UNIVERSITY OF COPENHAGEN
FACULTY OF SOCIAL SCIENCES
DEPARTMENT OF ECONOMICS

Essays in Macroeconomics:

Consumption Behavior, Price Dynamics, and Fiscal Spending

PhD Thesis

Rasmus Bisgaard Larsen

Supervisors: Emiliano Santoro and Søren Hove Ravn

Submitted on: January 31, 2020

Contents

Acknowledgements

As a then-mathemaphobic music conservatory dropout enrolling in an economics programme without much forethought, my road to writing this thesis has been long and winding. Along the way, I have shaken the mathemaphobia and learned that economics is not just an offshoot of sociology. I have also been fortunate to meet many inspiring people along the way to whom I owe a great deal of thanks.

First and foremost, I thank my supervisors, Emiliano Santoro and Søren Hove Ravn. Their expert guidance and unwavering support has been invaluable throughout the process of writing this thesis. I also owe many thanks to Søren for encouraging me to pursue a PhD degree, while I was writing my master's thesis under his supervision.

Two of the chapters in this thesis are co-authored. One chapter is co-authored with my office mate Christoffer Jessen Weissert. Working with Christoffer has been a great experience. I am grateful for our long discussions about research but also for him putting up with my idiosyncratic tastes in writing style and typesetting. Another chapter is co-authored with Emiliano and Søren. They invited me to work together on a project, which has become the third chapter of this dissertation. Doing research with them has taught me a lot.

I wish to thank the Department of Economics at the University of Copenhagen for funding my studies and providing a stimulating research environment. I also thank my former bosses at Kraka, Peter Mogensen and Jens Hauch. After handing in my master's thesis, I called Jens and explained to him my dire need of half a year of employment until vacant PhD scholarships were announced. He did not hesitate to offer me a job.

I owe special thanks to Morten O. Ravn for inviting me to visit University College London during the Fall of 2018. I am grateful for his warm hospitality and willingness to discuss my research with me.

I thank my fellow PhD students. Lunching, going out for beers and even running around Copenhagen with you has made the past three years infinitely more enjoyable. A special thank you goes out to the ragtag bunch of experts within the fields of labor, trade, migration, development, cultural economics, time series econometrics, environmental economics and macroeconomics that make up the PhD student body of building 26. Our (too long?) lunch and coffee breaks have not only been fun. They have also spawned thoughts that made it into this dissertation.

Lastly, I wish to thank my family and friends for unconditional love and support. No man is an island entire of itself.

Rasmus Bisgaard Larsen
Copenhagen, January 2020

Dansk introduktion

Denne ph.d.-afhandling består af tre selvstændige kapitler inden for makroøkonomi. De to første kapitler benytter sig af et stort scannerdatasæt fra USA. Datasættet indeholder detaljeret information om priser og salg fra detailforretninger såvel som omfattende information om et husholdningspanels varekøb. Ved hjælp af disse data undersøger det første kapitel, hvordan husholdningers indkomst påvirker kvaliteten af deres varekøb. Derudover analyseres det, hvordan dette påvirker deres opsparings- og forbrugsadfærd. Det andet kapitel undersøger, hvorledes regionale ændringer i offentligt forbrug påvirker de lokale detailpriser. Det tredje kapitel viser, at en ændring i det offentlige forbrug får boligpriserne til at bevæge sig i samme retning. Dette rationaliseres i en dynamisk stokastisk generel ligevægtsmodel med præferencer for variation og stigende skalaafkast.

Kapitel 1 – Quality and Consumption Basket Heterogeneity: Transitory Shocks and Implications for Consumption-Savings Behavior *med Christoffer Jessen Weissert*

Det første kapitel undersøger betydningen af forskelle i varekvalitet på tværs af husholdninger. Vi starter med at dokumentere, at husholdninger med højere indkomst forbruger varer af bedre kvalitet end fattigere husholdninger. Derudover viser vi, at en husholdning ikke kun øger sit forbrug men også kvaliteten af dens varekøb, når den modtager en midlertidig indkomstoverførsel fra staten. En husholdnings reaktion afhænger dog af dens indkomst. Mens tilbøjeligheden til at forbruge en andel af indkomstoverførslen afhænger negativt af husholdningens årlige indkomst, så følger sammenhængen mellem kvalitetsjusteringen og den årlige indkomst en omvendt U-form. Lavindkomsthusholdninger øger kvaliteten af deres varekøb en smule. Mellemindkomsthusholdninger øger kvaliteten af deres varekøb mest. Højindkomsthusholdninger justerer ikke kvaliteten af deres varekøb.

Disse resultater understøtter teorien om ikke-homotetiske præferencer: når husholdningers indkomst stiger, så vil de justere deres forbrug hen mod varer af højere kvalitet. Vi inkorporerer dette i en livscyklusmodel for husholdningsadfærd, hvori husholdninger oplever idiosynkratiske udsving i deres indkomst, som sammen med lånebegrænsninger tilskynder dem til at opbygge en formue til forsikring mod dårlige tider med lav indkomst (en såkaldt buffer-formue). Udover at vælge hvor mange penge, der skal bruges på forbrug og opsparing, så vælger husholdningerne også kvaliteten af varerne i deres forbrugsbundt. Efter at have kalibreret modellen med plausible fluktuationer i indkomst, viser vi, at den kan matche den negative sammenhæng mellem husholdningers årlige indkomst og deres tilbøjelighed til at forbruge en midlertidig stigning i indkomst med det samme i stedet for at opspare pengene. I en model med samme indkomstfluktuationer, men med homotetiske præferencer, er denne sammenhæng omvendt. Endeligt viser vi, at vores ikke-homotetiske model forudsiger den samme omvendte U-formede sammenhæng mellem en husholdnings årlige indkomst og dens justering af kvalitet som vi også fandt i data.

Kapitel 2 – Government spending and retail prices: Regional evidence from the United States Det andet kapitel viser, at detailforretninger øger deres priser i geografiske områder, hvor det offentlige forbrug stiger. En central udfordring ved at dokumentere dette er, at den offentlige sektor har en tendens til at dirigere sine udgifter hen mod områder, som oplever en ringe økonomisk udvikling. Det gør det svært at isolere den kausale effekt af offentligt forbrug. Er det offentligt forbrug i sig selv, som er årsag til den observerede prisudvikling eller blot de ringe økonomiske forhold?

Jeg identificerer den kausale effekt ved hjælp af to kilder til eksogen variation i offentligt forbrug. Først benytter jeg en stor amerikansk stimuluspakke, American Recovery and Reinvestment Act of 2009, som skabte variation i stimulustildelinger på tværs af stater. Nogle af bestemmelserne i stimuluspakken fordelte stimulus ud fra kriterier, som ikke var relaterede til den faktiske økonomiske udvikling i staterne. Som den anden kilde til eksogen variation udnytter jeg, at områder i USA er vedholdende og forskelligt eksponerede over for ændringer i de nationale militærudgifter. Når militærudgifterne stiger nationalt, vil områder der historisk er blevet tildelt mange militærkontrakter også modtage mange militærkontrakter nu uanset deres nuværende økonomiske forhold.

Prisændringer kan skyldes to ting. Enten ændres forretningernes marginale omkostninger eller også justerer de deres prismargin. Jeg argumenterer for, at den sidste årsag dominerer. For det første er den regionale variation i engrospriser af en utilstrækkelig størrelse til at kunne generere ændringer i marginale omkostninger. For det andet kontrollerer jeg for lønændringer og finder, at de ikke driver prisændringerne. Deraf konkluderer jeg, at detailforretningerne øger deres prismargin, når det lokale offentlige forbrug stiger. Dette er konsistent med teoretiske modeller, hvori forbrugernes prisfølsomhed påvirkes af deres indkomst eller beskæftigelse.

Kapitel 3 – House Prices, Increasing Returns, and Government Spending Shocks *med Søren Hove Ravn and Emiliano Santoro*

Boligprisers betydning for makroøkonomien har været diskuteret betydeligt i både den offentlige og akademiske debat efter den store recession. Ligeledes har finanspolitikkens rolle som et stabiliseringsværktøj. Det tredje kapitel forholder sig til begge af disse debatter ved at analysere, hvorledes boligpriser påvirkes af ændringer i det offentlige forbrug.

Vi estimerer en strukturel vektorautoregression og viser, at boligpriserne stiger, når der er et positivt stød til det offentlige forbrug. Estimaterne viser også, at reallønnen og totalfaktorproduktiviteten stiger.

Estimaterne er ikke konsistente med forudsigelserne fra en lang række af dynamisk stokastiske generelle ligevægtsmodeller. Tværtimod genererer disse modeller en negativ samvariation mellem boligpriser og det offentlige forbrug. Det skyldes, at en stigning i det offentlige forbrug følges af en stigning i skatter, som reducerer husholdningernes indkomst – den såkaldte negative formueeffekt. Det får husholdningerne til at reducere deres forbrug og arbejdsudbud, hvilket også sænker reallønnen. Ligeledes får formueeffekten boligefterspørgslen til at falde, hvilket sænker boligpriserne.

Vi konstruerer en dynamisk stokastisk generel ligevægtsmodel, som vender disse resultater på hovedet ved at overvinde den negative formueeffekt gennem stigende skalaafkast på et aggregeret niveau. Modellen har to lag i produktionskæden: en mellemproduktionssektor og et endeligt produktionslag. Mellemproduktionssektoren består af et variabelt antal virksomheder, der ind- eller udtræder af økonomien, når profitmuligheder opstår eller forvitrer.

Varerne produceret af denne sektor købes og kombineres af virksomhederne i det endelig produktionslag til et aggregeret forbrugsgode. Det endelige produktionslag har også præferencer for variation i mellemproduktionsvarer, hvilket sammen med ændringer i antallet af virksomheder i mellemproduktionssektoren skaber endogene variationer i totalfaktorproduktiviteten. Som et resultat deraf vil en stigning i efterspørgslen efter det endelige forbrugsgode skabt af et øget offentligt forbrug skabe nye profitmuligheder i mellemproduktionssektoren. Profitmulighederne skaber nye virksomheder, hvilket øger totalfaktorproduktiviteten og deraf reallønnen. Såfremt denne kanel er tilstrækkelig stærk, så vil den dominere den negative formueeffekt. Dette skaber en stigning i boligpriserne.

English introduction

This PhD dissertation consists of three self-contained chapters in the field of macroeconomics. The first two chapters rely a large scanner data set from the United States. The data set contains detailed pricing and sales information from retail stores as well as comprehensive information on purchases made by a panel of households. Using these data, the first chapter studies the effect of households' income on the quality of their purchases and the implications for their consumption-savings behavior. The second chapter investigates how regional changes in government spending affect local retail prices. The third chapter documents that a change in government spending causes house prices to move in the same direction, which is rationalized in a dynamic stochastic general equilibrium (DSGE) model with love of variety and increasing returns to scale.

Chapter 1 – Quality and Consumption Basket Heterogeneity: Transitory Shocks and Implications for Consumption-Savings Behavior *with Christoffer Jessen Weissert*

The first chapter investigates the importance of differences in the quality of consumption baskets across households. We begin by documenting that households with higher income consume products of better quality than poorer households. Moreover, we show that when a household receives a temporary transfer of money from the government, it will not only increase spending but also the quality of products purchased. This response, however, depends on the annual income of the household. While the propensity to spend out of the transfer is decreasing in the annual income of the household, the relationship between the change in quality and annual income displays an inverse U-shape. Low-income households increase the quality of their purchases a little. Middle-income households increase quality the most.

High-income households do not adjust their quality of purchases.

These findings support the theory of non-homothetic demand: when households' income changes, they tilt their spending towards products of higher quality. We incorporate this into a life-cycle buffer-stock model in which households experience idiosyncratic fluctuations in their income and face borrowing constraints that induce buffer-stock savings behavior. In addition to choosing how much to spend and save, households also choose the quality of products entering their consumption baskets. After calibrating the model with plausible fluctuations in income, we show that it is able to match the negative relationship between households' annual income and their propensity to consume out of a transitory income change. In a model in which households face the same fluctuations in income but have standard, homothetic preferences, the relationship is the exact opposite. Finally, we show that our non-homothetic model predicts the inverse U-shaped relationship between annual income and the change in quality as we also found in the data.

Chapter 2 – Government spending and retail prices: Regional evidence from the United States The second chapter documents that retailers charge higher prices in areas where government spending increases. A central challenge to documenting this is overcoming the fact that the government tends to direct its spending to areas that experience bad economic outcomes. This makes it difficult to isolate the causal effect of government spending. Are the observed changes in prices caused by spending itself or the poor economic conditions?

I identify the causal effect of government spending by exploiting two sources of exogenous variation. First, I use provisions within a large stimulus package, the American Recovery and Reinvestment Act of 2009, which generated variation in stimulus across states that was plausibly unrelated to the economic conditions in the states. Second, I exploit that areas in the United States are persistently exposed differently to changes in national military spending. When national military spending increases, the areas that historically received many military contracts will tend to receive many military contracts now irregardless of their current economic situation.

Price changes must be driven by a change in marginal costs or a change in markups. I provide evidence of the latter channel dominating. First, I argue that the regional variation in wholesale costs is likely insufficient to generate changes in marginal costs. Second, I show that wage growth cannot account for the change in prices. I conclude that retailers charge higher markups when government spending flows to the area in which they operate. These findings are consistent with theoretical models in which the price sensitivity of consumers change as their income or employment status change.

Chapter 3 – House Prices, Increasing Returns, and Government Spending Shocks *with Søren Hove Ravn and Emiliano Santoro*

The importance of house price fluctuations in shaping macroeconomic outcomes has received considerable attention in both the public and academic debate after the Great Recession. So has fiscal policy as a tool for stabilizing the economy. The third chapter speaks to both of these debates by analyzing how house prices move in response to fluctuations in government spending.

We estimate a structural vector auto regression and document that house prices increase as a result of a positive shock to fiscal spending. The estimates also show that real wages and total factor productivity (TFP) increase because of the fiscal spending shock.

These estimates are not consistent with the predictions by a large variety of DSGE models. By constrast, these models produce a negative comovement between house prices and government spending. The intuition underlying this feature goes as follows. When government spending rises so do taxes, which lowers the income of households (the negative wealth effect). This causes them to lower consumption and increase labor supply, which also lowers the real wage. Similarly, the negative wealth effect lowers demand for housing, which depresses prices.

We construct a model that is able to overturn these results by overcoming the negative wealth effect through increasing returns to scale at the aggregate level. The model has two layers of production: an intermediate goods sector and a final good sector. The intermediate goods sector consists of a variable number of firms that enter or exit the economy when profit opportunities emerge or disappear. Goods produced by these firms are purchased by the final good sector, which combines them into a final aggregate output. This sector also exhibits taste for variety, which together with the entry and exit of intermediate sector firms generates endogenous variations in TFP. As a result, an increase in demand through an expansion in government spending generates new profit opportunities for intermediate firms. When new firms enter the economy, TFP rises and exerts an upward pressure on real wages. If this channel is sufficiently strong, it overcomes the negative wealth effect on households' income and generates a positive response of house prices.

Chapter 1

Quality and Consumption Basket Heterogeneity: Transitory Shocks and Implications for Consumption-Savings Behavior

Quality and Consumption Basket Heterogeneity

Transitory Shocks and Implications for Consumption-Saving Behavior*

Rasmus Bisgaard Larsen† Christoffer J. 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 non-homothetic demand. To model this, we embed non-homothetic preferences over quantity and quality in an otherwise standard buffer-stock model. Contrary to the standard model, the non-homothetic 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 non-homothetic model also generates a more unequal wealth distribution, which is closer to the data.

*We wish to thank discussants Viktor Malein and Rasmus Kehlet Berg, and participants at the DGPE Workshop 2019 and internal seminars at Danmarks Nationalbank and University of Copenhagen for helpful comments and suggestions. We thank Nielsen for providing access to the Tax Rebate Survey 2008. Researchers own analyses calculated (or derived) based in part on data from The Nielsen Company (US), LLC and marketing databases provided through the Nielsen Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the Nielsen data are those of the researchers and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

†University of Copenhagen.

‡University of Copenhagen and Danmarks Nationalbank.

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 non-homothetic. 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 non-homothetic 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 (2019) and Faber and Fally (2017) 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 specification. 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 non-homothetic.

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. Argente and Lee (2019), Jaimovich et al. (2019a), Jaimovich et al. (2019b) and Michelacci et al. (2019)) 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 non-homotheticities in consumption. Except from these non-homotheticities, 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, and none have provided a rigorous, empirically documented microfoundation. 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 non-homothetic model. A feature of the model is that it allows us to match these moments whereas the standard model does not.[1] Strikingly, 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 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 non-homothetic 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 (2019) and Jaimovich et al. (2019a) show that households traded down in their consumption quality during the Great Recession, Jaimovich et al. (2019b) 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, 2019), 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 et al., 2019a,b). 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

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

propensities to consume out of transitory income shocks. Most related are the papers by Sahm et al. (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 et al. (2019). 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, 1992; Carroll, 1992). Some authors have incorporated non-homothetic preferences into these types of models through a bequest motive (Nardi, 2004; Straub, 2019) or by including wealth directly into the utility function (Carroll, 2000). 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 non-homotheticities are modeled as changing tastes for quality as in the work by Handbury (2019) and Faber and Fally (2017). 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 set 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 (2019) and Faber and Fally (2017), 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

$$U = \prod_m \left[\sum_{i \in G_m} \left(c_{mi} \varphi_{mi}(\xi, P) \right)^{\frac{\sigma-1}{\sigma}} \right]^{\alpha_m(P) \frac{\sigma}{\sigma-1}}, \tag{2.1}$$

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

$$\log \varphi_{mi}(\xi, P) = \gamma(\xi, P) \log \phi_{mi}, \tag{2.2}$$

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 (2.1) 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

$$\log \frac{x_{mi}}{x_{mk}} = (\sigma - 1)\left[\log \frac{\varphi_{mi}(\xi, P)}{\varphi_{mk}(\xi, P)} - \log \frac{\mathcal{P}_{mi}}{\mathcal{P}_{mk}}\right], \tag{2.3}$$

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 (2.3) 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 Appendix A, we show how we can represent equation (2.1) 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

$$\mathcal{P}_m(\xi, P) = \left(\sum_{i \in G_m} \mathcal{P}_{mi}^{1-\sigma} \varphi_{mi}(\xi, P)^{\sigma-1}\right)^{\frac{1}{1-\sigma}}, \tag{2.4}$$

by which we have that

$$U = \frac{X}{\mathcal{P}(\xi, P)} \prod_m \alpha_m(P)^{\alpha_m(P)} = \frac{X}{\mathcal{P}(\xi, P)} \cdot K(P). \tag{2.5}$$

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 non-homotheticities 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]

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 (2017) show, the preferences in equation (2.1) can be derived from aggregation of discrete-choice preferences across many households. Equivalently, the preferences in equation (2.1) hold at the household level in expectations in the logit model.[3]

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 (2.5). Let the problem of the household 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

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

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.

[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.

[3]These results mirror those of Anderson et al. (1987).

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.[4] 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, 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

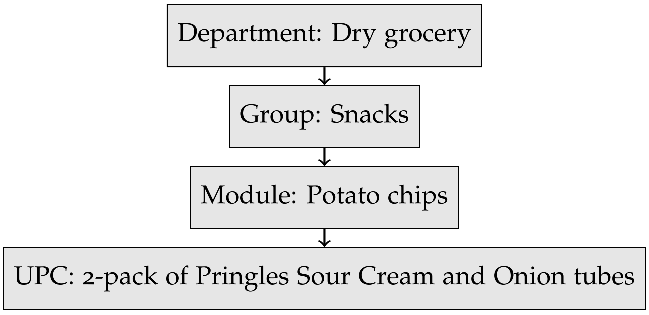
The RSD contains weekly pricing and quantity information at the product level from more than 90 retail chains across the contiguous United States. The data set covers approximately 3.2 million products – both food and non-food groceries – sold from over 35,000 different stores making up about half of 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

[4] It is helpful to remember that $\mathcal{P}(\xi, P)$ is the *quality-adjusted* price index. When $\mathcal{P}(\xi_1, P_1) > \mathcal{P}(\xi_2, P_1)$, it simply says that the *perceived* price of a given consumption basket is lower in the second period, thereby raising the marginal value of consuming that consumption basket. For example, a rich person may not experience the price of a bottle of champagne as the same as a poor person.

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.[5] 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 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

[5] 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.

even though they are two different products sold by two different retail chains.[6]

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.[7] 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 (2.2). 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.[8]

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.[9] 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 (2019) 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 price, $p_{i,m,s,c}$, using the prices of all stores selling the product within CBSA c.

[6]Dubé et al. (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 et al. (2018) estimate a negative effect of income on private-label purchase shares.

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

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

[9]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.

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:

$$q^j_{i,m,s,c} = \frac{\ln p_{i,m,s,c} - \ln \bar{p}_{m,s,c}}{\sigma_{m,s,c}} \tag{3.1}$$

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.[10]

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:

$$q^k_{i,m,c} = \frac{\ln p^u_{i,m,c} - \ln \bar{p}^u_{m,c}}{\sigma^u_{m,c}} \tag{3.2}$$

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

$$q^l_{b,m,c} = \frac{\ln p^u_{b,m,c} - \ln \bar{p}^u_{m,c}}{\sigma^b_{m,c}} \tag{3.3}$$

[10] A histogram of the quality index is shown in appendix figure B.1. We have also confirmed that the ranking of products are similar in other years as shown in Appendix C.

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 and quantity using The Nielsen Consumer Panel

The CPD is a household panel that includes 61,440 households in 2008.[11] 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.[12] 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

[11] 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.

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

of each product and indicate if each product was on sale or purchased using a coupon. Not all products purchased by the households are scanned and registered as individual products.[13] 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.[14] We then construct three aggregate quality measure, $Q^o_{h,t}$ for $o = j, k, l$, for the purchases in week t of household h as the expenditure-weighted averages of the quality of the households' purchases:

$$Q^o_{h,t} = \sum_m \sum_{i \in G_m} w_{i,h,t} q^o_{i,m,c(h)} \qquad \text{for } o = j, k, l \tag{3.4}$$

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 defined in section 3.1.

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, in one of the modules with multiple physical units for the two indices based on unit prices, or 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, appendix figure B.2 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).[15] Panel (a) shows that weekly spending and purchases of large products are positively correlated. Hence, when households increase

[13] Nielsen estimates that around 30 percent of household consumption is covered by the categories tracked in the data.

[14] 3,901 of the households do not live in a CBSA. We exclude these households from the data.

[15] This definition of large-sized products follows Nevo and Wong (2019).

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 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.[16] 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 appendix figure B.3. 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 spans 9 percent 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.

[16] Coibion et al. (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 the brand-based quality index.

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 earlier responses.[17] 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.[18]

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.

Number of goods purchased and household size in the ESP sample are roughly the

[17]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.

[18]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.

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 NCP panel of 2008.

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) and show the same summary statistics along with statistics for the ESP received within each group in table 2. 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.

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

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.

by heterogeneity in matching purchases to the quality indices across income.[19]

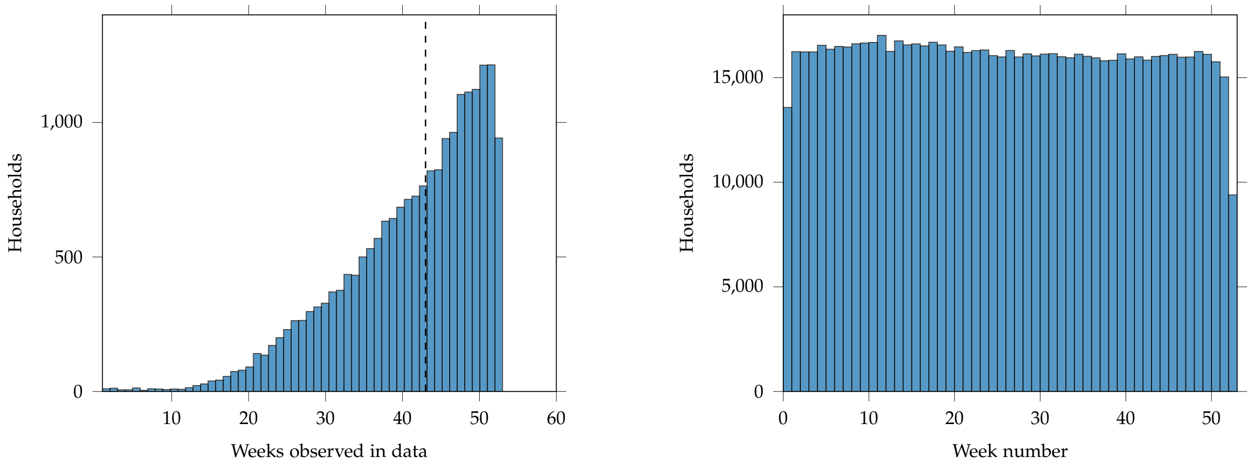
Our analysis of the effect of the ESP on quality is complicated by quality only being observed in weeks, where households actually purchase goods. Moreover, purchasing patterns across weeks are not random. Panel (a) in figure 2 shows the distribution of the number of weeks for which we observe purchases 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 appendix figure B.5. Panel (b) 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.

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) 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.

[19]Appendix figure B.4 shows average spending coverage for the income bins in our data. Coverage is approximately constant across the income bins.

Figure 2: Weekly purchasing patterns in the ESP sample

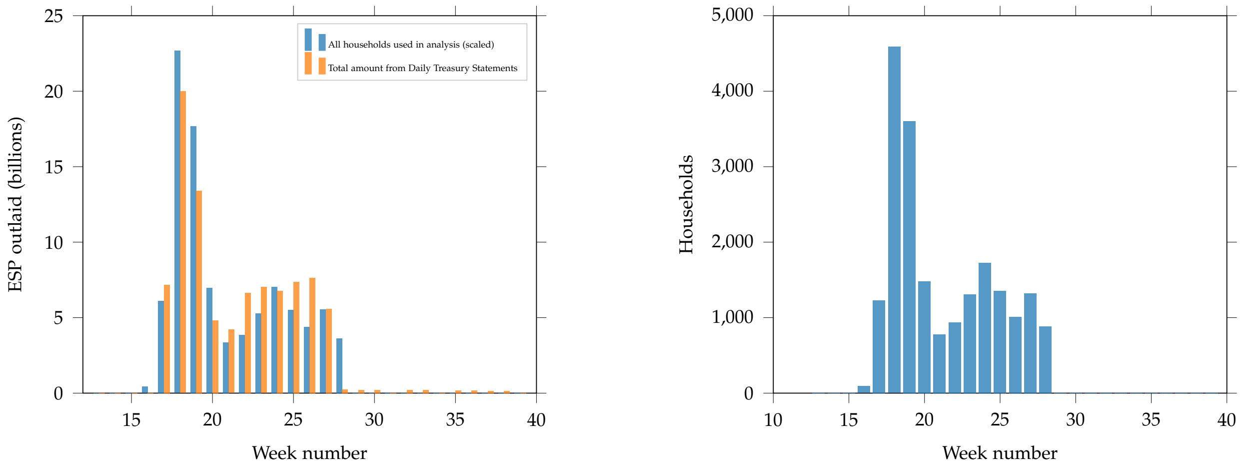


(a) Weekly observations per households

(b) Observations per week

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.

Figure 3: ESP disbursements in sample



(a) Total ESP disbursements by week

(b) Households receiving ESP by week

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.

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 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) 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., 2019b) as well as CPD data (Faber and Fally, 2017; Argente and Lee, 2019; Jaimovich et al., 2019b). The positive relationship holds across the spending and income distribution and is precisely estimated. When we control for other factors that might be correlated with both spending/income and quality – age, household size, race, CBSA of residence and education – the positive relationship is even more pronounced.

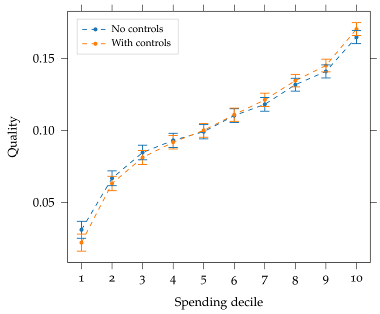
The relationship between quality, spending and income generally holds within the product groups. To show this, we divide the sample into expenditure and income quintiles and estimate the average quality of households' purchases for each quintile within product groups, g, using the following regression:[20]

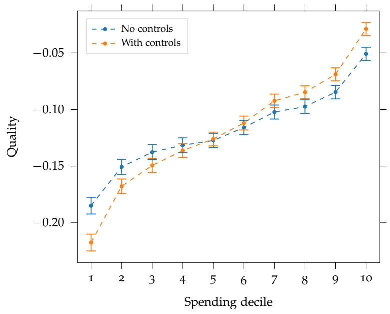
$$Q_{h,g} = \sum_{k=1}^{5} \beta_{g,k} \mathbf{1} \left\{ \text{Quintile}_h = k \right\} + \Gamma_g X_h + \varepsilon_{h,g} \tag{4.1}$$

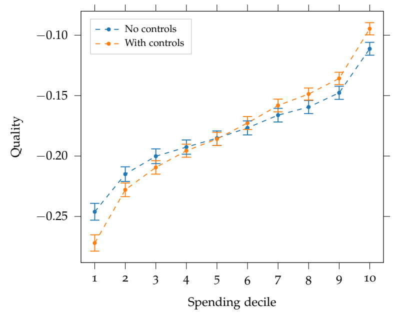
where X_h is as a vector of controls (household size, race, age brackets of the household heads, education of the household heads and CBSA of residence).

[20]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.

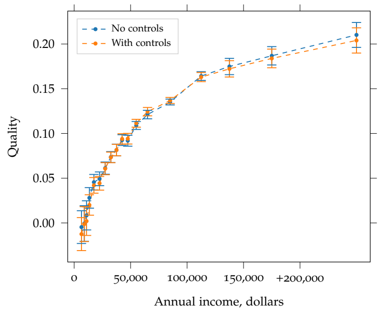
Figure 4: Quality across the spending and income distribution

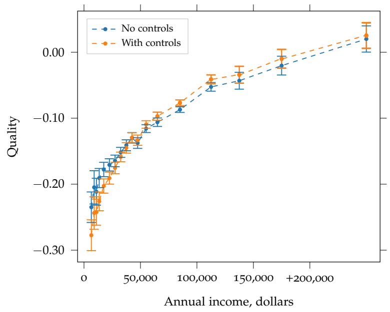


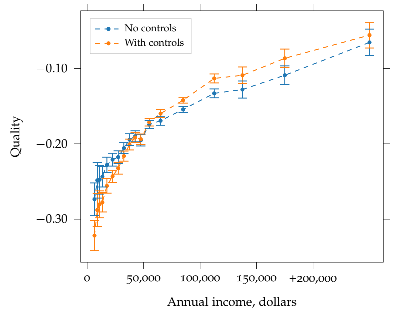




(a) Size-based index (b) Unit price-based index (c) Brand-based index





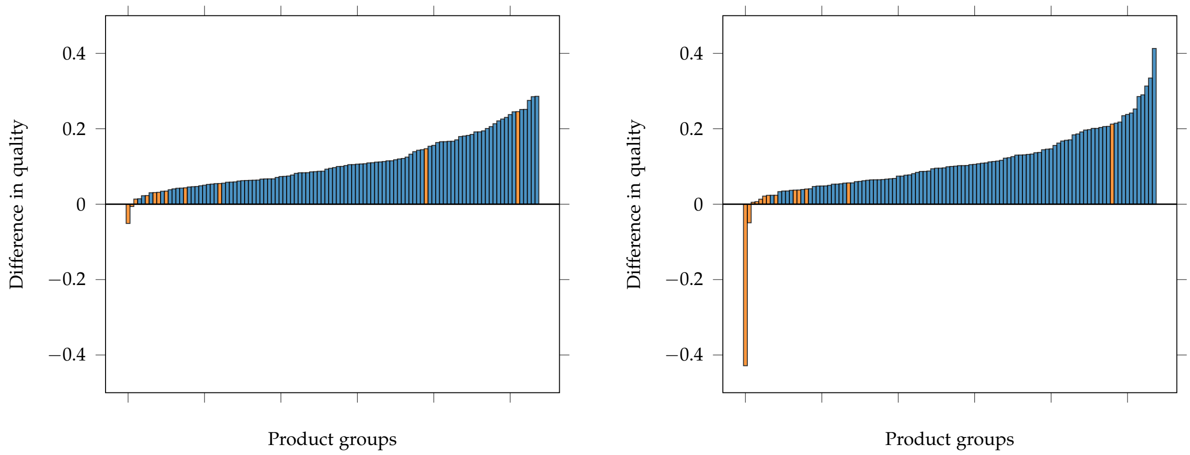


(d) Size-based index (e) Unit price-based index (f) Brand-based index

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.

$\beta_{g,k}$ in regression (4.1) 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.

Figure 5: Difference in quality by product group



(a) Difference in quality by spending quintiles (b) Difference in quality by income quintiles

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 (4.1)). 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.

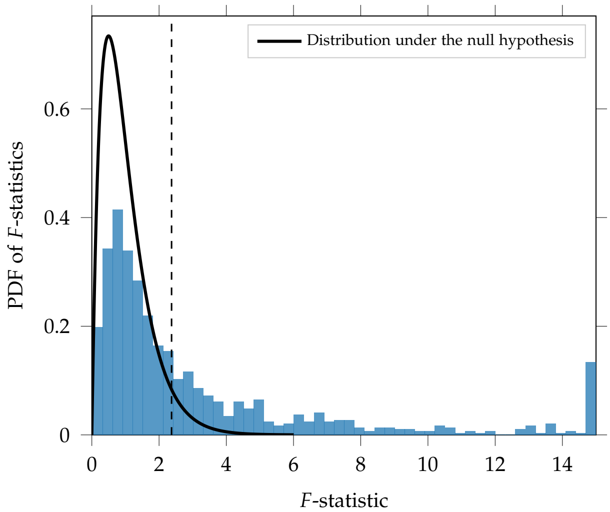
We also observe heterogeneity in the annual expenditure shares of the product modules across income. This is investigated by estimating equation (4.1) with expenditure shares of the modules as the dependent variable and testing that the shares are equal across quintiles, $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5$, 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.[21]

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.[22] Thus, spending on each module is not scaled up proportionally with total spending when comparing households across the income distribution. This implies

[21]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.

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

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 (4.1), 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.

that the parameters, $\alpha_n(P)$, pinning down the expenditure share for each module 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 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 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.[23]

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.[24] 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 to estimate the effect of the ESP on shopping behavior for household h in week t:

$$X_{h,t} = \mu_h + \eta_t + \sum_{s=-L}^{L'} \beta_s ESP_{h,t+s} + \varepsilon_{h,t} \tag{4.2}$$

where $X_{h,t}$ is one of the measures for quality in equation (3.4) or total spending in week

[23]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.

[24]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).

t by household h.[25]

$ESP_{h,t+s}$ 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.[26] 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 (4.2).

The regression includes two fixed effects. 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 in figure 4 above). Without

[25] 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.

[26] 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 and Jaravel, 2017; Abraham and Sun, 2018; Goodman-Bacon, 2018). We stick to the conventional OLS estimator with a flexible set of leads and lags.

household fixed effects, this would bias $\tilde{\beta}$ due to the changing composition of the sample (Borusyak and Jaravel, 2017). 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 having a missing quality variable.

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

$$X_{h,t} = \mu_h + \eta_{j,t} + \sum_j \sum_{s=-L}^{L'} \beta_{j,s} \mathbf{1}\left\{\text{Group}_h = j\right\} ESP_{h,t+s} + \varepsilon_{h,t} \tag{4.3}$$

where j index groups in the data.

This regression is identical to regression (4.2) 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 (4.2) 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 estimated using the Stata package REGHDFE for estimation of high-dimensional fixed effect models by Correia (2019).

Column 1 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 nondurables. According to Broda and Parker (2014), household-level spending in the CPD

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

	Spending	Size-based quality	Unit price-based quality	Brand-based quality
1 month before ESP	4.81***	0.65*	0.63*	0.80**
	(1.21)	(0.34)	(0.36)	(0.33)
Contemporaneous month	12.6***	1.14***	1.22***	1.09***
	(1.26)	(0.35)	(0.38)	(0.34)
2 months after ESP	6.04***	0.83**	0.95**	0.61*
	(1.23)	(0.35)	(0.38)	(0.34)
3 months after ESP	5.05***	0.40	0.67*	0.56*
	(1.17)	(0.34)	(0.37)	(0.34)
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 (4.2). 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.

is 35 percent of spending on non-durables or 19 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.[27]

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.

