)**.** *The weighting of the h-step ahead VAR and LP impulse responses that yields the smallest mean squared error is given by*

$$\alpha_h^* = \frac{\mathbb{E}[\beta_{LPC,h}]^2 - bias_{LP,h}\mathbb{E}[\beta_{LPC,h}] - \text{Cov}(\beta_{VAR,h}, \beta_{LPC,h})}{\mathbb{V}[\beta_{LPC,h}] + \mathbb{E}[\beta_{LPC,h}]^2}. \tag{11}$$

Proof. Using a similar MSE decomposition as before, we get that[3]

$$\begin{aligned} MSE_{LPCVAR,h} =& \mathbb{V}(\beta_{VAR,h}) + \alpha_h^2 \mathbb{V}(\beta_{LPC,h}) + \text{bias}_{LP,h}^2 + (1-\alpha_h)^2 \mathbb{E}[\beta_{LPC,h}]^2 - \\ & 2\text{bias}_{LP,h}(1-\alpha_h)\mathbb{E}[\beta_{LPC,h}] + 2\alpha_h \text{Cov}(\beta_{VAR,h}, \beta_{LPC,h}). \end{aligned} \tag{12}$$

Since MSE_{LPCVAR} is convex in α_h, the value that minimizes it, $\alpha_h^* \equiv \arg\min_{\alpha_h} MSE_{LPCVAR,h}$, is given by Equation (11). ☐

Equation (12) tells us how the mean squared error is affected by allowing for some correction of the VAR estimate. Equation (11) in Corollary 3 tells us what the optimal correction is. To see clearly that the optimal weight relates to the precision of the LP contribution, we will neglect

[3] See Section B.2 for a detailed derivation of Equation (12).

the (potential) bias of the LP as well as the covariance term. In this case we can write the optimal weight as

$$\alpha_h^* = \frac{\mathbb{E}[\beta_{LPC,h}]^2}{\mathbb{V}[\beta_{LPC,h}] + \mathbb{E}[\beta_{LPC,h}]^2} = \frac{1}{1 + \text{CV}_{LPC,h}^2}, \tag{13}$$

where CV_{LPC_h} is the coefficient of variation of the LPC estimate and thus measures the precision of the LPC estimate.

Equation (13) tells us that when the LPC parameter is precisely estimated (α^* close to 1), the LP estimate is preferred. On the other hand, when the LPC estimate is imprecisely estimated (α^* close to 0), the VAR estimate is preferred. As can be seen from Equation (13), whenever the squared coefficient of variation on the LPC estimate is different from zero but finite, the optimal weight α^* will lie between 0 and 1 and we suspect that this will be the case in many empirical applications. If we think of LP (complete utilization of all information, as the LPC and VAR estimates are orthogonal to each other) and VAR (optimal variance reduction) as lying on opposite ends of the bias-variance spectrum, Equation (13) suggests that impulse response functions should instead be estimated by utilizing the possibility to trade off the opposing factors instead of choosing one estimate over the other a priori.

In the following, we will base our approach on Equation (13) rather than Equation (11), as our emphasis is to provide a *practical* selection tool for empirical analyses: While the bias of the LP and the covariance term are unknown to the practitioner, the coefficient of variation is easily calculated and features as standard output of many statistical software programs. Further, we find both terms to be of small magnitude in our simulations. In the next two sections, we compare this approach to both VARs and LPs using two different simulation studies and demonstrate that this approach works well.

3 Example I: A univariate moving average DGP

3.1 Monte Carlo study

The DGP we use in our first MC study follows a $MA(25)$ model, i.e.

$$y_t = \left(1 + \sum_{j=1}^{25} \theta_j B^j\right)\varepsilon_t, \tag{14}$$

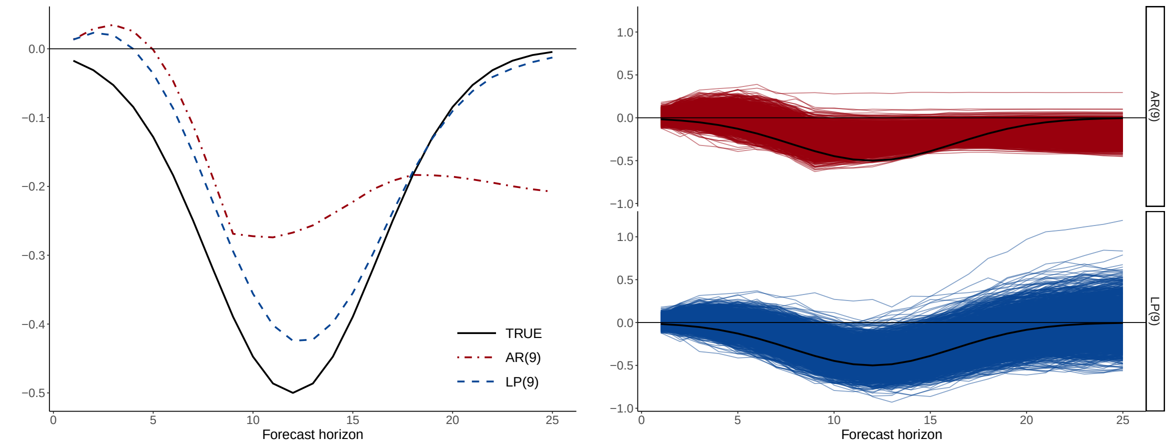
where the error terms are standard Gaussian, i.e. $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, and B is the backshift operator: $B^k x_t = x_{t-k}$. The cumulative impulse responses are given by the following set of coefficients, generated by the Gaussian basis function

$$\theta_{h,cum} = a \exp\left(-\frac{h-b}{c}\right) \text{ for } h = 1, \ldots 25, \tag{15}$$

with parameters $a = -0.5$, $b = 12$ and $c = 6$. The impulse responses from the DGP are depicted with a solid black line in panel (a) of Fig. 1.

We simulate 2,000 datasets with 500 observations for estimation (we use 500 observations as burn in). In Section 3.3 we look at datasets with 100, 250 and 750 observations to analyze the implications of smaller and larger datasets. We choose this range as the smallest datasets typically found in empirical macroeconomics have around 100 observations see e.g. Herbst and Johannsen, 2020 and the largest datasets have around 750 e.g. that used in Jordà, Singh and Taylor, 2020. We follow Jordà, Singh and Taylor, 2020 and estimate AR and LP models with respectively 2, 3, 9 and 12 lags on each dataset. We start out by looking at the $AR(9)$ and $LP(9)$ models to build intuition and keep things simple. In Section 3.2 we look at the other lag lengths.

Panel (a) of Fig. 1 shows the mean impulse responses for the $AR(9)$ and $LP(9)$ models over the 2,000 simulated datasets. The LP impulse responses are on average closer to the true impulse responses, resulting in a lower bias than the AR impulse responses. Panel (b) of Fig. 1 however shows that the lower bias comes at a higher variance cost and we see that for some datasets, the LP impulse responses are terribly off. Panel (a) and (b) of Fig. 1 together show the bias-variance trade-off the econometrician faces: the LP model is favorable in a pure bias sense but for a particular dataset the higher variance *can* result in very bad impulse responses.



(A) Mean impulse responses. (B) All 2,000 estimated impulse responses.

Figure 1. Estimated impulse responses from 2,000 datasets for LP and AR models with 9 lags. 500 observations.

How costly is the bias and variance in a mean squared error sense? In Figure 2 we decompose the MSE into its bias and variance terms. We clearly see what was alluded to in Figure 1: the AR model has lower variance and the LP has lower bias. In a MSE sense we also see that the LP model outperforms the AR model and that this happens due to the high bias cost. The right-most panel of Figure 2 shows the MSE error of the $LPCAR(9)$ model.

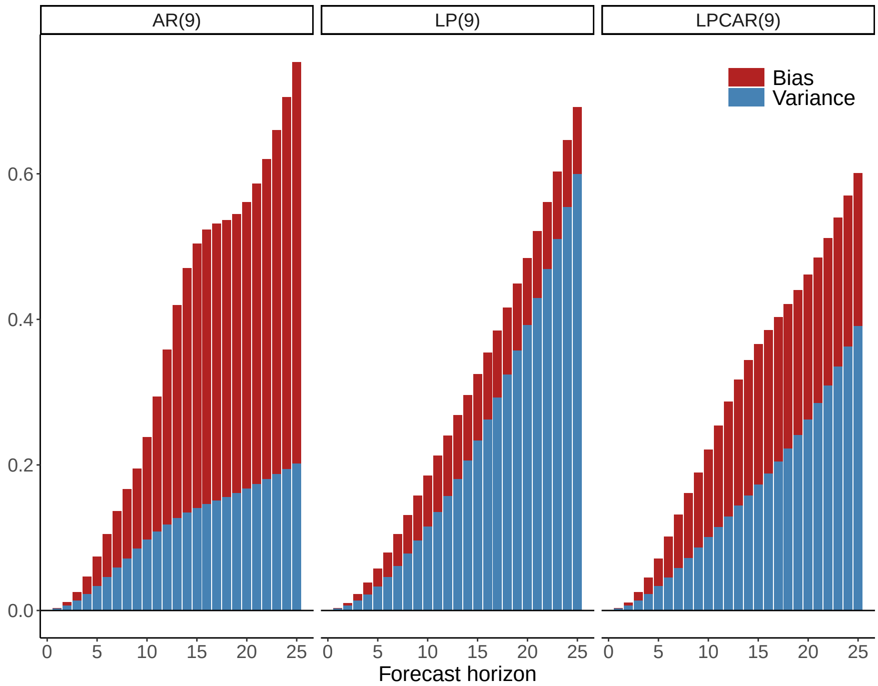


FIGURE 2. Bias-variance decomposition of cumulative mean squared error of AR, LP and LPCAR models with 9 lags. 500 observations.