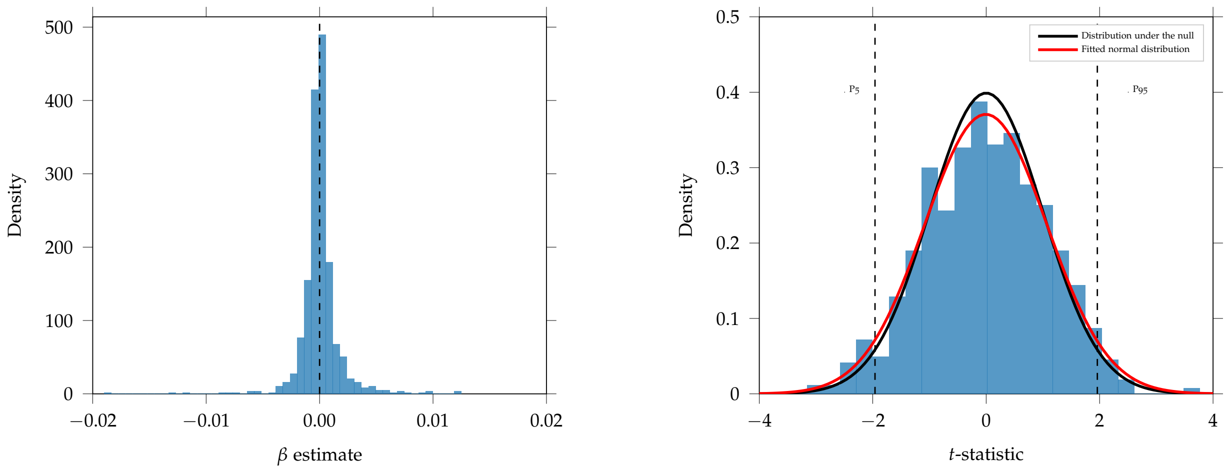
There are indications of a statistically significant effect on both spending and quality in the 4 weeks leading up to receiving the ESP. Due to the truly randomized timing of the ESP, the presence of effects on spending and quality prior to treatment does not invalidate the research design or yield biased estimates. Rather they reflect anticipation effects that are part of the treatment effect (Borusyak and Jaravel, 2017). However, as we

[27] 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 et al. (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).

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.

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 (4.2) 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.

Figure 7: Response of module-switching to the ESP



(a) Estimates for module-switching

(b) t-statistics for module-switching

Note: Panel (a) shows the distribution of estimates of $\tilde{\beta}$ in the 4 weeks after receiving an ESP from equation (4.2) 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.[28]

4.1.3 Heterogeneity of response to ESP by income

Next, we look at how the ESP effects differ by income.[29] 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 regression (4.3) by these groups and present the estimates in table 4.

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, 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.

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.[30]

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.

[28] 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.

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

[30] 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.

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	6.23**	0.21	1.17	0.75
	(2.76)	(0.90)	(0.98)	(0.88)
Contemporaneous month	19.7***	1.69*	2.36**	1.67*
	(2.93)	(0.93)	(0.99)	(0.89)
2 months after ESP	8.97***	0.60	1.75*	0.47
	(2.85)	(0.95)	(1.01)	(0.91)
3 months after ESP	7.86***	0.26	0.03	-0.52
	(2.64)	(0.91)	(0.98)	(0.89)
Week × household obs.	357,061	280,111	278,579	278,513
Households	6,737	6,734	6,735	6,735
	Income between $35,000 and $70,000			
1 month before ESP	5.83***	1.02**	0.56	0.70
	(1.81)	(0.51)	(0.55)	(0.49)
Contemporaneous month	12.9***	1.62***	1.55***	1.17**
	(1.86)	(0.53)	(0.57)	(0.52)
2 months after ESP	6.21***	1.54***	1.65***	0.90*
	(1.82)	(0.53)	(0.58)	(0.52)
3 months after ESP	5.11***	1.22**	1.92***	1.38***
	(1.79)	(0.52)	(0.56)	(0.51)
Week × household obs.	395,592	310,932	309,396	309,347
Households	7,464	7,458	7,458	7,458
	Income above $70,000			
1 month before ESP	3.22	0.45	0.26	0.87*
	(1.98)	(0.48)	(0.52)	(0.45)
Contemporaneous month	7.76***	0.25	0.034	0.60
	(2.06)	(0.49)	(0.54)	(0.48)
2 months after ESP	3.77*	0.14	-0.44	0.27
	(2.02)	(0.50)	(0.56)	(0.49)
3 months after ESP	3.31*	-0.39	-0.38	0.26
	(1.93)	(0.47)	(0.54)	(0.47)
Week × household obs.	316,622	244,427	243,269	243,247
Households	5,974	5,973	5,973	5,973

Notes: The table shows the estimates of $\tilde{\beta}$ from equation (4.3) split by income groups. 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.

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),

Table 5: Relative marginal propensities to consume

	Bottom-to-top	Bottom-to-middle	Middle-to-top
Relative MPC	2.46	1.51	1.64
	(1.05)	(0.45)	(0.69)

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.

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 et al. (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 et al. (2010) use the University of Michigan's monthly Survey of Consumers to study *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 et al. (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 regression (4.2) 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.[31] 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 appendix table B.1. 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 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 regression (4.2) without restricting coefficients and also formally test for pre-ESP effects, we follow Borusyak and Jaravel (2017). 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

[31] 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).

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.[32] 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. Appendix table B.2 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 3 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 regression (4.2) using all combinations of leads and lags 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 appendix figure B.6. 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 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.

[32] As emphasized by Borusyak and Jaravel (2017), 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).

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 et al. (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 (2019) 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.[33] We then divide households into income tertiles year-by-year based on the household size-adjusted income. Similar to the method by Dynan et al. (2004), we create the tertiles within the 9 age bins in the Nielsen data to account for age differences in income levels.[34] Finally, we use each household's modal value of these tertiles to group households. Appendix table B.3 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 regression (4.2) 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

[33]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).

[34]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.

fixed effects. Hence, these estimates reflect the average effects on spending and quality *within* product categories of the ESP. Note that this implies that the spending estimates will mechanically be smaller since they measure the average spending increase within each product category.[35] Appendix table B.4 presents 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 regression (4.2) 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, appendix table B.5 shows the results from regression (4.2) when we have included week $\times$ 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 appendix table B.5 is a weighted average of the disbursement method-specific ESP effects (Gibbons et al., 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.[36]

[35]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.

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

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 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.[37] 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 household receives 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

[37] 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.

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 (2.1) and V_t is the value function. Using equation (3.1), we can write the problem as

$$V_t(M_t, P_t, \xi_t) = \max_{X_t} \frac{(X_t \cdot f(\xi_t, P_t))^{1-\rho}}{1-\rho} + \beta \mathbb{E}_t[V_{t+1}(M_{t+1}, P_{t+1}, \xi_{t+1})]. \tag{5.1}$$

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

$$A_t = M_t - X_t, \tag{5.2}$$

$$M_{t+1} = RA_t + Y_{t+1}, \tag{5.3}$$

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

$$A_t \geq -\lambda P_t. \tag{5.4}$$

Households cannot leave life in debt and therefore

$$A_T \geq 0. \tag{5.5}$$

The income process is exogenous and given by

$$Y_{t+1} = \xi_{t+1} P_{t+1}, \qquad \xi_{t+1} = \begin{cases} \mu & \text{with prob. } \pi, \\ \frac{\varepsilon_{t+1} - \mu\pi}{1-\pi} & \text{else,} \end{cases} \tag{5.6}$$

$$P_{t+1} = GP_t \psi_{t+1}, \tag{5.7}$$

where

$$\log \psi_{t+1} \sim \mathcal{N}(-0.5\sigma_\psi^2, \sigma_\psi^2), \tag{5.8}$$

$$\log \varepsilon_{t+1} \sim \mathcal{N}(-0.5\sigma_\xi^2, \sigma_\xi^2). \tag{5.9}$$

The income process is similar to that found in e.g. Carroll (2019). 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

$$f(\xi_t, P_t) = \kappa e^{-\iota e^{-\delta \cdot \xi_t P_t}}, \tag{5.10}$$

which belongs to the class of sigmoid functions.[38] This function has some particularly nice features. In general, sigmoid functions follow an S-shape and in terms of equation (5.10), 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.[39] When calibrating the model, we thus also allow for the possibility that the standard buffer-stock model is favored. 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.[40]

In figure 8, we show 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 (2019) combined with an upper envelope algorithm as in Druedahl and Jørgensen (2017). In Appendix F, we provide details on the computational part.

5.2 Calibration

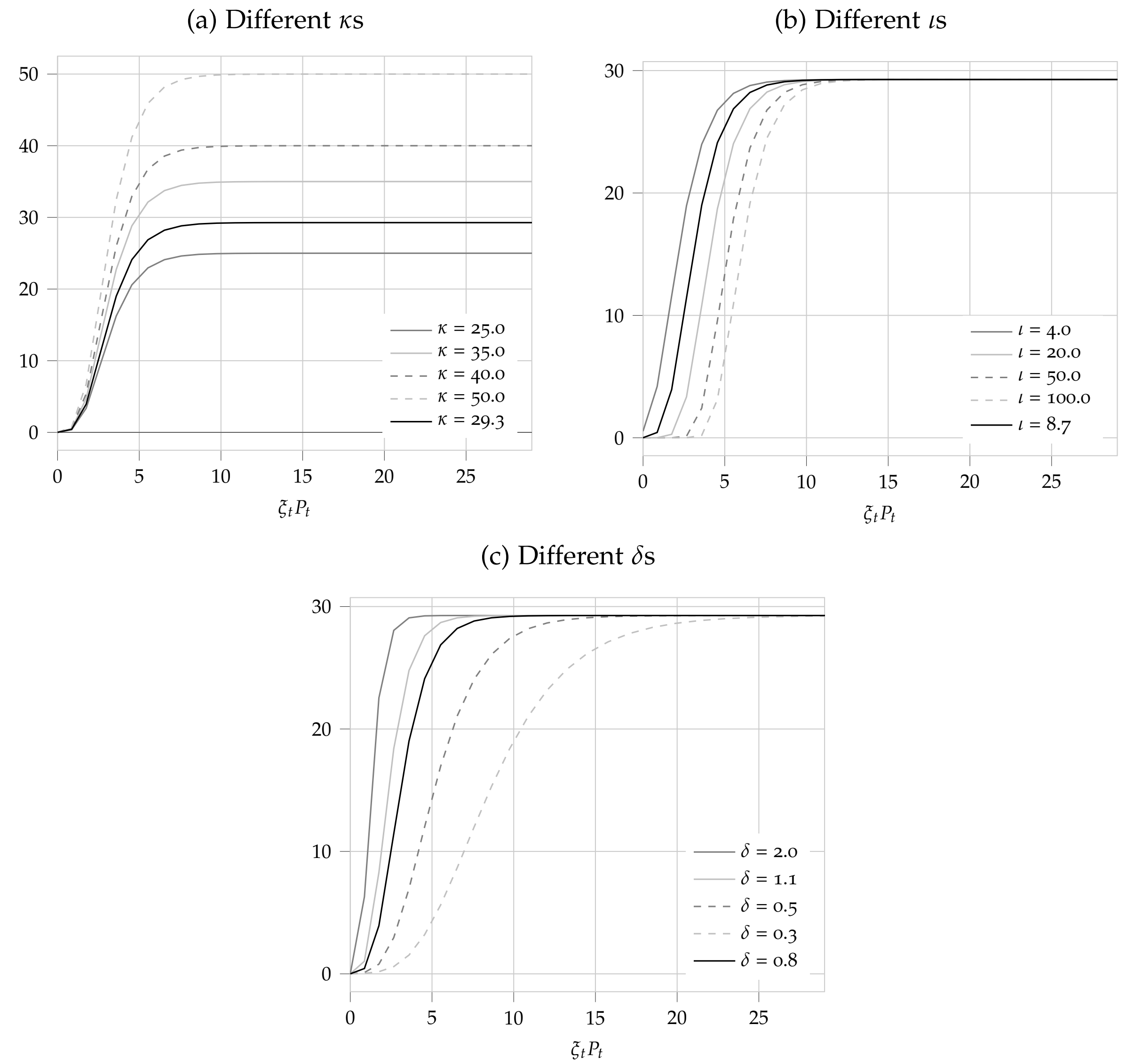
For the parts of the model which are specified as a standard buffer-stock model, we calibrate it in close alignment with the previous literature. Specifically, we use the exact same values for the standard parameters as in Carroll (2019). Households live in the

[38]Specifically, the function in equation (5.10) is the so-called Gompertz function.

[39]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 (2.5), $\kappa > 0$ is a restriction that the price index is positive.

[40]We also allow for the possibility, where either ι, δ or both are negative. We do, however, not find any support for this. Appendix 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 used 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.

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 pct. and with a probability of 0.5 pct. the household ends up in a zero-income event.

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

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 with non-homothetic preferences. The reason why the non-homothetic 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 in turn, 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 7: Calibration results for relative MPCs

		Model	
	Empirical	**Standard**	**Non-homothetic**
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

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

5.2.2 Discussion on relative MPC moments

This section goes into details about what the results in table 7 are driven by. Specifically, we shed some light on why the standard buffer-stock model predicts higher MPCs for richer households and why this is not the case in the non-homothetic model.

It is easiest to understand the properties of the non-homothetic model if we compare it with the standard model. As documented by e.g. Carroll (1997), households in this 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 non-homothetic 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.[41] From these, it is possible to infer the MPC from the gradient on a given point on the curve.

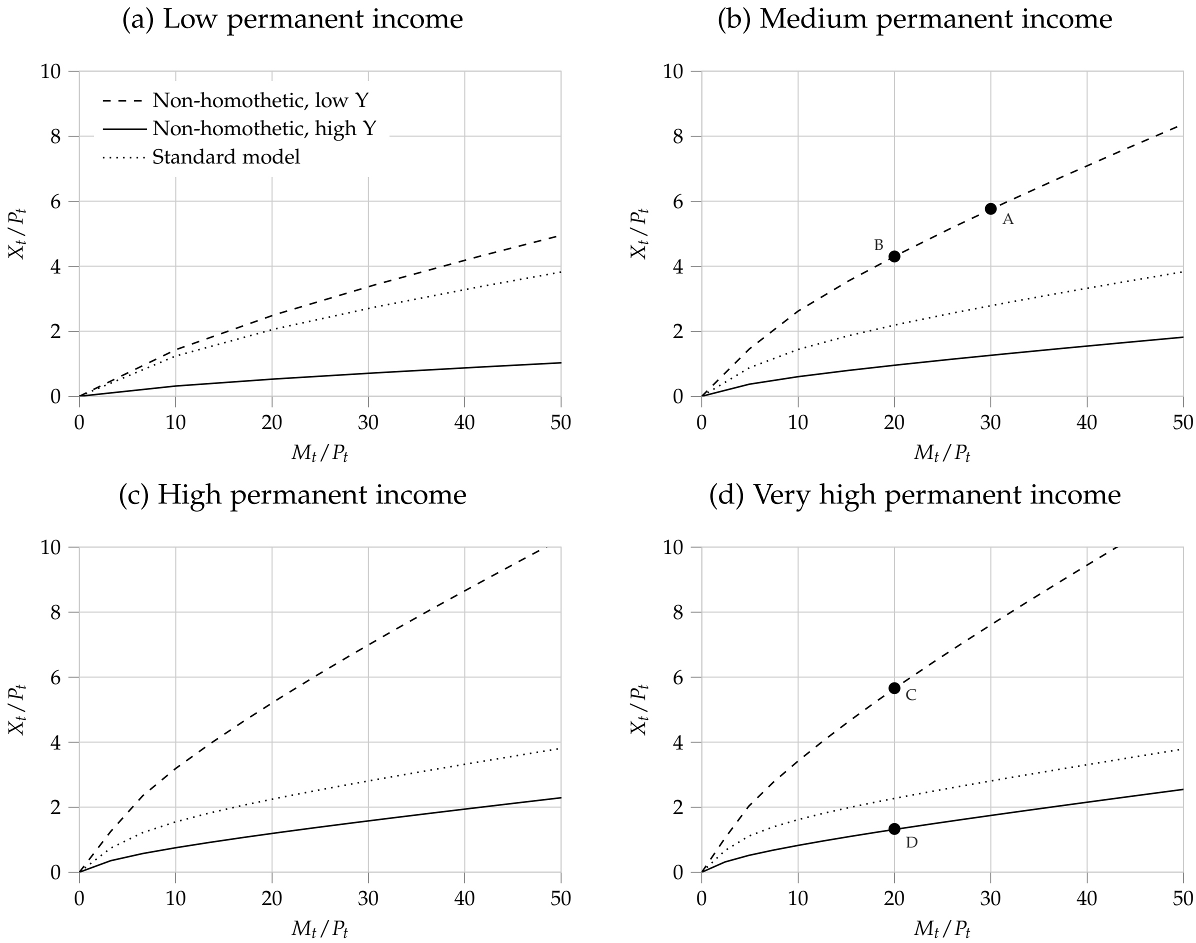
Figure 9 plots the policy functions for both models under various scenarios. In each of the four panels, we show how the policy function is affected by different levels of permanent income, ranging from a low level ($P = 0.86$) to a very high level ($P = 3.57$).[42] *Within* each panel, we show the policy function in the standard model (dotted line) and

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

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

the policy function in the non-homothetic 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 non-homothetic model, the policy functions vary with income. As we will argue now, the underlying reason why the non-homothetic model is able to match the relative MPC moments is exactly highlighted in figure 9.

Figure 9: Policy function in standard and non-homothetic model. Different levels of Y and P.



Notes: Policy functions are normalized by P. In panel 9a, $P = 0.86$. In panel 9b, $P = 1.74$. In panel 9c, $P = 2.65$. In panel 9d, $P = 3.57$. In all scenarios, low corresponds to $Y = 0.86$ and high to $Y = 2.66$.

Consider a household which has a medium level of permanent income but experienced a bad transitory income shock (low Y). Let this be illustrated by point A in panel 9b. Now, imagine 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. However, in the non-homothetic model, 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 9b to point C in panel 9d. 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 (5.3), 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 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 (2.4). 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. In Appendix H.1, we show that the requirement is

$$\sum_m \sum_{i \in G_m} A_m B_{mi} \frac{\partial \varphi_{mi}(\xi, P)}{\partial \xi} > 0. \tag{5.11}$$

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.[43] Lastly, coupling this with our study of relative demand of two goods in section 2, in particular equation (2.3), 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.

[43] 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.

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[44]

$$MQC^i \equiv \lim_{\Delta \downarrow 0} \frac{f(P_t^i, \xi_t^i + \Delta) - f(P_t^i, \xi_t^i)}{\Delta}, \tag{5.12}$$

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.

As figure 10 reveals, 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. In general, that is what figure 10 shows.[45] Thus, the poorest 1 pct. have a MQC of 0.5, the richest 1 pct. of the households have a MQC close to zero, while the lowest MQC among the middle-income households is 4.3.[46]

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 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

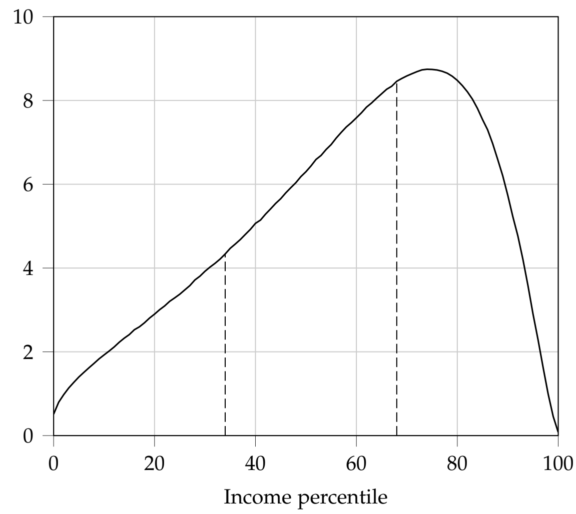
[44]Note, that we scale the shock so that all households receive the same windfall in absolute terms.

[45]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 appendix 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.

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

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. With that being said, the high dispersion of the MQC within the high-income group perhaps explains why the empirical estimates for this group is insignificant.

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 (5.12). That is, group 1 is the 1 pct. poorest households and group 100 is the 1 pct. 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 of, we investigate a result, which has been key to previous studies looking at the effects of non-homotheticities in dynamic consumption-saving models: the wealth distribution. As both highlighted in Carroll (2000) and latest in Straub (2019), non-homotheticities 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. However, in contrast to the previous literature, the foundation on which our non-homotheticities rely are completely different. In Straub (2019), the non-homotheticities come from a bequest motive. In Carroll (2000), 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 non-homotheticities come from the microfoundation outlined in section 2 and rigorously explored in section 4. To this end, we feel confident about our microfoundation and to corroborate previous findings, we should echo their results. Now, before proceeding, a small word of caution is warranted. 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 for a full-fledged general equilibrium model to speak to.

Table 8 reports the Gini coefficients of the two models and the corresponding Lorenz curves are shown in figure 11. In Appendix G, we also show the asset distributions in the two models. Wealth inequality is more than three times higher in the non-homothetic model compared to the standard model. The standard model provides a fairly equal wealth distribution with a wealth Gini of 0.15. In the non-homothetic model, the wealth Gini is 0.49. The differences in the wealth distributions in the two models is further highlighted in the three last columns of table 8. Here we show the wealth holdings of the bottom 50 pct., the top 10 pct. and the top 1 pct. In the standard model, the bottom 50 pct. hold 39.1 pct. of wealth. In the non-homothetic model, the bottom 50 pct. hold 17.7 pct. According to the World Income Inequality Database (WIID),[47] the bottom 50 pct. owned around 0 pct. of (net) wealth in the U.S. in 2014.[48] For the top 10 pct., they own 14.5 pct. of total wealth in the standard model, whereas in the non-homothetic model they own 36.5 pct. The WIID reports that this figure was 73 pct. in the U.S. in 2014. For the top 1 pct., they own 1.6 pct. in the standard model and 6.9 pct. in the non-homothetic model. The WIID reports that this was 38.6 pct. in the U.S. in 2014.

Why does the model with non-homothetic 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 non-homothetic model, their increased demand for high quality goods acts as a savings motive because they wish to continue consuming goods of high quality. This exacerbates wealth inequality.

[47] Latest version: UNU-WIDER, World Income Inequality Database (WIID4). See Piketty et al. (2018) for further details.

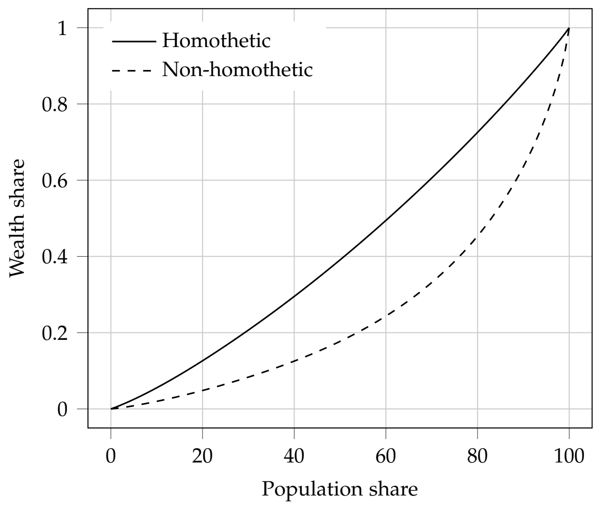
[48] Net personal wealth is defined as the total value of non-financial and financial assets (housing, land, deposits, bonds, equities, etc.) held by households, minus their debts in the WIID.

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.

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 non-homothetic preferences. In this model, households not only choose the quantity but also take into account the quality of their consumption. The non-homothetic 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 by Straub (2019) regarding non-homothetic preferences and savings behavior since the non-homothetic preferences in our model generate more wealth inequality than the standard model does.

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 non-homothetic 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 (2017). 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

Abraham, S. and Sun, L. (2018). Estimating Dynamic Treatment Effects in Event Studies With Heterogeneous Treatment Effects. *Unpublished*.

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

Argente, D. and Lee, M. (2019). Cost of Living Inequality during the Great Recessions. *Kilts Center for Marketing at Chicago Booth – Nielsen Dataset Paper Series*, 1-032.

Bils, M. and Klenow, P. J. (2001). Quantifying Quality Growth. *American Economic Review*, 91(4):1006–1030.

Borusyak, K. and Jaravel, X. (2017). Revisiting Event Study Designs. *Unpublished*.

Broda, C. and Parker, J. A. (2014). The Economic Stimulus Payments of 2008 and the aggregate demand for consumption. *Journal of Monetary Economics*, 68:S20–S36.

Broda, C. and Weinstein, D. E. (2010). Product Creation and Destruction: Evidence and Price Implications. *American Economic Review*, 100(3):691–723.

Campanale, C. (2018). Luxury consumption, precautionary savings and wealth inequality. *The B.E. Journal of Macroeconomics*, 18(1).

Carroll, C. (2006). The method of endogenous gridpoints for solving dynamic stochastic optimization problems. *Economics Letters*, 91(3):312–320.

Carroll, C. D. (1992). The Buffer-Stock Theory of Saving: Some Macroeconomic Evidence. *Brookings Papers on Economic Activity*, 2:61—156.

Carroll, C. D. (1997). Buffer-stock saving and the life cycle permanent income hypothesis. *The Quarterly Journal of Economics*, 112(1):1–55.

Carroll, C. D. (2000). Why Do the Rich Save So Much? In Slemrod, J. B., editor, *Does Atlas Shrug? The Economic Consequences of Taxing the Rich*. Harvard University Press.

Carroll, C. D. (2019). Theoretical Foundations of Buffer Stock Saving. *Unpublished*.

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

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.

Deaton, A. (1991). Saving and Liquidity Constraints. *Econometrica*, 59(5):1221–1248.

Deaton, A. (1992). Understanding Consumption. *Oxford: Oxford University Press*.

Dobkin, C., Finkelstein, A., Kluender, R., and Notowidigdo, M. J. (2018). The Economic Consequences of Hospital Admissions. *American Economic Review*, 108(2):308–352.

Druedahl, J. (2019). A Guide On Solving Non-Convex Consumption-Saving Models. *Unpublished*.

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*, 74:87 – 107.

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*, 37(1):22–53.

Dynan, K. E., Skinner, J., and Zeldes, S. P. (2004). Do the Rich Save More? *Journal of Political Economy*, 112(2):397–444.

Faber, B. and Fally, T. (2017). Firm Heterogeneity in Consumption Baskets: Evidence from Home and Store Scanner Data. NBER Working Paper No. 23101, National Bureau of Economic Research.

Feenstra, R. (1994). New Product Varieties and the Measurement of International Prices. *American Economic Review*, 84(1):157–77.

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

Goodman-Bacon, A. (2018). Difference-in-Differences with Variation in Treatment Timing. NBER Working Paper No. 25018, National Bureau of Economic Research.

Handbury, J. (2019). Are Poor Cities Cheap for Everyone? Non-Homotheticity and the Cost of Living Across U.S. Cities. NBER Working Paper No. 26574, National Bureau of Economic Research.

Jaimovich, N., Rebelo, S., and Wong, A. (2019a). Trading down and the business cycle. *Journal of Monetary Economics*, 102:96–121.

Jaimovich, N., Rebelo, S., Wong, A., and Zhang, M. B. (2019b). Trading Up and the Skill Premium. In Eichenbaum, M. S., Hurst, E., and Parker, J. A., editors, *NBER Macroeconomics Annual 2019, volume 34*. University of Chicago Press.

Jappelli, T. and Pistaferri, L. (2014). Fiscal Policy and MPC Heterogeneity. *American Economic Journal: Macroeconomics*, 6(4):107–136.

Johnson, D. S., Parker, J. A., and Souleles, N. S. (2006). Household Expenditure and the Income Tax Rebates of 2001. *The American Economic Review*, 96(5):1589–1610.

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

Kaplan, G. and Violante, G. L. (2014). A Model of the Consumption Response to Fiscal Stimulus Payments. *Econometrica*, 82(4):1199–1239.

Kueng, L. (2018). Excess Sensitivity of High-Income Consumers. *The Quarterly Journal of Economics*, 133(4):1693–1751.

Lewis, D., Melcangi, D., and Pilossoph, L. (2019). Latent Heterogeneity in the Marginal Propensity to Consume. In *Federal Reserve Bank of New York Staff Reports*, number 902.

Michelacci, C., Paciello, L., and Pozzi, A. (2019). The Extensive Margin of Aggregate Consumption Demand. *Unpublished*.

Misra, K. and Surico, P. (2014). Consumption, Income Changes, and Heterogeneity: Evidence from Two Fiscal Stimulus Programs. *American Economic Journal: Macroeconomics*, 6(4):84–106.

Nardi, M. D. (2004). Wealth Inequality and Intergenerational Links. *The Review of Economic Studies*, 71(3):743–768.

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

Parker, J. A. (2017). Why Don't Households Smooth Consumption? Evidence from a $25 Million Experiment. *American Economic Journal: Macroeconomics*, 9(4):153–183.

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*, 1(3):273–290.

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*, 103(6):2530–2553.

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

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*, 135(1):503–560.

Sahm, C. R., Shapiro, M. D., and Slemrod, J. (2010). *Household Response to the 2008 Tax Rebate: Survey Evidence and Aggregate Implications*, book 3, pages 69–110. The University of Chicago Press.

Shapiro, M. and Slemrod, J. (2003). Consumer Response to Tax Rebates. *American Economic Review*, 93:381–396.

Shapiro, M. and Slemrod, J. (2009). Did the 2008 Tax Rebates Stimulate Spending? *The American economic review*, 99:374–379.

Straub, L. (2019). Consumption, Savings, and the Distribution of Permanent Income. *Unpublished*.

Stroebel, J. and Vavra, J. (2019). House Prices, Local Demand, and Retail Prices. *Journal of Political Economy*, 127(3):1391–1436.

Verhoogen, E. A. (2008). Trade, Quality Upgrading, and Wage Inequality in the Mexican Manufacturing Sector. *Quarterly Journal of Economics*, 123(2):489–530.

Wachter, J. A. and Yogo, M. (2010). Why Do Household Portfolio Shares Rise in Wealth? *The Review of Financial Studies*, 23(11):3929–3965.

A Additional derivations for the theoretical setup

A.1 Price index and indirect utility function

The household solves the following problem

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

The Lagrangian is given by

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

from which it holds that

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

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

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

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

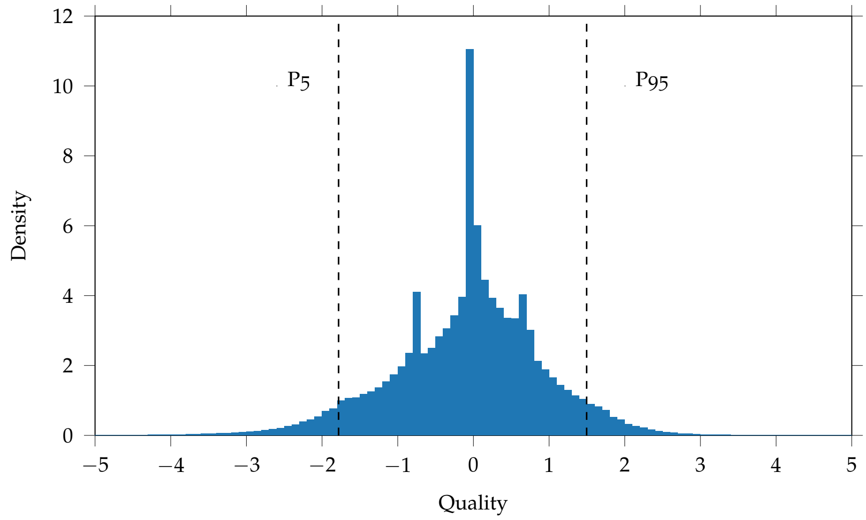
$$\begin{aligned} U &= \prod_n \left[\frac{X_m}{\mathcal{P}_m(\xi, P)}\right]^{\alpha_m(P)} \\ &= X\prod_m \left(\frac{\alpha_m(P)}{\mathcal{P}_m(\xi, P)}\right)^{\alpha_m(P)}. \end{aligned}$$

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

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

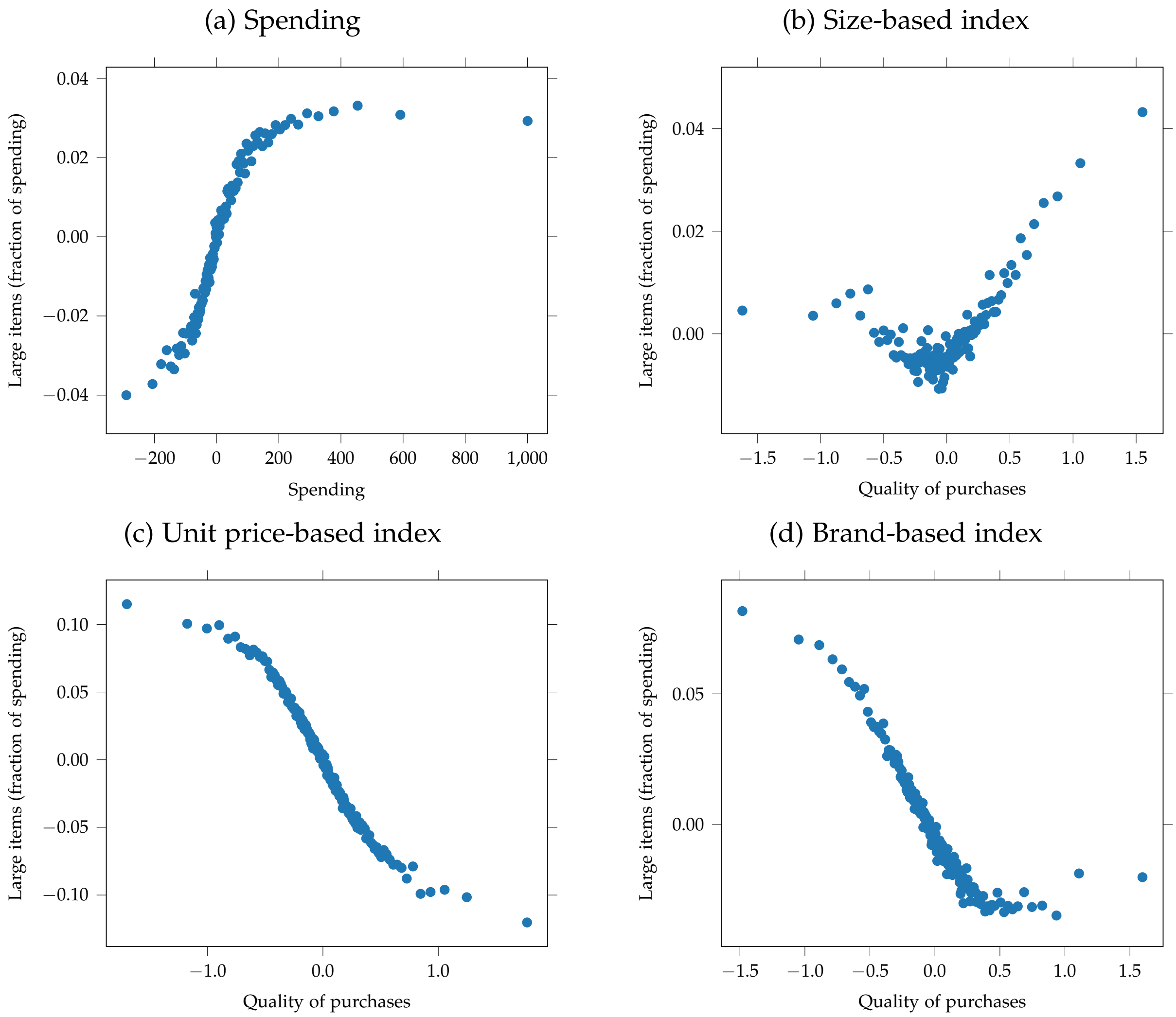
B Additional figures and tables

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

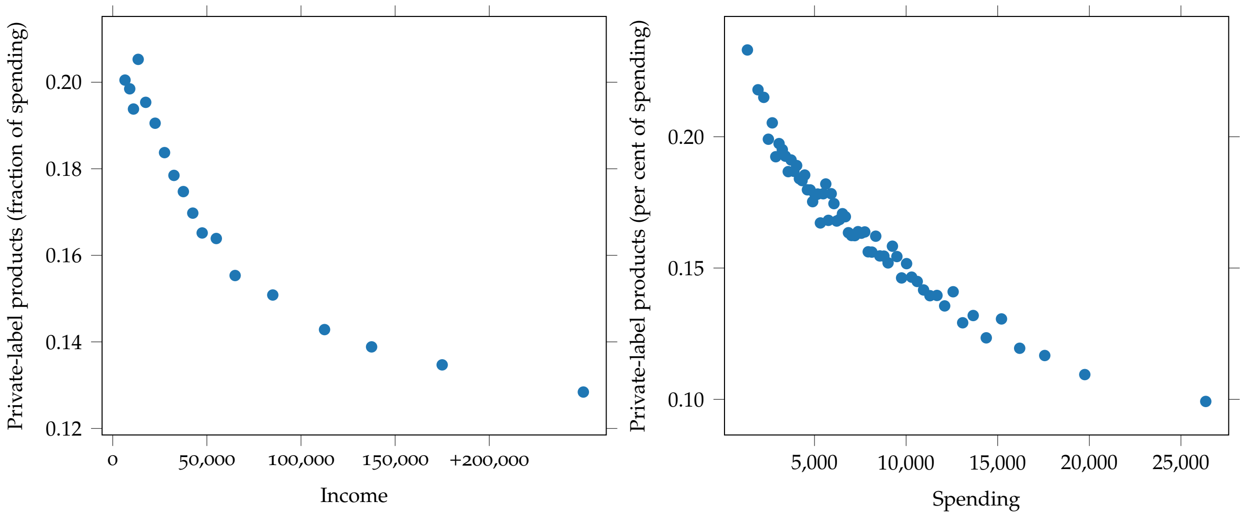


Notes: Binned scatter plots in which each point is the mean value within bins. y-axes display the fraction of weekly spending on products in the top 40 percent of the product size distribution within product modules. All variables have been residualized with household and week fixed effects.