For each estimated AR model we estimate the local projection corrected AR (LPCAR) impulse responses using the selection rule in Equation (13). From Figure 2 we see that the selection rule effectively deals with the bias-variance trade-off between the AR and LP model. In summary, the LPCAR model has a lower bias than the AR model and a lower variance than the LP model. On the cost side, the LPCAR model has a higher bias than the LP model and a higher variance than the LP model, but the overall MSE is lowest for the LPCAR model.

Are these results impressive? In the next two subsections we will show and discuss a particular appealing feature of the LPCAR model that we believe to ratify its worth: while the AR and LP models alternate between being the worst performing estimator the LPCAR model always follows the best of the two. We finish this section by looking at how the LPCAR model mixes between the AR and LP models.

To discuss the LPCAR method in action, we consider *how much* LP correction our selection rule imply for the AR models. Figure 3 depicts the mean α_h^* for each estimated model over the 2,000 simulated datasets. The results confirm the visual findings from Figure 1. First, the α_h^*s increase immediately at the point where $h > p$, i.e. where the forecast horizon exceeds the lag length of the AR models. This is the point where the mean impulse response of the AR model starts diverting from the true impulse responses (i.e. where the misspecification due to truncation bias is most severe). Second, the α_h^*s follow a wave shape which mimics the bias gap between the AR and LP models shown in panel (a) of Figure 1. Since we look at the mean α_h^*s this is natural. We also see that α_h^* tends to decrease at longer horizons, reflecting the fact that the variance in the LP starts weighing heavily. The shaded area in Figure 3 marks the are between the 10^{th} and 90^{th} α_h^* percentile.[4] The shaded areas first and foremost reveals that the selection rule is able to pick up the peculiarities of each dataset and almost not LP correct the VAR impulse response or almost correct it entirely. In conclusion, this indicates that the simple approach suggested in Section 2.1 works as intended, including the ability to add dataset- and horizon-specific information to the impulse responses when the data warrants it.

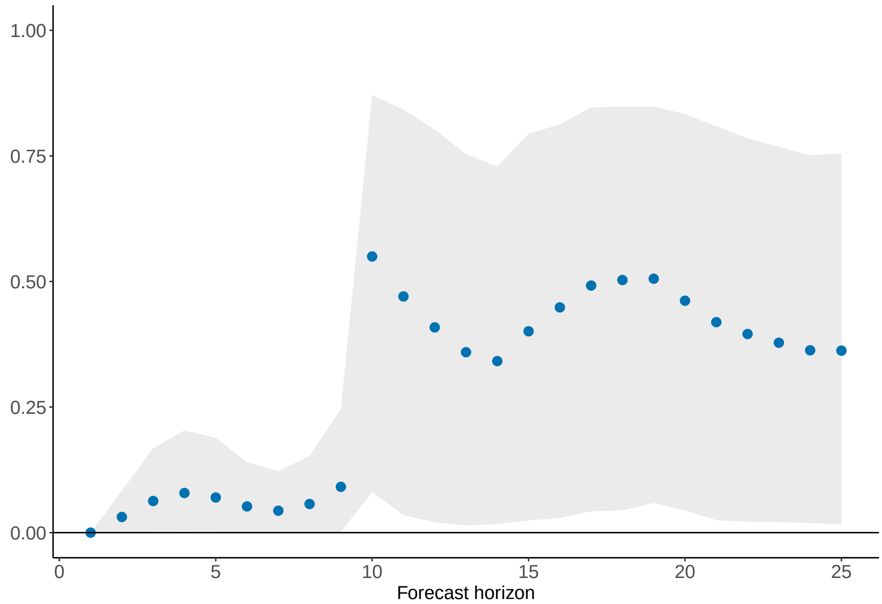


FIGURE 3. Mean α_h^* for the LPCAR model with 9 lags. 500 observations.

Notes. The shaded area marks the area between the 10^{th} and 90^{th} percentile of all α_h^*s across the 2,000 datasets. The minimum α_h^* is close to zero at all forecast horizons whereas the maximum hovers around 0.50 when $h \leq p$ and around 0.95 after.

3.2 Different lag lengths

What does the lag length imply for the performance of the estimators? In this section we compare the performance of the AR, LP and LPCAR models with respectively 2, 3, 9 and 12 lags. We keep the other details of the Monte Carlo study in the previous section fixed.

[4] The minimum α_h^* across all datasets hovers around zero at all forecast horizons whereas the maximum α_h^* hovers around 0.5 when the lag length is less than or equal to the forecast horizon and around 0.95 after.

Figure 4 shows the bias-variance decomposition of the MSE for all the models considered. We see that the AR models suffer a great bias cost relative to the LP models when the lag length is 2 and 3. The LPCAR models in this case follow the LP models and do not obtain the big MSE of the AR but instead gets a low MSE as the LP. Second, we see that as the lag length increases the AR model start outperforming the LP model. When the AR model starts outperforming the LP model, we see that the LPCAR model moves from lying close to the LP model to lying closer to the AR model (this is also clear from Figure 5). As we shall demonstrate throughout the following sections, the LPCVAR estimator consistently chooses the best performing estimator among VAR and LP and follows it up until the point where it no longer is better. We take this result as a strong indication of our selection rule working as intended. In panel (c) of Figure 5 we show the mean α_h^*s for the different LPCAR models. The figure shows that for low lag lengths (2 and 3) the LPCAR model corrects the AR model by 70 pct. at its peak whereas the maximum correction for lag length 12 is around 45 pct. This thus clarifies that for low lag lengths, the LPCAR model is closer to the LP model and for higher lag lengths the LPCAR model is closer to the AR model. We also see the same feature as discussed above: the α_h^*s jump at the point where $h > p$.

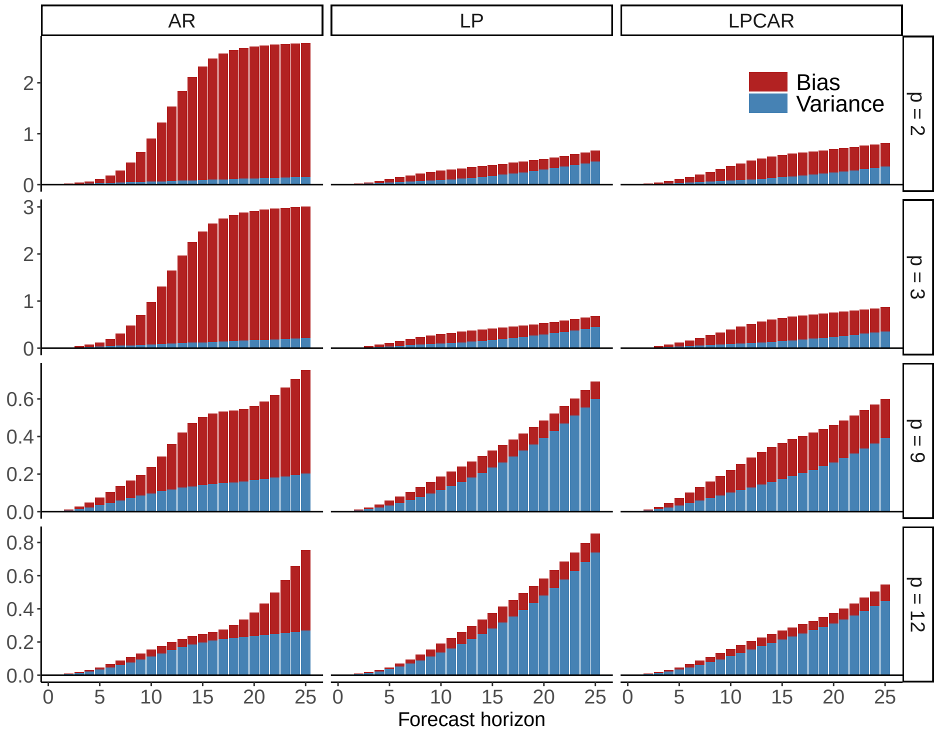


Figure 4. Bias-variance decomposition of the cumulative mean squared error for AR, LP and LPCAR models with 2, 3, 9 and 12 lags. 500 observations.