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

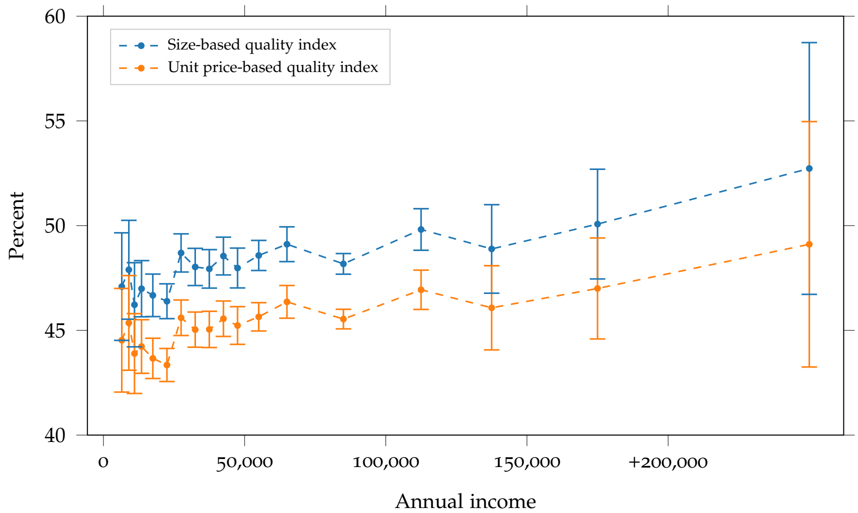
(a) Income and private-label products

(b) Spending and private-label products



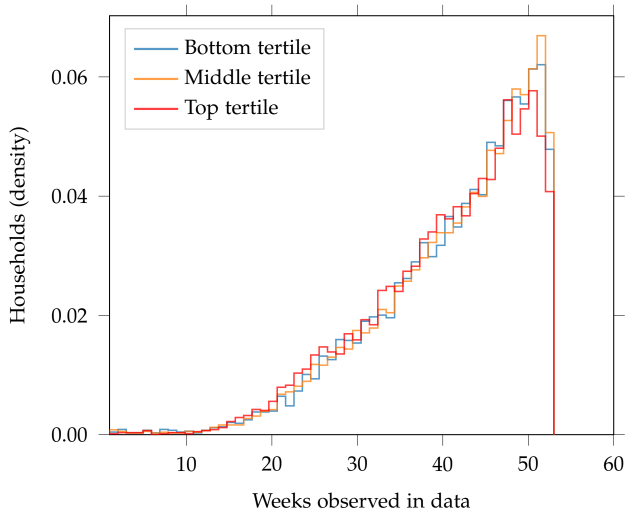
Notes: Panel (a) plots the average expenditure share of private-label products within the midpoint of income bins. Panel (b) is a binned scatter plot of annual spending against the expenditure share of private label products.

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 $\times$ 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 (4.2) 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.

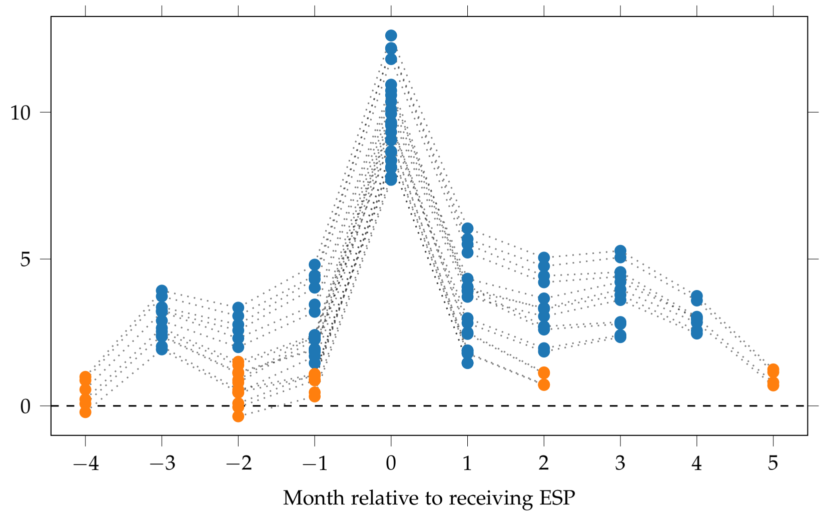
Table B.2: Robustness: ESP estimates from non-constrained regression

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

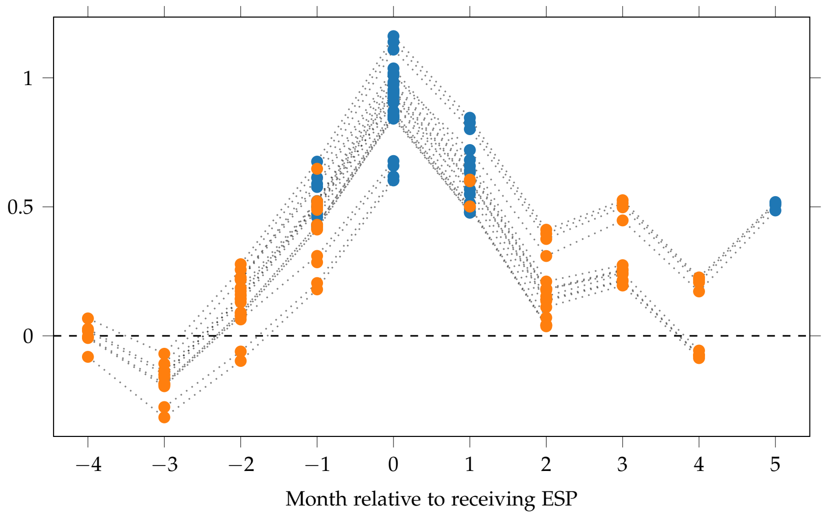
Notes: The table shows the estimates of $\tilde{\beta}$ from equation (4.2) when the coefficients are not constrained. The 1 and 16 weeks lead coefficients are normalized to zero. The sample is balanced around ESP receipt. Estimates from regressions with a quality measures as the dependent variable have 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.

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

(a) Spending estimates



(b) Quality estimates (size-based)



Notes: Figures plot the estimates from equation (4.2) 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 (4.3). 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.

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 (4.2) 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.

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 $\times$ 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 (4.2) when week fixed effects are replaced with disbursement method $\times$ 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.

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

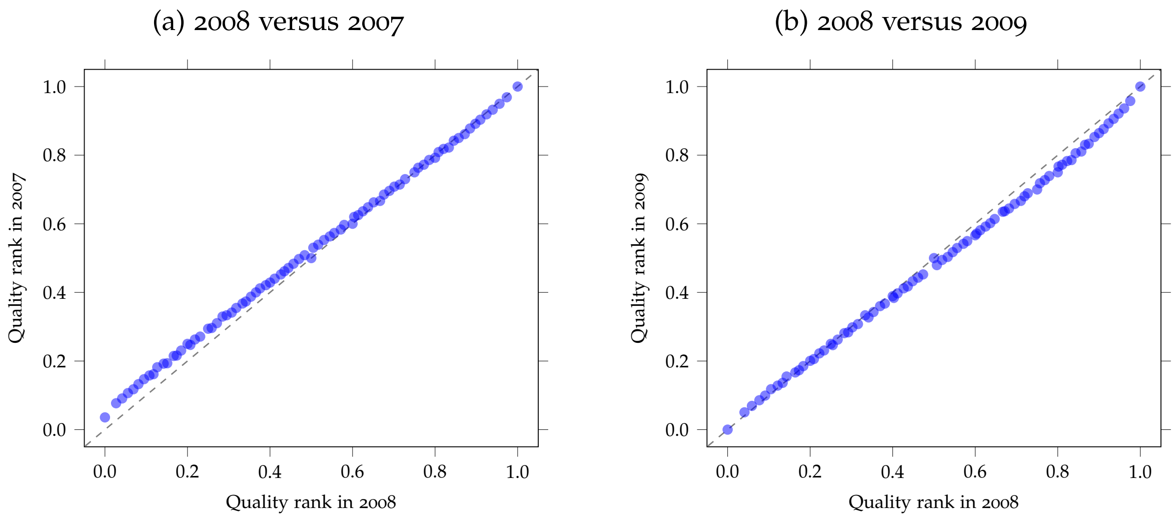
$$\frac{\text{rank}_{i,m,s,c} - 1}{N_{m,s,c} - 1} \tag{C.1}$$

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.[49] 35 percent of the households in our sample report that they do not have access to liquid

[49]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".

wealth. We now look at how households' quality response differ by their access to liquidity. Regression (4.3) is estimated according to this split, and the results are presented in table D.1.

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 (4.3) 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

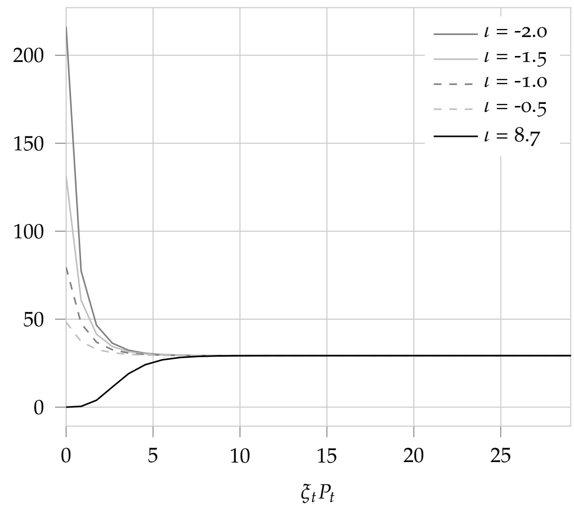
$$f(\xi_t, P_t) = \kappa e^{-\iota e^{-\delta \cdot \xi_t P_t}}. \tag{E.1}$$

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.

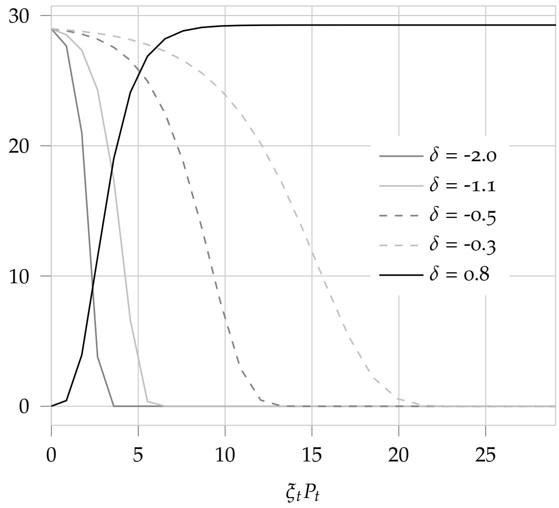


Note: The black line shows the calibrated $f()$ function used in the solution of the model.

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.



Note: 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 (2019) 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).[50] 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.

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

Following Druedahl (2019), 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.[51]

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.[52]

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

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

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. Specifi-

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

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

cally, we have that

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

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

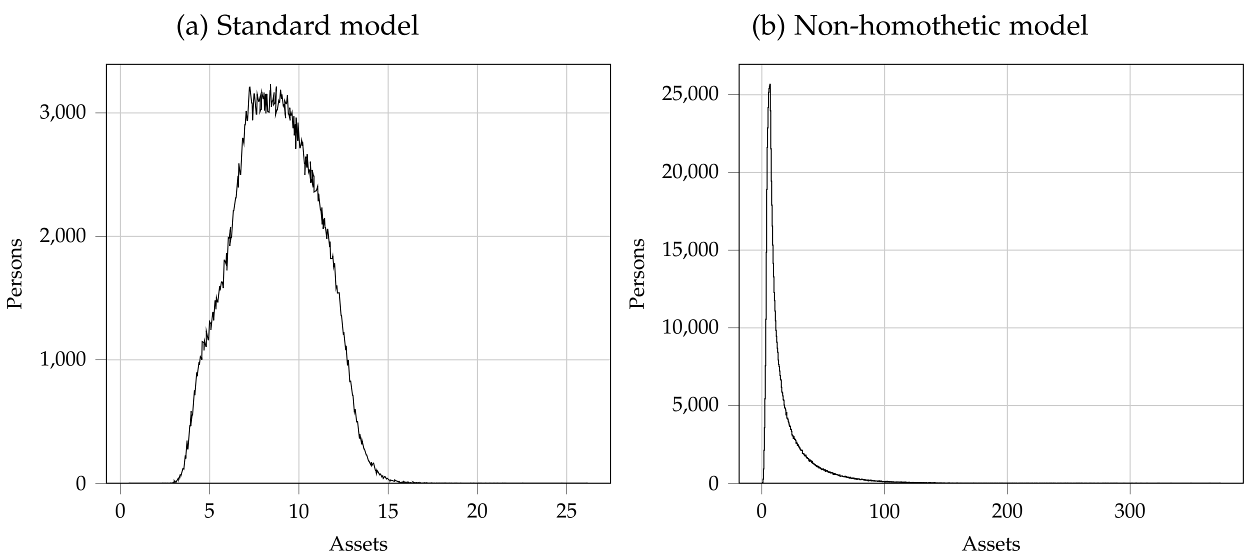
$$X_t = F(A_t, \xi_t, P_t; \xi_{t+1}, \psi_{t+1}), \tag{F.6}$$

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

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

G Asset distributions in the two models

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



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:

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

Now, the requirement is that $\frac{\partial f(\xi, P)}{\partial P} > 0$. Using this, equation (I) gives us

$$\begin{aligned} &\sum_m \alpha_m(P) \frac{1}{\mathcal{P}_m(\xi, P)^{1-\sigma}} \cdot \sum_{i \in G_m} \mathcal{P}_{mi}(\xi, P)^{1-\sigma} \varphi_{mi}(\xi, P)^{\sigma-2} \frac{\partial \varphi_{mi}(\xi, P)}{\partial \xi} > 0 \\ &= \sum_m \sum_{i \in G_m} A_m B_{mi} \frac{\partial \varphi_{mi}(\xi, P)}{\partial \xi} > 0, \end{aligned}$$

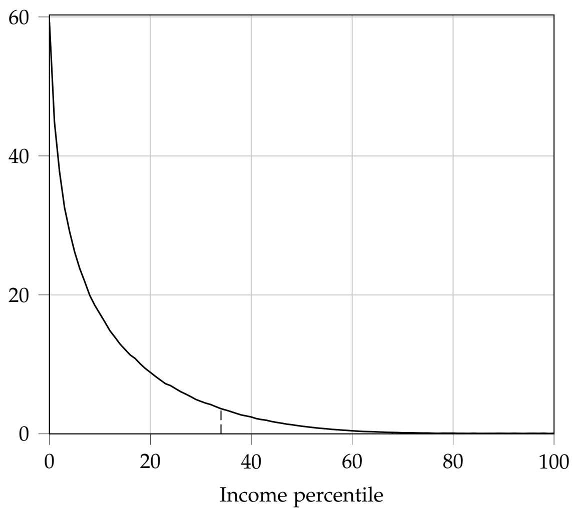
where

$$A_m \equiv \alpha_m(P) \frac{1}{\mathcal{P}_m(\xi, P)^{1-\sigma}} \geq 0, \qquad \text{and} \qquad B_{mi} \equiv \mathcal{P}_{mi}(\xi, P)^{1-\sigma} \varphi_{mi}(\xi, P)^{\sigma-2} > 0.$$

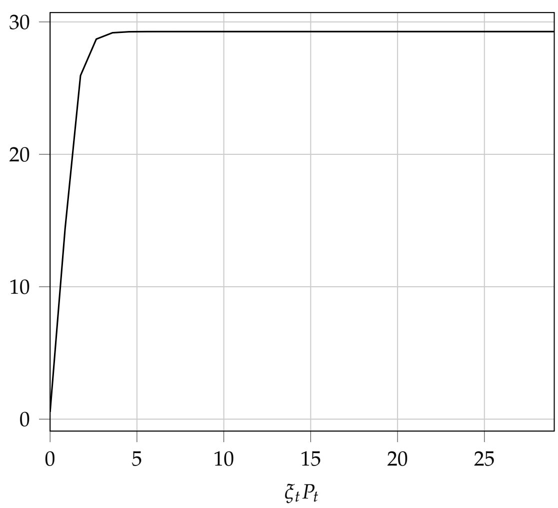
H.2 MQC with alternative parameter values

Figure H.1: MQC and $f()$ function with $\kappa = 50$, $\iota = 4$ and $\delta = 4$.

(a) MQC.



(b) $f()$ function.



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.

76

Chapter 2

Government spending and retail prices: Regional evidence from the United States

Government spending and retail prices: Regional evidence from the United States*

Rasmus Bisgaard Larsen†

Abstract

I study the effects of local government spending on local retail prices using retail scanner data from the United States. Spending shocks are identified with two sources of regional variation: spending components from the American Recovery and Reinvestment Act of 2009 and Department of Defense contracts. Estimates from both sources show that retail prices respond positively to an increase in government spending. I provide evidence, which indicate that this cannot be accounted for by changes in marginal costs. This suggests that retailers charge higher markups following an increase in government spending, which runs counter to the predictions by standard sticky-price models.

*I would like to thank discussants Lars Other and Emil Holst Partsch, and the participants of the DGPE Workshop 2019, the RGS Doctoral Conference 2019 in Bochum, and the internal macro and PhD seminars at the University of Copenhagen for helpful comments and suggestions. Researcher own analyses calculated (or derived) based in part on data from The Nielsen Company (US), LLC and marketing databases provided through the Nielsen Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the Nielsen data are those of the researcher and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

†Department of Economics, University of Copenhagen.

1 Introduction

The inflationary effects of a government spending shock play a central role in how the shock is transmitted to consumption in the textbook New Keynesian model. One transmission channel works through intertemporal substitution and can have large effects on consumption at the zero lower bound. If an increase in government spending raises (expected) inflation and the nominal interest rate is stuck at zero, the real interest rate falls and households shift consumption towards the present. This channel can overcome the negative wealth effect that tends to crowd out consumption.[1] Inflation can also affect consumption through other channels than intertemporal substitution. For example, if households have different marginal propensities to consume, inflation may also affect aggregate consumption through revaluation of nominal balance sheets and households' exposure to real interest rates (Auclert, 2019). What causes price changes – a pass-through from changes in marginal costs or variations in markups – is also important for the transmission of spending shocks (Hall, 2009).

In this paper, I analyze how changes in regional government spending in the United States affect local retail prices. Since official high-quality price indices are not available at the regional level, I construct price indices for the period of 2007 to 2017 using retail scanner data from Nielsen. Thus, the paper is focused on government spending's effect on the prices of a subset of goods purchased by households: food and non-food groceries. The product categories included in the data correspond to around 13 % of total consumption expenditures in the Consumption Expenditure Survey.[2] About two thirds of the products in the scanner data are classified as food or beverage, which receives a weight of 7-9 % in the Bureau of Labor Statistics' (BLS's) consumer prices indices.[3]

I rely on two sources of regional variation in federally financed government spending: the American Recovery and Reinvestment Act of 2009 (henceforth, the ARRA) and

[1] Christiano et al. (2011) provides a thorough exposition of this argument.

[2] Although groceries constitute a small share of total consumption, their prices can be important in forming households' inflation expectations since grocery prices are observed frequently in households' daily lives. D'Acunto et al. (2019) link household scanner data from Nielsen with a survey on inflation expectations and show that the household-level price index is a strong predictor for inflation expectations. The correlation between price changes and inflation expectations is even stronger for goods that households purchase more frequently.

[3] There is no one-to-one correspondence between the product categories in the Nielsen data to the expenditure categories defined by the BLS. The remaining third of the products in the Nielsen data are non-food groceries such as personal care products and general merchandise (e.g. batteries, electronics, office supplies, etc.).

Department of Defense (DoD) contracts. The key empirical challenge to analyzing the effects of fiscal policy is that the regional allocation of government spending is unlikely to be orthogonal to local economic conditions that also affect retail prices. Areas experiencing poor economic outcomes were prone to receiving relatively more aid under the ARRA, while the allocation of military spending is notoriously political (Nakamura and Steinsson, 2014). Depending on the nature of the shocks hitting local economies, this will bias the ordinary least squares (OLS) relationship between changes in retail prices and government spending. I handle this by identifying exogenous variation in both sources of government spending using instrumental variables (IV) strategies.

The analysis of the ARRA is done in a purely cross-state setting, and I identify cross-state variation in the ARRA following the approach by Chodorow-Reich et al. (2012), Wilson (2012) and Dupor and Mehkari (2016) among others. This approach exploits that some provisions within the ARRA legislation distributed funds based on allocation schemes that were plausibly exogenous to local economic conditions. This allows me to construct instruments that isolate spending related to these provisions of the ARRA.

For the DoD contracts, I move to a panel data analysis and aggregate the contracts to a measure of DoD spending at the level of core-based statistical areas (CBSAs). I identify annual within-CBSA variation in DoD spending with an instrument popular in regional analyses: the Bartik-type instrument.[4] Because of persistent heterogeneity in CBSAs' sensitivity to national changes in DoD spending, some CBSAs will always receive a larger share of military spending than others. When national DoD spending fluctuates, these CBSAs experience larger changes in local DoD spending irregardless of their current, local economic conditions. The Bartik instrument isolates this persistent component of changes in local DoD spending and is constructed as each CBSA's pre-sample share of DoD spending interacted with national changes in DoD spending.

The two sources of spending variation point in the same direction. Consistent with the traditional wisdom regarding the inflationary effects of government spending, I find that government spending causes a local increase in retail prices. The estimates show that an increase in ARRA spending of 1 % of GDP over two years increases retail prices by 0.6-0.9 % relative to other states over the same period. The estimates for the DoD spending are one order of magnitude smaller but still significant and qualitatively similar. They show that prices increase by 0.06 % over two years within the CBSA when DoD spending

[4] A large number of papers across many fields in economics use a Bartik-type or shift-share research design. For recent applications within macroeconomics see for example Saiz (2010), Nakamura and Steinsson (2014), Oberfield and Raval (2014), Guren et al. (2018), Auerbach et al. (2019a).

increases by 1 % of GDP. For both sources of spending, OLS estimates are lower than the IV estimates, which is consistent with an allocation of federal funds to areas that experienced relatively worse economic outcomes due to negative demand type shocks. I also find that prices eventually revert back to their trend, which is consistent with theory since areas in the United States are part of a currency union.

Prior research has typically found small and positive but not very statistically significant effects of regional fiscal shocks on prices (Canova and Pappa, 2007; Nakamura and Steinsson, 2014; Dupor et al., 2018; Auerbach et al., 2019b). None of these papers use scanner data, however, and are somewhat limited by data availability. Dupor et al. (2018) focus on estimating consumption multipliers but also look at the inflationary effects of ARRA spending using regional personal consumption expenditure (PCE) deflators and find no significant effects on prices at the metropolitan statistical area (MSA) level.[5] Auerbach et al. (2019b) use the GDP deflator as well as housing rental prices as proxies for consumer prices and identify CBSA-level shocks using DoD spending.[6] They find small effects on the GDP deflator but strong effects on rental prices. Nakamura and Steinsson (2014) take a brief look at the inflationary effects of government spending using cross-regional variations in DoD spending but only find small and insignificant positive effects. Canova and Pappa (2007) identify fiscal shocks with sign restrictions on state-level vector auto regressions, which also allows them to estimate state-specific impulse response functions. They find that the average effect of government spending on prices is positive and hump-shaped. However, the response is quite heterogeneous across states: prices even fall in some states.[7]

The literature on the national inflationary effects of government spending is larger. While two recent studies show that inflation declines when government spending in-

[5]PCE deflators are constructed using a multi-step procedure (see Bureau of Economic Analysis (2016)) and probably an imperfect measure of changes in regional prices. The BLS does not collect retail prices for all MSAs leading them to impute prices in some MSAs using prices from other MSAs. Housing cost data, on the other hand, are collected at a more granular level.

[6]The Bureau of Economic Analysis (BEA) deflate CBSA-level GDP by applying national price indices to current dollars values of CBSA-level GDP at the industry level (Bureau of Economic Analysis, 2015). Thus, changes in the regional GDP deflator will reflect changes in industry composition or national prices rather than local prices.

[7]Both Nakamura and Steinsson (2014) and Canova and Pappa (2007) rely on state-level CPI series constructed by Del Negro (1998) covering the period 1969-1995. Nakamura and Steinsson (2014) extend the series through 2006 by multiplying the US aggregate CPI with population-weighted cost of living indices from the American Chamber of Commerce Realtors Association.

creases (DAlessandro et al., 2019; Jørgensen and Ravn, 2018), the evidence is in general rather mixed as illustrated by Jørgensen and Ravn's (2018) survey. Some authors report that prices fall, others report that they increase. My findings are also related to a well-known puzzle regarding the depreciation of the real exchange rate following a fiscal expansion spending shock as documented in a number of papers (Kim and Roubini, 2008; Monacelli and Perotti, 2010; Ravn et al., 2012; Forni and Gambetti, 2016). U.S. states and CBSAs are small open economies belonging to a currency union. Hence, my results stand in contrast to the exchange rate puzzle since they can be interpreted as an appreciation of the state or CBSA-level real exchange rate (i.e. prices) relative to the rest of US following a local increase in government spending.[8]

An additional prediction by the textbook New Keynesian model is that the inflationary effects of government spending are due to a rise in firms' marginal costs when they increase production. When prices are sticky, this results in a countercyclical markup. Hall (2009) argues that this feature is essential for delivering an empirically reasonable output multiplier in New Keynesian models. I assess this prediction by investigating whether the observed increase in retail prices from DoD spending can be accounted for by rising marginal costs or not. Wholesale costs make up around three quarters of total retail costs, while labor costs make up 12-14 % according to the Census's national estimates on retailer costs from the Annual Retail Trade Statistics. Thus, variations in wholesale and labor costs are the most likely drivers of changes in marginal costs. First, I argue that spatial variations in wholesale costs are likely too small to drive changes in marginal costs. Second, I analyze if the rise in retail prices can be accounted for by higher labor costs. I find that wage growth in the retail sector does not explain the increase in retail prices. Third, I control for common changes in retail prices at the retail chain level and still find a significant, positive increase in retail prices following the spending shock. Lastly, I control for changes in the number of retailers per capita, which does not eliminate the price response either. In summary, these findings point towards a local procyclical response of markups in the retail sector to government spending.

There is not much evidence on the regional markup response to government spending shocks but my results are in line with the recent findings by Stroebel and Vavra (2019) concerning retail prices and house price movements. They document that retail prices respond positively to a house price increase and argue that this occurs because

[8] A recent study by Ferrara et al. (2019) re-investigates the exchange rate puzzle by embedding Ramey (2016)'s series on defense news shocks in a proxy-SVAR. They find that the real exchange rate appreciates following an increase in government spending. Moreover, prices rise, while consumption falls.

of positive wealth effects on homeowners, which reduce their shopping effort causing retailers to raise markups. Anderson et al. (2018) also document that gross margins in the retail sector – defined as sales minus the cost of goods sold – are higher in areas with higher household income, which they attribute to geographic differences in product assortment. These findings are rationalized in a model in which households buy a bundle products that are less substitutable as their income increases. As a result of the decrease in substitutability, retailers' market power and thereby markups also increase with income. Hence, although the economic mechanisms differ, both Stroebel and Vavra (2019) and Anderson et al. (2018) emphasize the role of market power in shaping markup fluctuations.

My findings are informative for understanding the transmission of government spending shocks since a procyclical markup is difficult to reconcile with markup variations caused solely by sticky prices.[9] This also has implications beyond just business cycle modeling. Most importantly, it suggests that government spending is not only able to affect aggregate demand but also inefficiencies associated with retailers' desired markups. In addition, a vast literature studies optimal monetary policy in the context of sticky price models that generate countercyclical movements in markups conditional on demand shocks (see e.g. the seminal work by Woodford (2003)). In the flexible price equilibrium of these models, however, markups and the degree of inefficiency stemming from them are constant. This implies that shocks to the economy shift output levels in the flexible-price and efficient equilibria to the same extent, which makes complete inflation stabilization optimal. Once markups also vary in the flexible-price equilibrium – as my results suggest they do – complete price stability is no longer optimal (Woodford, 2003).

When interpreting my findings, however, it is important to keep in mind that the estimates capture the response of local retail prices to a local government spending shock. This result cannot necessarily be translated into an aggregate closed-economy response. As is well-known in cross-sectional studies, time fixed effects absorb any common variation in policy and fundamentals that affect the dependent variable (Nakamura and Steinsson, 2018). This includes the monetary policy reaction to changes in government

[9]Evidence of procyclical markups does not imply that prices are not sticky (Stroebel and Vavra, 2019). If desired markups are procyclical, while the markup variations caused by sticky prices are countercyclical, the realized markup measured in data depends on the relative strength of these two opposing forces.

spending as well as common taxation across areas.[10]

Other economic forces are also at play at the national relative to the regional level. For example, products in the Nielsen data are often produced outside the local market, and the law of one price across US regions is probably not a bad approximation at the wholesale level. Thus, from the point of view of the individual region, wholesale goods are an import, and local economic conditions do not affect import prices. At the national level, however, government spending shocks may cause changes in wholesale prices that are passed onto consumers. Indeed, Anderson et al. (2018) show that while wholesale prices vary little spatially, they are highly procyclical nationally.

This paper is related to the literature on the regional effects of government spending and especially the literature concerning the ARRA and DoD spending. As summarized by Chodorow-Reich (2019), this literature has mostly focused on labor market and output effects. Authors tend to find large, positive effects on employment and output relative to those typically found in national-level estimates. The evidence on price effects is relatively scant, however, as mentioned above. Hence, I contribute to this literature by providing evidence on the regional effects on prices as well as markups.

My findings are also related to an emerging literature using scanner data to infer regional price dynamics such as the papers by Anderson et al. (2018), Dubé et al. (2018), Stroebel and Vavra (2019), Gagnon and López-Salido (forthcoming), Beraja et al. (2019), DellaVigna and Gentzkow (2019), Hitsch et al. (2019), Coibion et al. (2015) and Renkin et al. (2019). These papers study different types of regional shocks in the United States and also reach different conclusions regarding the extent to which local economic conditions can affect local prices. For example, Coibion et al. (2015) do not find much reaction of local price-setting to changes in the local unemployment rate, while Gagnon and López-Salido (forthcoming) find that large changes in demand associated with labor conflicts, mass population displacement, and shopping sprees around major snowstorms and hurricanes have only small effects on supermarket prices. On the other hand,

[10] The spending shocks in this paper are federal spending shocks such that the residents of an area experiencing an increase in government spending only pay a negligible fraction of it. Thus, the negative wealth effect associated with government spending due to increased taxation is absorbed by time fixed effects. Nakamura and Steinsson (2014) show, however, that shutting off the wealth effect in an open economy New Keynesian model does not affect output multipliers much. On the other hand, Farhi and Werning (2016) find that the output effects of outside-financed government spending compared to self-financed spending can become larger when there are many hand-to-mouth consumers and prices are not too flexible. Finally, closed-economy multipliers depend critically on the monetary policy reaction (Nakamura and Steinsson, 2014).

Stroebel and Vavra (2019) find rather large effects of house prices on retail prices, while Renkin et al. (2019) find that state-level increases in minimum wages are fully passed through onto retail prices.

The remainder of the paper is organized as follows. Section 2 describes the data. The results are presented in section 3, while section 4 discusses whether the results concerning DoD spending estimates are driven by changes in markups or marginal costs. Finally, I conclude in section 5.

2 Data description

I combine data from several sources to construct a panel data set of U.S. states excluding Alaska, Hawaii and District of Columbia as well as a panel of core-based statistical areas (CBSAs).

2.1 The retail price index

Regional retail price indices are constructed using Nielsen Retail Scanner Data from the Kilts Center for Marketing. The data set contains weekly pricing and quantity information from 2006 until 2017 at the product level from more than 90 retail chains across the contiguous United States (around 30,000-35,000 grocery, drug, mass merchandiser, and other stores).[11] Weekly prices are only recorded for products with positive sales within the week. The data set covers approximately 3.2 million products – both food and non-food groceries – identified by their Unique Product Code (UPC), which are grouped into slightly fewer than 1,100 product modules. Data are recorded at the point-of-sale, which can be matched with geographic identifiers for the individual stores down to the ZIP code level. This allows me to construct regional indices. Similar to Beraja et al. (2019) and Stroebel and Vavra (2019), I consider UPC-store pairs as individual goods, henceforth indexed by g, when constructing the indices.

The price index is constructed in two steps similar to the method used by Beraja et al. (2019).[12] First, I calculate chained Laspeyres price indices for each of the product

[11] Nielsen estimates that as of the end 2011 the data covered about half of all sales from food and drug stores and a third of all sales from mass merchandisers. The coverage of convenience and liquor stores is much lower.

[12] The two-step procedure is also used by the BLS and reduces the computational burden considerably since the entire Nielsen data set takes up around 6 terabytes of disk space.

modules indexed by m for each geographic area i (either a state or a CBSA). Let t denote the quarter and $y(t)$ the year that quarter t belongs to. The Laspeyres price index, $P_{m,i,t}$, for product module m in quarter t and area i is then constructed as follows:

$$P_{m,i,t} = P_{m,i,t-1} \cdot \frac{\sum_{g \in m} p_{g,i,t} q_{g,i,y(t)-1}}{\sum_{g \in m} p_{g,i,t-1} q_{g,i,y(t)-1}} \tag{2.1}$$

where $p_{g,i,t}$ is the quantity-weighted average price of the good g in the product module m sold in area i in the quarter t, and $q_{g,i,y(t)-1}$ is the quantity sold of good g in product module m in area i in the previous year, $y(t) - 1$.

I use the previous year's quantities as weights for two reasons. First, updating the weights each year implies that the basket of goods is continually updated, which takes into account that products enter and exit the market. By contrast, a fixed base index can only be calculated using goods that are sold in the base year. Second, high-frequency, chain-linked indices have a tendency to produce chain drift, which occurs due to quantity shifts around large price changes of individual goods (de Haan and van der Grient, 2011). Using yearly quantities as weight ameliorates the chain drift problem by reducing quantity fluctuations in the individual goods.

In the second step, I construct the aggregate area-level price index (henceforth, the Nielsen price index), $P_{i,t}$, for area i at quarter t as a geometric revenue-weighted average of the growth of the product module indices:

$$P_{i,t} = \prod_m \left(\frac{P_{m,i,t}}{P_{m,i,t-1}} \right)^{\frac{w_{m,i,y(t)} + w_{m,i,y(t)-1}}{2}} \tag{2.2}$$

where $w_{m,i,y(t)}$ and $w_{m,i,y(t)-1}$ are the revenue shares of module m in area i in the current and previous year respectively.

Some goods and stores do not enter the data set in all quarters. This can be result of missing sales in those quarters, UPCs entering and leaving the market, or stores opening and closing. I handle this issue by following the method by Beraja et al. (2019) and only include UPC-store pairs that are sold in all quarters of year $y(t)$ and the fourth quarter of the previous year, $y(t) - 1$, when computing $P_{m,i,t}$. Similarly, I leave out product modules that enter or exit when constructing $P_{i,t}$.

This handling of missing observations might seem unsatisfactory. Consider, for example, what happens if a regional increase in unemployment causes households in that area to stop purchasing certain high-priced goods. This will imply that the price change of those goods do not enter the price index. However, it is worthwhile stressing that variations over time in the indices are driven by location-specific price *changes*. That is, cross-regional differences in product availability, stores, sales quantities and price levels do not affect the indices unless the goods or stores exhibit different price trends (Stroebel and Vavra, 2019). However, one concern might be that households in a region substitute towards high inflation goods over time or permanently tend to consume high inflation goods relative to other regions. In both cases, this will result in an increase in the regional index relative to other regions. Although these changes in the index reflect actual changes in the cost of living, I construct two alternative indices to investigate my results' sensitivity to cross-area variations in consumption composition. In the first alternative index, the weights $q_{g,i,y(t)-1}$, $w_{m,i,y(t)-1}$ and $w_{m,i,y(t)}$ are replaced with the nationwide annual quantities and revenues of the UPC and module respectively. The second index is a fixed-base index in which the weights for quantities and revenues are fixed at their initial values.[13] Individual goods are still defined as UPC-store pairs, and the remaining calculations are unchanged for both alternative indices.

Temporary sales are prevalent in high frequency data (Kehoe and Midrigan, 2015). If a temporary sale causes households to hoard the good and not buy it in the following period, the price decrease will enter the index, while the following price increase will not, thereby systematically biasing the index downwards. Given that the index is constructed at a quarterly frequency, however, the bias is limited compared to indices constructed at a higher frequency.