3.3 Different number of observations

How much does the number of observations matter? In this section we look at the performance of the estimators when there are 100, 250, 500 and 750 observations available for estimation. As we have seen above the bias of the AR models with few lags is very exaggerated and the variance cost of the LP model when 500 observations are available becomes negligible next to this. We also saw that the LPCAR models followed the LP model in this case. With fewer observations we should expect the variance cost to be greater and we shall now demonstrate that the LPCAR model will tend to the AR models instead.

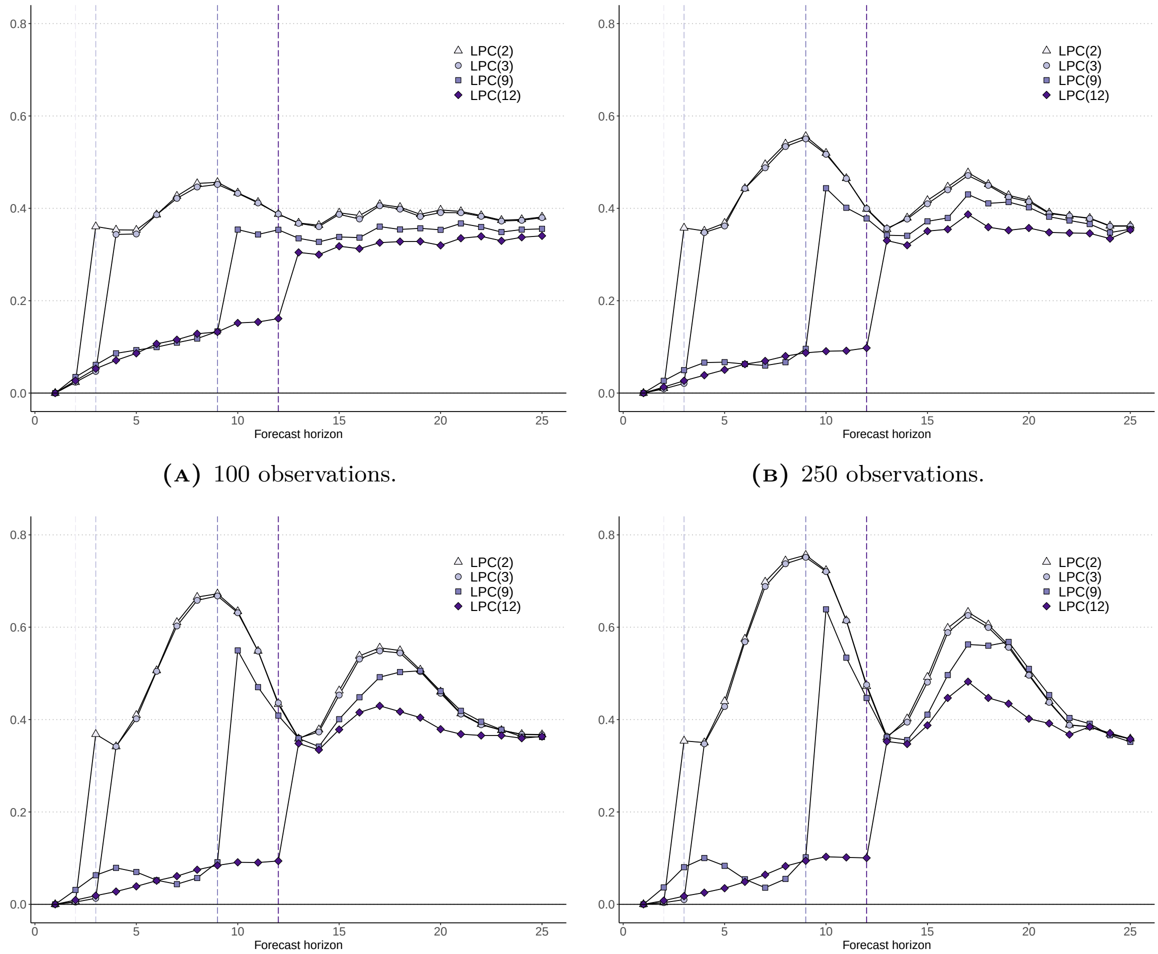
We start by looking at the α_h^*s. Panel (a) of Figure 5 shows that the maximum correction of the AR estimates when there are only 100 observations is 46 pct. for the $AR(2)$ and $AR(3)$ models. In panel (d) we see that with 750 observations the AR estimates are instead corrected with 75 pct. Another feature Figure 5 highlights is that when the forecast horizon exceeds the lag length, the LPCAR corrects the AR impulse responses by more or less the same factor and as the forecast horizon increases the correction factor seems to converge. This happens for two reasons. First off, because the bias of the AR impulse responses start kicking in when the forecast horizon exceeds the lag length. Secondly, because the variance cost of the AR does not increase excessively when the lag length increases. That the bias cost kicks in and the variance cost does not increase excessively holds across scenarios with different number of observations and is evident from both Figures 2 and 4.

In Figure 6 we look at the bias-variance decomposition of the cumulative mean squared error at forecast horizon 25. In the upper-left corner we see the performance of the AR, LP and LPCAR models estimated with 2 lags on datasets with 100 observations. We see that the bias cost of the AR model is high enough to make it it the worst performing estimator. We also see that the variance cost of the LP model is the highest among the three estimators. The LPCAR model mixes the bias and variance cost, as also shown previously, to make it the best performing estimator in the case with 2 and 3 lags and 100 observations.

Moving from left to right in Figure 6 we see that the bias of all models decrease and is replaced with a higher variance cost. This implies that the AR model starts performing better and the LP model performs worst. As also highlighted in Figure 5, the LPCAR model lies closer to the AR model in this case and hence disregards the LP due to its high variance.

Moving from top to bottom in Figure 6 we see that the variance of all models decrease. The low variance makes the LP perform very good and the LPCAR model therefore lies closer to the LP.

Finishing off at the lower-right corner we see that all models perform very well but we still see that the AR and LP trade off bias and variance very differently: the AR model has the highest



(A) 100 observations. **(B)** 250 observations.

(C) 500 observations. **(D)** 750 observations.

Figure 5. Mean α_h^* of AR, LP and LPCAR models with 2, 3, 9 and 12 lags in cases with 100, 250, 500 and 750 observations across 2,000 simulated datasets.

Notes. The dashed vertical lines indicate the lag lengths used in the different models and hence shows the point where the forecast horizon equals the lag length.

MSE due to its bias being high. The LP model has the second highest MSE due to its variance being high. The LPCAR model mixes bias and variance and performs best when the models are estimated with 12 lags on 750 observations.

What can we take away from these exercises? First we see that the LPCAR always outperforms either the LP or AR model. Second, when it is not the best performing estimator it still closely follows the estimator that performs best. In 9 out of the 16 cases considered in Figure 6, however, the LPCAR model is the best performing estimator. For practitioners this is a particular strong quality of the LPCAR model because it is always difficult to assess whether bias or variance plays the largest role in the specific analysis at hand. The LPCAR model with the simple

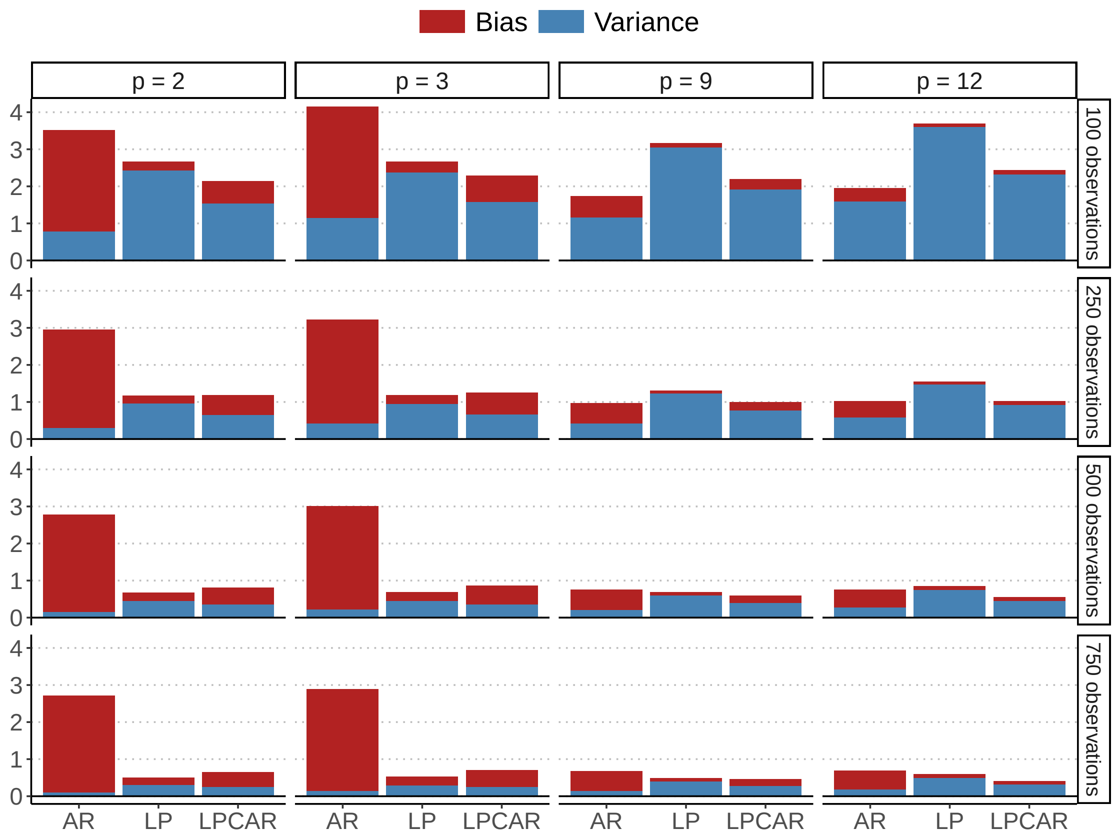


Figure 6. Bias-variance decomposition of the cumulative mean squared error at $H = 25$ of AR, LP and LPCAR models with 2, 3, 9 and 12 lags and 100, 250, 500 and 750 observations across 2,000 simulated datasets.