Comparing the retail price index with official price indices

Table 1 presents some summary statistics for the Nielsen price index at both the state and the CBSA level. The cross-state average increase in the Nielsen price index from the first quarter of 2007 until the last quarter of 2017 was 19 percent, while the minimum and maximum increase over the same period was 13.5 and 26.5 percent, respectively. The growth of the index at the CBSA level is slightly lower on average and more dispersed. For comparison, the CPI for all goods and the CPI for food-at-home released by the BLS both grew approximately 21 percent nationally over the same period.

To assess the validity of the Nielsen index, I compare it with indices released by the BLS. The closest available index with respect to product types is the food-at-home index, which is available at the national level and for 25 metropolitan statistical areas

[13] As mentioned above, one downside of using the fixed base index is that it only includes products that are observed in all quarters of the sample period.

Table 1: Summary statistics for the Nielsen price index

	Mean	Median	S.d.	Min	Max
			State-level index		
Percentage growth in Nielsen index, 2007-17	0.190	0.188	0.026	0.135	0.265
Standard deviation of annual inflation rate, 2008-17	0.021	0.021	0.002	0.015	0.026
			CBSA-level index		
Percentage growth in Nielsen index, 2007-17	0.174	0.170	0.031	0.074	0.292
Standard deviation of annual inflation rate, 2008-17	0.018	0.019	0.003	0.010	0.027

Notes: The summary statistics cover 48 states and 352 CBSAs. The first and third lines show cross-region statistics for the growth in the Nielsen price index from the first quarter of 2007 until the last quarter of 2017. The second and fourth lines show cross-region statistics for the standard deviation of the quarter-to-quarter annual inflation rate from the first quarter of 2008 until the last quarter of 2017.

(MSAs) at different frequencies – annual, bi-annual and monthly – in the sample period of 2007-2017.

I construct the Nielsen index for the 25 MSAs as well as a national version covering the contiguous U.S. Panel (a) in figure 1 shows the national food-at-home index together with the national Nielsen index, while panel (b) plots the half-year to half-year annual growth rates from 2007 until 2017 in the food-at-home index against the same growth rates of the Nielsen index for the 25 MSAs. When comparing the two indices, we should, however, keep in mind that the food-at-home index covers a narrower set of goods.[14] The sampling and measurement error in the BLS indices at the MSA level is also substantially higher than at the national level.[15]

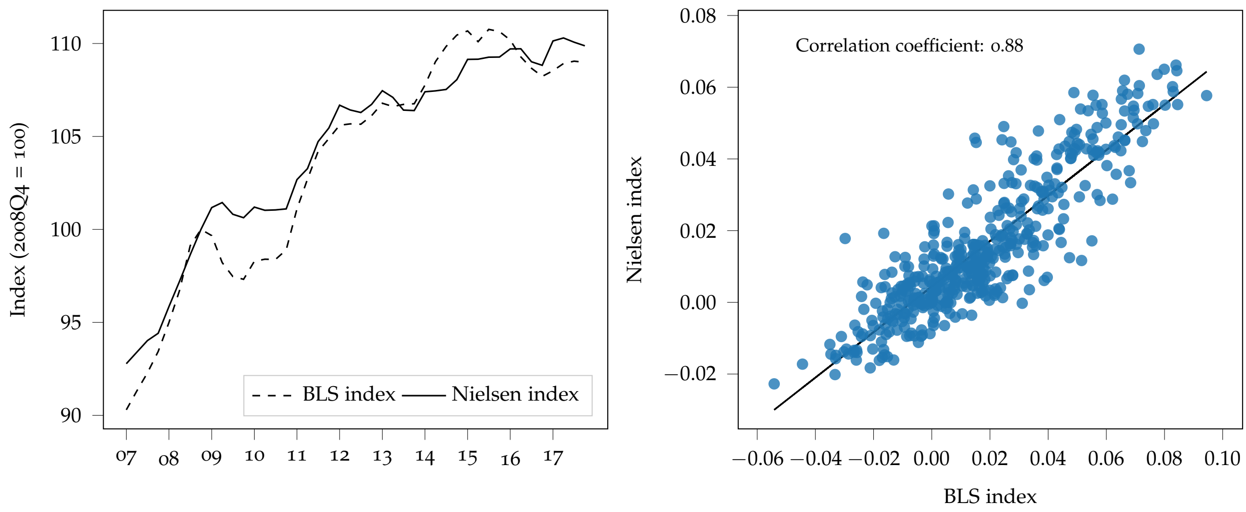
Although the national Nielsen index is less volatile than its food-at-home counterpart in panel (a), the overall behavior of the two indices is the same. Panel (b) shows that the correlation between the MSA-level growth rates in the BLS and Nielsen indices is positive with a correlation coefficient of 0.88.[16] In summary, figure 1 shows that the Nielsen index is not only able to capture the same dynamics, but also cross-regional differences, as the BLS food-at-home index.

[14]In addition to food, the Nielsen data includes non-food items sold in grocery stores (health and beauty aids, non-food groceries and general merchandise).

[15]https://www.bls.gov/cpi/questions-and-answers.htm.

[16]Stroebel and Vavra (2019) find a correlation coefficient of the same magnitude when comparing the growth of a price index constructed using retail scanner data from the company IRI Worldwide with the BLS food-at-home price index. The IRI data set covers fewer products and stores than the Nielsen data set but also a different sample period (2001-2011).

Figure 1: Nielsen index versus BLS food-at-home index



(a) National comparison

(b) MSA-level comparison

Notes: Panel (a) shows the national version of the Nielsen price index together with the BLS food-at-home price index over the period 2007-2017. The BLS index has been transformed into quarterly values by quarterly averaging the monthly series. Panel (b) plots the half-year to half-year annual growth rates in the BLS index against the same growth rates in the Nielsen index for 25 MSAs.

2.2 Government spending data

ARRA stimulus data

Detailed data on ARRA spending by state were available at the now defunct recovery.gov website, which published funding obligations and outlays by departments of the federal government in weekly Financial and Activity Reports from April 2009 until the end of 2013. Although the website is still accessible through archival websites, its functionality and data accessibility is limited. Hence, I use the series on ARRA outlays in the data set accompanying the article by Chodorow-Reich (2019). My baseline specification uses cumulative ARRA outlays through the end of 2010 normalized by GDP in 2008.

ARRA instruments

The data used to construct the three ARRA instruments come from multiple sources.

I use cumulative Department of Transportation (DoT) obligations under the ARRA until the end of 2010 provided in the data set accompanying the article by Wilson (2012). This captures most of the DoTs ARRA expenditures since around 90% of all DoT spending was obligated at this point (Wilson, 2012). Data on miles of federal highway, vehicle-miles traveled on federal-aid highways, payments into the federal highway trust fund and Federal Highway Administration (FHWA) obligation limitations are from the FHWA's publication Highway Statistics.

Medicaid spending in 2007 is from the publication Data Compendium published by the Centers for Medicare & Medicaid Services.

The Dupor and Mehkari (2016) narrative instrument was collected from the data set accompanying the article by Chodorow-Reich (2019). This variable differs from the original variable by Dupor and Mehkari (2016) since Chodorow-Reich (2019) uses agency-reported spending instead of recipient-reported spending when creating the variable.

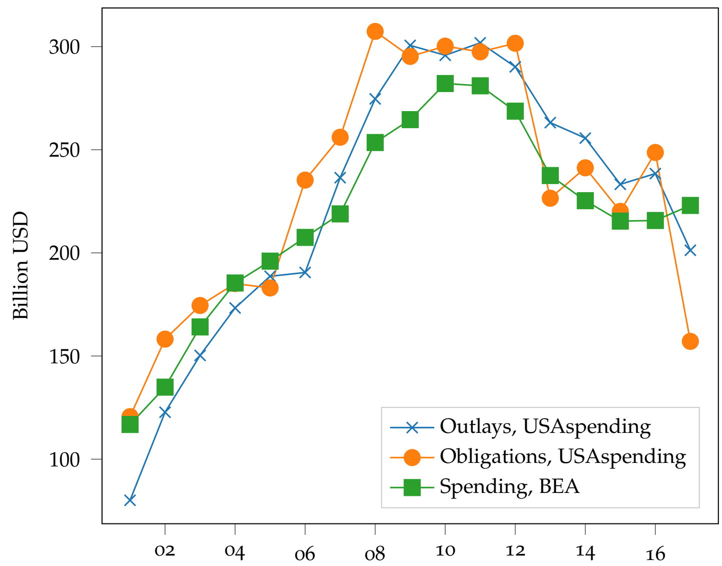
Department of Defense data

I collected data from USAspending.gov covering all DoD prime contracts signed from October 2000 through 2017. The website contains information on the primary place of work performance (ZIP code) and industry classification of the contractor as well as the duration and total dollar amount obligated in the contract. The data set also contains terminated contracts (de-obligated amounts).

The data source has also been used by Demyanyk et al. (2018) and Auerbach et al. (2019a), and I follow their approach to clean the raw data. First, I remove terminated contracts by matching a terminated contract with the original contract if a de-obligated amount of a contract falls within 0.5 % of another contract and both contracts have the same contractor ID and zip code. If this is the case, I remove both contracts. This removes 3.9 % of the contracts from the data. Second, I remove contracts that terminate after 2021, which removes an additional 0.1 % of the contracts. Although I use changes in obligations in the analysis, I also construct a proxy for actual outlays by dividing the obligated amount evenly among the months of the contract (e.g. a contract of $240,000 running in the first half year of 2009 will result in $40,000 being outlaid to each month from January through June).

Figure 2 evaluates the data by plotting national obligations and spending constructed using the USAspending data as described above together with intermediate goods and services purchased for national defense from BEA's NIPA tables. The two USAspending series roughly match the BEA data in terms of both magnitude and yearly movements.

Figure 2: Military spending according to USAspending and BEA data



Notes: The blue and yellow lines are annual outlays and obligations constructed using USAspending.gov data. The green line is "Intermediate goods and services purchased" in the BEA's NIPA Table 3.11.5, "National Defense Consumption Expenditures and Gross Investment by Type."

2.3 Additional data

Employment and wage data are from the BLS. I use non-farm employees from the Current Employment Statistics as the measure of employment in the analysis of the ARRA. Wage, employment and establishment data used in the analysis of DoD spending are from the Quarterly Census of Employment and Wages. The two-digit employment shares used in section 3.2 are from the Census' County Business Patterns.

GDP, population and personal income data are from the Regional Economic Accounts produced by the BEA. Population density data are from the Census.

Several house prices indices exist. I use the Federal Housing Finance Agency's all-transactions house price index for single-family homes in the analysis of the ARRA. The instruments for regional house price sensitivities by Gyourko et al. (2008), Saiz (2010) and Guren et al. (2018) used in the analysis of DoD spending are from the supplementary data sets to these authors' articles.

The estimate of state-level tax benefits under the ARRA are from the data set accompanying the article by Wilson (2012). The estimated tax benefits include tax benefits from the Making Work Pay tax cuts and the tax benefits received from the ARRAs increase of the threshold at which the Alternative Minimum Tax becomes binding. The tax benefits are normalized by population as of December 2008.

3 Empirical analysis

The main challenge to identifying the effects of regional government spending shocks is that spending is disproportionately directed towards regions that experience economic downturns. This will bias OLS estimates from a regression of regional price changes on changes in regional government spending. Furthermore, it is not straightforward whether the bias is positive or negative since a shock causing a drop in employment or output can cause retail prices to move in either direction depending on the nature of the shock.[17]

Other authors have documented a relationship between local economic conditions and retail prices. Beraja et al. (2019) show how state-level demand and supply shocks induced changes in state-level retail prices during the Great Recession, while Stroebel and Vavra (2019) document that local house price movements cause changes in local retail prices. Their findings imply that spending directed towards areas with relatively higher unemployment rates and larger drops in house prices would bias the estimates from a regression of retail price changes on changes in local government spending. Thus, I rely on two complementary IV approaches to solve the identification issue.

The first approach uses cross-state variation in federal spending in the ARRA. Since stimulus was especially directed toward states, which were hit the hardest by the recession, the geographical allocation was endogenous. I handle this issue by exploiting provisions within the ARRA legislation that generated cross-state variation in the allocation of federal spending that were exogenous to local economic outcomes. However, a disadvantage of this approach is that it solely relies on cross-state variation since each state was only exposed to one shock.

The second approach uses CBSA-level variation in DoD contracts as a measure of government spending. Contrary to the ARRA identification method, the DoD approach offers time series variation, which allows me to construct a panel data set and analyze within-CBSA variation in government spending. Additionally, the cross-sectional dimen-

[17]Consider a negative supply shock. This reduces employment but increases prices, thereby biasing the OLS estimate of the price response upward. Conversely, a negative demand shock would bias the OLS estimate of the price response downward. When it comes to estimating the output or employment effects of government spending, however, the bias should be negative in both cases since output and employment move in the same direction for both supply and demand shocks.

sion of the data is much larger. As in the case with ARRA spending, there is reason to believe that the regional allocation of DoD contracts is endogenous. I rely on a Bartik-style instrument measuring the CBSA-level exposure to national DoD spending movements to isolate movements in DoD spending that are orthogonal to changes in local economic conditions.

3.1 Evidence from the American Recovery and Reinvestment Act

I begin by relating price changes to a state-level government spending shock induced by the ARRA. The overall scope and elements of this stimulus package were proposed by then President-Elect Barack Obama in December 2008 amid concerns that the economy was sliding into a prolonged and severe recession. The act was enacted shortly thereafter in February 2009 and consisted of spending, transfers and tax cuts of around $800 billion. I focus on the spending component, which was concentrated in the first couple of years of the program with around three quarters of total spending outlaid by the end of 2010 (Chodorow-Reich, 2019). Although the ARRA offers limited time series variation, it has been analyzed extensively in the literature estimating regional government spending multipliers.

The estimate of the price response is obtained from the following regression for the quarter t response of retail prices to a government spending shock, G_i, in state i occurring in the first quarter of 2009:

$$\pi_{i,t} = \alpha_t + \beta_t X_i + \gamma_t G_i + \epsilon_{i,t} \qquad \text{for } t = 1, 2, \ldots, T \tag{3.1}$$

where $\pi_{i,t}$ is the growth of the log-detrended Nielsen price index for state i from the last quarter of 2008 to period t, G_i is cumulative ARRA-related spending through the end of 2010 normalized by GDP in 2008, and X_i is vector of controls observed in periods before the ARRA was passed. α_t are time fixed effects that capture the impact in quarter t of common fundamentals and policy on all states such as the national business cycle or monetary policy.

I construct the log-detrended Nielsen price index for state i as the exponential function of the residuals from the regression $\ln p_{i,t} = \tilde{\alpha}_i + \tilde{\gamma}_i \cdot t + \tilde{\epsilon}_{i,t}$. This allows for a state-specific trend in the price index. As shown in appendix figure A.1, there is some heterogeneity in the trend estimates with annual growth rates in the trend ranging from 1 to 2 %

The parameters $\Gamma = (\gamma_1, \gamma_2, \ldots, \gamma_T)$ are the object of interest: it is the impulse response function of retail prices to the spending shock, G_i, in one state relative to other states. Each state is only exposed to one spending shock occurring simultaneously across states. Therefore, the estimate of Γ is driven by the cross-state evolution of prices, not the within-state price changes following multiple shocks. Additionally, the stimulus shock, G_i, is cumulative ARRA spending through 2010 so the shock can be interpreted as a news shock concerning ARRA spending directed towards a state in the first two years of the program. Although this ignores the evolution of actual spending, capturing the news about future government spending rather than the actual spending itself is important for inferring the effects of government spending (Ramey, 2011).

I deal with the endogeneity issue by instrumenting for G_i with three instruments: Medicaid spending in 2007, a measure of exogenous ARRA spending by the Department of Transportation (DoT), and a narrative instrument for exogenous ARRA spending. These instruments have previously been used by Chodorow-Reich et al. (2012), Wilson (2012) and Dupor and Mehkari (2016) among others to estimate employment multiplier effects of the ARRA. Chodorow-Reich (2019) also combine these three instruments to identify the ARRA's effect on employment and output. The common idea behind the instruments is that they rely on provisions within the ARRA that generated cross-state variation in the allocation of federal spending that were exogenous to local economic outcomes. Together they constitute about $123.5 billions of the around $800 billions of funds contained in the ARRA.

The Medicaid instrument

The Medicaid spending instrument by Chodorow-Reich et al. (2012) exploits that the ARRA set up an $87 billion State Fiscal Relief Fund to aid states in paying Medicaid expenses. This was one of the first ARRA components to affect the economy with aid payments flowing to states already by March 2009. Before the ARRA was passed, the federal government reimbursed 50 to 83 % of states' Medicaid expenses according to state-specific reimbursement rates (the Federal Medical Assistance Percentages, or FMAP). The reimbursement scheme is set up such that poorer states – which tend to have larger Medicaid expenses – have a higher FMAP. Each state's rate was recalculated every fiscal year based on a three-year trailing average of the state's per capita personal income relative to the national average. With the passage of the ARRA, states' FMAP could not decrease from their 2008 level, and all states' FMAP was raised by 6.2 percentage points.

Thus, states with larger Medicaid expenses prior to the ARRA received a larger transfer from the State Fiscal Relief Fund, and Medicaid expenses in 2007 can be used as an instrument for this transfer. Although state governments could use the transfer for any purpose, Chodorow-Reich et al. (2012) argue that states did not retain the transfer on their budgets.[18]

The DoT instrument

Around three quarters of the DoT's $40 billion ARRA funding was allocated to the Federal Highway Administration (FHWA). 50 % of these funds were divided among states based on a pre-existing allocation formula, which is a weighted average of three factors measured with a three-year lag: miles of federal-aid highway, vehicle-miles traveled on federal-aid highways, and payments into the federal highway trust fund. The remaining 50 % of the funds were allocated in proportion to the 2008 FHWA obligation limitations on funding for each state's Federal-Aid Highway Programs. Hence, an instrument can be constructed by taking the fitted values from a regression of total DoT ARRA funding on the miles of federal-aid highway in 2006, estimated vehicle-miles traveled on federal-aid highways in 2006, estimated payments into the federal highway trust fund in 2006, and FHWA obligation limitations in 2008.[19]

Dupor and Mehkari instrument

Dupor and Mehkari (2016) take the logic of the two other instruments to the extreme by classifying components of the ARRA that were exogenous to local economic conditions using a narrative approach similar to that of Romer and Romer (2010) and Ramey (2011). They found these components by reading through the provisions of the ARRA as well as the federal codes and regulations used to allocate stimulus. If the criteria for allocation of funds were plausibly exogenous to the local business cycle, they deemed the funds as exogenous. For example, one exogenous component is around $6 billion of funds that was authorized to urbanized areas (UZAs) to improve public transit capital under the Capital Transit Assistance program. UZAs with a population between 50,000 and

[18]Additionally, the positive employment effects estimated by Chodorow-Reich et al. (2012) are concentrated in sectors relying on state funds – the state and local government, health and education sectors – which suggest that the transfers were used to avoid state program cuts.

[19]Standard errors in the first and second-stage regressions should in principle be adjusted for the use of fitted values as an instrument. However, I do not since the four factors explain almost all of the variation in DoT spending.

199,999 persons received funds based on their population and population density, while UZAs with a higher population received funds according to a number of factors such as bus passenger miles and bus revenue vehicle miles. After adding up all exogenous components, the instrument identifies $21.5 billion by the end of 2010.

I normalize these three instruments by GDP to match the normalization of the dependent variable, G_i, and collect them into a vector, Z_i. The first-stage regression is then given by:

$$G_i = \kappa + \lambda Z_i + \delta X_i + \mu_i \tag{3.2}$$

The exclusion restriction implies that, conditional on the controls, the instruments should only affect retail prices through government spending:

$$E\left[Z_i \epsilon_{i,t} | X_i\right] = 0 \qquad \text{for } t = 1, 2, \ldots, T \tag{3.3}$$

Not only should contemporaneous observed shocks to retail prices be independent of the instruments at the passage of the ARRA but all subsequent shocks should as well. As highlighted by Chodorow-Reich (2019), this might be an unrealistic assumption as the estimation horizon grows since additional spending, transfers and tax reductions of $709 billion were enacted after the ARRA's passage. Most of the outlays under these additional fiscal measures happened in 2011 and 2012, and some directly expanded ARRA features such as the Medicaid relief component, which accounted for 12 % of the additional fiscal support (Council of Economic Advisers, 2014).[20] In principle, this invalidates the exclusion restriction for all t and biases γ_t if agents could anticipate the spatial distribution of the additional fiscal support before its enactment. However, if the spatial distribution was unanticipated by the agents of the economy, the estimates of γ_t before 2011 are still unbiased. After 2011, it is not clear in which way the bias works. If states that already received relatively more spending according to the instruments also received more additional stimulus as was the case for the Medicaid relief component, the bias should have the same sign as γ_t before 2011. If states that initially received relatively little spending subsequently received more additional stimulus, the bias should have the opposite sign.

Controls, X_i, are included to either help predict retail prices after 2008 or account for particular features of the ARRA that might be a threat to identification. I include growth

[20] Around half of the additional fiscal support consisted of tax cuts and incentives, which should not bias the estimates by itself if the cross-state impact of these cuts and incentives was independent of the instruments. Aid to directly impacted individuals and public investment outlays each made up about 20 % of the subsequent fiscal support.

in retail prices from the first quarter of 2007 until the last quarter of 2008 to control for pre-ARRA price growth. I also control for economic conditions by including GDP growth from 2007 to 2008, employment in December 2008 as well as the employment change from the peak of the pre-ARRA national business cycle in December 2007 until December 2008.[21] In addition, I control for house price growth between 2003 and 2007 due to the importance of house prices in explaining regional business cycles during the 2000s as highlighted by Mian et al. (2013) and Mian and Sufi (2014) among others together with link from house prices to retail prices documented by Stroebel and Vavra (2019). Lastly, I control for two specific features of the ARRA following Wilson (2012). First, I include his estimates of the ARRA household tax benefits normalized by population as of December 2008. Second, I include the change from 2005 to 2006 in the three-year trailing average of personal income per capita. This variable is included since a state's FMAP could increase beyond the common increase of 6.2 percentage points in 2009 if the three-year trailing average of personal income per capita decreased from 2005 to 2006.

The first-stage estimates from equation (3.2) are presented in appendix table A.1 in which the four columns show the estimates from different specifications of the regression. The Medicaid and narrative instruments are positively related to ARRA spending and each of them enter the regression significantly. The DoT instrument is positively related to ARRA spending although the relationship is not very significant. To assess the strength of the instruments, I calculate the effective F-statistic by Montiel Olea and Pflueger (2013), which is robust to heteroskedastic errors when using multiple instruments. With 3 instruments and 1 endogenous variable, Stock and Yogo (2005) provide a value of 12.83 for the F-statistic below which the instruments could be weak. The F-statistics shown at the bottom of the table are well above this value and in the range of 57-89. Hence, there does not seem to be any weak instrument issues, which would bias the IV estimator towards OLS.

The estimates for the coefficients on the control variables hint at spending being directed toward states that were hit harder by the recession. The employment level and change enter negatively, while pre-ARRA retail price growth enters the regression negatively but not very significantly. To get a sense of the geographical distribution of spending, the heat map of the United States in appendix figure A.2 shows predicted spending according to the first-stage regression without any controls. Predicted spending does show some regional concentration. States in the New England region as well as North-

[21] Both employment variables are normalized by population in December 2008.

Table 2: The effect of ARRA spending on retail prices

	Dependent variable: Price growth from 2008Q4 to 2010Q4							
	(1)		(2)		(3)		(4)	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
ARRA spending normalized by GDP	0.470*	0.769***	0.403	0.597**	0.477	0.696**	0.666**	0.924***
	(0.263)	(0.267)	(0.296)	(0.265)	(0.306)	(0.275)	(0.313)	(0.287)
Emp. rate, Dec. 2008			−0.006	0.006	0.009	0.024	0.019	0.036
			(0.044)	(0.043)	(0.043)	(0.040)	(0.043)	(0.040)
Emp. change, Dec. 2007-08			0.232**	0.217**	0.282**	0.273**	0.287**	0.281**
			(0.096)	(0.095)	(0.125)	(0.128)	(0.120)	(0.123)
GDP growth, 2007-08			−0.046	−0.049	−0.064*	−0.070*	−0.073*	−0.082**
			(0.030)	(0.032)	(0.038)	(0.039)	(0.040)	(0.042)
House price growth, 2003-07					0.008	0.009	0.003	0.004
					(0.009)	(0.009)	(0.009)	(0.009)
Retail price growth, 2007-08					−0.088	−0.098	−0.042	−0.044
					(0.114)	(0.117)	(0.111)	(0.115)
Tax benefits per capita							0.009*	0.010**
							(0.005)	(0.005)
Change in avg. personal income, 2005-06							0.147**	0.180**
							(0.075)	(0.083)
Constant	−0.038***	−0.044***	−0.026	−0.042	−0.037	−0.056	−0.069	−0.094**
	(0.006)	(0.006)	(0.045)	(0.042)	(0.044)	(0.041)	(0.046)	(0.043)
RMSE × 100	0.891	0.903	0.838	0.842	0.825	0.830	0.799	0.806
p-value of Hansen *J*-statistic	-	0.386	-	0.195	-	0.127	-	0.291
N	48	48	48	48	48	48	48	48

Notes: The table presents the OLS and IV estimates from the second-stage regression $\pi_i = \alpha + \beta X_i + \gamma G_i + \epsilon_i$. π_i is the growth of the log-detrended Nielsen price index from the fourth quarter of 2008 to the fourth quarter of 2010, G_i are cumulative ARRA outlays through 2010 normalized by annual GDP of 2008 and X_i is a vector of controls. The endogenous variable, G_i, has been instrumented for using the following three instruments: Medicaid spending in 2007, the DoT instrument, and the Dupor and Mehkari (2016) narrative instrument. All instruments are normalized by GDP. Heteroskedasticity-robust standard errors are shown in parentheses. ***, ** and * denote significance at the 0.01, 0.05 and 0.1 level respectively.

and Midwestern states (e.g. Montana and the Dakotas) tended to receive more stimulus, while states on the west coast received less stimulus on average.

ARRA estimates

Table 2 shows the OLS and IV estimates from the second-stage regression (3.1) in the fourth quarter of 2010 using four different specifications. As explained above, this is the last quarter for which the estimates are very unlikely to be biased.

The IV estimates for the effect of ARRA spending on retail prices, γ_t, range from 0.60 to 0.92. Expressed differently, an increase of ARRA spending of 1 % of initial GDP relative to other states caused a relative increase in retail prices of 0.60-0.92 % over the next two years. The estimate is relatively stable across the four different models presented, while the p-values for the Hansen (1982) J-statistic testing the overidentifying restrictions are above conventional significance values. Thus, the J-statistics fail to reject the null of exogeneity of the instruments, which strengthens the validity of the exclusion restriction.

The OLS estimates for the effect of the ARRA on prices are lower than their IV counterparts for all four specifications. A similar downward bias of OLS estimates for ARRA employment multipliers has been found by Wilson (2012), Dupor and Mehkari (2016) and Chodorow-Reich (2019) among others. It also suggests that spending was directed disproportionately toward states experiencing poor economic outcomes, which were driven primarily by negative demand shocks.

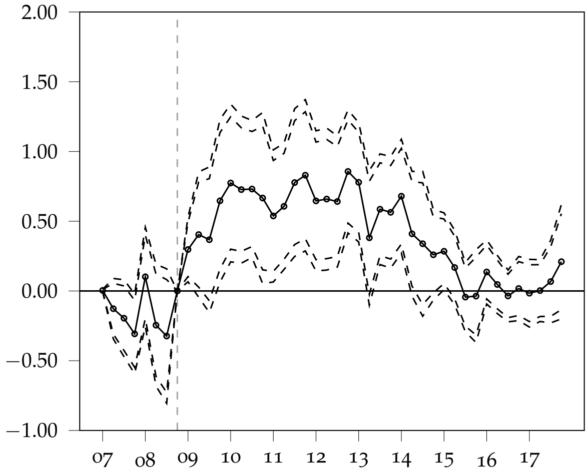
There is no clear relationship between economic outcomes prior to the ARRA and subsequent retail price growth after controlling for ARRA spending. Pre-ARRA GDP growth enters the regression negatively, while the coefficient on year-to-year change in employment is positive. The coefficient on retail price growth before the ARRA is slightly negative but not estimated with much precision.

The point estimates in table 2 only show the effect of the spending shock after two years. I now turn to the dynamic response of prices over the entire sample period of 2007-2017, which are shown in figure 3. OLS estimates are shown in panel (a), while panel (b) plots the IV estimates. Their 90 and 95 % confidence bands are based on heteroskedasticity-robust standard errors and indicated by dashed lines. As explained above, I believe that the estimates may not be unbiased after 2010 but I nonetheless present the estimates for the post-2010 period.

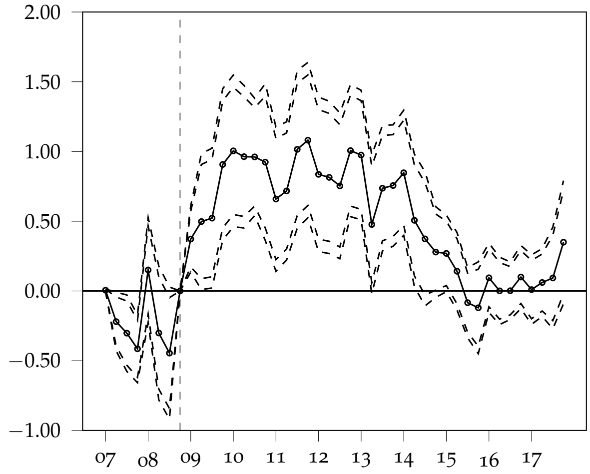
The estimates in figure 3 show that local ARRA spending caused an increase in local retail prices relative to other states. The point estimates increase to around 1 after 1 year following the spending shock and revert toward zero again about 4 years after the spending shock. Similar to the estimates reported in table 2, the OLS estimates are lower than the IV estimates for the period after the ARRA. In periods prior to the ARRA, the estimates are close to zero, which indicates that there is no correlation between pre-ARRA price movements and ARRA spending conditional on the control variables.

Although the price index eventually reverts back to trend, it takes 6 years so the response is rather long-lived. However, the estimates after 2010 might be biased by the additional fiscal measures enacted after the ARRA and as discussed above, it is not clear in which direction the bias works. I have tested the overidentifying restriction quarter-by-quarter for all estimates of γ_t using the Hansen (1982) $J-$statistic. While the statistics'

Figure 3: The dynamic effect of ARRA spending on retail prices



(a) OLS estimates



(b) IV estimates

Notes: The solid lines show the estimates of $\Gamma = (\gamma_1, \gamma_2, \ldots, \gamma_T)$ from the regression $\pi_{i,t} = \alpha_t + \beta_t X_i + \gamma_t G_i + \epsilon_{i,t}$ for $t = 1, 2, \ldots, T$. $\pi_{i,t}$ is the growth of the log-detrended Nielsen price index relative to the fourth quarter of 2008, G_i is cumulative ARRA outlays through 2010 normalized by GDP in 2008 and X_i is the vector of controls used in column 7-8 in table 2. The endogenous variable, G_i, has been instrumented for using Medicaid spending in 2007, the DoT instrument, and the Dupor and Mehkari (2016) narrative instrument. All instruments are normalized by GDP. The left panel shows the IV estimates, while the OLS estimates are shown in the right panel. Dashed lines indicate 95 and 90 % confidence bands calculated using heteroskedasticity-robust standard errors. Vertical lines indicate the enactment of the ARRA.

p-values are above any conventional value before 2011, they occasionally drop below 0.1 thereafter. This indicates that the impulse response estimates after 2010 are indeed biased. Nonetheless, other authors have found very persistent effect of shocks on relative prices in the US. For example, Cecchetti et al. (2002) use CPI data for 19 US cities over the period of 1918-1995 to show that the half-life of price level convergence between the cities is around 9 years. Canova and Pappa (2007) also find a hump-shaped response of the price level to a state-financed government spending shock, which first becomes insignificant at the 68 % level after 6 years.

When interpreting the impulse response estimates at a longer horizon, we should also keep the limitations of the econometric framework in mind. Since the regression does not include any post-ARRA explanatory variables, the error term will tend to accumulate over the horizon and widen the confidence band. This also induces a high amount of autocorrelation in the impulse response estimates. I have explored this feature using a bootstrap, and the correlation is especially present for estimates closer to each other in time.[22] Hence, the confidence bands reported in figure 3 likely understate the uncertainty regarding the possible trajectories for the long run impulse response function.

Robustness

I have explored the robustness of the estimates. The results are reported in Appendix B.

First, I have explored whether or not states that received more stimulus also had higher inflation rates before the enactment of the ARRA using the CPI index constructed by Nakamura and Steinsson (2014). Although this would not necessarily invalidate the exclusion restriction, it could suggest that the results are spurious and reflect time-invariant inflation rate differentials not picked up by the control variables. Appendix figure B.1 shows that there is no systematic relationship between inflation dynamics during the period of 1971-2006 and ARRA spending.

Second, I have subjected the baseline estimates in table 2 to various robustness checks. These are shown in table B.1 and figure B.2. Rows 2 and 3 show the estimates when normalizing ARRA spending by population or personal income instead of GDP. The estimates are also positive and significant when using these alternative normalizations. Rows 4-6 show the estimates when only using one instrument at a time. Row 7 includes the squared and cubed pre-ARRA trend in retail prices, while row 8 includes bi-annual inflation rates in the years 2000 through 2006 based on the Nakamura and Steinsson (2014) CPI data. Neither affect the estimates much. Row 9 replaces the price index with the alternative price index in which all UPCs and product modules receive a weight corresponding to the annual national quantities and revenues (that is, weights are identical across states, not time) to investigate if the baseline estimates reflect product-switching toward high inflation products. The IV estimate is slightly lower. Row 10 presents estimates when using a fixed-base Laspeyres price index in which weights are fixed at 2007 quantities and revenues. This almost doubles the price response but prices still converge back to trend as seen in panel (j) of appendix figure B.2. Finally, I include fixed effects for the 4 Census regions in row 11. Although the point estimate for the last quarter of 2010 becomes insignificant, the dynamic response of prices is qualitatively similar to the baseline estimates and significant in some periods as shown in panel (k) of appendix figure B.2.

Third, I show an added variable plot for the price growth from the fourth quarter of 2008 to the fourth quarter of 2010 against ARRA spending in figure B.3. The figure

[22] A heatmap of the correlation coefficients are shown appendix figure A.3.

shows no indications of outlier-driven estimates.

3.2 Evidence from Department of Defense spending shocks

The analysis of the ARRA in the previous section looked at a spending shock that only varies across states and not over time. Thus, the estimates were driven by the behavior of retail prices across states when they were exposed to different spending shocks. I now turn to a panel data setting, which has the benefit of allowing me to control for the average growth of retail prices of each CBSA and estimate the response of prices *within* CBSAs as they are exposed to government spending shocks.[23]

I use changes in military spending as a source of variation in government spending. This approach has been used extensively in the literature estimating national multipliers going back to Ramey and Shapiro (1998) and applied to a regional setting by Nakamura and Steinsson (2014), Demyanyk et al. (2018), Auerbach et al. (2019a), Auerbach et al. (2019b) and Auerbach et al. (2020) among others. Of course, military spending has little direct effect on the retail sector. Only 0.2 % of the obligations in the DoD contract data go to food and beverage stores. Instead, the bulk of military spending goes towards manufacturing and professional, scientific, and technical services. Thus, the estimates do not capture direct demand effects on the retail sector but instead indirect effects such as Keynesian income multiplier effects or factor demand effects.

I use the following regression to estimate the h-years ahead response of retail prices to a change in government spending over the same horizon:[24]

$$\pi_{i,t+h} = \alpha_{i,h} + \eta_{t+h} + \beta_h \frac{\sum_{k=1}^{h} (G_{i,t+k} - G_{i,t})}{Y_{i,t}} + \varepsilon_{i,t+h} \tag{3.4}$$

The dependent variable, $\pi_{i,t+h}$, is retail price inflation in CBSA i from year t to year $t+h$, while $\frac{\sum_{k=1}^{h}(G_{i,t+k}-G_{i,t})}{Y_{i,t}}$ is the cumulative change in military spending over the same period normalized by GDP of the CBSA in year t. To be clear, the military spending variable is measured using obligations, not the outlay proxy described in section 2, since this should capture anticipation effects following the arguments by Ramey (2011). The regression is equivalent to the regression used in the analysis of the ARRA except that I include year fixed effects, η_{t+h}, to account for national inflation trends, while CBSA

[23] The analysis is done at the CBSA level instead of the state level to get more statistical power.

[24] I experimented with using a quarterly instead of yearly specification. However, there is significant seasonality in the DoD data, which leads me to prefer the yearly specification.

fixed effects, $\alpha_{i,h}$, controls for the CBSA-specific horizon h inflation trend over the sample period.