selection rule in Equation (13) will make sure to provide a data-driven approach that resolves the bias-variance trade-off in an easy and strong performing way.

4 Example II: Approximating the infinite-order VAR representation of a DSGE model

How does the selection rule perform in more complex environments? In this section we revisit Ravenna (2007) and simulate data from a DSGE model. Subsequently, we estimate impulse responses using vector autoregressions, local projections and local projection corrected vector autoregressions in a multivariate setting.

We consider the setup in this section for two reasons. First, we use the DSGE model as a (stylized) representation of an economy that practitioners could be interested in estimating impulse responses of. Second, the parametrization of structural models are increasingly being carried out by moment matching instead of full information maximum likelihood (see for example

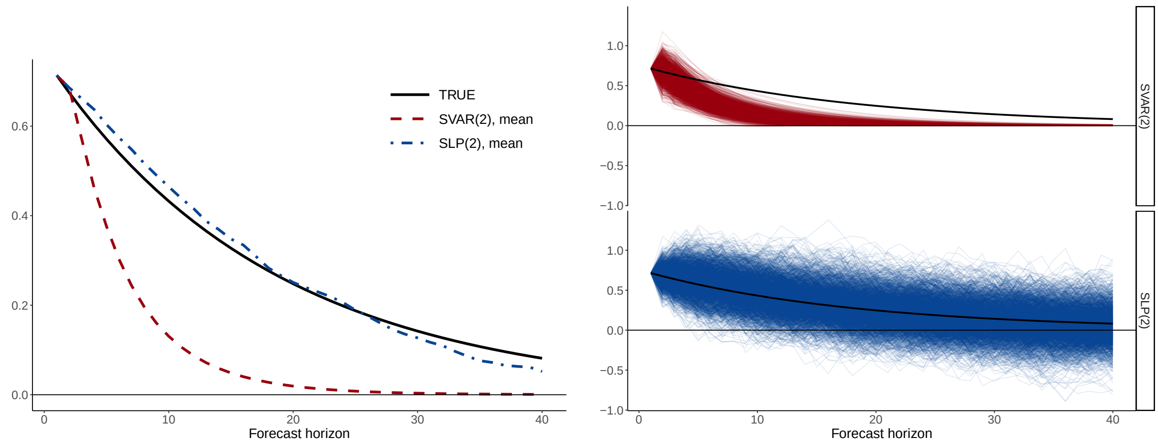
Christiano, Eichenbaum and Trabandt, 2016). Impulse responses of central variables to the shocks of interest naturally suggest themselves as relevant moments in this regard, thus raising the question of which impulse response estimator to use. Third, the setup lies at the heart of the discussion on whether vector autoregressions or local projections should be used to estimate impulse responses. In a nutshell, the use of vector autoregressions implies an assumption that the DGP follows a VAR process. This assumptions is typically justified when working with DSGE models since Structural VAR models will often contain enough information to retrieve one or several (if not all) shocks from the underlying DSGE model (Forni and Gambetti, 2014). However, as has been known at least since Ravenna (2007) and Fernández-Villaverde, Rubio-Ramírez, Sargent and Watson (2007), when a subset of the endogenous state variables in DSGE models is unobservable to the econometrician, the true representation of the DSGE may be instead be a VARMA model. Given that the invertibility conditions are satisfied, this in turn implies that the DSGE model can be represented by an infinite-order VAR model - and that a finite-order VAR model may or may not be a good representation of the true DGP (i.e. depending on the extent of the resulting truncation bias). Jordà (2005) raises the same point, referring to studies by Cooley and Dwyer (1998) who also highlight the VARMA representation of RBC models, and argues that when using finite-order VAR models to estimate impulse responses, these impulse response estimates will suffer from misspecification biases.

4.1 Monte Carlo study

We estimate impulse responses based on the RBC model with indivisible labor developed in Hansen, 1985. The model characterizes an economy where hours worked, the gross real interest rate, consumption and output are endogenous control variables, capital stock is an endogenous state variable and technology and labor supply shocks are exogenous control variables. We assume that an econometrician lives in an economy characterized by the DSGE model and that she wants to estimate the response of hours worked to technology shocks.

The true impulse response of hours worked to the technology shock is shown in panel (a) of Figure 7. The entire economy has a finite-order VAR representation, which makes the choice of using a VAR to estimate the impulse response adequate. The econometrician, however, only has data on hours worked and output. This is sufficient to investigate the link that she is interested in but as Ravenna, 2007 points out, when the capital stock is unobserved by the econometrician, the DGP instead has an infinite-order VAR representation. With the limitation of a finite sample of 400 observations, the econometrician however has to estimate a finite-order VAR. As in Ravenna, 2007 we assume that the econometrician uses a $VAR(2)$ model. We furthermore assume that the contemporaneous responses of hours worked and output to any shock is correctly identified. The econometrician thus investigates the structural impulse responses. We add capital "S" to any model name abbreviation whenever needed for clarity.

Using the $VAR(2)$ model leads to a considerable bias in the impulse response estimates, as shown in panel (a) of Figure 7. The figure clearly shows that the mean estimated impulse responses of the $VAR(2)$ models across the 2,000 simulated datasets do not accurately capture the true impulse responses of the model except for at the shortest horizons (in panel (a) of figure Figure 8 we compute and show the bias over all horizons of the impulse responses).



(A) Mean impulse responses. (B) All impulse responses.

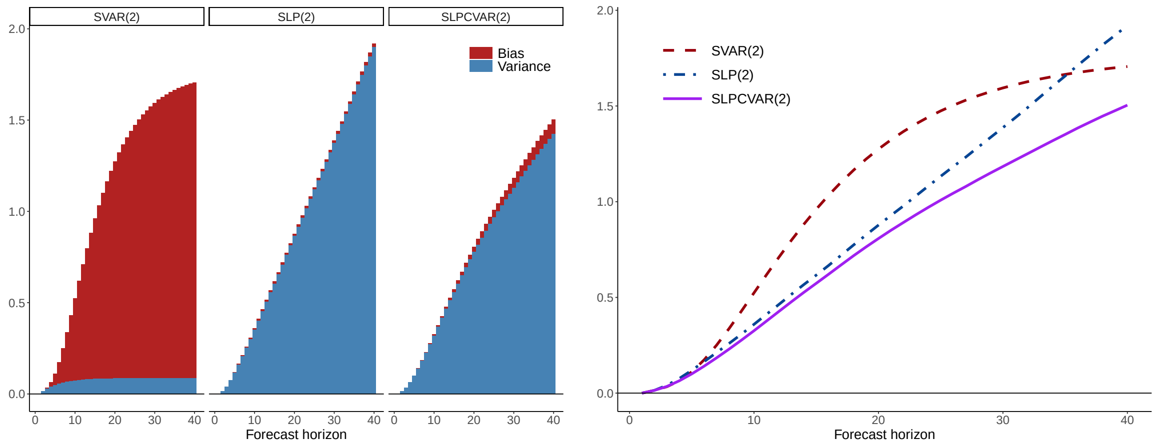
Figure 7. $VAR(2)$ and $LP(2)$ impulse responses of hours to a technology shock in the Hansen, 1985 model. 400 observations. 2,000 simulations.

Notes. The solid black line is the true impulse response function. Capital "S" in front of the model names is short for "Structural" and denotes that the econometrician correctly identifies the contemporaneous responses of hours worked and output and thus estimates structural impulse responses.

On the contrary, the mean estimate of the $LP(2)$ model does quite well across all horizons. Does this mean that the $LP(2)$ is to be preferred over the $VAR(2)$ model in this specific application? Not necessarily: While the mean estimates across many simulations are close to the true impulse responses, what also becomes clear, however, is that a higher variance cost has to be payed at the dispense of the smaller bias. This is not visible from the plot of the mean impulse responses in panel (a) of Figure 7, but in panel (b) we depict all the estimated impulse responses. From panel (b) of Figure 7 it is clear that the $LP(2)$ model has a much larger variance than the $VAR(2)$ model. Combining the insights from both panels in Figure 7 we get an idea of the bias-variance trade-off when using VARs and LPs.

Next, we assess the bias and variance cost in a MSE sense. Panel (a) of Figure 8 shows the size of the bias and variance terms across different horizons. From the figure it is clear that the $LP(2)$ model does much better than the $VAR(2)$ in terms of bias.

Panel (b) of Figure 8 shows the cumulative MSEs. The figure shows that in the short and medium run the bias in the VAR is costlier than the larger variance in the LP. In the longer run, however, the bias of the VAR reduces and so does the MSE. The cumulative MSE therefore



(A) Bias and variance decomposition.

(B) Cumulative MSEs.

Figure 8. Mean squared error and bias-variance decomposition of estimated hours response to a technology shock. 400 observations. 2,000 simulations.

Notes. Capital "S" in front of the model names is short for "Structural" and denotes that the econometrician correctly identifies the contemporaneous responses of hours worked and output and thus estimates structural impulse responses.

settles in the long run. In the long run the VAR is outperforming the LP in a MSE sense. The shift in how the two estimators perform relative to each other tells that one could gain from a mixed estimate where, the LP is picked in the short and medium run but as the LP variance starts weighing heavier than the VAR bias, the mixed estimate should shift from the $LP(2)$ to the $VAR(2)$ model. This is exactly what the $LPCVAR(2)$ model does.

Panel (b) of Figure 8 shows that the $LPCVAR(2)$ model is able to balance off the bias-variance trade-off better than both the $VAR(2)$ and $LP(2)$ models. In Figure 9 we show the mean α_h^* across the 2,000 simulations. First we once again see that for $h \leq p$, the $LPCVAR(2)$ model picks the $VAR(2)$ and from Figure 8 it is clear why: neither the $VAR(2)$ nor the $LP(2)$ model suffer from bias and their variance losses are almost indistinguishable. Hence, the econometrician will not gain from LP correcting the VAR impulse responses. When $h = p + 1$, α_h^* jumps up because the bias of the VAR starts kicking in and the LP becomes a better alternative. We then see that the α^*s increase to the point where the bias of the VAR is at its peak. At this point the LPCVAR corrects the VAR estimate with almost 70 pct. of the LP estimate. Lastly, as the bias of VAR starts decreasing but the variance of the LP prevails, the α_h^*s starts decreasing and the LPCVAR starts reverting back to the VAR estimates. In Panel (b) of Figure 8 we see that the mixing of the VAR and LP results in a cumulative MSE that is better than the two extremes in both the short, medium and long run.

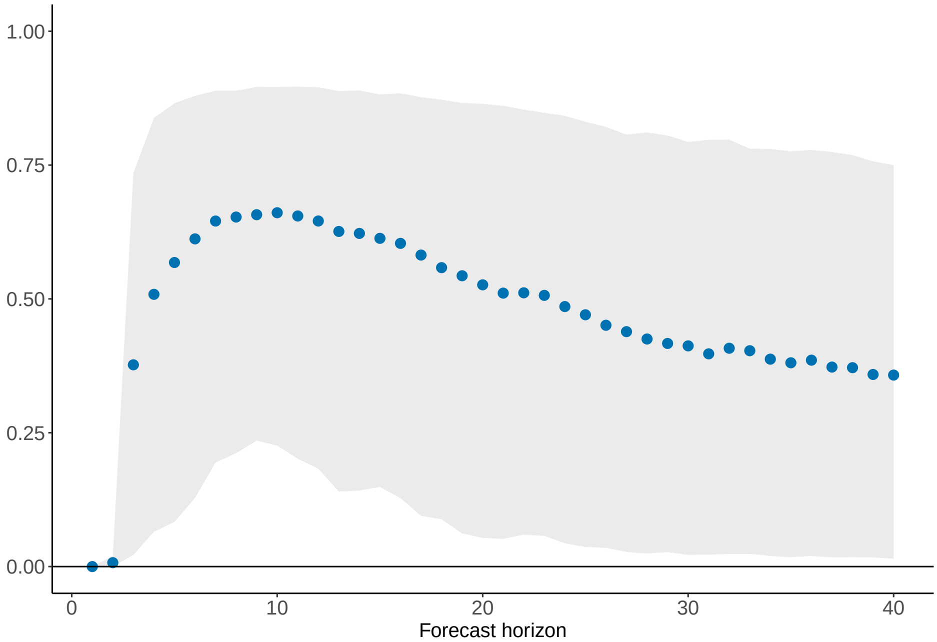


Figure 9. Mean α_h^*s. 400 observations. 2,000 simulations.

Notes. The shaded area marks the area between the 10th and 90th percentile of all α_h^*s across the 2,000 datasets. The minimum α_h^* is close to zero at all forecast horizons whereas the maximum hovers around 0.15 when $h \leq p$ and around 0.95 after.

5 Conclusion

We provide a solution to an inherent problem researchers face when estimating impulse responses in practise: local projections and vector autoregressions trade off bias and variance differently and this implies that the two methods often times yield different impulse response estimates. The problem is to answer "What impulse response do we rely on?" We introduce a data-driven selection rule that navigates the bias-variance trade-off and yields an appropriate mix of the local projection and vector autoregression impulse responses.

One of the main benefits of using our suggested selection rule is that it allows the researcher to choose a mix of LPs and VARs when estimating impulse responses instead of choosing one estimator a priori. Since the weight put on the LP contribution is based on the squared coefficient of variation, the optimal weight will typically lie strictly between 0 and 1 - and we suspect that this will be the case in many empirical applications as well.

An important, and perhaps surprising, detail is that mixing the local projection and vector autoregression impulse responses is not an ad hoc solution. We show instead that this follows naturally from a new insight provided in the paper: local projection impulse responses can be derived *from* VARs and the local projection impulse response may be viewed as a correction to the VARs'. This insight reduces the question of which impulse response to rely on to "How much correction of the VAR should we allow for?"

Our solution to the problem is to introduce a simple and easily implementable data-driven selection rule that exactly answers how much one should local projection correct the VAR impulse response. The performance of the selection rule that we introduce relies on two terms not being too large: the bias of the local projection estimator and the covariance between the VAR impulse response and the correction of the LP. Proponents of LPs typically highlight the low bias of the estimator and in the Monte Carlo studies conducted in this paper we do not find it to be large either. We furthermore only find a negligible size of the covariance term and in general we find that the selection rule works well across all Monte Carlo studies.

As the LP Contribution is orthogonal to the VAR estimates of the IRF, one can think of LPs (emphasis on using all available information) and VARs (emphasis on variance reduction) as lying on opposite ends of a bias-variance spectrum. The message of this paper is thus that impulse response functions should be estimated by optimally navigating the trade-off on this bias-variance spectrum instead of using only one type of estimator and our selection rule is a simple and easily implementable tool that researchers can take up to do so.

References

Barnichon, R. and Brownlees, C. (2019). "Impulse response estimation by smooth local projections." *The Review of Economics and Statistics*, vol. 101(3), pp. 522–530 (cited on page 125).

Brugnolini, L. (2018). "About Local Projection Impulse Response Function Reliability." *CEIS Research Paper*, vol. 16 (cited on pages 123, 126).

Choi, C.-Y. and Chudik, A. (2019). "Estimating impulse response functions when the shock series is observed." *Economics Letters*, vol. 180, pp. 71–75 (cited on page 126).

Christiano, L. J., Eichenbaum, M. S. and Trabandt, M. (2016). "Unemployment and business cycles." *Econometrica*, vol. 84(4), pp. 1523–1569 (cited on page 139).

Cooley, T. F. and Dwyer, M. (1998). "Business cycle analysis without much theory A look at structural VARs." *Journal of Econometrics*, vol. 83, pp. 57–88 (cited on page 140).

Fernández-Villaverde, J., Rubio-Ramírez, J. F., Sargent, T. J. and Watson, M. W. (2007). "ABCs (and Ds) of understanding VARs." *American Economic Review*, vol. 97(3), pp. 1021–1026 (cited on page 140).

Forni, M. and Gambetti, L. (2014). "Sufficient information in structural VARs." *Journal of Monetary Economics*, vol. 66, pp. 124–136 (cited on page 140).

Hansen, G. D. (1985). "Indivisible Labor and the Business Cycle." *Journal of Monetary Economics*, vol. 16, pp. 19 (cited on pages 140, 141).

Herbst, E. P. and Johannsen, B. K. (2020). "Bias in Local Projections." *Finance and Economics Discussion Series*, vol. 2020 (cited on page 133).