β_h is an estimate of the percentage change in retail prices over h years as a result of a cumulative change in military spending over the same period of 1 % of initial GDP. However, due to the political nature of military spending, there is reason to believe that it flows disproportionately towards areas that experience economic downturns, which would bias this estimate. I handle this issue by using an instrument building on the Bartik (1991) intuition: the change in CBSA-level military spending is instrumented by the national change in military spending over the same period interacted with the CBSA's average share of national military spending over the pre-sample period.[25] The first-stage of the IV regression is then given by

$$\frac{\sum_{k=1}^{h}\left(G_{i,t+k}-G_{i,t}\right)}{Y_{i,t}} = \tilde{\alpha}_{i,h} + \tilde{\eta}_{t+h} + \tilde{\beta}_h s_i \times \frac{\sum_{k=1}^{h}\left(G_{t+k}^{nat}-G_{t}^{nat}\right)}{Y_{i,t}} + \mu_{i,t+h} \tag{3.5}$$

where G_t^{nat} is national military spending in year t and s_i is CBSA i's average annual share of national military obligations in the pre-sample period of 2002-2006.

The exclusion restriction for the instrument requires that national changes in military spending interacted with the pre-sample CBSA share of military spending only affects retail prices through its impact on changes in military spending in the CBSA. In other words, the CBSAs are exposed differently to national changes in military spending. The Bartik instrument identifies the effect of local effect of military spending on retail prices by assuming that retail price changes are only affected by the differential exposure through its effect on changes in local military spending. This exclusion restriction is weaker than the restriction assumed in studies of national government spending, where military spending needs to be exogenous to the national business cycle (Nakamura and Steinsson, 2014).

As mentioned by Nakamura and Steinsson (2014), a threat to exclusion restriction would be if the federal government increased national DoD spending because CBASs that have previously received many DoD contracts experienced poor economic outcomes. Another threat to identification is if changes in national military spending are correlated

[25]Nakamura and Steinsson (2014) use an alternative approach, where the first-stage regression is local changes in DoD spending regressed on the national changes in DoD spending interacted with a regional dummy. This constructs an instrument for each region and is equivalent to instrumenting using historical sensitivities to DoD spending. Given the relatively short panel I use and the many instruments the Nakamura and Steinsson (2014) approach would produce – one for each CBSA – I use the simpler Bartik-approach instead.

with some unobserved aggregate factor in the time series *and* that the CBSAs' exposure to national military spending is also correlated with their exposure to the aggregate factor. For example, a trade shock might affect the local price movements of CBSAs differently because of differences in industry composition. Again, this is only a threat to identification if exposure to the trade shock is correlated with the DoD spending shares as well as movements in national DoD spending.[26]

Although these threats to identification are not directly testable, I have regressed the DoD spending shares used to construct the instrument on various CBSA characteristics that could independently influence retail movements through correlation with the error term in regression (3.4). The coefficients and R^2 statistics from separate regressions using standardized variables are reported in table 3.

Table 3: Correlations with Department of Defense obligations shares

Variable	Estimate		R^2	CBSAs
Wharton Regulation Index	0.090**	(0.043)	0.017	258
Saiz (2010) instrument	−0.158***	(0.044)	0.053	258
Guren et al. (2018) instrument	0.196**	(0.083)	0.017	373
Population density, weighted	0.361**	(0.103)	0.147	828
Grocery stores per capita	0.041*	(0.023)	0.002	865
Two-digit industry employment shares, 2006	–		0.170	865
Product-quarter obs. per store	0.126***	(0.049)	0.015	865
Nielsen stores per capita	−0.047**	(0.022)	0.002	865
Product-quarter obs. per capita	0.036*	(0.02)	0.001	865

Notes: The table presents estimates from separate cross-sectional regressions of the average annual share of DoD obligations during 2002-2006 on CBSA characteristics. All variables are standardized by their standard deviation. The Wharton Regulation Index by Gyourko et al. (2008), the Saiz (2010) instrument and the Guren et al. (2018) instrument are from the supplementary data sets to their articles. Population density is the Census's population-weighted density as of 2000, while grocery store data is from the QCEW. Variables in the last three rows are constructed using Nielsen data and average population over the sample period from Census. Heteroskedasticity-robust standard errors are shown in parentheses. ***, ** and * denote significance at the 0.01, 0.05 and 0.1 level respectively.

The first three rows in table 3 show the coefficients from regressions of the spending share on three measures of exposure to aggregate house price movements: the Wharton Regulation Index, the Saiz (2010) instrument, and the Bartik-like instrument by Guren et al. (2018).[27] Even though the coefficients are statistically significant, the R^2 values are

[26] This concern echoes the message of Goldsmith-Pinkham et al. (2018) who show that Bartik instruments are equivalent to using a GMM estimator with shares as instruments.

[27] The Wharton Regulation Index by Gyourko et al. (2008) is based on a survey on regulation of residential building, the Saiz (2010) instrument is mainly based on land unavailability, and the Guren et al. (2018) instrument is constructed using historical sensitivities to regional house prices.

all quite low (0.02-0.05) so the three variables explain very little of the variation in the spending shares. Thus, differential exposure to house price movements does not seem to be systematically related to exposure to aggregate changes in DoD spending.

One might worry that retailer competition varies systematically with spending share, which can influence price movements. I use population density and the number of grocery stores per capita as crude proxies for factors that affect local retailer competition. The results from separate regressions of the spending share on average number of grocery stores per capita in the years 2007 through 2017 and population density according to the Census's 2000 estimates are shown in rows 4 and 5.[28] Both of these variables are positively correlated with the spending shares. However, grocery stores per capita explain almost none of the variation in spending shares (the R^2 is 0.002). On the contrary, the R^2 for population density is 0.147 so this factor does explain a non-negligible part of the variation in the spending shares. I address this below in a robustness check and show that my results are not driven by differential price fluctuations between high or low population density CBSAs.

Row 6 shows the R^2 from a regression of spending shares on two-digit NAICS industry employment shares in 2006. These shares explain 17 % of the variation in spending shares, which is primarily due to the two industries information and professional, scientific, and technical services. I show below that my results are robust to controlling for variation associated with industry composition in the cross-section.

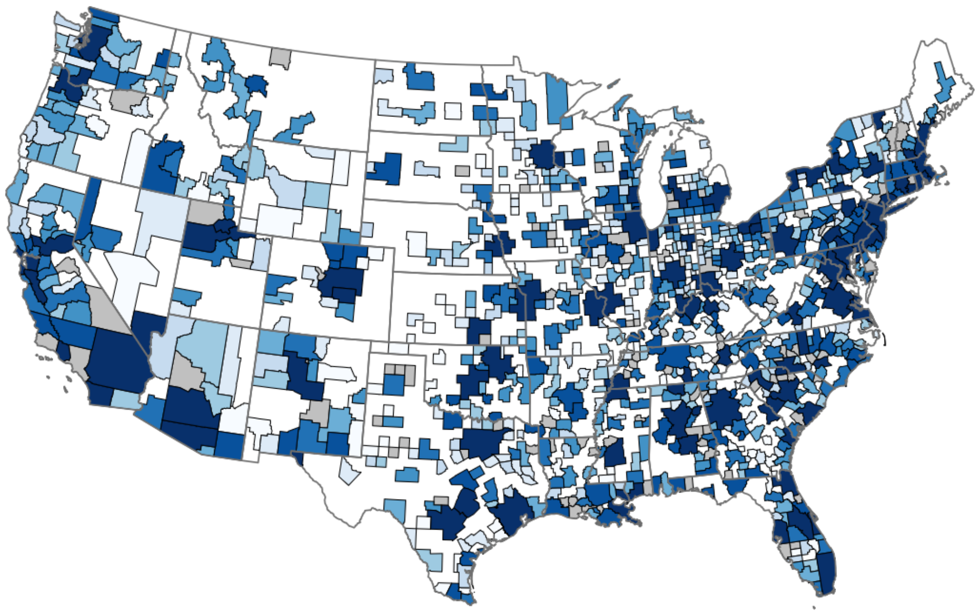
Because the number of products and stores entering the price index differs between CBSAs, measurement error in the price index could also affect estimates if it is systematically correlated with the spending share. I have assessed this using some proxies for measurement error in the price index. These proxies are 1) the average number of quarterly product observations per store entering the price index, 2) the average number of stores in the Nielsen data per capita, and 3) the average number of quarterly product observations entering the price index per capita. None of these variables capture much of the variation in the DoD spending shares as shown in rows 7 through 9.

Lastly, I show a heat map of the spending shares in figure 4 to get a sense of the geographic distribution of spending shares.

The map shows no clear geographic clustering of the spending shares. High and low spending shares are distributed more or less evenly across the entire country at a

[28]Grocery store data are from the QCEW instead of the Nielsen data since the Nielsen data set's coverage of stores is not geographically uniform.

Figure 4: Department of Defense spending shares



Notes: This heat map shows plots the geographic distribution of the average annual DoD spending share over the period 2002-2006. Darker colors represent a higher share. A grey color indicates no spending.

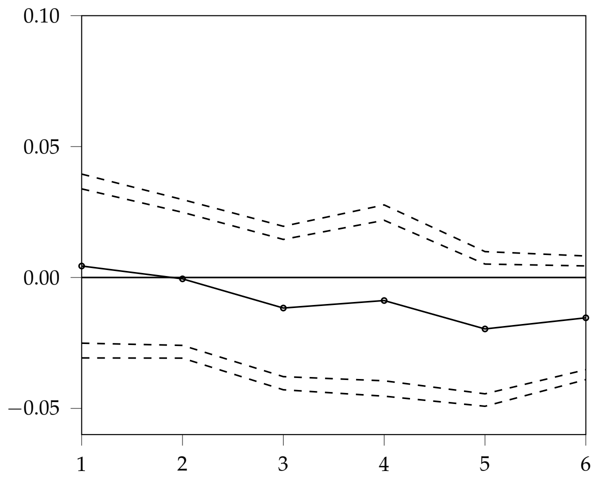
broad level. As an example, the three CBSAs with the highest share of DoD obligations – Washington-Arlington-Alexandria, Los Angeles-Long Beach-Anaheim and Dallas-Fort Worth-Arlington – are in DC/Virginia, California and Texas.

Department of Defense results

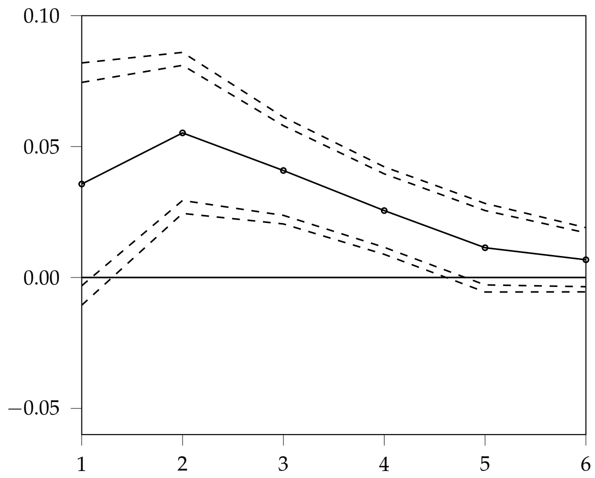
Figure 5 shows the estimates of β_h over 6 years. Panel (a) plots the OLS estimates, while panel (b) plots the IV estimates. Each h-horizon regression is estimated separately, and the dashed lines indicate the 90 and 95 % confidence bands based on the pointwise standard errors for β_h. The error term, $\epsilon_{i,t+h}$, is clustered by CBSA to account for within-CBSA serial correlation. I exclude CBSAs with fewer than 10 stores in the Nielsen data to limit measurement error in the dependent variable (this is 24 CBSAs out of 376) and winsorize local changes in DoD spending at the 1 % level by year to account for outliers.

The IV estimates in panel (b) show that the response of retail prices is hump-shaped, statistically significant and peaks around 0.06 after 2 years before it reverts to zero. This is equivalent to an increase of 0.06 % in retail prices when military spending increases by 1 % of GDP relative to its average value. Note that the estimate is one order of magnitude

Figure 5: Response of retail prices to military spending



(a) OLS estimates



(b) IV estimates

Notes: The figure shows the estimates of β_h from regression (3.4) using a panel of 352 CBSAs. Panel (a) plots the OLS estimates, while IV estimates are plotted in panel (b). Dashed lines indicate 90 and 95 % confidence bands calculated based on standard errors clustered by CBSA.

lower than the two-year estimate from the ARRA analysis in section 3.1, which was in the range of 0.9-1.2 depending on which controls were included in the regression. For the case of income multipliers, Demyanyk et al. (2018) also find that the effects of DoD spending increase with the size of the geographic unit. This might be because spending is more likely to leak into other areas as the geographic unit becomes smaller.

To get a sense of the magnitude of the peak estimates in relation to the data, the standard deviation of the two-year cumulative changes in DoD spending relative to GDP within CBSA and within year is 0.022 over the sample period. The standard deviation of the within-CBSA and within-year inflation rate over the same horizon is 0.9 %.[29] Thus, a typical change in DoD spending within a CBSA causes an increase in prices of $0.06 \cdot 0.022 = 0.13\%$ or about 14 % of the typical growth in prices within the CBSA.

The instrument is strong as shown in appendix table A.2. The estimates of β_h fluctuate around 1, which is consistent with a change in national DoD causing a roughly proportional increase in local DoD spending on average. The period-by-period first-stage Kleibergen-Papp F-statistics are generally around 100, while within R^2 values show that the instrument explains around 10-20 % of the within-CBSA variation in DoD spend-

[29]The standard deviations were calculated using data residualized with CBSA and year dummies.

ing.[30] The instrument becomes weaker at the 6-year horizon, however, where the test statistic drops to around 29.

Contrary to the IV estimates, the OLS estimates fluctuate around zero with relatively tight error bands. As mentioned previously, it is not clear a priori in what direction the OLS estimates should be biased. The bias in this case is, however, negative as is also the case for the ARRA estimates.

Robustness

Table A.3 presents a number of robustness checks of the results. First, I have normalized the spending variable by personal income and population instead of GDP and report the results in columns 2 and 3. Since population and personal income are available for more CBSAs, this increases the number of observations included in the regression with around 100 CBSAs. The price response estimates from these specifications are qualitatively similar to the specification using the GDP normalization although there is small increase in the estimates at the 6-year horizon.

Next, I estimate the regression using the two alternative price index, where the weights are either equal to national instead of CBSA-level quantities or fixed at their initial values in 2007. This addresses the worry that the results are not driven by actual price changes but differences in consumption composition across CBSAs. Column 4 and 5 shows that the price response is larger for both of these alternative indices compared to the baseline results. Moreover, the response of the fixed base index is not hump-shaped but monotonically decreasing over time. This could be because of consumption switching towards lower-inflation goods, which contributes to dampen the initial price response in the baseline estimates.

The estimates in column 6 address the positive correlation between population density and the spending shares mentioned above. This column reports estimates from a regression in which the CBSAs are divided into deciles based on average population density over the sample period and then density decile $\times$ year fixed effects are included. Hence, the estimates are identified off movements within deciles. This has little effect on the estimates except that they become more imprecise (albeit still significant).

Column 7 includes the two-digit industry employment shares interacted with year dummies. This controls for differential inflation rates associated with industry composi-

[30]The Kleibergen-Papp F-statistic is equal to the Montiel Olea and Pflueger (2013) effective F-statistic when there is only 1 instrument.

tion. Although the 1-year horizon estimate from this regression is roughly zero, estimates at longer horizons are more or less unaffected.

Lastly, I check the influence of measurement error in the dependent variable and outliers in spending changes. I include the additional 24 CBSAs with fewer than 10 retailers and report the results in column 8, which has little effect on the estimates. Column 9 shows the estimates from a regression in which I remove the 10 CBSAs with the highest share of DoD spending in the pre-sample period (the CBSAs are listed in table A.4). The price response is slightly larger but qualitatively similar. Column 10 reports the estimates when using non-winsorized changes in DoD spending, while the estimates in column 9 are from a regression in which I exclude winsorized observations. The estimates from the former are a little lower than those from the baseline regression, while the estimates from the latter are slightly larger.[31]

4 Marginal costs or markups

The results in the previous section provided evidence of a positive response of local retail prices to changes in government spending. The response can either be driven by changing markups or a pass-through of marginal costs. This section shows that the response is likely to be driven primarily by changing markups. I focus on the DoD analysis because of the panel data framework and the larger number of observations, which gives me more statistical power.

Unfortunately, the Nielsen data set does not include any measures of retailer costs so I rely on other sources of data. To get an idea of the average cost structure for the type of stores in the Nielsen data, I use the Census's national estimates of retailers' costs, which are released every five years in the Annual Retail Trade Statistics. I break down the costs for food and beverage stores into categories following Renkin et al. (2019) and show the cost shares for 2007, 2012 and 2017 in table 4.

The cost structure is largely stable over time. Wholesale costs make up the lion's share of costs: around three quarters of the retailers' total costs stem from wholesale costs. Labor costs is the second largest cost component and constitute 12-13 % of costs. The remaining costs are mostly building and equipment expenses, which are likely fixed at shorter horizons.

[31]The first-stage becomes weaker when using non-winsorized changes in DoD spending (F-statistics are in range of 10-20).

Table 4: Retailers' cost structure

	Wholesale	Labor	Building and equipment	Other costs
		Share of total costs		
2017	76.2	13.3	4.6	5.8
2012	77.2	12.2	4.3	6.2
2007	76.8	12.4	4.4	6.4

Notes: The table presents a breakdown of food and beverage stores' costs based on the Annual Retail Trade Survey. Labor costs include salaries, fringe benefits, and commissions. Wholesale costs are defined as annual purchases minus the year-to-year change in inventories. Building and equipment costs include rents, purchases of equipment, utilities, and depreciation. Other costs include all remaining operating costs (purchased services, taxes, transportation, etc.)

The cost shares discipline how much different input factors can contribute to changes in marginal costs. To see this, consider a retailer using N different production inputs to produce the good Y with a constant-returns-to-scale Cobb-Douglas production function:

$$Y\left(X_1, X_2, \dots, X_N\right) = \prod_{i=1}^{N} X_i^{\alpha_i} \tag{4.1}$$

where $X_1, X_2, \dots, X_N$ are the factor inputs and $\sum_{i=1}^{N} \alpha_i = 1$ ensures constant returns to scale in the production function.

If retailers minimize costs, marginal costs, MC, are given by

$$MC = \Phi \prod_{i=1}^{N} W_i^{\alpha_i} \tag{4.2}$$

where Φ is a function of the parameters and W_i is the price of input i.

α_i is equal to the cost share of input factor i for the firm. Thus, the percentage change in marginal costs as a result of a change in government spending is equal to a weighted average of the percentage change in factor prices, where the weights are the cost shares:

$$\frac{\partial MC}{\partial G}\frac{1}{MC} = \frac{1}{MC}\sum_{i=1}^{N}\frac{\partial MC}{\partial W_i}\frac{\partial W_i}{\partial G} = \sum_{i=1}^{N}\alpha_i\frac{\partial W_i}{\partial G}\frac{1}{W_i} \tag{4.3}$$

4.1 Wholesale costs

I begin by looking at wholesale costs, which make up the bulk of costs. According to equation (4.3) an increase in wholesale costs by $0.06/0.75 = 0.08\%$ over 2 years as a result of an increase in DoD spending corresponding to 1% of GDP would account for the entire increase in retail prices in figure 5.

Goods in the retail sector are rarely produced locally so local wholesale costs should be insensitive to local shocks (Stroebel and Vavra, 2019). Existing evidence from the U.S. retail sector also suggest that wholesale costs vary little geographically. Using Nielsen PromoData which consists of wholesale costs for quarters in the years 2006 through 2012 from one grocery wholesaler in each of 32 geographical areas, Stroebel and Vavra (2019) show that wholesale costs only vary little geographically. For 26 of the 32 markets, the average wholesale costs are within 1 % of the national average, while the most expensive area is 2.9 % above the national average and the least expensive area is 2.3 % below the national average. Additionally, the wholesale price of 78 % of the goods is exactly equal to the modal price. They find similar estimates when using data from a large national retail chain. A similar conclusion is reached by DellaVigna and Gentzkow (2019) who find no relationship between local consumer income and wholesale costs when analyzing store-level data from a large U.S. retail chain.

In summary, the findings by Stroebel and Vavra (2019) and DellaVigna and Gentzkow (2019) indicate that regional variations in wholesale prices should not be able to account for regional variation in retail prices. However, if chains are segmented geographically and they face similar wholesale costs – for example, due to using the same suppliers and selling the same private-label products – or face common storage and transport costs, my estimates above could be due to pass-through from higher wholesale costs at the chain level. I investigate this formally by constructing price indices for each chain within each CBSA using the procedure described in section 2.1 and estimating equation (3.4) at the CBSA $\times$ chain level:

$$\pi_{i,c,t+h} = \alpha_{i,c,h} + \eta_{c,t+h} + \beta_h \frac{\sum_{k=1}^{h} (G_{i,t+k} - G_{i,t})}{Y_{i,t}} + \varepsilon_{i,c,t+h} \tag{4.4}$$

where $\pi_{i,c,t+h}$ is the h-year inflation rate from period t to $t+h$ of chain c operating in CBSA i. As in equation (3.4), $G_{i,t}$ is DoD spending in the CBSA, where chain c operates, which is normalized by GDP in the CBSA.

The regression includes CBSA $\times$ chain fixed effects, $\alpha_{i,c,h}$, controlling for the average inflation rate of each chain within each CBSA as well as year $\times$ chain fixed effects, $\eta_{c,t+h}$. By controlling for the average inflation rate within each chain c using year $\times$ chain fixed effects, the regression controls for common shocks to stores within a chain and identifies β_h off price movements by stores belonging to the same chain but placed in different CBSAs.

The estimates of β_h from regression (4.4) are presented in table 5. Columns 1-2 show the OLS and IV estimates with year fixed effects, which are equivalent to disaggregated

CBSA-level estimates. Columns 3-4 show the same estimates but including year × chain fixed effects, while columns 5-6 add year × chain × Census division fixed effects to control for common regional shocks to chains. Standard errors are two-way clustered at the CBSA × year and CBSA × chain level to allow for serial correlation of the error term within each chain-CBSA pair as well as intra-year correlation of chains within CBSAs.[32]

Table 5: Controlling for common shocks to retail chains

	Year FE		Year X chain FE		Year X chain X Census division FE	
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	IV	OLS	IV	OLS	IV
1-year horizon	0.015*	0.047***	0.0090*	0.032***	-0.0013	0.022***
	(0.009)	(0.016)	(0.005)	(0.010)	(0.003)	(0.007)
2-year horizon	0.0099	0.10***	0.0059	0.060***	0.00031	0.051***
	(0.009)	(0.032)	(0.004)	(0.018)	(0.003)	(0.016)
3-year horizon	0.013	0.073***	0.0078	0.034***	-0.00046	0.028**
	(0.011)	(0.022)	(0.005)	(0.010)	(0.003)	(0.012)
4-year horizon	0.014	0.086***	0.0068	0.040***	0.00036	0.033**
	(0.010)	(0.028)	(0.005)	(0.014)	(0.003)	(0.015)
5-year horizon	0.0024	0.048***	0.0063	0.027**	0.0049**	0.019**
	(0.008)	(0.015)	(0.005)	(0.011)	(0.002)	(0.009)
6-year horizon	-0.0023	0.095	0.0016	0.039	0.0028	0.024
	(0.007)	(0.060)	(0.004)	(0.030)	(0.002)	(0.021)
Chain × CBSA pairs	1690	1690	1690	1690	1690	1690
CBSAs	376	376	376	376	376	376

Notes: The figure presents the OLS and IV estimates of β_h from equation (4.4). The regression only includes chain × CBSA pairs that are present in the entire period of 2007-2017, and local changes in DoD spending are winsorized using the same bounds as in the CBSA-level regression. Columns 1-2 show the estimates when including year fixed effects, columns 3-4 show the estimates when including year × chain fixed effects, and columns 5-6 show the estimates when including year × chain × Census division fixed effects. All regressions include chain × CBSA fixed effects. Standard errors are shown in parenthesis and two-way clustered at the chain × CBSA and year × CBSA levels. *, **, and *** denotes significance at the 10, 5, and 1 percent level respectively.

The estimates from the store-level regression when including year fixed effects in columns 1 and 2 are larger than those the CBSA-level regression. However, the price response is also hump-shaped and significant with a peak response at the 2-year horizon. Including year × chain fixed effects instead of year fixed effects reduces the estimates by 30 to 55 % but they are still significant. Controlling for the average inflation rate for each chain within each Census division in columns 5-6 reduces the estimates slightly more.

These results might seem surprising in light of the results by DellaVigna and Gentzkow

[32]Since the only regressor in the regression affect all chains within a CBSA equally, it is natural to also cluster by CBSAs × year to account for systematic over- or underprediction within a CBSA by the model (Cameron and Miller, 2015).

(2019) who find that chains charge nearly uniform prices across stores in the Nielsen data. However, their results are based on comparing the prices of frequently sold products, which is a much smaller set of products than what is included in the indices I construct. My price indices also allow for regional variation in the products sold. Finally, DellaVigna and Gentzkow (2019) do find that some chains vary prices by region.

4.2 Labor costs and retailer entry/exit

Next, I control for factors that could affect labor costs for retail stores as well as changes in the number of retailers. Using equation (4.3), the wage would need to grow by around 0.5 % to account for the entire increase in prices since labor costs only make up 12-13 % of costs. Although DoD spending should not directly affect wages in the retail sector, it can raise the local, average wages through effects on labor demand. Indeed, Auerbach et al. (2019a) find evidence of DoD spending increasing wage earnings not only in industries exposed to DoD spending either directly or through backward linkages but also in other local industries. In addition, DoD spending could poach workers from other industries, including the retail sector. Both higher wages and labor poaching can pass through to higher retail prices.

Column 2 in table 6 controls for the percentage change in the average wage in the retail sector, while column 3 controls for the growth in retail sector employment. Wage changes have no effect on the baseline estimates, while controlling for employment growth reduces the price effect slightly.[33] Lastly, I control for changes in the number of retailers per capita since this could affect pricing decisions through competition effects. However, this has no effect on the estimates.

4.3 Explaining procyclical markups

The results above indicate that local markups react positively to a local government spending shock since marginal costs do not change sufficiently. This finding runs counter to the textbook New Keynesian model. In this type of model, the realized markup is countercyclical conditional on a demand shock due to sticky prices that prevent firms

[33] Estimates of the regression using retail sector wage growth or retail employment growth as dependent variables show no significant effect of DoD spending on wages but a negative effect on employment growth. This indicates that spending does reallocate some labor away from the retail sector. Estimates for average, weekly wages and employment in all industries reveal significant, positive effects on overall wages and employment.

Table 6: Controlling for labor costs and retailer entry/exit

	(1) Baseline	(2) Control for wage growth	(3) Control for change in emp. share	(4) Control for retailer entry/exit
1-year horizon	0.021	0.021	0.017	0.021
	(0.024)	(0.024)	(0.024)	(0.024)
2-year horizon	0.047***	0.047***	0.039***	0.046***
	(0.012)	(0.012)	(0.011)	(0.012)
3-year horizon	0.038***	0.038***	0.032***	0.038***
	(0.009)	(0.008)	(0.009)	(0.009)
4-year horizon	0.025***	0.025***	0.018***	0.025***
	(0.007)	(0.007)	(0.007)	(0.007)
5-year horizon	0.0074	0.0072	0.00085	0.0074
	(0.008)	(0.008)	(0.008)	(0.008)
6-year horizon	0.0011	0.00077	0.0037	0.0011
	(0.004)	(0.004)	(0.005)	(0.004)
CBSAs	229	229	229	229

Notes: The figure shows the estimates of β_h from regression (4.4). The regression only includes the 229 CBSAs with no missing data in the QCEW due to disclosure issues. Column 1 shows the baseline estimates, which are the same of those presented in figure 5 but only with 229 CBSAs. Column 2 controls for the growth in average, weekly wages in the retail sector, column 3 controls for the change in the retail sector employment and column 4 controls for the change in the number of retailers per capita. All regressions include CBSA and year fixed effects, and standard errors are clustered at the CBSA level. Standard errors are shown in parentheses. *, **, and *** denotes significance at the 10, 5, and 1 percent level respectively.

from adjusting prices to the level at which markups are at the constant, desired value. The desired markup, in turn, reflects households' time-invariant elasticity of substitution between goods.[34] Hall (2009) states that there is weak empirical support for falling markups in response to higher government spending, while he also argues that the procyclical behavior of advertising is inconsistent with a countercyclical markup (Hall, 2014). Similarly, Nekarda and Ramey (2019) have revisited the markup cyclicality literature with updated data and methods. They conclude that markups are either procyclical or acyclical irrespective of the type of shock hitting the economy. Lastly, Anderson et al. (2018) study markup fluctuations in the retail sector and find that they are mildly procyclical.

Other authors have provided explanations for why markups might be procyclical. One strand of literature focuses on changes in consumers' price sensitivity. Stroebel and Vavra (2019) present evidence of homeowners becoming less price sensitive as house

[34] Estimated New Keynesian models traditionally attribute a large share of inflation movements to fluctuations in desired markups or cost-push shocks (e.g. Smets and Wouters (2007) and Justiniano et al. (2010)). However, these movements in desired markups are exogenous and do not arise endogenously in the models.

prices increase because of positive wealth effects. This could cause retailers' to raise markups. This mechanism is specific to house price shocks but one could think of a similar pricing mechanism for regional government spending shocks, where an increase in federal spending decreases households' price sensitivity on average through positive effects on employment and income. For example, Aguiar et al. (2013) find that unemployed workers in the American Time Use Survey spend more time shopping relative to the employed, while Kaplan and Menzio (2015) find that households in the Nielsen HomeScan data for which at least one of the household heads is unemployed pay 1-4 % less for the same basket of goods compared to fully employed households because they shop at a larger number of stores. Kaplan and Menzio (2016) model this difference in consumer search behavior of the employed and unemployed workers and show how it creates countercyclical movements in market power of retailers, which results in procyclical markups.

A different mechanism is suggested by Anderson et al. (2018). They use scanner data on retail and wholesale prices from a large retailer and find a positive cross-county correlation between income and markups, which is mostly attributed to differences in product assortment rather than deviations from uniform pricing. They rationalize these findings in a model with endogenous product assortment in which consumers buy goods of a higher quality and at a higher markup as their income increases.

5 Concluding remarks

This paper studied how government spending affects retail prices by analyzing two spending shocks: the state-level spending shock embedded in the ARRA of 2009 and CBSA-level DoD spending shocks. Both analysis revealed a positive effect of government spending on retail prices. The ARRA estimates for the two-year response of retail price growth to an increase of 1 % of GDP in government spending over the same period were in the range of 0.6-0.9 %, while the DoD estimates were more modest in size (around 0.06 %).

I showed that the inflationary effects of the DoD spending shocks are unlikely to be driven by changes in the marginal costs faced by retailers. This points towards a procyclical markup, which stands in contrasts to the predictions of a standard sticky-price model. As discussed above, these findings are more in line with a strand of literature on the countercyclical nature of households' shopping intensity although none of these articles are directly concerned with the effects of government spending. Whether government spending actually does affects households' price sensitivity is of course an empirical question. Hence, further research is needed to understand the mechanism that can cause procyclicality of markups in reaction to demand shocks.

Finally, this paper only studied inflationary effects in the retail sector. This sector is unlikely to be affected directly by government spending shocks. Instead, I have argued that my estimates capture indirect effects operating through variable markups. My results are silent on how price-setting and costs are affected for firms supplying goods and services directly to the government. The military contract data used in this paper, however, would be suitable for such an analysis.

References

Aguiar, M., Hurst, E., and Karabarbounis, L. (2013). Time Use During the Great Recession. *American Economic Review*, 103(5):1664–1696.

Anderson, E., Rebelo, S., and Wong, A. (2018). Markups Across Space and Time. NBER Working Paper No. 24434.

Auclert, A. (2019). Monetary Policy and the Redistribution Channel. *American Economic Review*, 109(6):2333–2367.

Auerbach, A., Gorodnichenko, Y., and Murphy, D. (2019a). Local Fiscal Multipliers and Fiscal Spillovers in the United States. NBER Working Paper No. 25457.

Auerbach, A., Gorodnichenko, Y., and Murphy, D. (2019b). Macroeconomic Frameworks. NBER Working Paper No. 26365.

Auerbach, A., Gorodnichenko, Y., and Murphy, D. (2020). Effects of Fiscal Policy on Credit Markets. NBER Working Paper No. 26655.

Bartik, T. J. (1991). *Who Benefits from State and Local Economic Development Policies?* W.E. Upjohn Institute.

Beraja, M., Hurst, E., and Ospina, J. (2019). The Aggregate Implications of Regional Business Cycles. *Econometrica*, 87(6):1789–1833.

Bureau of Economic Analysis (2015). GDP by Metropolitan Area Methodology.

Bureau of Economic Analysis (2016). Real Personal Income and Regional Price Parities.

Cameron, A. C. and Miller, D. L. (2015). A Practitioner's Guide to Cluster-Robust Inference. *Journal of Human Resources*, 50(2):317–372.

Canova, F. and Pappa, E. (2007). Price Differentials in Monetary Unions: The Role of Fiscal Shocks. *The Economic Journal*, 117(520):713–737.

Cecchetti, S. G., Mark, N. C., and Sonora, R. J. (2002). Price Index Convergence Among United States Cities. *International Economic Review*, 43(4):1081–1099.

Chodorow-Reich, G. (2019). Geographic Cross-Sectional Fiscal Spending Multipliers: What Have We Learned? *American Economic Journal: Economic Policy*, 11(2):1–34.

Chodorow-Reich, G., Feiveson, L., Liscow, Z., and Woolston, W. G. (2012). Does State Fiscal Relief During Recessions Increase Employment? Evidence from the American Recovery and Reinvestment Act. *American Economic Journal: Economic Policy*, 4(3):118–145.

Christiano, L., Eichenbaum, M., and Rebelo, S. (2011). When Is the Government Spending Multiplier Large? *Journal of Political Economy*, 119(1):78–121.

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

Council of Economic Advisers (2014). Economic Report of the President.

D'Acunto, F., Malmendier, U., Ospina, J., and Weber, M. (2019). Exposure to Daily Price Changes and Inflation Expectations. NBER Working Paper No. 26237.

DAlessandro, A., Fella, G., and Melosi, L. (2019). Fiscal Stimulus with Learning-by-Doing. *International Economic Review*, 60(3):1413–1432.

de Haan, J. and van der Grient, H. A. (2011). Eliminating chain drift in price indexes based on scanner data. *Journal of Econometrics*, 161(1):36–46.

Del Negro, M. (1998). Aggregate Risk Sharing Across US States and Across European Countries. *Unpublished*.

DellaVigna, S. and Gentzkow, M. (2019). Uniform Pricing in U.S. Retail Chains. *The Quarterly Journal of Economics*, 134(4):2011–2084.

Demyanyk, Y., Loutskina, E., and Murphy, D. (2018). Fiscal Stimulus and Consumer Debt. *The Review of Economics and Statistics*, 101(4):728–741.

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*, 37(1):22–53.

Dupor, B., Karabarbounis, M., Kudlyak, M., and Mehkari, M. S. (2018). Regional Consumption Responses and the Aggregate Fiscal Multiplier. *Federal Reserve Bank of St. Louis, Working Paper 2018-004A*.

Dupor, B. and Mehkari, M. S. (2016). The 2009 Recovery Act: Stimulus at the extensive and intensive labor margins. *European Economic Review*, 85:208–228.

Farhi, E. and Werning, I. (2016). Fiscal Multipliers: Liquidity Traps and Currency Unions. In Taylor, J. and Uhlig, H., editors, *Handbook of Macroeconomics*, volume 2, pages 2417–2492. Elsevier.

Ferrara, L., Metelli, L., Natoli, F., and Siena, D. (2019). Questioning the puzzle: fiscal policy, real exchange rate and inflation. *Unpublished*.

Forni, M. and Gambetti, L. (2016). Government spending shocks in open economy VARs. *Journal of International Economics*, 99:68–84.

Gagnon, E. and López-Salido, D. (forthcoming). Small Price Responses to Large Demand Shocks. *Journal of the European Economic Association*.

Goldsmith-Pinkham, P., Sorkin, I., and Swift, H. (2018). Bartik Instruments: What, When, Why, and How. NBER Working Paper No. 24408.

Guren, A., McKay, A., Nakamura, E., and Steinsson, J. (2018). Housing Wealth Effects: The Long View. NBER Working Paper No. 24729.

Gyourko, J., Saiz, A., and Summers, A. (2008). A New Measure of the Local Regulatory Environment for Housing Markets: The Wharton Residential Land Use Regulatory Index. *Urban Studies*, 45(3):693–729.

Hall, R. E. (2009). By How Much Does GDP Rise If the Government Buys More Output? *Brookings Papers on Economic Activity*, 40(2 (Fall)):183–249.

Hall, R. E. (2014). What the Cyclical Response of Advertising Reveals about Markups and other Macroeconomic Wedges. *Unpublished*.

Hansen, L. P. (1982). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, 50(4):1029–1054.

Hitsch, G., Hortaçsu, A., and Lin, X. (2019). Prices and Promotions in U.S. Retail Markets: Evidence from Big Data. NBER Working Paper No. 26306.

Jørgensen, P. L. and Ravn, S. H. (2018). The Inflation Response to Government Spending Shocks: A Fiscal Price Puzzle? *Unpublished*.