Jordà, O. (2005). "Estimation and Inference of Impulse Responses by Local Projections." *American Economic Review*, vol. 95, pp. 161–182 (cited on pages 125, 126, 128, 140).

Jordà, O., Singh, S. R. and Taylor, A. M. (2020). "The Long-Run Effects of Monetary Policy." (Cited on pages 126, 130, 133).

Judd, K. and Small, M. (2000). "Towards long-term prediction." *Physica D: Nonlinear Phenomena*, vol. 136, pp. 31–44 (cited on page 125).

Kilian, L. and Kim, Y. J. (2011). "How Reliable Are Local Projection Estimators of Impulse Responses?" *Review of Economics and Statistics*, vol. 93, pp. 1460–1466 (cited on pages 123, 126, 130).

Li, D., Plagborg-Møller, M. and Wolf, C. K. (2021). "Local Projections vs. VARs: Lessons From Thousands of DGPs." (Cited on pages 123, 126).

Marcellino, M., Stock, J. H. and Watson, M. W. (2006). "A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series." *Journal of Econometrics*, vol. 135, pp. 499–526 (cited on page 123).

Meier, A. (2005). *How Big is the Bias in Estimated Impulse Responses? A Horse Race between VAR and Local Projection Methods.* (Cited on pages 123, 126).

Miranda-Agrippino, S. and Ricco, G. (2021). "The transmission of monetary policy shocks." *American Economic Journal: Macroeconomics*, vol. 13(3), pp. 74–107 (cited on page 125).

Plagborg-Møller, M. and Wolf, C. K. (2021). "Local projections and VARs estimate the same impulse responses." *Econometrica*, vol. 89(2), pp. 955–980 (cited on pages 123, 125).

Ramey, V. A. (2016). "Chapter 2 - Macroeconomic Shocks and Their Propagation." *Handbook of macroeconomics* (pp. 71–162). Elsevier. (Cited on page 123).

Ravenna, F. (2007). "Vector autoregressions and reduced form representations of DSGE models." *Journal of Monetary Economics*, vol. 54, pp. 2048–2064 (cited on pages 139, 140).

Sims, C. A. (1980). "Macroeconomics and Reality." *Econometrica*, vol. 48, pp. 1 (cited on page 125).

Taieb, S. B. and Hyndman, R. J. (2012). *Recursive and direct multi-step forecasting: The best of both worlds* (Working Paper). (Cited on page 125).

Timmermann, A. (2006). "Chapter 4 - Forecast Combinations." *Handbook of Economic Forecasting* (pp. 135–196). Elsevier. (Cited on page 125).

A Proof of Proposition 4

Proof. Write the $VAR(p)$ projection as

$$\boldsymbol{y}_t = \boldsymbol{P}_t\boldsymbol{B} + \boldsymbol{v}_t, \tag{A.1}$$

where $\boldsymbol{B} \equiv [\boldsymbol{B}_1, \boldsymbol{B}_2, \dots, \boldsymbol{B}_p]'$ is a $(kp \times k)$ dimensional coefficient matrix and $\boldsymbol{v}_t$ is a vector of residuals.

By stacking the last $n = N - p$ observations in the matrix $\boldsymbol{Y}$ such that it is a $(n \times k)$ dimensional matrix with variables in the columns and time observations in rows and stacking the corresponding n observations of $\boldsymbol{P}_t$ in the matrix $\boldsymbol{P}$ such that it is a $(n \times kp)$ dimensional matrix, we have that

$$\boldsymbol{Y} = \boldsymbol{P}\boldsymbol{B} + \boldsymbol{V}, \tag{A.2}$$

where $\boldsymbol{V}$ is a $(n \times k)$ dimensional matrix of residuals.

The fitted VAR model can be obtained from the projected values of $\boldsymbol{Y}$ onto $\boldsymbol{P}$, given by

$$\text{proj}_{\boldsymbol{P}}\boldsymbol{Y} = \boldsymbol{P}(\boldsymbol{P}^\top\boldsymbol{P})^{-1}\boldsymbol{P}^\top\boldsymbol{Y} = P_{\boldsymbol{P}}\boldsymbol{Y} = \boldsymbol{P}\hat{\boldsymbol{B}} \equiv \hat{\boldsymbol{Y}}^{VAR}. \tag{A.3}$$

Note that $P_{\boldsymbol{P}} \equiv \boldsymbol{P}(\boldsymbol{P}^\top\boldsymbol{P})^{-1}\boldsymbol{P}^\top$ is the VAR projection matrix, $\hat{\boldsymbol{B}} \equiv (\boldsymbol{P}^\top\boldsymbol{P})^{-1}\boldsymbol{P}^\top\boldsymbol{Y}$ is the projection coefficient matrix (which coincides with the OLS estimates) and $\boldsymbol{P}$ is the projector.[5]

A.1 *h*-step ahead predictions and impulse-responses

The VAR model assumes that the relationship in Equation (A.1) holds and the h-step ahead predicted values are

$$\hat{\boldsymbol{y}}_{t+h}^{VAR} = \boldsymbol{P}_{t+h}\hat{\boldsymbol{B}}, \qquad \boldsymbol{P}_{t+h} = \begin{cases} \left[\hat{\boldsymbol{y}}_t^{VAR}, \boldsymbol{y}_{t-1}, \dots, \boldsymbol{y}_{t-p+1}\right] & \text{if } h = 1, \\ \left[\hat{\boldsymbol{y}}_{t+h-1}^{VAR}, \dots, \hat{\boldsymbol{y}}_t^{VAR}, \boldsymbol{y}_{t-1}, \dots, \boldsymbol{y}_{t-p+h}\right] & \text{if } 1 < h < p, \\ \left[\hat{\boldsymbol{y}}_{t+h-1}^{VAR}, \dots, \hat{\boldsymbol{y}}_{t+h-p}^{VAR}\right] & \text{if } h \geq p. \end{cases} \tag{A.4}$$

As in the main text, assume that the economy is hit by a shock at time t such that the variables move by $\boldsymbol{s}$ on impact of the shock. Let $\tilde{\boldsymbol{y}}_t$ denote the time t response of the shock such that

[5] When we think of $\hat{\boldsymbol{Y}}^{VAR}$ as the predicted values of $\boldsymbol{Y}$ from a regression, $\boldsymbol{P}$ can be thought of as the predictor. Since $\hat{\boldsymbol{B}} = (\boldsymbol{P}^\top\boldsymbol{P})^{-1}\boldsymbol{P}^\top\boldsymbol{Y}$ is equal to the OLS estimate of $\boldsymbol{B}$, the last equality, $\hat{\boldsymbol{Y}}^{VAR} = \boldsymbol{P}\hat{\boldsymbol{B}}$, defines $\boldsymbol{P}$ as a predictor in a regression sense.

$\tilde{\boldsymbol{y}}_t = \boldsymbol{y}_t + \boldsymbol{s}$. Define the time $t+h$ response of the economy to the time t shock as

$$IR(h, \boldsymbol{s}) \equiv \tilde{\boldsymbol{y}}_{t+h} - \boldsymbol{y}_{t+h}, \tag{A.5}$$

from which the usual $VAR(p)$ impulse-response follows:

$$I\hat{R}_{VAR}(h, \mathbf{s}; \mathbf{P}_t) = \boldsymbol{s}\hat{\boldsymbol{B}}^h_{VAR,1}. \tag{A.6}$$

where $\boldsymbol{B}^h_{VAR,i}$ is as in the main text.

Since the VAR does not update the coefficient matrix, the VAR IRF relies completely on the fitted model in Equation (A.3). If, however, the VAR model in Equation (A.1) is a misspecification of the true DGP, there is a reason to believe that the h-step ahead predicted values will be off and due to the re-use of fitted values in next-period predictions (see Equation (A.4) again) far-ahead predictions will be even worse off.[6] Nothing, however, prevents us from evaluating whether the VAR h-step predictions are off.

A.2 Local projection corrected vector autoregressions

Suppose we want to evaluate the first H predicted values of the VAR model and correct them if there is a systematic in-sample error in the prediction values.

Define the two $H \times k$ matrices $\mathbf{Z}_t \equiv (\hat{\mathbf{y}}^{VAR}_{t+1}, \ldots, \hat{\mathbf{y}}^{VAR}_{t+H})^\top$ and $\mathbf{F}_t \equiv (\mathbf{y}_{t+1}, ..., \mathbf{y}_{t+H})^\top$ containing the H predicted and realized future values of $\mathbf{y}_t$, respectively. Further, we define the matrix $\mathbf{R}_t \equiv \mathbf{F}_t - \mathbf{Z}_t$ and let $\mathbf{r}_{ht}$ denote the h'th row of $\mathbf{R}_t$.

Consider the linear projection of the VAR h-step ahead prediction errors onto the subspace $\boldsymbol{P}_t$

$$\mathbf{r}_{ht} = \mathbf{P}_t\mathbf{C}^{LPC}_{t+h} + \mathbf{u}_{ht}. \tag{A.7}$$

Make the same stacking of observations as before and introduce the following notation to keep track of things: $\boldsymbol{Y}_h$ denotes the h-step ahead values of $\boldsymbol{Y}$, $\boldsymbol{P}_{h-1}$ denotes the h-step ahead predictor, $\boldsymbol{C}_h$ denotes the h-step ahead projection coefficients.

The h-step ahead projected forecast errors the VAR model will make is given by

$$\hat{\boldsymbol{\mathcal{R}}}_h = \boldsymbol{P}_0(\boldsymbol{P}_0^\top\boldsymbol{P}_0)^{-1}\boldsymbol{P}_0^\top\boldsymbol{\mathcal{R}}_h = P_{\boldsymbol{P}_0}\boldsymbol{\mathcal{R}}_h = \boldsymbol{P}_0\boldsymbol{C}^{LPC}_{h+1} \tag{A.8}$$

[6] Notice how the lag length plays an important role in how many predicted vs. actual values that are being used for h-step predictions.

The second last equality tells us that the LPC uses the same projection matrix as in the VAR projection in Equation (A.3) but projects the VAR forecast errors, $\boldsymbol{\mathcal{R}}_h$, rather than $\boldsymbol{Y}_1$. The last equality tells us that the LPC, equivalently, uses the same predictor, $\boldsymbol{P}_0$, as in the VAR proejction but a) uses it to predict the forecast errors of the VAR model with the projection coefficients in $\boldsymbol{C}_{h+1}^{LPC}$ and b) maintains the predictor at all horizons rather than changing it as the VAR does.

Notice that the VAR h-step ahead predictions can be written on a form akin to a projection by letting $\boldsymbol{P}_{h-1}$ be a $(n \times kp)$ matrix with row values as in Equation (A.4):

$$\hat{\mathbf{Y}}_h^{VAR} = \boldsymbol{P}_{h-1}\hat{\boldsymbol{B}}, \tag{A.9}$$

and write the h-step ahead VAR prediction in Equation (A.9) as

$$\hat{\mathbf{Y}}_h^{VAR} = \boldsymbol{P}_{h-1}(\boldsymbol{P}_0^\top \boldsymbol{P}_0)^{-1}\boldsymbol{P}_0^\top \mathbf{Y}_1 \equiv P_{\boldsymbol{P}_h}^{VAR}\mathbf{Y}_1, \tag{A.10}$$

where $P_{\boldsymbol{P}_h}^{VAR} \equiv \boldsymbol{P}_{h-1}(\boldsymbol{P}_0^\top \boldsymbol{P}_0)^{-1}\boldsymbol{P}_0^\top$ denotes the "pseudo" VAR projection matrix. We use this for brevity but note that P_h^{VAR} is not a projection matrix.[7]

Use this and the second equality in Equation (A.8) to write the h-step ahead local projection corrected VAR prediction as

$$\begin{aligned}\hat{\mathbf{Y}}_h^{LPCVAR} =& P_{\boldsymbol{P}_h}^{VAR}\mathbf{Y}_1 + P_{\boldsymbol{P}_0}\boldsymbol{\mathcal{R}}_h \\ =& P_{\boldsymbol{P}_h}^{VAR}\mathbf{Y}_1 + P_{\boldsymbol{P}_0}(\mathbf{Y}_h - \hat{\mathbf{Y}}_h^{VAR}) \\ =& P_{\boldsymbol{P}_h}^{VAR}\mathbf{Y}_1 + P_{\boldsymbol{P}_0}\mathbf{Y}_h - P_{\boldsymbol{P}_0}P_{\boldsymbol{P}_h}^{VAR}\mathbf{Y}_1.\end{aligned} \tag{A.11}$$

We will now use that $P_{\boldsymbol{P}_0}$ is a projection matrix and therefore is idempotent and furthermore that a once projected matrix is equal to itself if projected once more with the same projection matrix:

$$P_{\boldsymbol{P}_0}\hat{\mathbf{Y}}_h^{LPCVAR} = P_{\boldsymbol{P}_0}P_{\boldsymbol{P}_0}\mathbf{Y}_h = P_{\boldsymbol{P}_0}\mathbf{Y}_h = \hat{\mathbf{Y}}_h^{LPCVAR}. \tag{A.12}$$

It now follows from Equation (A.11), the idempotent property of $P_{\boldsymbol{P}_0}$ and Equation (A.12) that the h-step ahead local projection corrected VAR prediction is equal to the h-step ahead LP

[7] $P_{\boldsymbol{P}_h}^{VAR}$ is not idempotent.

projection:

$$\begin{aligned}
\hat{\mathbf{Y}}_h^{LPCVAR} =& P_{\boldsymbol{P}_0}\hat{\mathbf{Y}}_h^{LPCVAR} && \text{(by (A.12))}\\
=& P_{\boldsymbol{P}_0}P_{\boldsymbol{P}_h}^{VAR}\mathbf{Y}_1 + P_{\boldsymbol{P}_0}P_{\boldsymbol{P}_0}\mathbf{Y}_h - P_{\boldsymbol{P}_0}P_{\boldsymbol{P}_0}P_{\boldsymbol{P}_h}^{VAR}\mathbf{Y}_1. \\
=& P_{\boldsymbol{P}_0}P_{\boldsymbol{P}_h}^{VAR}\mathbf{Y}_1 + P_{\boldsymbol{P}_0}\mathbf{Y}_h - P_{\boldsymbol{P}_0}P_{\boldsymbol{P}_h}^{VAR}\mathbf{Y}_1 && \text{(by idempotency)}\\
=& P_{\boldsymbol{P}_0}\mathbf{Y}_h \\
=& \hat{\mathbf{Y}}_h^{LP}. && \text{(A.13)}
\end{aligned}$$

To derive the h-step ahead local projection corrected VAR impulse-responses, write Equation (A.11) on "prediction" form instead:

$$\hat{\mathbf{Y}}_h^{LPCVAR} = P_{h-1}\hat{\boldsymbol{B}} + \boldsymbol{P}_0\hat{\boldsymbol{C}}_h^{LP} \tag{A.14}$$

It then also follows that the h-step ahead LP impulse response is equal to the h-step ahead Local Projection Corrected VAR impulse response:

$$I\hat{R}_{LPC}(h,\boldsymbol{s};\boldsymbol{P}) = \boldsymbol{s}\hat{\boldsymbol{B}}_{VAR,1}^h + \boldsymbol{s}\hat{\boldsymbol{C}}_h^{LPC} = \boldsymbol{s}\hat{\boldsymbol{C}}_h^{LP} = I\hat{R}_{LP}(h,\boldsymbol{s};\boldsymbol{P}) \tag{A.15}$$

□

B Mean squared error decompositions

B.1 LP MSE decomposition

This appendix section derives the decomposition of the MSE of the h-step ahead LP impulse response. The MSE of the h-step ahead impulse response is given by Equation (7) in the main text and re-iterated here:

$$\begin{aligned}
MSE_{LP,h} =& \mathbb{E}\left[(\beta_{LP,h} - \beta_{TRUE,h})^2\right]\\
=& \mathbb{E}\left[(\beta_{VAR,h} + \beta_{LPC,h} - \beta_{TRUE,h})^2\right]. && \text{(Equation (7) in main text)}
\end{aligned}$$

For ease of notation, suppress the h subscipt.

Add and substract $\mathbb{E}[\beta_{VAR}]$ and $\mathbb{E}[\beta_{LPC}]$ in Equation (7) to get

$$MSE_{LP} = \mathbb{E}\left[(\beta_{VAR} - \mathbb{E}[\beta_{VAR}] + \mathbb{E}[\beta_{VAR}] + \beta_{LPC} - \mathbb{E}[\beta_{LPC}] + \mathbb{E}[\beta_{LPC}] - \beta_{TRUE})^2\right].$$

Define the following variables

$$A \equiv \beta_{VAR} - \mathbb{E}[\beta_{VAR}], \;\; B \equiv \beta_{LPC} - \mathbb{E}[\beta_{LPC}], \;\; C \equiv \mathbb{E}[\beta_{VAR}] + \mathbb{E}[\beta_{LPC}] - \beta_{TRUE},$$

and notice that

$$MSE_{LP} = \mathbb{E}\left[(A + B + C)^2\right] = \quad \mathbb{E}[A^2] + \mathbb{E}[B^2] + \mathbb{E}[C^2] + 2\mathbb{E}[AB] + 2\mathbb{E}[AC] + 2\mathbb{E}[BC].$$