Justiniano, A., Primiceri, G. E., and Tambalotti, A. (2010). Investment shocks and business cycles. *Journal of Monetary Economics*, 57(2):132–145.

Kaplan, G. and Menzio, G. (2015). The Morphology of Price Dispersion. *International Economic Review*, 56(4):1165–1206.

Kaplan, G. and Menzio, G. (2016). Shopping Externalities and Self-Fulfilling Unemployment Fluctuations. *Journal of Political Economy*, 124(3):771–825.

Kehoe, P. and Midrigan, V. (2015). Prices are sticky after all. *Journal of Monetary Economics*, 75:35–53.

Kim, S. and Roubini, N. (2008). Twin deficit or twin divergence? Fiscal policy, current account, and real exchange rate in the U.S. *Journal of International Economics*, 74(2):362–383.

Mian, A., Rao, K., and Sufi, A. (2013). Household Balance Sheets, Consumption, and the Economic Slump. *The Quarterly Journal of Economics*, 128(4):1687–1726.

Mian, A. and Sufi, A. (2014). What Explains the 2007-2009 Drop in Employment? *Econometrica*, 82(6):2197–2223.

Monacelli, T. and Perotti, R. (2010). Fiscal Policy, the Real Exchange Rate and Traded Goods. *The Economic Journal*, 120(544):437–461.

Montiel Olea, J. L. and Pflueger, C. (2013). A Robust Test for Weak Instruments. *Journal of Business & Economic Statistics*, 31(3):358–369.

Nakamura, E. and Steinsson, J. (2014). Fiscal Stimulus in a Monetary Union: Evidence from US Regions. *American Economic Review*, 104(3):753–792.

Nakamura, E. and Steinsson, J. (2018). Identification in Macroeconomics. *Journal of Economic Perspectives*, 32(3):59–86.

Nekarda, C. and Ramey, V. (2019). The Cyclical Behavior of the Price-Cost Markup. *Unpublished*.

Oberfield, E. and Raval, D. (2014). Micro Data and Macro Technology. NBER Working Paper No. 20452.

Ramey, V. A. (2011). Identifying Government Spending Shocks: It's all in the Timing. *The Quarterly Journal of Economics*, 126(1):1–50.

Ramey, V. A. (2016). Defense news shocks, 1889-2015: Estimates based on news sources. *Unpublished*.

Ramey, V. A. and Shapiro, M. D. (1998). Costly capital reallocation and the effects of government spending. *Carnegie-Rochester Conference Series on Public Policy*, 48:145–194.

Ravn, M. O., Schmitt-Grohé, S., and Uribe, M. (2012). Consumption, government spending, and the real exchange rate. *Journal of Monetary Economics*, 59(3):215–234.

Renkin, T., Montialoux, C., and Siegenthaler, M. (2019). The Pass-through of Minimum Wages into US Retail Prices: Evidence from Supermarket Scanner Data. *Unpublished*.

Romer, C. D. and Romer, D. H. (2010). The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks. *American Economic Review*, 100(3):763–801.

Saiz, A. (2010). The Geographic Determinants of Housing Supply. *Quarterly Journal of Economics*, 125(3):1253–1296.

Smets, F. and Wouters, R. (2007). Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach. *American Economic Review*, 97(3):586–606.

Stock, J. H. and Yogo, M. (2005). Testing for weak instruments in linear IV regression. In Andrews, D. W. K. and Stock, J. H., editors, *Identification and Inference for Econometric Models*, pages 80–108. Cambridge University Press.

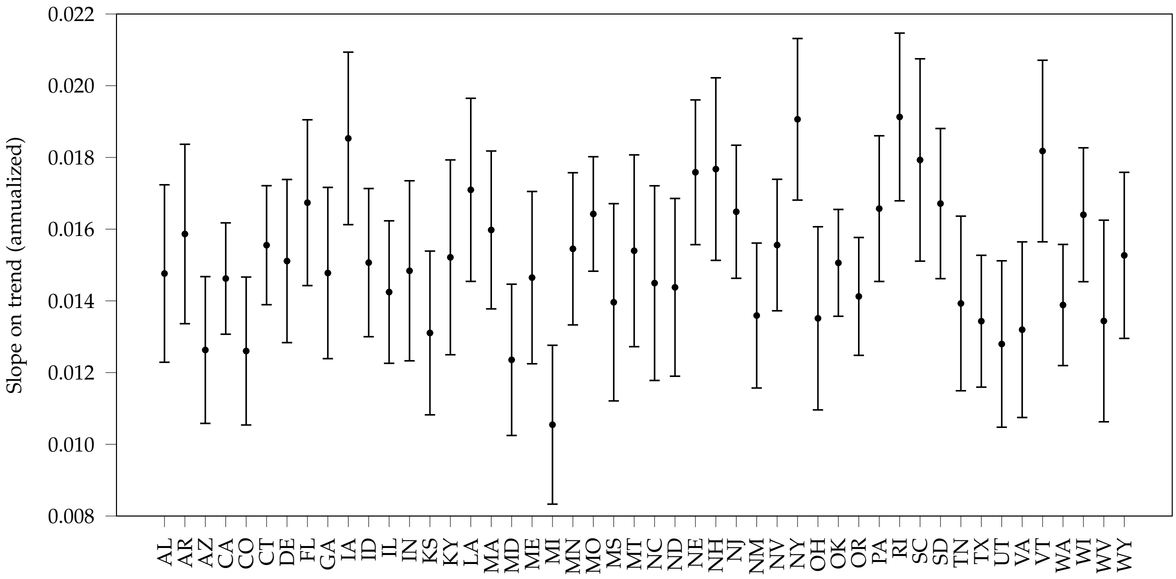
Stroebel, J. and Vavra, J. (2019). House Prices, Local Demand, and Retail Prices. *Journal of Political Economy*, 127(3):1391–1436.

Wilson, D. J. (2012). Fiscal Spending Jobs Multipliers: Evidence from the 2009 American Recovery and Reinvestment Act. *American Economic Journal: Economic Policy*, 4(3):251–282.

Woodford, M. (2003). *Interest and Prices*. Princeton University Press.

A Additional figures and tables

Figure A.1: Slope estimates for state-specific trends (ARRA analysis)



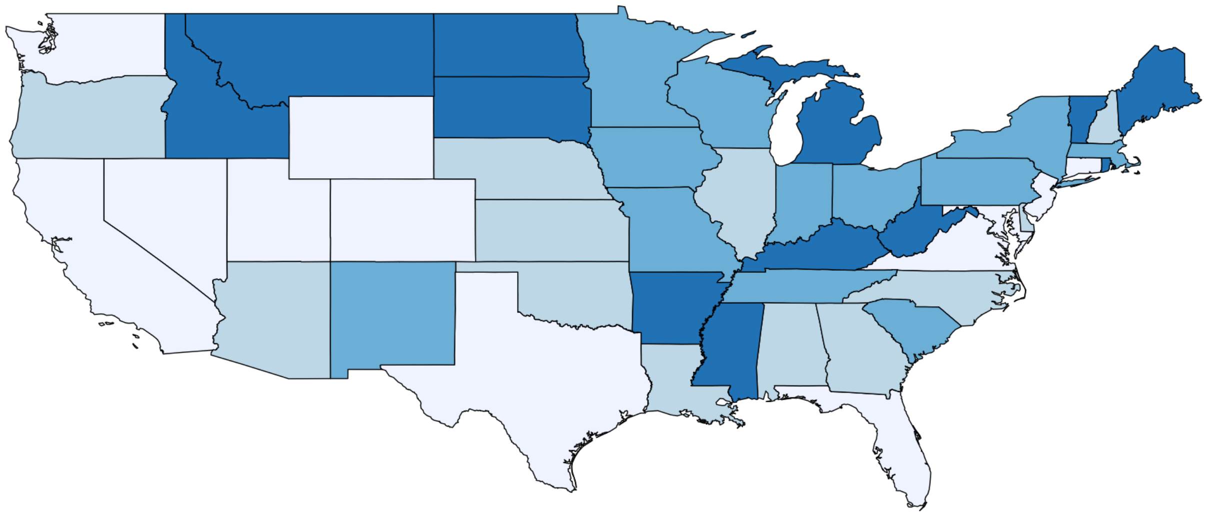
Notes: This figure plots the state-specific estimates of $\tilde{\gamma}_i$ from the regression $\ln p_{i,t} = \tilde{\alpha}_i + \tilde{\gamma}_i \cdot t + \tilde{\epsilon}_{i,t}$. The estimates have been annualized by multiplying with 4 such that the figure shows the state-specific average annual growth rate in the price index. Error bars indicate 95 % confidence bands calculated using heteroskedasticity-robust standard errors.

Table A.1: First-stage estimates (ARRA analysis)

	Dependent variable: Cumulative ARRA outlays through 2010 (normalized by GDP in 2008)			
	(1)	(2)	(3)	(4)
Instruments (all normalized by GDP)				
DoT instrument	−0.148	0.626	0.731*	0.676
	(0.481)	(0.387)	(0.396)	(0.455)
Narrative instrument	5.793***	5.266***	5.159***	4.919***
	(1.005)	(0.640)	(0.637)	(0.797)
Medicaid spending in 2007	0.275***	0.260***	0.263***	0.269***
	(0.041)	(0.035)	(0.037)	(0.037)
Controls				
Emp. rate, Dec. 2008		−0.027***	−0.028***	−0.028***
		(0.007)	(0.008)	(0.009)
Emp. change, Dec. 2007-08		−0.0403*	−0.0558**	−0.055**
		(0.022)	(0.024)	(0.022)
GDP growth, 2007-08		−0.001	0.001	0.003
		(0.009)	(0.008)	(0.009)
House price growth, 2003-07			−0.002	−0.001
			(0.001)	(0.002)
Retail price growth, 2007-08			−0.029	−0.035*
			(0.018)	(0.021)
Tax benefits per capita				−0.001
				(0.001)
Change in avg. personal income, 2005-06				−0.023
				(0.029)
Constant	0.006***	0.029***	0.031***	0.035***
	(0.001)	(0.007)	(0.007)	(0.008)
Effective *F*-statistic on excluded instruments	56.81	89.01	83.81	66.65
R^2	0.798	0.892	0.897	0.900
N	48	48	48	48

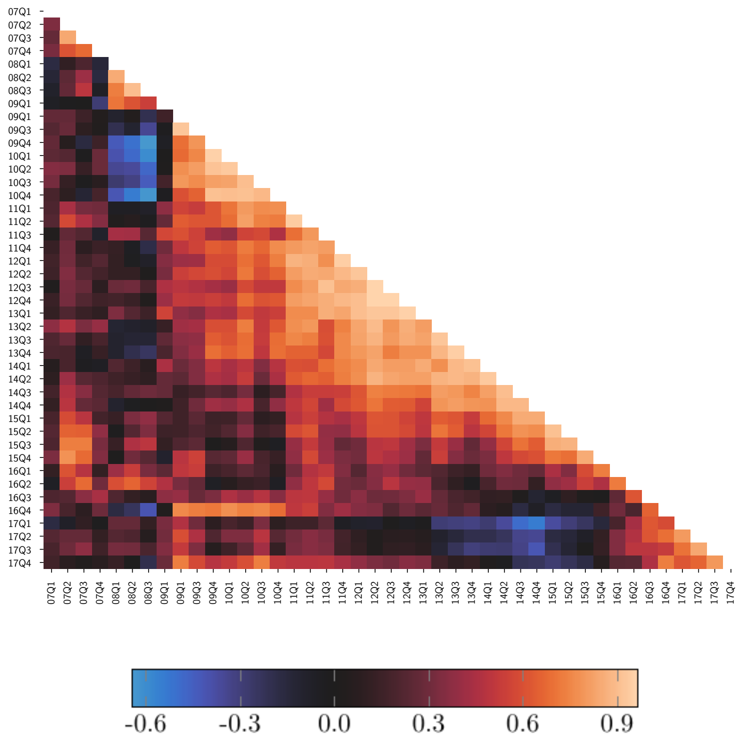
Notes: The figure shows the estimates from the first-stage regression $G_i = \kappa + \lambda Z_i + \delta X_i + \mu_i$. The reported *F*-statistic is the effective *F*-statistic by Montiel Olea and Pflueger (2013). Heteroskedasticity-robust standard errors are shown in parentheses. *, **, and *** denotes significance at the 10, 5, and 1 percent level respectively.

Figure A.2: Distribution of predicted ARRA spending



Notes: This heat map shows predicted ARRA spending normalized by GDP according to the first-stage regression without controls, X_i. Darker colors represent more spending.

Figure A.3: ARRA analysis: Heatmap of bootstrapped correlations between IV estimates



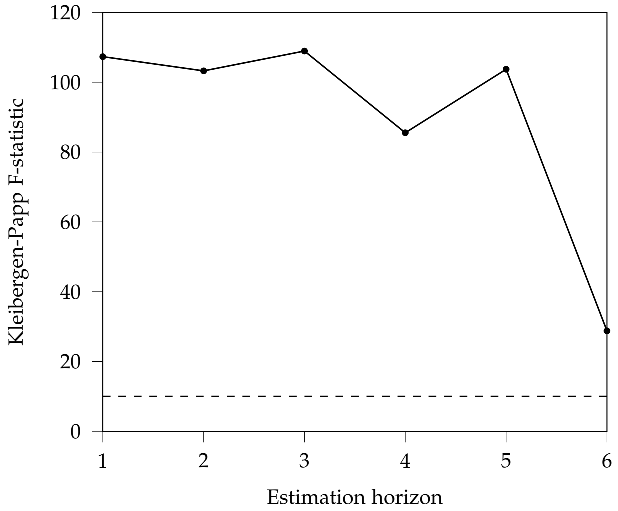
Notes: This heat map shows the correlation coefficients of the IV estimates of $\Gamma = (\gamma_1, \gamma_2, \ldots, \gamma_T)$ from the second-stage regression $\pi_{i,t} = \alpha_t + \beta_t X_i + \gamma_t G_i + \epsilon_{i,t}$. The correlations were computed using 500 iterations from a cluster bootstrap that for each bootstrap iteration draws 48 states with replacement and estimates the entire set of IV estimates, Γ.

Table A.2: First-stage estimates (Department of Defense analysis)

	Dependent variable:					
	Cumulative increase in local DoD spending normalized by initial GDP					
	1 year	2 year	3 year	4 year	5 year	6 year
Bartik instrument	1.02***	0.99***	1.14***	1.12***	0.99***	1.06***
	(0.10)	(0.10)	(0.11)	(0.12)	(0.10)	(0.20)
F-statistic	107.33	103.26	108.94	85.53	103.72	28.79
Within R^2	0.15	0.11	0.18	0.20	0.21	0.17

Notes: The table presents the estimates of $\tilde{\beta}_h$ from regression (3.5) with 352 CBSAs and over the period 2007-2017. Kleibergen-Papp F-statistics and the within R^2 statistics are presented in the bottom rows. Regressions include CBSA and year fixed effects, and standard errors are clustered at the CBSA level. Standard errors are shown in parentheses. *, **, and *** denotes significance at the 10, 5, and 1 percent level respectively.

Figure A.4: Kleibergen-Papp F-statistic in Department of Defense analysis



Notes: The figure shows the Kleibergen-Papp F-statistic from the first-stage regression in the DoD analysis. The horizontal, dashed line indicates the rule-of-thumb value of 10 above which instruments are not weak.

Table A.4: CBSAs with highest Department of Defense obligations share

CBSA name	Share of national DoD spending, 2002-06
Washington-Arlington-Alexandria, DC-VA-MD-WV	0.137
Los Angeles-Long Beach-Anaheim, CA	0.066
Dallas-Fort Worth-Arlington, TX	0.064
St. Louis, MO-IL	0.036
Boston-Cambridge-Newton, MA-NH	0.032
San Diego-Carlsbad, CA	0.025
Virginia Beach-Norfolk-Newport News, VA-NC	0.024
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	0.023
New York-Newark-Jersey City, NY-NJ-PA	0.022
Phoenix-Mesa-Scottsdale, AZ	0.021

Notes: The table shows the average, annual share of national DoD obligations during the period of 2002-2006 for the 10 CBSAs with the highest shares.

Table A.3: Robustness: Department of Defense spending analysis

	(1) Baseline	(2) Normalized by income	(3) Normalized by population	(4) Equal weights price index	(5) Fixed base price index	(6) Pop. density × year FEs	(7) Industry shares × year dummies	(8) Include all CBASs	(9) Remove top 10 DoD exposure CBSAs	(10) Non-winsorized	(11) Drop instead of winsorizing
1-year horizon	0.036	0.034	0.080	0.063***	0.079**	0.030	−0.0071	0.041*	0.037	0.023	0.048
	(0.024)	(0.030)	(0.072)	(0.018)	(0.032)	(0.029)	(0.026)	(0.024)	(0.025)	(0.018)	(0.040)
2-year horizon	0.055***	0.039**	0.095**	0.076***	0.062**	0.061**	0.058***	0.059***	0.062***	0.034***	0.084***
	(0.016)	(0.018)	(0.045)	(0.015)	(0.028)	(0.026)	(0.021)	(0.016)	(0.016)	(0.013)	(0.026)
3-year horizon	0.041***	0.029**	0.069**	0.053***	0.045***	0.053***	0.045***	0.044***	0.046***	0.026***	0.055**
	(0.010)	(0.012)	(0.030)	(0.011)	(0.017)	(0.019)	(0.015)	(0.011)	(0.011)	(0.008)	(0.021)
4-year horizon	0.026***	0.012	0.030	0.033***	0.022	0.037**	0.031**	0.028***	0.029***	0.017**	0.038**
	(0.009)	(0.010)	(0.025)	(0.011)	(0.016)	(0.018)	(0.013)	(0.009)	(0.009)	(0.007)	(0.016)
5-year horizon	0.011	0.0058	0.014	0.018**	0.018	0.018	0.0023	0.013	0.013	0.0070	0.029**
	(0.009)	(0.011)	(0.027)	(0.008)	(0.014)	(0.017)	(0.012)	(0.009)	(0.009)	(0.006)	(0.014)
6-year horizon	0.0068	0.015**	0.037**	0.0057	0.0013	0.0075	0.00021	0.0080	0.0072	0.0061	0.0082
	(0.006)	(0.007)	(0.018)	(0.005)	(0.010)	(0.013)	(0.010)	(0.006)	(0.006)	(0.006)	(0.011)
CBSAs	352	449	449	352	352	346	376	342	352	352	352

Notes: The table shows estimates from robustness checks to the estimates of β_h from regression (3.4). Column 1 are the baseline estimates shown in figure 5. Column 2 normalizes the change in DoD obligations by initial personal income. Column 3 normalizes the change in DoD obligations by population and scales estimates by 100. Column 4 uses a price index with CBSA-level weights set to the their national quantities. Column 5 uses a fixed-base Laspeyres index with weights fixed at quantities and revenues as of 2007. Column 6 includes population density decile × year fixed. Column 7 includes two-digit NAICS shares interacted with year dummies. Column 8 includes CBSAs with under 10 store observations in the regression. Column 9 excludes the 10 CBSAs with the highest pre-sample share of DoD obligations. Column 10 uses the non-winsorized change in DoD spending. Column 11 excludes observations with a winsorized change in DoD spending. All specifications include CBSA and year fixed effects and use error terms clustered by CBSA. Standard errors are shown in parentheses. *, **, and *** denotes significance at the 10, 5, and 1 percent level respectively.

B ARRA analysis: Robustness

This appendix presents robustness checks to the estimates from the ARRA analysis.

B.1 Pre-ARRA inflation dynamics

One worry is that states receiving more ARRA spending might have had higher inflation rates not captured by the state-specific linear trend before the enactment of the ARRA. If this is the case, the estimate for the response of retail prices to spending could be spurious. Unfortunately, the Nielsen price index is only available for the period of 2007-2017, which is insufficient to uncover any such long-run differences in price dynamics prior to the ARRA.

Instead, I rely on annual CPI data used by Nakamura and Steinsson (2014). Their CPI series cover the period of 1969-2006 and are imputed from several sources. From 1969 until 1995 they use a CPI series constructed by Del Negro (1998). After 1995 they construct state-level indices by multiplying the U.S. aggregate CPI index with population-weighted cost of living indices from the American Chamber of Commerce Realtors Association.

I estimate 36 versions of regression (3.1) in which I replace the dependent variable with the two-year inflation rate, $\pi_{i,t}^{NM}$, from the Nakamura and Steinsson (2014) data set from 1971 until 2006:

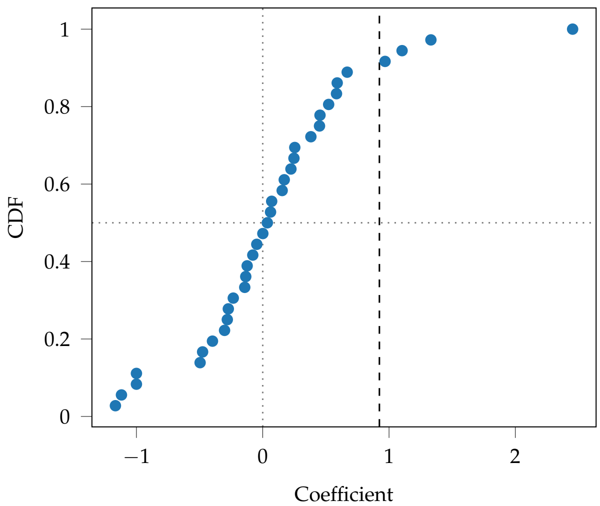
$$\pi_{i,t}^{NM} = \alpha_t + \gamma_t G_i + \beta_t X_i + \epsilon_{i,t} \qquad \text{for } t = 1971, 1972, \ldots, 2006 \tag{B.1}$$

where the full set of controls and instruments from the baseline regression are included.

The empirical cumulative distribution function for γ_t using the 36 placebo IV estimates are shown in figure B.1. The estimate for the response of retail prices in the fourth quarter of 2010 is indicated by the vertical, dashed line.

If the estimates presented in section 3 reflect a spurious correlation between the average inflation rate and ARRA spending, the placebo estimates would be positive on average. This is not the case: the median placebo estimate is close to zero. These findings are also in line with the placebo analysis by Chodorow-Reich et al. (2012) for ARRA employment multipliers. Only four of the placebo estimates are above the retail price response estimate of 0.9. Keep in mind, however, that the magnitude of the estimates are not directly comparable since the placebo estimates are based on CPI data, which cover a broader set of goods.

Figure B.1: Placebo estimates for 1971-2006 (ARRA analysis)



Notes: Dots show the cumulative distribution function for the IV estimates of γ_t from the regressions $\pi_{i,t}^{NM} = \alpha_t + \gamma_t G_i + \beta_t X_i + \epsilon_{i,t}$ for $t = 1971, 1972, \ldots, 2006$, where $\pi_{i,t}^{NM}$ is the two-year CPI inflation rate constructed by Nakamura and Steinsson (2014). I use the same full set of controls and instruments as for the regression results presented in section 3.

B.2 Additional robustness checks

Table B.1 presents the OLS and IV estimates of γ_t from regression (3.1) at the fourth quarter of 2010 when subjecting the estimates to various robustness checks. All estimates include the full set of controls used in columns 7-8 of table 2. Figure B.2 shows the full dynamic response of prices for all of these specifications.

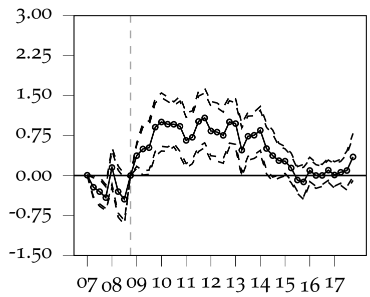
Figure B.3 shows the added variable plot for price growth from the fourth quarter of 2008 to the fourth quarter of 2010 against ARRA spending. The y-axis variable is the residuals from a regression of price growth on all controls excluding ARRA spending, while the x-axis spending variable is the residuals from a regression of predicted ARRA spending from the first-stage regression on all controls. The figure shows that the estimate is not driven by outliers.

Table B.1: Robustness checks (ARRA analysis)

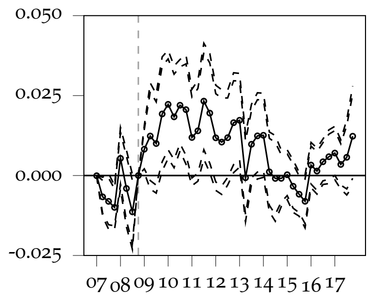
	Specification	OLS	IV
(1)	Baseline	0.666**	0.924***
		(0.313)	(0.287)
(2)	Spending normalized by population	0.004	0.013**
		(0.004)	(0.006)
(3)	Spending normalized by personal income	0.393	0.832**
		(0.348)	(0.353)
(4)	Only Medicaid instrument	0.666**	1.240***
		(0.313)	(0.392)
(5)	Only DoT instrument	0.666**	0.915*
		(0.313)	(0.496)
(6)	Only narrative instrument	0.666**	0.711**
		(0.313)	(0.281)
(7)	Control for squared and cubed pre-trend	0.621**	0.866***
		(0.304)	(0.282)
(8)	Control for bi-annual CPI inflation 2000-06	0.590**	0.795***
		(0.248)	(0.245)
(9)	Use equal weights price index	0.618***	0.702***
		(0.006)	(0.194)
(10)	Use fixed-base price index	1.341***	1.669***
		(0.309)	(0.332)
(11)	Include region dummies	0.243	0.358
		(0.277)	(0.317)

Notes: The table presents the OLS and IV estimates from the regression $\pi_i = \alpha + \beta X_i + \gamma G_i + \epsilon_i$. π_i is the growth of the retail price index from the last quarter of 2008 to the last quarter of 2010, G_i are cumulative ARRA outlays through 2010 normalized by GDP in 2008, and X_i is a vector of controls. The endogenous variable, G_i, has been instrumented for using the three instruments described in section 3.1: Medicaid spending in 2007, the DoT instrument and the Dupor and Mehkari (2016) narrative measure. In the regression using the equal weights and fixed-base price indices as a dependent variable (row 9-10), the pre-ARRA trend in retail prices is measured using the equal weight and fixed-base price indices respectively. Heteroskedasticity-robust standard errors are shown in parentheses. ***, ** and * denote significance at the 0.01, 0.05 and 0.1 level respectively.

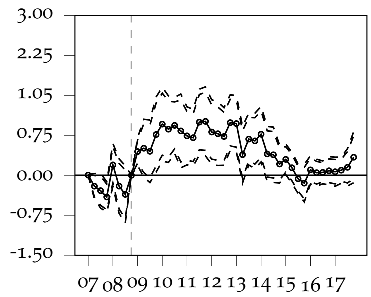
Figure B.2: Dynamic IV estimates from robustness checks (ARRA analysis)



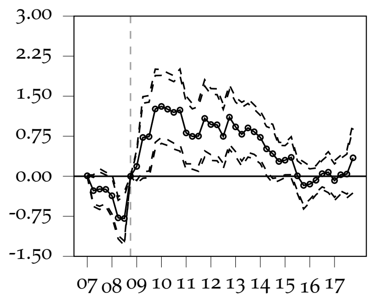
(a) Baseline



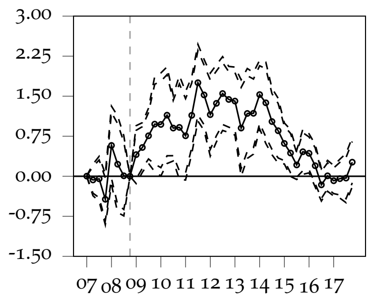
(b) Normalize ARRA by pop.



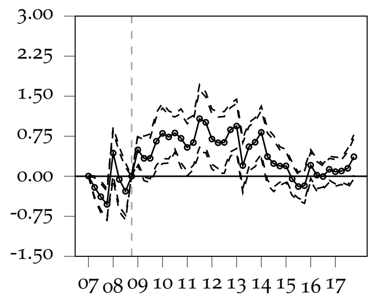
(c) Normalize ARRA by inc.



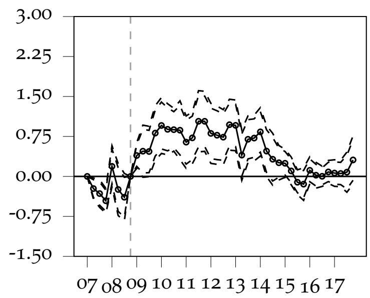
(d) Medicaid inst. only



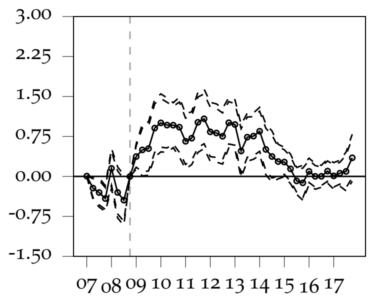
(e) DoT inst. only



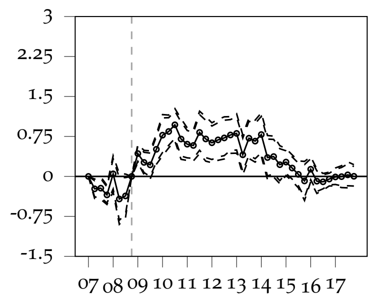
(f) Narrative inst. only



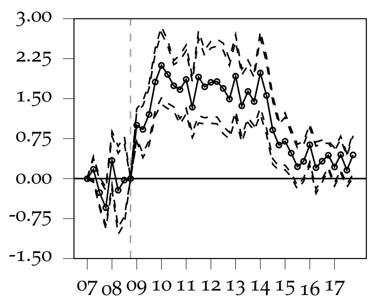
(g) Include squared and cubed pre-trend



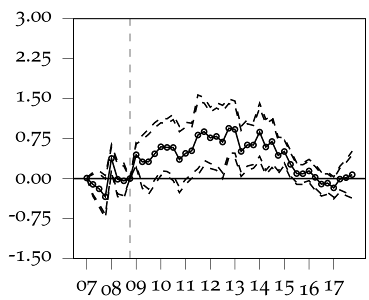
(h) Control for bi-annual CPI inflation 2000-06



(i) Use equal weights price index



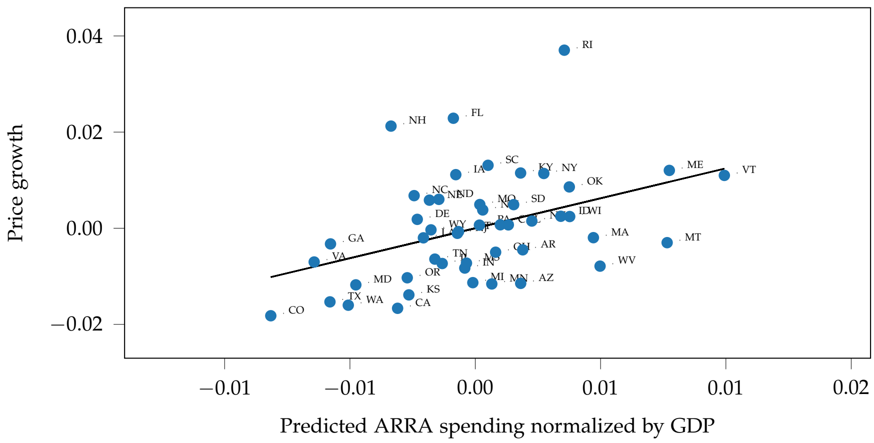
(j) Use fixed-base price index



(k) Include region dummies

Notes: The figures show IV estimates over the entire sample period for the robustness checks also provided in table B.1. All regressions include the full set of control included in column 4 of table 2. 90 % and 95 % confidence bands based on heteroskedasticity-robust standard errors are indicated by dashed lines. The vertical dashed lines indicate the enactment of the ARRA.

Figure B.3: Price growth against government spending (ARRA analysis)



Notes: The figure shows the added variable plot for price growth against ARRA spending from regression (3.1). The y-axis variable is the residuals from a regression of price growth from the fourth quarter of 2008 to the fourth quarter of 2010 on all controls excluding ARRA spending, while the x-axis spending variable is the residuals from a regression of predicted ARRA spending from the first-stage regression on all controls.

Chapter 3

House Prices, Increasing Returns, and Government Spending Shocks

House Prices, Increasing Returns, and the Effects of Government Spending Shocks*

Rasmus Bisgaard Larsen† Søren Hove Ravn‡

Emiliano Santoro§

Abstract

We report extensive evidence indicating that U.S. house prices persistently increase in the face of positive shocks to fiscal spending. In sharp contrast with these findings, though, house prices are found to fall in a large variety of dynamic general equilibrium models embedding the collateral channel in lender-borrower economies, as pointed out by Khan and Reza (2017). This inconsistency rests on the negative wealth effect induced by a concurrent increase in the present value of lump-sum taxes, which contracts Ricardian households' (i.e., lenders) nondurable consumption and, due to their negative comovement with Ricardian households' marginal utility of consumption, house prices. To address this problem, we devise a model embedding a lender-borrower relationship with two layers of production: a final-good, fully competitive sector, and a monopolistically competitive intermediate goods sector. Combining endogenous entry in the intermediate goods sector with a certain degree of "taste for variety" generates increasing returns to scale that, conditional on a positive shock to fiscal spending, overcome the negative wealth effect experienced by Ricardian households, ultimately increasing their nondurable consumption and,

*The authors wish to thank for their comments Vivien Lewis and Ivan Petrella. The usual disclaimer applies.

†Department of Economics, University of Copenhagen, Øster Farimagsgade 5, Bld. 26, 1353 Copenhagen, Denmark. *E-mail*: rbl@econ.ku.dk.

‡Department of Economics, University of Copenhagen, Øster Farimagsgade 5, Bld. 26, 1353 Copenhagen, Denmark. *E-mail*: Soren.Hove.Ravn@econ.ku.dk.

§Department of Economics, University of Copenhagen, Øster Farimagsgade 5, Bld. 26, 1353 Copenhagen, Denmark. *E-mail*: Emiliano.Santoro@econ.ku.dk.

thus, house prices. Introducing taste for variety à la Benassy (1996) allows us to disentangle the taste parameter from the degree of market power, so that we are able to generate an increase in house prices while obtaining plausible estimates of the markup.

1 Introduction

The interplay between house prices and borrowers' capacity to access credit has been widely recognized as an important channel of the monetary transmission mechanism (see, e.g., Iacoviello (2005)), and has proven to be a crucial aspect of the deleveraging episodes following credit-fueled expansions, as observed during the Great Recession. In this respect, the last financial crisis has also fostered a renewed interest in the effects of fiscal policy as a tool for attenuating the effects of the recession and sustaining the subsequent recovery. The scope of this paper is to unveil the transmission of shocks to fiscal spending shocks on house prices, and to produce a theoretical model capable of framing this channel.

Khan and Reza (2017) have recently reported structural Vector Autoregression (VAR) evidence that real house prices persistently rise after a positive government spending shock. In sharp contrast with their findings, though, they highlight that real house prices fall in a large variety of dynamic stochastic general equilibrium (DSGE, hereafter) models embedding the collateral channel in lender-borrower economies, in the vein of Kiyotaki and Moore (1997) and Iacoviello (2005). In fact, it is possible to show that any model in which a Ricardian household participates in the housing market, even as the only type of household in the economy, will feature this type of property. Why is this the case? As originally highlighted by Barsky et al. (2007) to explain the counterfactual negative comovement between durable and nondurable goods consumption in the face of a monetary policy shock, the problem lies in the fact that, from the perspective of the lender—typically a Ricardian household—housing features an approximately constant shadow value. Two key features lead to this property. First, the marginal utility of housing depends on the stock of housing, which is weakly affected by changes in the flow of housing. Second, temporary shocks—as those to government spending— exert little influence on the future marginal utility of housing.[1] Following an increase in government spending, the present value of disposable income drops, thus raising the shadow value of lenders' income, and reducing their consumption. Since the shadow value of housing remains approximately constant, the relative price of housing must track the behavior of nondurable consumption.

To overcome this structural limitation, we devise a model embedding a lender-borrower

[1]In this respect, housing preference shocks represent an exception, as they feature directly in the housing Euler equation, thus breaking the direct link between the house price and the marginal utility of consumption. See, e.g., Iacoviello and Neri (2010) or Liu et al. (2013).

relationship with two layers of production: A final-good, fully competitive sector, and a monopolistically competitive intermediate goods sector. Combining endogenous entry in the intermediate goods sector with a certain degree of 'taste for variety' generates increasing returns to scale at the aggregate level that overcome the negative wealth effect induced by an increase in fiscal spending (financed either through a tax hike or an increase in government debt).[2] How is this possible? To address this question, it is instructive to examine the labor market equilibrium. An expansion in government spending typically leads to an increase in labor supply, at given factor prices. In a standard economy with no entry, holding the number of intermediate good producers fixed would consequently lead to a fall in the real wage, thus exacerbating the fall in the present value of disposable income. With free entry, instead, enhanced profit opportunities determine an increase in the number of intermediate producers: Despite output is a constant returns function to the primary factors of production—for a given measure of intermediates—changes in the number of firms shift the relationship between output and the production factors endogenously, as in Devereux et al. (1996), so that total factor productivity (TFP, hereafter) increases. The strength of this channel depends on two mutually reinforcing factors: The elasticity of substitution in the demand for intermediate goods, and the degree of love for variety. If the endogenous response of TFP to the fiscal stimulus is strong enough, the increase in the real wage can lead to a substitution out of leisure and into consumption, for both borrowers and—for the sake of generating a positive response of house prices—lenders.