Looking at each term separately, we have that

$$\begin{aligned}
\mathbb{E}[A^2] =& \mathbb{E}[(\beta_{VAR} - \mathbb{E}[\beta_{VAR}])^2] = \mathbb{V}(\beta_{VAR}),\\
\mathbb{E}[B^2] =& \mathbb{E}[(\beta_{LPC} - \mathbb{E}[\beta_{LPC}])^2] = \mathbb{V}(\beta_{LPC}),\\
\mathbb{E}[C^2] =& \mathbb{E}[(\mathbb{E}[\beta_{VAR}] + \mathbb{E}[\beta_{LPC}] - \beta_{TRUE})^2] = \text{bias}^2_{LP},\\
\mathbb{E}[AB] =& \mathbb{E}[(\beta_{VAR} - \mathbb{E}[\beta_{VAR}])(\beta_{LPC} - \mathbb{E}[\beta_{LPC}])] = \text{Cov}(\beta_{VAR}, \beta_{LPC}),\\
\mathbb{E}[AC] =& \mathbb{E}[(\beta_{VAR} - \mathbb{E}[\beta_{VAR}])(\mathbb{E}[\beta_{VAR}] + \mathbb{E}[\beta_{LPC}] - \beta_{TRUE})]\\
=& \mathbb{E}[\beta_{VAR}\mathbb{E}[\beta_{VAR}]] + \mathbb{E}[\beta_{VAR}\mathbb{E}[\beta_{LPC}]] - \mathbb{E}[\beta_{VAR}\beta_{TRUE}] -\\
& \mathbb{E}[\mathbb{E}[\beta_{VAR}]\mathbb{E}[\beta_{VAR}]] - \mathbb{E}[\mathbb{E}[\beta_{VAR}]\mathbb{E}[\beta_{LPC}]] + \mathbb{E}[\mathbb{E}[\beta_{VAR}]\beta_{TRUE}] = 0,\\
\mathbb{E}[BC] =& \mathbb{E}[(\beta_{LPC} - \mathbb{E}[\beta_{LPC}])(\mathbb{E}[\beta_{VAR}] + \mathbb{E}[\beta_{LPC}] - \beta_{TRUE})]\\
=& \mathbb{E}[\beta_{LPC}\mathbb{E}[\beta_{VAR}]] + \mathbb{E}[\beta_{LPC}\mathbb{E}[\beta_{LPC}]] - \mathbb{E}[\beta_{LPC}\beta_{TRUE}] -\\
& \mathbb{E}[\mathbb{E}[\beta_{LPC}]\mathbb{E}[\beta_{VAR}]] - \mathbb{E}[\mathbb{E}[\beta_{LPC}]\mathbb{E}[\beta_{LPC}]] + \mathbb{E}[\mathbb{E}[\beta_{LPC}]\beta_{TRUE}] = 0,
\end{aligned}$$

and the standard MSE decomposition of the LP impulse response estimate given in Equation (8) in the main text follows directly:

$$MSE_{LP} = \mathbb{V}[\beta_{VAR}] + \mathbb{V}[\beta_{LPC}] + \text{bias}^2_{LP} + 2\text{Cov}[\beta_{VAR}, \beta_{LPC}].$$

B.2 LPCVAR MSE decomposition

The MSE of the h-step ahead LPCVAR impulse response is given by

$$\begin{aligned} MSE_{LPCVAR,h} =&\mathbb{E}\left[(\beta_{LPC,h} - \beta_{TRUE,h})^2\right] \\ =&\mathbb{E}\left[(\beta_{VAR,h} + \alpha_h\beta_{LPC,h} - \beta_{TRUE,h})^2\right]. \end{aligned} \tag{B.1}$$

For ease of notation, suppress the h subscript.

Add and substract $\mathbb{E}[\beta_{VAR}]$ and $\alpha\mathbb{E}[\beta_{LPC}]$ to get

$$MSE_{LPCVAR} = \mathbb{E}\left[(\beta_{VAR} - \mathbb{E}[\beta_{VAR}] + \mathbb{E}[\beta_{VAR}] + \alpha\beta_{LPC} - \alpha\mathbb{E}[\beta_{LPC}] + \alpha\mathbb{E}[\beta_{LPC}] - \beta_{TRUE})^2\right]$$

Define the following variables

$$A \equiv \beta_{VAR} - \mathbb{E}[\beta_{VAR}], \;\; B \equiv \alpha(\beta_{LPC} - \mathbb{E}[\beta_{LPC}]), \;\; C \equiv \mathbb{E}[\beta_{VAR}] + \alpha\mathbb{E}[\beta_{LPC}] - \beta_{TRUE},$$

and notice that

$$MSE_{LPCVAR} =\mathbb{E}\left[(A+B+C)^2\right] = \quad \mathbb{E}[A^2] + \mathbb{E}[B^2] + \mathbb{E}[C^2] + 2\mathbb{E}[AB] + 2\mathbb{E}[AC] + 2\mathbb{E}[BC].$$

Looking at each term separately, we have that

$$\begin{aligned} \mathbb{E}[A^2] =&\mathbb{E}[(\beta_{VAR} - \mathbb{E}[\beta_{VAR}])^2] = \mathbb{V}(\beta_{VAR}), \\ \mathbb{E}[B^2] =&\alpha^2\mathbb{E}[(\beta_{LPC} - \mathbb{E}[\beta_{LPC}])^2] = \alpha^2\mathbb{V}(\beta_{LPC}), \\ \mathbb{E}[C^2] =&\mathbb{E}[(\mathbb{E}[\beta_{VAR}] + \alpha\mathbb{E}[\beta_{LPC}] - \beta_{TRUE})^2] = \text{bias}^2_{LPCVAR}, \\ \mathbb{E}[AB] =&\alpha\mathbb{E}[(\beta_{VAR} - \mathbb{E}[\beta_{VAR}])(\beta_{LPC} - \mathbb{E}[\beta_{LPC}])] = \alpha\text{Cov}(\beta_{VAR}, \beta_{LPC}), \\ \mathbb{E}[AC] =&\mathbb{E}[(\beta_{VAR} - \mathbb{E}[\beta_{VAR}])(\mathbb{E}[\beta_{VAR}] + \alpha\mathbb{E}[\beta_{LPC}] - \beta_{TRUE})] \\ =&\mathbb{E}[\beta_{VAR}\mathbb{E}[\beta_{VAR}]] + \mathbb{E}[\beta_{VAR}\alpha\mathbb{E}[\beta_{LPC}]] - \mathbb{E}[\beta_{VAR}\beta_{TRUE}] - \\ &\mathbb{E}[\mathbb{E}[\beta_{VAR}]\mathbb{E}[\beta_{VAR}]] - \mathbb{E}[\mathbb{E}[\beta_{VAR}]\alpha\mathbb{E}[\beta_{LPC}]] + \mathbb{E}[\mathbb{E}[\beta_{VAR}]\beta_{TRUE}] = 0, \\ \mathbb{E}[BC] =&\mathbb{E}[\alpha(\beta_{LPC} - \mathbb{E}[\beta_{LPC}])(\mathbb{E}[\beta_{VAR}] + \alpha\mathbb{E}[\beta_{LPC}] - \beta_{TRUE})] \\ =&\mathbb{E}[\alpha\beta_{LPC}\mathbb{E}[\beta_{VAR}]] + \mathbb{E}[\alpha^2\beta_{LPC}\mathbb{E}[\beta_{LPC}]] - \mathbb{E}[\alpha\beta_{LPC}\beta_{TRUE}] - \\ &\mathbb{E}[\alpha\mathbb{E}[\beta_{LPC}]\mathbb{E}[\beta_{VAR}]] - \mathbb{E}[\alpha^2\mathbb{E}[\beta_{LPC}]\mathbb{E}[\beta_{LPC}]] + \mathbb{E}[\alpha\mathbb{E}[\beta_{LPC}]\beta_{TRUE}] = 0 \end{aligned}$$

It then follows that

$$MSE_{LPCVAR} = \mathbb{V}(\beta_{VAR}) + \alpha^2 \mathbb{V}(\beta_{LPC}) + \text{bias}^2_{LPCVAR} + 2\alpha \text{Cov}(\beta_{VAR}, \beta_{LPC}). \tag{B.2}$$

Finally realise that

$$\begin{aligned} \text{bias}_{LPCVAR} =& \mathbb{E}[\beta_{VAR}] + \alpha \mathbb{E}[\beta_{LPC}] - \beta_{TRUE} \\ =& \mathbb{E}[\beta_{VAR}] + \mathbb{E}[\beta_{LPC}] - \beta_{TRUE} - (1-\alpha)\mathbb{E}[\beta_{LPC}] \\ =& \text{bias}_{LP} - (1-\alpha)\mathbb{E}[\beta_{LPC}], \end{aligned}$$

and substitute this into Equation (B.2) to get the expression for the decomposition of the MSE of the LPCVAR impulse response estimate in Equation (12) in the main text:

$$\begin{aligned} MSE_{LPCVAR} =& \mathbb{V}(\beta_{VAR}) + \alpha^2 \mathbb{V}(\beta_{LPC}) + \text{bias}^2_{LP} + (1-\alpha)^2 \mathbb{E}[\beta_{LPC}]^2 - \\ & 2\text{bias}_{LP}(1-\alpha)\mathbb{E}[\beta_{LPC}] + 2\alpha \text{Cov}(\beta_{VAR}, \beta_{LPC}). \end{aligned}$$