Following Christiano et al. (2005), among others, we validate the model by matching its impulse responses to the empirical responses of a structural VAR that includes TFP and a measure of the real wage in the set of endogenous variables originally considered by Khan and Reza (2017). In line with our model, both variables increase following a positive shock to fiscal spending. In fact, matching the real wage response helps us obtain a large increase in TFP, which is key to generate a crowding-in effect on private consumption and, thus, an increase in house prices. Moreover, our estimation strategy allows us to obtain independent estimates of the parameter controlling the taste for variety and the elasticity of substitution in the demand of intermediate goods. Unlike Devereux et al. (1996)—where obtaining a crowding-in effect of fiscal spending shocks typically requires an average mark-up that is substantially higher that what observed in the data—our strategy returns estimates of the average mark-up that are in line with the data.

[2] As shown by Devereux et al. (1996), this effect could be attained by simply accounting for endogenous entry, though at the cost of obtaining implausibly high degrees of market power.

Related literature Our paper builds on a literature that stresses the role of endogenous firm entry, variety effects, and monopolistic competition in propagating business cycles.

Devereux et al. (1996) develop a real business cycle model close to the one studied here, and analyze the implications for the propagation of government spending shocks. They show that endogenous entry can generate consumption crowding-in through increasing returns stemming from taste for variety. However, the parameter controlling the strength of the variety effect is tied down by the inverse of the elasticity of substitution in the Dixit and Stiglitz (1977) CES aggregator. For plausible values of the markup, this implies that the degree of increasing returns is too weak to generate consumption crowding-in, unless labor supply is extremely elastic (Bilbiie, 2011).

To overcome this limitation, we use a variant of the CES function with generalized love for variety studied by Benassy (1996).[3] This function disentangles market power from love of variety such that a strong degree of increasing returns to scale is possible without requiring an implausibly high markup. Several recent papers have also used this specification of the CES function to analyze the implications of endogenous entry and product variety for optimal fiscal policy (Chugh and Ghironi, 2011), optimal monetary policy (Bilbiie et al., 2014), inefficiencies related to endogenous product variety under monopolistic competition (Bilbiie et al., 2019), and the transmission of monetary policy (Bilbiie, forthcoming). Welfare implications and optimal policy, however, depend on the strength of the variety effects.

There is scant evidence on the plausible values for the parameter governing the extent of love of variety (Chugh and Ghironi, 2011; Bilbiie et al., 2019). Similarly, Lewis and Poilly (2012) analyze the transmission of monetary policy in the presence of variety effects and stress that the parameter is poorly identified when using the CES function with generalized love of variety. Contrary to this model, however, our model includes a housing market, and the strength of the variety effect is important for the model's ability to generate a rise in house prices after a positive government spending shock.

Besides generating endogenous TFP fluctuations, firm entry also affects firm competition in our model. Since more firms generate stronger competition, procyclical entry gives rise to countercyclical movements in the markup. This competition effect interacts with the variety effect since markups also determine firms' entry decision. In this regard, this paper is related to the work of Jaimovich (2007), Jaimovich and Floetotto (2008), Lewis and Poilly (2012), and Lewis and Winkler (2017).

[3]This variant of the CES function was also studied in the working paper version of Dixit and Stiglitz (1977).

Structure The paper proceeds as follows. Section 2 reports extensive evidence based on both a structural VAR and regional data on the response of U.S. house prices in the face of shocks to fiscal spending. In Section 3 we outline the details of the model. Section 4 describes our calibration and estimation. In Section 5 we report and discuss model dynamics. Section 6 concludes.

2 Empirical evidence

In this section we provide empirical evidence to support the claim that increases in government spending have a positive effect on house prices in the United States. We first rely on a structural VAR approach, following the tradition of most of the empirical literature on the aggregate effects of government spending shocks. To corroborate our VAR-based findings, we then resort to a more disaggregated analysis by studying how local house prices respond to regional differences in the stance of federal spending.

2.1 Aggregate evidence: A structural VAR model

We begin by setting up a structural VAR model for the U.S. economy. Since the work of Blanchard and Perotti (2002), this approach has served as the workhorse for most empirical analyses of the dynamic effects of changes in fiscal policy. To account for the risk that government spending shocks identified using the Cholesky decomposition proposed by Blanchard and Perotti (2002) may suffer from anticipation effects, we use the survey-based forecast errors of the growth rate of government spending computed by Auerbach and Gorodnichenko (2012) to identify truly unexpected shocks. This is done by including these forecast errors (denoted FE_t) in our VAR model. We begin with a relatively parsimonious VAR model featuring six variables other than FE_t.[4] The variables are: Real government consumption and investment (G_t), real GDP (Y_t), real private consumption (C_t), real net tax revenues (T_t), real mortgage debt (D_t), and the real house price (Q_t). All variables are in log per capita terms (except the house price, which is only in logs). The construction of the net tax measure follows that of Khan and Reza (2017), and the variable is included to account for the overall stance of fiscal policy. We use the Median Sales Price of Houses Sold, which is constructed by the U.S.

[4]In the impulse-response matching exercise we perform in Section 4, we augment the VAR model with the real wage and total factor productivity, as these are crucial to the mechanism we are going to embed in our DSGE model.

Census Bureau, and we deflate it using the GDP deflator.[5] The data sample is 1966:Q4-2010:Q3, as dictated by the availability of FE_t from Auerbach and Gorodnichenko (2012). Additional details regarding our data are provided in Appendix A.

We estimate a VAR model with four lags, a constant, and linear and quadratic time trends. The ordering of the variables is the following:

$$\mathbf{X}_t = \begin{bmatrix} FE_t & G_t & Y_t & C_t & T_t & D_t & Q_t \end{bmatrix}'.$$

This ordering reflects our identification strategy: The forecast errors are ordered first in the system, as these are assumed to be orthogonal to the economy in the sense that they do not respond to any of the other variables within-quarter. This allows us to recover a truly unexpected shock to government spending. We follow Auerbach and Gorodnichenko (2012) and order government spending immediately after FE_t, while the ordering of the remaining variables turns out to be of little importance for the results.

In Figure 1 we present the impulse responses to a positive government spending shock normalized to 1 percent, along with 68 and 90 percent confidence bands, obtained using the delta method with 2000 replications. Following such a shock, we observe a hump-shaped increase in government spending itself, as well as in output and private consumption. These responses are largely in line with existing studies; see, e.g., Blanchard and Perotti (2002), Galí et al. (2007), or d'Alessandro et al. (2019). Moreover, we observe persistent and statistically significant increases in the real house price and in mortgage debt. In other words, fiscal stimulus has a clear, expansionary effect on the housing market.[6] These findings are in line with those presented by Khan and Reza (2017). They are also consistent with recent evidence reported by Auerbach et al. (2020), who find that government spending has a stimulative effect on credit markets.

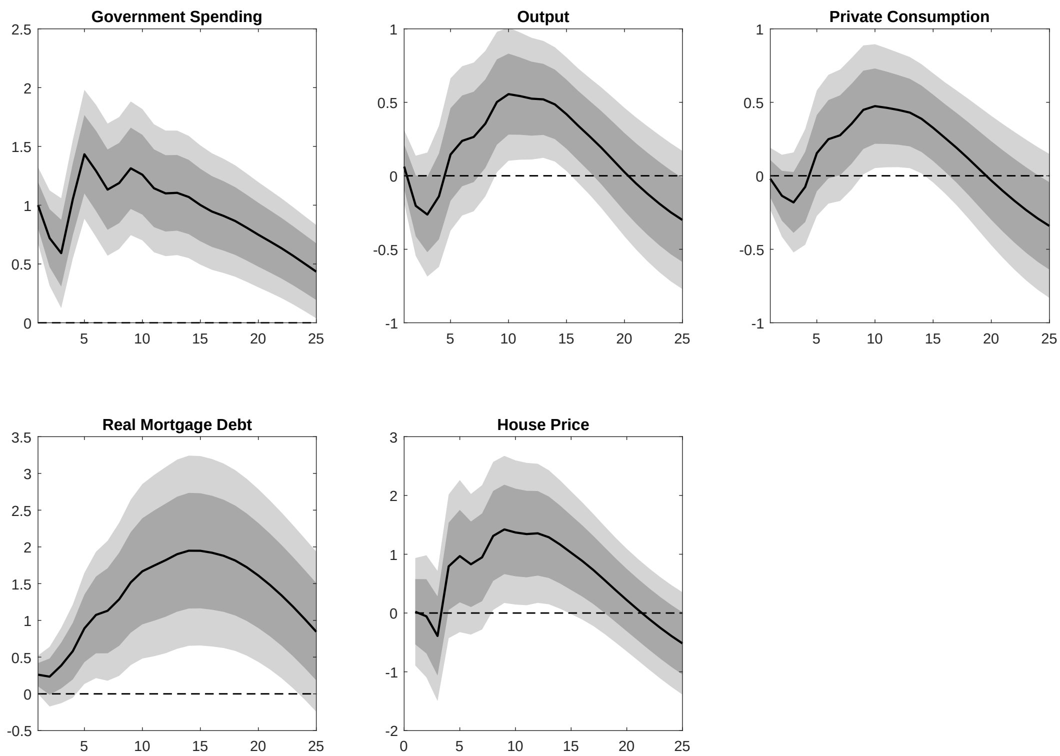
2.2 Regional evidence

The next step consists of studying how a change in federal government spending in a given city relative to another affects relative house price movements between the two

[5]Other popular house price indices, such as the Case-Shiller National Home Price Index or the All-Transactions House Price Index, are only available for shorter samples. Since government spending shocks have been found to be much smaller and less persistent since around 1980 (e.g., Bilbiie et al. (2008)), we prioritize the availability of a long data sample.

[6]In Appendix A we report impulse responses obtained using a traditional Cholesky decomposition. In this case, the data sample can be extended to 1960:Q1-2017:Q2. This exercise confirms all the findings above.

Figure 1: Dynamic effects of a government spending shock



Notes: The figure shows the effects of a shock to government spending (normalized to 1 percent) identified using forecast errors. The black line represents the estimated response, while the grey areas indicate the 68 and 90 percent confidence bands.

cities. We do so by following the approach of Nakamura and Steinsson (2014) and Auerbach et al. (2019), where military procurement is used as a source of regional variation in spending.

Our analysis relies on Department of Defense (DoD, hereafter) contract data from the website USAspending.gov, covering the years of 2001 through 2018. This website contains information on individual prime contracts signed between companies and the DoD, which we aggregate up to the core-based statistical area (CBSA, hereafter) level to get a variable for all DoD contracts obligated annually to each CBSA. We refer to this variable as DoD spending. Additional information on the data and the aggregation procedure is described in Appendix A.2.1. To measure local house prices, we use the Freddie Mac House Price Index, while we normalize DoD spending by local activity using GDP from the BEA. The final panel data set covers 380 CBSAs from 2001 through

2018 at an annual frequency.

We estimate the following regression of house price growth in CBSA i over h years on the cumulative change in DoD spending over the same horizon (normalized by initial GDP):

$$\frac{Q_{i,t+h} - Q_{i,t}}{Q_{i,t}} = \alpha_{i,h} + \eta_{t+h} + \beta_h \frac{\sum_{k=1}^{h} (G_{i,t+h} - G_{i,t})}{Y_{i,t}} + \varepsilon_{i,t+h}, \tag{2.1}$$

where $Q_{i,t}$ is the house price index, $G_{i,t}$ is DoD spending and $Y_{i,t}$ is GDP, $\alpha_{i,h}$ is a CBSA fixed effect that controls for CBSA-specific trends in house prices, while the time fixed effect, η_{t+h}, controls for common, national variation in house prices.[7] All variables are measured in nominal values, though we obtain similar results when using the CBSA-level GDP deflator.[8]

The coefficient of interest is β_h, which measures the growth in house prices over h years when DoD spending grows by 1 % of initial GDP over the same time horizon. The OLS estimate of β_h, however, is likely biased since military contracts tend to flow disproportionately to areas that experience relative bad economic outcomes because of political factors influencing the allocation of contracts (Nakamura and Steinsson, 2014). Together with the importance of house prices in driving regional business cycles over the sample period, this will bias β_h. Moreover, Auerbach et al. (2019) argue that some contracts are anticipated by firms and will not induce actual changes in production, wages and employment. This would attenuate the OLS estimate of β_h toward zero.

We deal with the potential bias by instrumenting the change in local DoD spending with a Bartik (1991) instrument: the change in national DoD spending interacted with the CBSA's average share of national DoD spending over the sample period. This instrument identifies the effect of spending on house prices by relating the changes in DoD spending to the CBSAs' persistent and differential exposure to changes in national military spending.[9] That is, when the federal government expands military spending, some CBSAs tend to receive more DoD contracts than others because they are systematically

[7]The CBSA-level normalized cumulative change in DoD spending is winsorized at the 1 % level by year. Non-winsorized estimates are somewhat smaller in magnitude but qualitatively similar as shown in Appendix A.2.3.

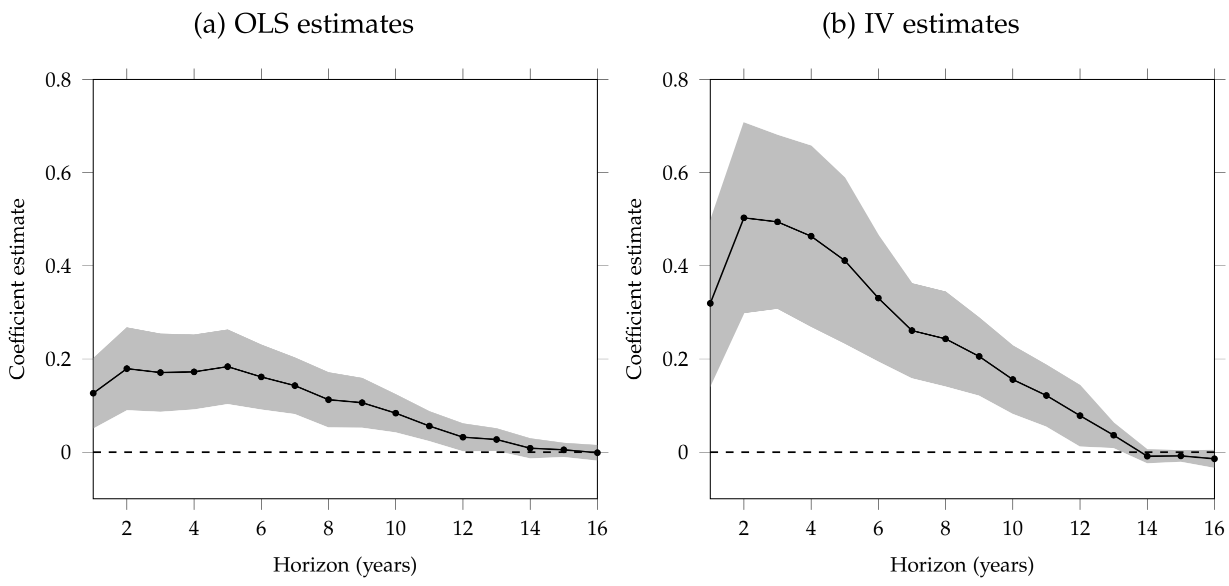
[8]We use nominal values since there are no official statistics that measure cross-regional differences in prices well. The BEA construct CBSA-level GDP deflators by applying national price indices to current dollar values of CBSA-level GDP at the industry level (Bureau of Economic Analysis, 2015). Hence, these statistics do not capture cross-regional differences in prices but instead differences in industry composition.

[9]The instrument is quite strong. The period-by-period first-stage Kleibergen-Papp F-statistics are shown in Figure A.3 and are all well above the rule-of-thumb value of 10.

more exposed to changes in military spending. This systematic component of changes in local DoD spending is isolated by the instrument. The identifying assumption is that there are no confounding factors affecting local house prices that are correlated with the CBSAs' exposure to changes in military spending, both in the cross section as well as the national change in military spending in the time series.

We estimate regression (2.1) separately for $h = 16$ horizons and present the estimates in figure 2. OLS estimates are shown in the panel (a), while IV estimates are shown in the panel (b). Standard errors are heteroskedasticity-robust and clustered at the CBSA level to account for within-CBSA correlation of the error term. 95 percent confidence bands based on the point estimate standard errors are indicated by the grey areas.

Figure 2: The regional response of house prices to military spending



Notes: The figure shows the estimates of β_h from regression (2.1) using an annual panel of 380 CBSAs covering the period 2001-2018. Panel (a) plots the OLS estimates, while IV estimates are plotted in panel (b). Grey areas indicate 95 percent confidence bands constructed using heteroskedasticity-robust standard errors clustered by CBSA.

The IV estimates are in line with the results from the structural VAR analysis. House prices' response to government spending follow a hump-shaped pattern in which they peak in the second year before slowly reverting back to trend after 14 years. In terms of magnitude, the peak estimate of 0.5 is equivalent to a relative increase in house prices of 0.5 % over two years, as a result of a cumulative increase in spending of 1 % of initial GDP over the same horizon. The OLS estimates also follow a hump-shaped pattern but are biased toward zero.

These estimates are robust to a number of alternative specifications of (2.1), as reported in Appendix A.2.3. Specifically, we present results from regressions with real variables, alternative normalizations of DoD spending changes, a proxy for DoD outlays instead of obligations, and controls for differential house price movements associated with potential confounding factors. In addition, we examine the robustness of our results to outliers.

3 The model

We develop a real business cycle model with heterogeneous agents. The economy is populated by two types of households differentiated by their discount factors: Impatient households have a lower discount factor than patient households, and can borrow up to a share of the present value of their housing stock. This implies that patient households act as lenders. Both household types work, consume non-durables and accumulate housing. Patient households also accumulate capital that is rented to firms producing intermediate goods.

Production of non-durables and investment goods occurs in a two-layer production sector, in the vein of Rotemberg and Woodford (1992), Devereux et al. (1996) and Jaimovich (2007), among others. The first production layer consists of a continuum of sectors of measure one. Each sector contains a finite number of firms producing differentiated sector-specific goods using capital and labor as inputs, while firms enter and exit the sectors until a zero-profit condition is satisfied. The differentiated goods enter as imperfect substitutes in an aggregate sectoral good used in the second production layer. That layer consists of a representative firm combining the continuum of aggregate sectoral goods to produce a final good to be sold to the households.

3.1 Households

The economy is populated by two groups of households, each consisting of a continuum of unit mass. Both household types derive utility from nondurable consumption, C_t^j, housing, H_t^j, and the fraction of time devoted to labor, N_j^t, where $j \in \{b, l\}$ indexes impatient-borrowing and patient-lending household-specific variables, respectively. Each type of household maximizes the following life-time utility function:

$$E_0 \left\{ \sum_{t=0}^{\infty} \left(\beta^j\right)^t \left[\frac{\left(C_t^j - h^j C_{t-1}^j\right)^{1-\sigma_c}}{1-\sigma_c} + \Upsilon^j \frac{\left(H_t^j\right)^{1-\sigma_h}}{1-\sigma_h} - \Psi^j \frac{\left(N_t^j\right)^{1+\psi}}{1+\psi} \right] \right\}, \tag{3.1}$$

where $\beta^l > \beta^b$ are the discount factors. This difference in impatience implies that patient households will act as lenders to the impatient households. In addition, $\sigma_c \geq 0$ and $\sigma_h \geq 0$ are the coefficients of relative risk aversion for consumption and housing, respectively, ψ is the inverse Frisch elasticity, and $h^j \in [0;1[$ measures the degree of internal habit formation in consumption, while $\Upsilon^j > 0$ and $\Psi^j > 0$ are the utility weights on housing and labor, respectively.

Impatient households choose consumption, housing, labor and borrowing subject to their budget constraint and a loan-to-value constraint:

$$C_t^b + q_t H_t^b + R_{t-1} B_{t-1}^b = w_t^b N_t^b + B_t^b + q_t H_{t-1}^b - \tau_t^b, \tag{3.2}$$

$$B_t^b \leq \gamma B_{t-1}^b + (1-\gamma) m \frac{q_{t+1} H_t^b}{R_t}, \tag{3.3}$$

where q_t is the price of housing in units of consumption, B_t^b is the stock of debt held at the end of period t, R_t is the gross real interest rate on debt between period t and $t+1$, w_t^b is the real wage of impatient households, and τ_t^b is a lump-sum tax.

The borrowing constraint in equation (3.3) states that the impatient households can partially borrow up to a fraction $m \in [0;1]$ of the present value of their housing stock at the beginning of the next period, as in Kiyotaki and Moore (1997). As in Guerrieri and Iacoviello (2017), we allow for inertia in the dynamics of mortgage debt, as measured by $\gamma \in [0;1[$. We assume that shocks to the economy are sufficiently small such that the borrowing constraint always holds with equality.

Impatient households' behavior is described by the following first-order conditions for consumption, housing, labor, and debt, respectively:

$$\lambda_t^b = \left(C_t^b - h^b C_{t-1}^b\right)^{-\sigma_c} - \beta^b h E_t \left\{ \left(C_{t+1}^b - h^b C_t^b\right)^{-\sigma_c} \right\}, \tag{3.4}$$

$$q_t \lambda_t^b = \Upsilon^b \left(H_t^b\right)^{-\sigma_h} + \beta^b E_t \left\{ \lambda_{t+1}^b q_{t+1} \right\} + E_t \left\{ \mu_t^b m (1-\gamma) \frac{q_{t+1}}{R_t} \right\}, \tag{3.5}$$

$$w_t^b \lambda_t^b = \Psi^b \left(N_t^b\right)^{\psi^b}, \tag{3.6}$$

$$\lambda_t^b + \beta^b \gamma E_t \left\{ \mu_{t+1}^b \right\} = \mu_t^b + \beta^b E_t \left\{ \lambda_{t+1}^b R_t \right\}, \tag{3.7}$$

where λ_t^b and μ_t^b are the multipliers on the budget and borrowing constraints, respectively.

Patient households choose consumption, housing, labor, capital, investment, and bond holdings subject to a budget constraint:

$$C_t^l + q_t H_t^l + I_t + B_t^l = w_t^l N_t^l + q_t H_{t-1}^l + R_{t-1} B_{t-1}^l + r_t^k K_{t-1} - \tau_t^l, \tag{3.8}$$

where I_t is investment in capital, B_t^l are one-period bonds at the end of period t, w_t^l is the real wage of patient households, r_t^k is the real rental rate of capital and τ_t^l is a lump-sum tax. We assume that capital rented to the firms evolves according to the following law of motion:

$$K_t = K_{t-1}(1-\delta) + I_t (1-\Phi_t), \tag{3.9}$$

where $\delta \in [0;1]$ denotes the depreciation rate and $\Phi_t = \frac{\phi}{2}\left(\frac{I_t}{K_{t-1}} - \delta\right)^2 \frac{K_{t-1}}{I_t}$ are convex capital adjustment costs, with $\phi > 0$.

The first-order conditions with respect to C_t^l, H_t^l, N_t^l, B_t, K_t and I_t are

$$\lambda_t^l = \left(C_t^l - h^l C_{t-1}^l\right)^{-\sigma_c} - \beta^l h^l E_t \left\{ \left(C_{t+1}^l - h^l C_t^l\right)^{-\sigma_c} \right\}, \tag{3.10}$$

$$q_t \lambda_t^l = \Upsilon^l \left(H_t^l\right)^{-\sigma_h} + \beta^l E_t \left\{ \lambda_{t+1}^l q_{t+1} \right\}, \tag{3.11}$$

$$w_t^l \lambda_t^l = \Psi^l \left(N_t^l\right)^{\varphi^l}, \tag{3.12}$$

$$\lambda_t^l = E_t \left\{ \lambda_{t+1}^l \beta^l R_t \right\}, \tag{3.13}$$

$$q_t^k = \beta^l E_t \left\{ \frac{\lambda_{t+1}^l}{\lambda_t^l} \left[r_{t+1}^k + q_{t+1}^k \left((1-\delta) - \phi \left(\frac{I_{t+1}}{K_t} - \delta \right) \left(\frac{1}{2} \left(\frac{I_{t+1}}{K_t} - \delta \right) - \frac{I_{t+1}}{K_t} \right) \right) \right] \right\}, \tag{3.14}$$

$$q_t^k = \left[1 - \phi \left(\frac{I_t}{K_{t-1}} - \delta \right) \right]^{-1}, \tag{3.15}$$

where λ_t^b is the multiplier on the budget constraint and q_t^k is the relative price of capital in terms of consumption.

3.2 Production

Production occurs in two stages. A first layer of intermediate good firms produces distinct intermediate goods using capital rented from the patient households and labor supplied by both household types. There exists a continuum of sectors indexed by $j \in [0;1]$, with each of these sectors consisting of $F_t(j)$ intermediate good firms. These firms sell their goods to a representative final good firm in a monopolistic competitive market subject to free entry. Second, the final good firm transforms the intermediate goods into aggregate sectoral goods, $\{Q_t(j)\}_{j=0}^{1}$, which in turn are aggregated into a final good, Y_t, that is sold to the households in a perfectly competitive market.

3.2.1 Final goods production

The final good, Y_t, is produced by a representative firm using a CES production function that aggregates a continuum of measure one aggregate sectoral goods:

$$Y_t = \left[\int_0^1 Q_t(j)^{\omega} dj\right]^{\frac{1}{\omega}}, \qquad \omega \in]0;1[. \tag{3.16}$$

Each intermediate good sector consists of $F_t(j) > 1$ firms producing differentiated goods that are aggregated into a sectoral goods using the following aggregation function proposed by Benassy (1996):

$$Q_t(j) = F_t(j)^{\tau+\frac{\rho-1}{\rho}} \left[\sum_{i=1}^{F_t(j)} m_t(j,i)^{\rho}\right]^{\frac{1}{\rho}} \qquad \rho \in]0;1[, \tag{3.17}$$

where $m_t(j,i)$ is the output of firm i in sector j.

The production function in equation (3.17) is a generalization of the Dixit and Stiglitz (1977) CES aggregation function that disentangles the variety effect from the elasticity of substitution across inputs. The variety effect is measured by $\tau \geq 0$ and implies that as the number of intermediate firms within a sector increases, the sectoral aggregate good increases for a given input of intermediate goods. If $\tau = -\frac{\rho-1}{\rho}$, the function reduces to the Dixit-Stiglitz function in which the variety effect is tied to the elasticity of substitution, while $\tau = 0$ implies that the variety effect is eliminated.

The final good firm's demand for each sectoral aggregate good, $Q_t(j)$, is given by the following standard demand function:

$$Q_t(j) = \left(\frac{p_t(j)}{P_t}\right)^{\frac{1}{\omega-1}} Y_t, \tag{3.18}$$

where $p_t(j)$ is the price index for the sector j aggregate good and $P_t = \left[\int_0^1 p_t(j)^{\frac{\omega}{\omega-1}} dj\right]^{\frac{\omega-1}{\omega}}$ is the price index of the final good.

Similarly, the demand for good $m_t(j,i)$ follows from solving the final good firm's cost minimization problem and is given by

$$m_t(j,i) = \left(\frac{p_t(j,i)}{p_t(j)}\right)^{\frac{1}{\rho-1}} \left(\frac{p_t(j)}{P_t}\right)^{\frac{1}{\omega-1}} \frac{Y_t}{\left(F_t(j)^{\tau+\frac{\rho-1}{\rho}}\right)^{\frac{\rho}{\rho-1}}}, \tag{3.19}$$

where $p_t(j,i)$ is the price of $m_t(j,i)$, and the sectoral price index is equal to

$$p_t(j) = \frac{1}{F_t(j)^{\tau+\frac{\rho-1}{\rho}}} \left[\sum_{i=1}^{F_t(j)} p_t(j,i)^{\frac{\rho}{\rho-1}}\right]^{\frac{\rho-1}{\rho}}. \tag{3.20}$$

Firms sell the final good to the households in a competitive market, which implies that their price is equal to marginal costs:

$$P_t = \left(\int_0^1 p_t(j)^{\frac{\omega}{\omega-1}} dj\right)^{\frac{\omega-1}{\omega}}. \tag{3.21}$$

3.2.2 Intermediate goods production

Each intermediate good, $m_t(j,i)$, is produced using capital and labor according to the following constant-returns-to-scale production technology for firm i in each sector j:

$$m_t(j,i) = z_t k_{t-1}(j,i)^{\mu} \left[\left(n_t^b(j,i)\right)^{\alpha} \left(n_t^l(j,i)\right)^{1-\alpha}\right]^{1-\mu} - \varphi, \qquad \alpha,\mu \in]0;1[, \tag{3.22}$$

where $\varphi > 0$ is a fixed cost of production, $k_{t-1}(j,i)$ is firm-level capital input, $n_t^b(j,i)$ is firm-level labor input from the impatient households, $n_t^l(j,i)$ is firm-level labor input from the patient households, and z_t is economy-wide productivity.

Firms sell the intermediate good to the final good firms in a monopolistic competitive market within each sector. In doing so, they take their effect on the sectoral price index, $p_t(j)$, but not the final good price, P_t, into account following Jaimovich (2007). Thus, the elasticity of demand for the intermediate firm according to the demand curve (3.19) and the price index (3.20) is

$$\varepsilon_{m_t(j,i)} = \frac{1}{\rho-1} + \left(\frac{1}{\omega-1} - \frac{1}{\rho-1}\right)\left(\frac{p_t(j,i)}{p_t(j)F_t(j)^{\tau}}\right)^{\frac{\rho}{\rho-1}} \frac{1}{F_t(j)}. \tag{3.23}$$

We assume that the elasticity of substitution within sectors is higher than the elasticity of substitution across sectors, $\frac{1}{1-\omega} < \frac{1}{1-\rho}$.[10] This implies that if an individual firm increases its price, $p_t(j,i)$, relative to the sectoral price index adjusted for the variety effect,

[10] This assumption is consistent with the evidence by Broda and Weinstein (2006), who show that as product categories are disaggregated, varieties become increasingly substitutable.

$p_t(j)F_t(j)^{\tau}$, the elasticity of demand increases since the demand for the aggregate sectoral good falls through the firm's effect on the sectoral price index.[11] The strength of this competition effect is decreasing in the number of firms, $F_t(j)$, within the sector.

The elasticity of demand in equation (3.23) results in firms setting prices at the following markup over marginal costs:

$$x_t(j,i) = \frac{\varepsilon_{m_t(j,i)}}{1+\varepsilon_{m_t(j,i)}}, \tag{3.24}$$

which is a decreasing function of the number of firms and converges to the standard constant markup $\frac{1}{\rho}$ as $F_t(j) \to \infty$ and to $\frac{1}{\omega}$ as $F_t(j) \to 1$. Hence, the markup is bounded between $\frac{1}{\rho}$ and $\frac{1}{\omega}$.

Firms buy labor and capital inputs in competitive markets so their factor demand curves are given by

$$r_t^k = \mu \frac{p_t(j,i)}{P_t} \frac{z_t k_{t-1}(j,i)^{\mu} \left[\left(n_t^b(j,i)\right)^{\alpha} \left(n_t^l(j,i)\right)^{1-\alpha}\right]^{1-\mu}}{x_t(j,i) k_{t-1}(j,i)}, \tag{3.25}$$

$$w_t^b = (1-\mu)\alpha \frac{p_t(j,i)}{P_t} \frac{z_t k_{t-1}(j,i)^{\mu} \left[\left(n_t^b(j,i)\right)^{\alpha} \left(n_t^l(j,i)\right)^{1-\alpha}\right]^{1-\mu}}{x_t(j,i) n_t^b(j,i)}, \tag{3.26}$$

$$w_t^l = (1-\mu)(1-\alpha) \frac{p_t(j,i)}{P_t} \frac{z_t k_{t-1}(j,i)^{\mu} \left[\left(n_t^b(j,i)\right)^{\alpha} \left(n_t^l(j,i)\right)^{1-\alpha}\right]^{1-\mu}}{x_t(j,i) n_t^l(j,i)}, \tag{3.27}$$

while cost minimization by the firms results in the following cost function:

$$C_t(j,i) = A \left(r_t^k\right)^{\mu} \left(w_t^b\right)^{\alpha(1-\mu)} \left(w_t^l\right)^{(1-\alpha)(1-\mu)} \left(\frac{m_t(j,i)+\varphi}{z_t}\right), \tag{3.28}$$

where $A = \frac{1}{(1-\mu)^{1-\mu}(1-\alpha)^{(1-\alpha)(1-\mu)}\mu^{\mu}\alpha^{\alpha(1-\mu)}}$.

We assume that firms can enter and exit the sectors freely. They do so until profits are driven to zero, which results in the following free entry condition:

$$\frac{p_t(j,i)}{P_t} m_t(j,i) = A \left(r_t^k\right)^{\mu} \left(w_t^b\right)^{\alpha(1-\mu)} \left(w_t^l\right)^{(1-\alpha)(1-\mu)} \left(\frac{m_t(j,i)+\varphi}{z_t}\right). \tag{3.29}$$

Combining the free entry condition with the cost function in equation (3.28) and the pricing schedule $\frac{p_t(j,i)}{P_t} = x_t(j,i) \cdot \frac{\partial C_t(j,i)}{\partial m_t(j,i)}$ pins down each firm's production as a function of fixed costs and the markup:

$$m_t(j,i)\left(x_t(j,i)-1\right) = \varphi. \tag{3.30}$$

[11] Notice from (3.20) that the sectoral price index is not equal to an average of the individual firms' prices due to the variety effect.

3.2.3 Symmetric-firm equilibrium

Intermediate firms face identical technology, entry costs and demand curves for their good. Thus, we focus on a symmetric equilibrium in which the number of firms is equalized across sectors and all firms set identical prices and produce the same quantity of output using the same input of productions factors: $\forall(j,i) \in [0;1] \times [1, F_t(j)] : F_t(j) = F_t$, $p_t(j,i) = p_t$, $x_t(j,i) = x_t$, $m_t(j,i) = m_t$, $k_{t-1}(j,i) = k_{t-1}$, $n_t^b(j,i) = n_t^b$, $n_t^l(j,i) = n_t^l$. In addition, market clearing in the capital and labor markets implies that $k_{t-1} = \frac{K_{t-1}}{F_t}$, $n_t^b = \frac{N_t^b}{F_t}$ and $n_t^l = \frac{N_t^l}{F_t}$.

Letting the final good price act as numeraire and using the cost function in equation (3.21) together with the price index for $p_t(j)$ in equation (3.20) allows us to express the relative price p_t as a function of the number of firms:

$$p_t = F_t^{\tau}. \tag{3.31}$$

Moreover, setting $m_t(j,i) = m_t$ in (3.17) results into

$$Y_t = F_t^{1+\tau} m_t. \tag{3.32}$$

Equations (3.31) and (3.32) yield two insights about the variety effect. First, the relative price of an intermediate good to final goods, p_t, is increasing in the number of firms. This results from the increased variety lowering marginal costs for the final goods firm, thereby lowering the price of the final good relative to intermediate goods. Second, a larger number of intermediate firms increases final goods output more than one-for-one, for given firm-level production. Thus, there are increasing returns to the number of firms, while production technology at the intermediate firm level is constant-returns-to-scale.

Identical price setting results in the elasticity of demand being

$$\varepsilon_t = \frac{1}{\rho - 1} + \left(\frac{1}{\omega - 1} - \frac{1}{\rho - 1}\right)\frac{1}{F_t}. \tag{3.33}$$

Inserting this into (3.24) gives a markup that is decreasing in the number of firms:

$$x_t = \frac{(1-\omega) F_t - (\rho - \omega)}{\rho (1-\omega) F_t - (\rho - \omega)}. \tag{3.34}$$

Factor market clearing requires that $F_t k_{t-1} = K_{t-1}$, $F_t n_t^b = N_t^b$ and $F_t n_t^l = N_t^l$. Inserting this into the production function (3.22) together with the free entry condition (3.30) returns:

$$F_t = \frac{x_t - 1}{x_t \phi} z_t K_{t-1}^{\mu} \left[\left(N_t^b\right)^{\alpha} \left(N_t^l\right)^{1-\alpha}\right]^{1-\mu}. \tag{3.35}$$

We can use this equation together with (3.32) to write output as

$$Y_t = \frac{1}{x_t}\left(\frac{x_t-1}{x_t\phi}\right)^{\tau} z_t \left(K_{t-1}^{\mu}\left[\left(N_t^b\right)^{\alpha}\left(N_t^l\right)^{1-\alpha}\right]^{1-\mu}\right)^{1+\tau}, \tag{3.36}$$

from which it is clear that setting $\tau = 0$ implies a constant returns to scale technology. Combining (3.35) and (3.36) yields the number of firms as a function of output and the markup:

$$F_t = \left(\frac{x_t-1}{\varphi}\right)^{\frac{1}{1+\tau}} Y_t^{\frac{1}{1+\tau}}, \tag{3.37}$$

from which we see that the number of firms is procyclical, while markups are countercyclical since the markup is decreasing in the number of firms.

Finally, we solve for the values of r_t^k, w_t^b and w_t^l that are consistent with factor market clearing, (3.31) and (3.34):

$$r_t^k = \frac{\mu}{x_t}\left(\frac{x_t-1}{x_t\varphi}\right)^{\tau}\left[z_t\left(K_{t-1}\right)^{\mu-\frac{1}{1+\tau}}\left(\left(N_t^b\right)^{\alpha}\left(N_t^l\right)^{1-\alpha}\right)^{1-\mu}\right]^{1+\tau}, \tag{3.38}$$

$$w_t^b = \frac{(1-\mu)\alpha}{x_t}\left(\frac{x_t-1}{x_t\varphi}\right)^{\tau}\left[z_t\left(K_{t-1}\right)^{\mu}\left(\left(N_t^b\right)^{\alpha-\frac{1}{(1+\tau)(1-\mu)}}\left(N_t^l\right)^{1-\alpha}\right)^{1-\mu}\right]^{1+\tau}, \tag{3.39}$$

$$w_t^l = \frac{(1-\mu)(1-\alpha)}{x_t}\left(\frac{x_t-1}{x_t\varphi}\right)^{\tau}\left[z_t\left(K_{t-1}\right)^{\mu}\left(\left(N_t^b\right)^{\alpha}\left(N_t^l\right)^{1-\alpha-\frac{1}{(1+\tau)(1-\mu)}}\right)^{1-\mu}\right]^{1+\tau}. \tag{3.40}$$

These equations depict the relationship between factor prices and factor inputs that are consistent with factor market clearing and zero profits, as in Devereux et al. (1996). Importantly, they show that although there are decreasing returns at the firm level, increasing returns can occur at the aggregate level, once firm entry and exit is taken into account. For example, consider an increase in the number of hours. This increases output at the firm level but tends to lower the marginal return on labor. New firms enter, however, as profit opportunities arise. This tends to increase the marginal return on labor as the relative price of intermediate goods becomes higher through the variety effect depicted in (3.31). If the variety effect is sufficiently strong, the latter effect will dominate the former and the real wage goes up.

3.2.4 Endogenous TFP variations

Given its central role in the mechanism we unveil, it is important to spend a few words on the determinants of TFP. To this end, we combine the expression for aggregate output in (3.36) together with the expression for the number of firms in (3.35):

$$TFP_t = \frac{Y_t}{K_{t-1}^{\mu}\left[\left(N_t^b\right)^{\alpha}\left(N_t^l\right)^{1-\alpha}\right]^{1-\mu}} = \frac{F_t^{\tau} z_t^{1-\tau}}{x_t}. \tag{3.41}$$

The entry and exit of firms results in endogenous procyclical TFP variations through two channels: i) a competition effect, and ii) a variety effect. As emphasized by Jaimovich and Floetotto (2008), the competition effect implies that TFP is affected by the response of the markup to changes in the number of firms. To see this, abstract from productivity shocks (i.e., set $z_t = 1$) in (3.41) and consider an increase in the number of firms fostered by a fiscal expansion, which lowers the markup through more intense competition. In turn, this induces firms to increase production to cover their fixed cost. Hence, the ratio of both capital and labor to the fixed cost increase, and so does TFP. The second channel through which the entry and exit of firms affects TFP is the variety effect: Keeping aggregate capital and labor fixed, a higher number of firms increases aggregate output in the final goods sector, as implied by (3.32), thus resulting into a higher TFP. Intuitively, the capacity of both channels increases in τ, and so does the response of TFP.

3.3 Fiscal policy

Government spending follows an autoregressive process:

$$G_t = (1-\gamma_g)\bar{G} + \gamma_g G_{t-1} + \epsilon_{g,t}, \qquad \epsilon_{g,t} \sim N\left(0, \sigma_g^2\right). \tag{3.42}$$

We consider an economy in which the government runs a balanced budget period-by-period, while the households pay a fixed share of the transfers corresponding to their labor income share:

$$\tau_t^b = \alpha G_t, \tag{3.43}$$

$$\tau_t^l = (1-\alpha)\, G_t. \tag{3.44}$$

3.4 Market clearing

The market clearing conditions for non-durables and durables are

$$Y_t = C_t + G_t + I_t, \tag{3.45}$$

$$C_t = C_t^b + C_t^l, \tag{3.46}$$

$$H = H_t^l + H_t^b, \tag{3.47}$$

where H is a fixed stock of housing in the economy.

Lastly, the bond market clears when patient household lending equals impatient household borrowing:

$$B_t^l = B_t^b. \tag{3.48}$$

4 Estimation and calibration

We split the parameters of the model into two groups. The first group of parameters are calibrated, while the second group is estimated using indirect inference.

4.1 Calibration

The vector $\omega_1 = \{\alpha, \beta^b, \beta^l, \delta, \mu, m, \theta, \omega, \rho\}$ contains the parameters that we choose to calibrate. We set the income share of borrowers to $\alpha = 0.21$, in line with the estimate of Iacoviello and Neri (2010). The discount factors of borrowers and lenders are set to $\beta^b = 0.97$ and $\beta^l = 0.99$, respectively, as in Jensen et al. (2018). The depreciation rate of capital is set at $\delta = 0.03$, while the income share of capital is set to $\mu = 0.25$. These values imply ratios of investment to output and of capital to output of 0.19 and 6.2, respectively, both of which are slightly higher than historical averages for the U.S. economy. We set the loan-to-value ratio m to 0.85, as in Iacoviello and Neri (2010). The share of government spending to output, denoted θ, is set to 0.24 in line with the average value over the sample we consider in our VAR model. We then turn to the parameters governing the elasticity of substitution within and across sectors, ρ and ω. We set $\rho = 0.9$ and $\omega = 0.75$ in order to obtain elasticities of substitution of 10 within-sector and 4 across sectors, respectively. We collect the calibrated parameters in Panel A of Table 1.[12]

4.2 Estimation strategy

The remaining parameters are estimated by impulse-response matching, as in Christiano et al. (2005) and Iacoviello (2005), among others. This is done by matching the model-implied impulse responses to a government spending shock to the empirical responses

[12] We also need to set values for the parameters measuring the (dis)utility weights of labor and housing. We set Y to ensure a ratio of housing wealth to output of 1.45 at the annual frequency, as in Jensen et al. (2018). The weight on labor disutility only affects the scale of the economy, and is simply set to 1. These values are of little importance for our results.

presented in Figure 1. We collect in $\omega_2 = \{\phi, \sigma^c, \sigma^h, h_b, h_l, \psi, \gamma, \gamma_G, \sigma_g, \tau, x\}$ the parameters to be estimated. Let $\Gamma(\omega_2)$ denote the model-implied impulse responses, which are functions of the parameters, while $\widehat{\Gamma}$ denotes the corresponding empirical estimates from our VAR model. We obtain the vector of parameter estimates $\widehat{\omega}_2$ by solving:

$$\widehat{\omega}_2 = \arg\min_{\omega_2} \left(\Gamma(\omega_2) - \widehat{\Gamma}\right)' W \left(\Gamma(\omega_2) - \widehat{\Gamma}\right). \tag{4.1}$$

The weighting matrix W is a diagonal matrix with the sample variances of the VAR-based impulse responses along the diagonal. Effectively, this means that we are attaching higher weights to those impulse responses that are estimated most precisely. We match impulse responses for the five variables reported in Figure 1, plus the wage rate and TFP, which we now include in our structural VAR model, using the responses during the first 25 quarters after the shock.[13,14]

4.2.1 Estimation results

We report the estimated parameter values in Panel B of Table 1, as well as the associated standard errors, which are computed using an application of the delta method, as described, e.g., in Hamilton (1994). We first note that most parameters take on values that are generally in line with the existing literature. The degree of habit formation is lower than what is found in most studies. The degree of inertia in mortgage debt is slightly lower than the estimate of Guerrieri and Iacoviello (2017) of 0.7. The estimate of ψ implies a Frisch elasticity of around 3, which is not uncommon in business cycle models with flexible prices. A distinctive trait of our estimates is that the data seem to favor the role of the variety effect over that of the competition effect. In fact, the steady-state markup, x, is estimated very closely to the lower bound given by $\frac{1}{\rho} = 1.11$. As for the estimate of τ, instead, we obtain a value of 4.925, which is higher than what most of the literature typically considers, although very little empirical evidence exists about this parameter (Chugh and Ghironi, 2011; Bilbiie et al., 2019).[15]

[13]For these series, we use the real compensation per hour in the nonfarm business sector and the non-utilization-adjusted productivity measure of Fernald (2014). See Appendix A for details. We include the net tax measure described in Section 2 along with the consumer price index and the short-term nominal interest rate in the VAR model to control for the stance of fiscal and monetary policy.

[14]We implement a penalty function to drive the procedure away from areas of the parameter space for which the model has no unique and determinate solution.

[15]Moreover, it is rather imprecisely estimated, in line with Lewis and Poilly (2012).

Table 1: Parameter values

Panel A: Calibrated parameters

Parameter	Description	Value
β^l	Discount factor, lenders	0.99
β^b	Discount factor, borrowers	0.97
μ	Capital share of production	0.25
δ	Capital depreciation rate	0.03
α	Income share of impatient households	0.21
θ	Ratio of government spending to output	0.24
m	Loan-to-value ratio of borrowers	0.85
ρ	Substitution parameter within sectors	0.9
ω	Substitution parameter across sectors	0.75

Panel B: Estimated parameters

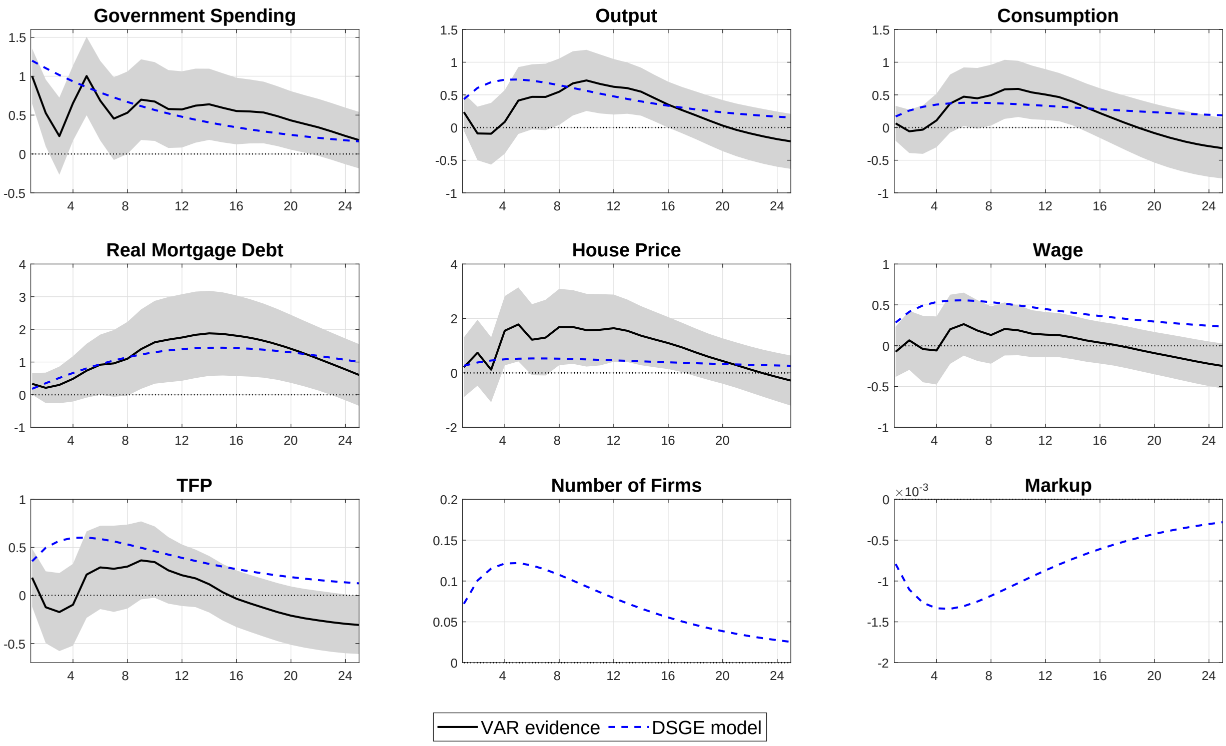
Parameter	Description	Value
ϕ	Capital adjustment cost parameter	4.272 (0.082)
σ_c	Curvature in utility of consumption	1.430 (0.688)
σ_h	Curvature in utility of housing	0.293 (0.457)
h^l	Habit formation, lenders	0.305 (1.191)
h^b	Habit formation, borrowers	0.027 (1.497)
ψ	Inverse Frisch elasticity	0.339 (0.615)
γ	Inertia of mortgage debt	0.556 (0.162)
τ	Love for variety parameter	4.925 (5.554)
x	Steady-state value of markup	1.122 (0.022)
γ_G	Persistence of government spending shock	0.920 (0.020)
σ_g	Std. dev. of government spending shock	0.11 (0.001)

Note: The standard errors of the estimated parameters are reported in brackets.

5 Model dynamics

We report the estimated impulse response functions from the model in Figure 3, alongside their empirical counterparts from the VAR model. The estimated model is able to match the sign of the responses of all variables, and the model-implied responses mostly remain within the confidence bands of the VAR model. Quantitatively, however, the response of the house price in the model falls short of that in the data. Regarding the other variables, the model overestimates the increase in the wage rate and, to a smaller extent, in TFP, while the response of consumption appears too smooth. The increase in TFP implied by our model is in line with recent empirical studies by d'Alessandro et al. (2019) and Jørgensen and Ravn (2018), among others.

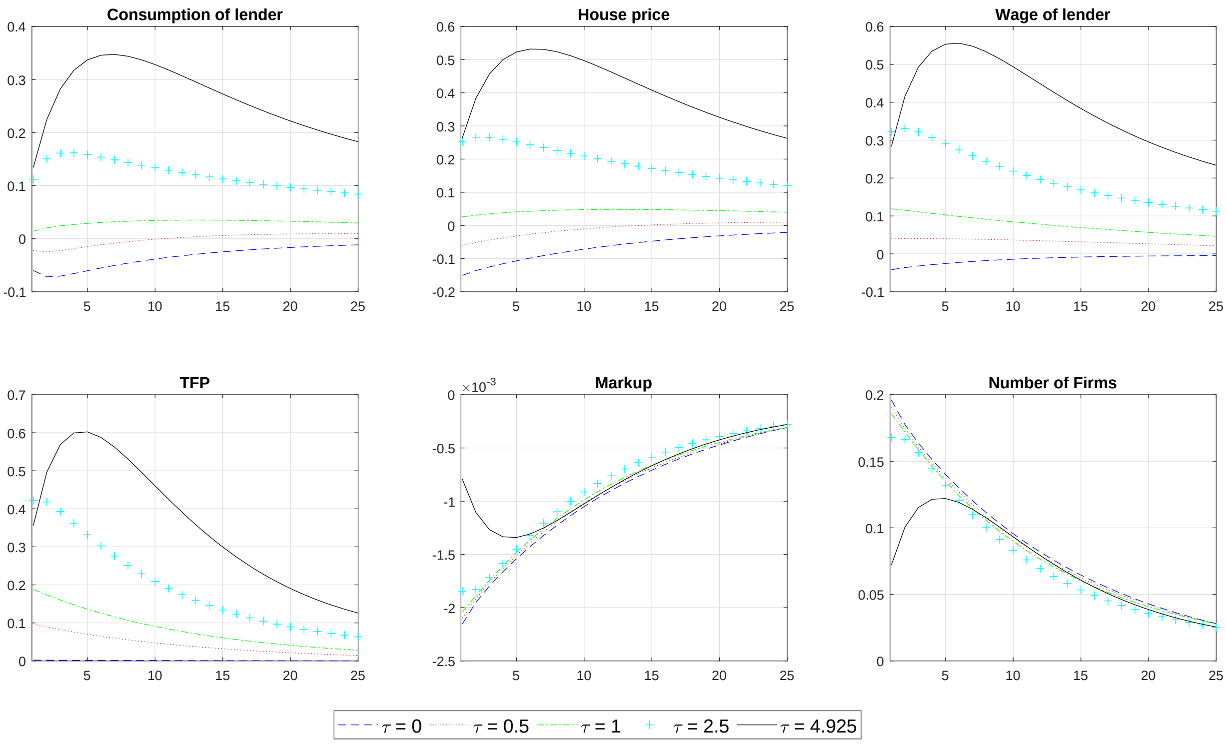
Figure 3: Estimated effects of a government spending shock



Notes: The figure shows the effects of a shock to government spending. Black line: VAR model. Grey areas: 90 percent confidence bands from VAR model. Dashed blue line: Estimated DSGE model.

To inspect the mechanism behind the increase house prices, Figure 4 reports the response of some selected variables in both the calibrated economy and some alternative economies featuring lower or no taste for variety (keeping all other coefficients at the calibrated/estimated value). As discussed in Section 3.2.4, the positive response of TFP is magnified by a positive degree of taste for variety, which amplifies the effect on TFP of the increase in the number of firms, as compared with what happens under $\tau = 0$. For a sufficiently high τ, this reflects into an outward shift in the demand for labor that counteracts the drop in labor supply, ultimately leading to a rise in the real wage.[16] Otherwise, under $\tau = 0$ the contraction in labor supply dominates and the real wage drops.

Figure 4: Effects of a government spending shock for different values of τ



Notes: The figure shows the effects of a shock to government spending for various values of the love-of-variety parameter τ. Dashed line: $\tau = 0$. Dotted line: $\tau = 0.5$. Dashed-dotted line: $\tau = 1$. Crossed line: $\tau = 2.5$. Solid line: $\tau = 4.925$ (estimated value). All other parameters are kept at their baseline values.

Why is this important for house prices? To address this question, we should focus on the responses of q_t and C_t^l. It is immediate to recognize that, under $\tau = 0$, the drop in the real wage is associated with a simultaneous fall in the nondurable consumption of lenders—the usual crowding-out effect of fiscal spending induced by an increase in the present value of lump-sum taxes. Otherwise, under a sufficiently high τ we observe a crowding-in effect, induced by a rise in household income that offsets the negative

[16] In fact, the TFP amplification also reflects into a marked increase in the rental rate of capital, which adds to the upward movement in the real wage, ultimately increasing patient households' income.

wealth effect due to the expected future fiscal tightening. The next logical step consists of recognizing that movements in patient households' consumption are tightly connected to those in the house price. This is true under either value of τ. To see why this is the case, consider patient households' Euler equation for housing (3.11), which may be solved forward to yield an expression for their shadow value of housing:

$$\lambda_t^l q_t = \mathrm{E}_t \left\{ \sum_{t=i}^{\infty} \left(\beta^l\right)^i \Upsilon^l \left(H_{t+i}^l\right)^{-\sigma_h} \right\} \equiv \Lambda_t. \tag{5.1}$$

Since housing does not depreciate, H_t^l is effectively an "idealized durable" according to Barsky et al. (2007): This means that the intertemporal elasticity of substitution in housing demand is close to infinite. As a result, any short-term movements in H_t^l—as those generated by a temporary shock to fiscal spending—will affect the right-hand side of (5.1) relatively little, given that β^l is close to one. So, it is possible to approximate

$$\lambda_t^i q_t = \Lambda_t \approx \Lambda. \tag{5.2}$$

This equilibrium condition confines movements in the price of housing to mirror movements in λ_t^l, i.e. the marginal utility of patient households' nondurable consumption. This is confirmed by our numerical analysis.

A key lesson we retrieve from this property is that, in fact, any model where a Ricardian household participates in the housing market—even as the only type of household in the economy—may be able to generate a conditional expansion in house prices, to the extent that it features a decline in the Ricardian households' marginal utility of nondurable consumption.[17] This is also the reason why the alternative frameworks considered by Khan and Reza (2017) are not able to reproduce a conditional increase in house prices. The solution does not rely on breaking the approximately constant shadow value of housing from the perspective of the lenders, but rather on inducing a positive response of their nondurable consumption by overcoming the negative wealth effect. This is the case, for instance, in Devereux et al. (1996), whose model does not contemplate housing, while featuring increasing returns to scale at the aggregate level of production, as in our framework. A key difference from our setting and Devereux et al. (1996), though, is that introducing taste for variety à la Benassy (1996) allows us to disentangle the taste parameter from the markup in the economy. Unlike our predecessors, this choice permits us to

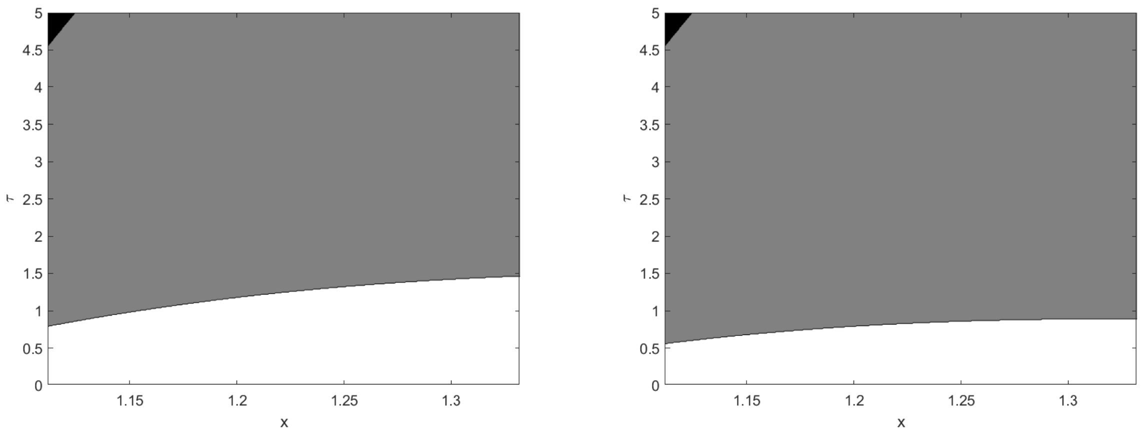
[17]More generally, in any model in which a Ricardian household participates in the housing market, this agent effectively determines the movements of the house price. To overcome this property, some recent studies of house-price dynamics exclude such agents from the housing market; see, e.g., Ferrero (2015) or Garriga et al. (2019).

estimate an empirically plausible markup, ultimately leading to qualitative results that are in line with the empirical evidence.

To shed additional light on the interplay between the taste parameter and the degree of market power, we find it useful to consider Figure 5. In the left panel, we report the combinations of the taste for variety parameter, τ, and the steady-state markup, x, for which the model generates an increase in the house price on impact, holding all other parameters fixed at the values reported in Table 1. In the right panel, we focus on the cumulative response of the house price. The main message from the two panels is the same: An increase in the house price obtains for a wide range of parameter combinations, and requires a sufficiently high degree of taste for variety, as measured by τ. Observe that a high level of the steady-state markup pushes up the minimum value of τ required to obtain an increasing house price. A high steady-state markup implies that fixed costs are relatively high, and there are relatively few firms with a lot of market power within each sector. As a result, firm entry is less responsive to aggregate shocks, and so are markups. Since firm entry is a key driver of increasing returns in our model economy, we need a stronger taste for variety to obtain a sizable increase in TFP required to bring about an equilibrium increase in the nondurable consumption of patient households, and thus of the house price.

Figure 5: House price response for different parameter combinations

(a) Impact response (b) Cumulative response



Notes: The figure shows the model outcomes for different combinations of the parameter values of τ and x. The grey (white) area indicates parameter combinations where the model produces an increase (a decline) in the house price in response to a government spending shock. The black area indicates combinations for which the model does not have a unique and determinate solution. The left panel considers the impact response of the house price, while the right panel considers the cumulated response over 25 periods.

6 Concluding remarks

In this paper we propose a dynamic general equilibrium economy in which we are able to reproduce an increase in house prices following a positive shock to fiscal spending, as observed in extensive empirical evidence. The key mechanism in our model—which rests on the combination of endogenous entry with a certain degree of taste for variety—generates increasing returns to scale at the aggregate level that overcome the negative wealth effect induced by an increase in fiscal spending. In economies that do not share this property, as those considered by Khan and Reza (2017), fiscal expansions are ultimately responsible for a drop in the nondurable component of Ricardian households' consumption, whose movements are tightly connected to those in house prices—as it is generally the case for any type of shock that does not exert a direct impact on the shadow value of housing (see Barsky et al. (2007)). By generating a crowding-in effect on Ricardian households' nondurable consumption, we are able to induce an increase in house prices following a fiscal expansion. A key feature of our modeling strategy, which consists of disentangling the taste for variety from the degree of market power, is that we may generate a conditional increase in house prices, while estimating an empirically plausible markup.

While our estimated model is able to account for the qualitative responses of all variables under consideration, it produces a smaller increase in the house price than that observed in the data. In future work, we plan to augment the model with additional propagation mechanisms, such as variable capital utilization and nominal rigidities, so as to investigate whether an estimated medium-scale version of the model may improve on the quantitative match of the data. Furthermore, we are currently investigating various ways to shed more light on what constitutes an empirically realistic degree of taste for variety.

References

Auerbach, A., Gorodnichenko, Y., and Murphy, D. (2019). Local Fiscal Multipliers and Fiscal Spillovers in the United States. *NBER Working Papers*, 25457.

Auerbach, A. J. and Gorodnichenko, Y. (2012). Measuring the Output Responses to Fiscal Policy. *American Economic Journal: Economic Policy*, 4(2):1–27.

Auerbach, A. J., Gorodnichenko, Y., and Murphy, D. (2020). Effects of fiscal policy on credit markets. *NBER Working Papers*, 26655.

Barsky, R. B., House, C. L., and Kimball, M. S. (2007). Sticky-price models and durable goods. *American Economic Review*, 97(3):984–998.

Bartik, T. J. (1991). *Who Benefits from State and Local Economic Development Policies?* W.E. Upjohn Institute.

Benassy, J.-P. (1996). Taste for variety and optimum production patterns in monopolistic competition. *Economics Letters*, 52(1):41–47.

Bilbiie, F. O. (2011). Nonseparable Preferences, Frisch Labor Supply, and the Consumption Multiplier of Government Spending: One Solution to a Fiscal Policy Puzzle. *Journal of Money, Credit and Banking*, 43(1):221–251.

Bilbiie, F. O. (forthcoming). Monetary Neutrality with Sticky Prices and Free Entry. *The Review of Economics and Statistics*.

Bilbiie, F. O., Fujiwara, I., and Ghironi, F. (2014). Optimal monetary policy with endogenous entry and product variety. *Journal of Monetary Economics*, 64:1–20.

Bilbiie, F. O., Ghironi, F., and Melitz, M. J. (2019). Monopoly Power and Endogenous Product Variety: Distortions and Remedies. *American Economic Journal: Macroeconomics*, 11(4):140–174.

Bilbiie, F. O., Meier, A., and Müller, G. J. (2008). What accounts for the changes in us fiscal policy transmission? *Journal of Money, Credit and Banking*, 40(7):1439–1470.

Blanchard, O. and Perotti, R. (2002). An empirical characterization of the dynamic effects of changes in government spending and taxes on output. *Quarterly Journal of Economics*, 117(4):1329–1368.

Broda, C. and Weinstein, D. E. (2006). Globalization and the Gains From Variety. *Quarterly Journal of Economics*, 121(2):541–585.

Bureau of Economic Analysis (2015). GDP by Metropolitan Area Methodology.

Christiano, L. J., Eichenbaum, M., and Evans, C. L. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy*, 113(1):1–45.

Chugh, S. and Ghironi, F. (2011). Optimal Fiscal Policy with Endogenous Product Variety. *NBER Working Papers*, 17319.

d'Alessandro, A., Fella, G., and Melosi, L. (2019). Fiscal stimulus with learning-by-doing. *International Economic Review*, 60(3):1413–1432.

Devereux, M. B., Head, A. C., and Lapham, B. J. (1996). Monopolistic competition, increasing returns, and the effects of government spending. *Journal of Money, Credit and Banking*, 28(2):233.

Dixit, A. K. and Stiglitz, J. E. (1977). Monopolistic Competition and Optimum Product Diversity. *American Economic Review*, 67(3):297–308.

Fernald, J. (2014). A quarterly, utilization-adjusted series on total factor productivity.

Ferrero, A. (2015). House Price Booms, Current Account Deficits, and Low Interest Rates. *Journal of Money, Credit and Banking*, 47(S1):261–293.

Galí, J., López-Salido, J. D., and Vallés, J. (2007). Understanding the effects of government spending on consumption. *Journal of the european economic association*, 5(1):227–270.

Garriga, C., Manuelli, R., and Peralta-Alva, A. (2019). A macroeconomic model of price swings in the housing market. *American Economic Review*, 109(6):2036–72.

Guerrieri, L. and Iacoviello, M. (2017). Collateral constraints and macroeconomic asymmetries. *Journal of Monetary Economics*, 90:28–49.

Guren, A., McKay, A., Nakamura, E., and Steinsson, J. (2018). Housing Wealth Effects: The Long View. NBER Working Paper No. 24729.

Hamilton, J. D. (1994). *Time series analysis*, volume 2.

Iacoviello, M. (2005). House Prices, Borrowing Constraints, and Monetary Policy in the Business Cycle. *American Economic Review*, 95(3):739–764.

Iacoviello, M. and Neri, S. (2010). Housing Market Spillovers: Evidence from an Estimated DSGE Model. *American Economic Journal: Macroeconomics*, 2(2):125–164.

Jaimovich, N. (2007). Firm dynamics and markup variations: Implications for sunspot equilibria and endogenous economic fluctuations. *Journal of Economic Theory*, 137(1):300–325.

Jaimovich, N. and Floetotto, M. (2008). Firm dynamics, markup variations, and the business cycle. *Journal of Monetary Economics*, 55(7):1238–1252.

Jensen, H., Ravn, S. H., and Santoro, E. (2018). Changing credit limits, changing business cycles. *European Economic Review*, 102:211–239.

Jørgensen, P. L. and Ravn, S. H. (2018). The inflation response to government spending shocks: A fiscal price puzzle? Working Paper, University of Copenhagen.

Khan, H. and Reza, A. (2017). House prices and government spending shocks. *Journal of Money, Credit and Banking*, 49(6):1247–1271.

Kiyotaki, N. and Moore, J. (1997). Credit cycles. *Journal of Political Economy*, 105(2):211–248.

Lewis, V. and Poilly, C. (2012). Firm entry, markups and the monetary transmission mechanism. *Journal of Monetary Economics*, 59(7):670–685.

Lewis, V. and Winkler, R. (2017). Government spending, entry, and the consumption crowding-in puzzle. *International Economic Review*, 58(3):943–972.

Liu, Z., Wang, P., and Zha, T. (2013). Land-price dynamics and macroeconomic fluctuations. *Econometrica*, 81(3):1147–1184.

Nakamura, E. and Steinsson, J. (2014). Fiscal Stimulus in a Monetary Union: Evidence from US Regions. *American Economic Review*, 104(3):753–792.

Rotemberg, J. J. and Woodford, M. (1992). Oligopolistic Pricing and the Effects of Aggregate Demand on Economic Activity. *Journal of Political Economy*, 100(6):1153–1207.

Saiz, A. (2010). The Geographic Determinants of Housing Supply. *Quarterly Journal of Economics*, 125(3):1253–1296.

A Appendix to the empirical analysis

This appendix contains additional details on the data used in the SVAR and cross-CBSA analyses, as well as some robustness checks.

A.1 Appendix to the structural VAR analysis

A.1.1 Data used in the structural VAR analysis

All data used in the baseline specification of our SVAR model—with the exception of the forecast errors of Auerbach and Gorodnichenko (2012)—are taken from the Federal Reserve Economic Data (FRED) database. The series are described in detail below with series names in FRED indicated in brackets:

G_t: Government consumption expenditure and gross investment (GCEC1, seasonally adjusted, Chained 2009 \$).

Y_t: Real Gross Domestic Product (GDPC1, seasonally adjusted, Chained 2009 \$).

C_t: Real Personal Consumption Expenditures (PCECC96, seasonally adjusted, Chained 2009 \$).

T_t: Government current tax receipts (W054RC1Q027SBEA) + Government income receipts on assets (W058RC1Q027SBEA) + Government current transfer receipts (W060RC-1Q027SBEA) - Government current transfer payments (A084RC1Q027SBEA) - Government interest payments (A180RC1Q027SBEA) - Government subsidies (GDISUBS).[18] All series are seasonally adjusted. We convert from nominal to real terms using the GDP deflator (GDPDEF).

D_t: Home mortgages (liabilities) of households and nonprofit organizations from the Flow of Funds (HMLBSHNO). We convert the series to real terms using the GDP deflator.

Q_t: Median Sales Price of Houses Sold for the United States (MSPUS). We convert the series to real terms using the GDP deflator.

The first five series are converted to per capita terms using the Census Bureau Civilian Population (All Ages) estimates, which we also collect from the FRED database (POP). We then take logs of all variables. Finally, we use the following series of "narrative" shocks to government spending:

FE_t: Forecast error of government spending, computed as the difference between forecasts (obtained from the Greenbook data of the Federal Reserve Board combined

[18] Since the series turns negative at some points in time, we add a constant to it before taking logs.

with the Survey of Professional Forecasters) and the actual, first-release data for the growth rate of government spending. We obtain the series directly from Auerbach and Gorodnichenko (2012).

In the VAR model used for impulse-response matching, we also use:

Raw Total Factor Productivity series constructed by the Federal Reserve Bank of San Francisco based on the methodology of Fernald (2014).[19]

Nonfarm Business Sector: Real Compensation Per Hour (COMPRNFB, Seasonally Adjusted, 2012=100).

Nominal interest rate on 3-month Treasury Bills (TB3MS).

Consumer Price Index for All Urban Consumers: All Items (CPIAUCSL, seasonally adjusted, 2009=100).

A.1.2 Robustness of the VAR results

The figure below reports the impulse responses to a positive government spending shock (normalized to 1 percent) identified using the Cholesky decomposition. As can be seen from the figure, all results from the analysis in the main text are confirmed.

A.2 Appendix to the regional analysis

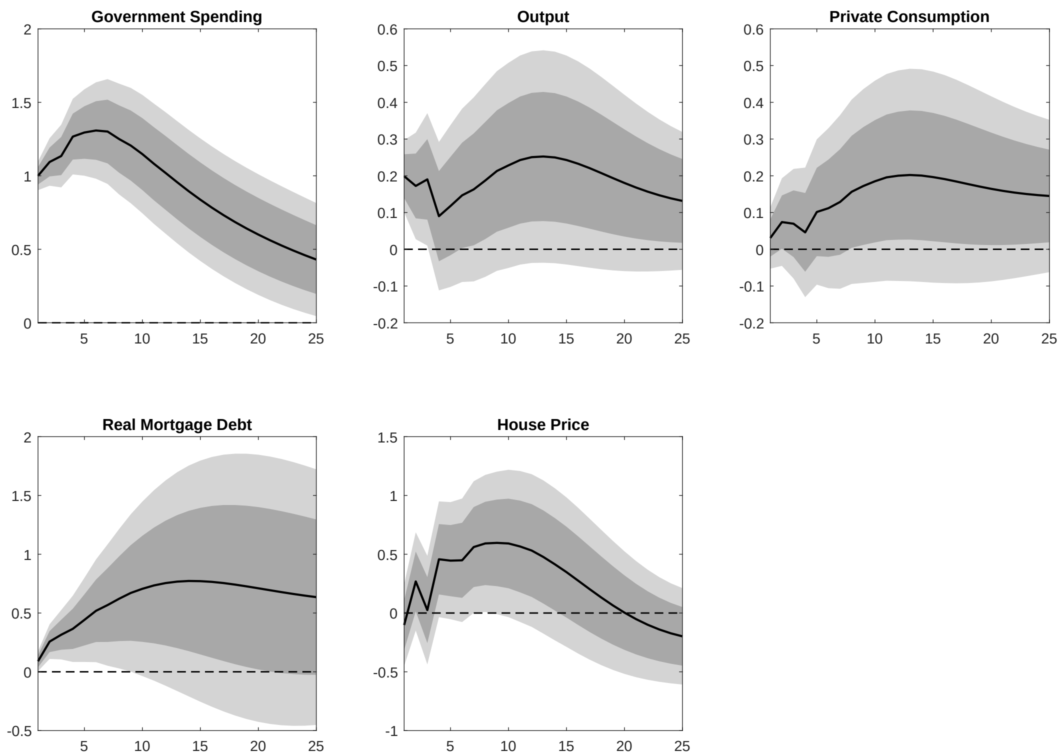
A.2.1 Data used in the regional analysis

We collected data on contracts signed by firms with the Department of Defense from USAspending.gov to construct the data used in section 2.2. The data cover all DoD prime contracts signed from 2001 through 2018 including terminated contracts. The data set does not contain information on the timing of actual outlays to contractors but it does contain information on the duration and total dollar amount obligated per contract. Additionally, the data set contains the name of the contractor and the primary place of work performance at the ZIP code level.

The raw data is cleaned using the same approach as Auerbach et al. (2019). First, we match a terminated contract with its original contract if a de-obligated dollar amount falls within 0.5 % of dollars obligated in another contract and both contracts have the same contractor ID and ZIP code. These matched obligations and de-obligations are

[19]The data can be collected from https://www.frbsf.org/economic-research/indicators-data/total-factor-productivity-tfp/

Figure A.1: Dynamic effects of a government spending shock: Cholesky decomposition



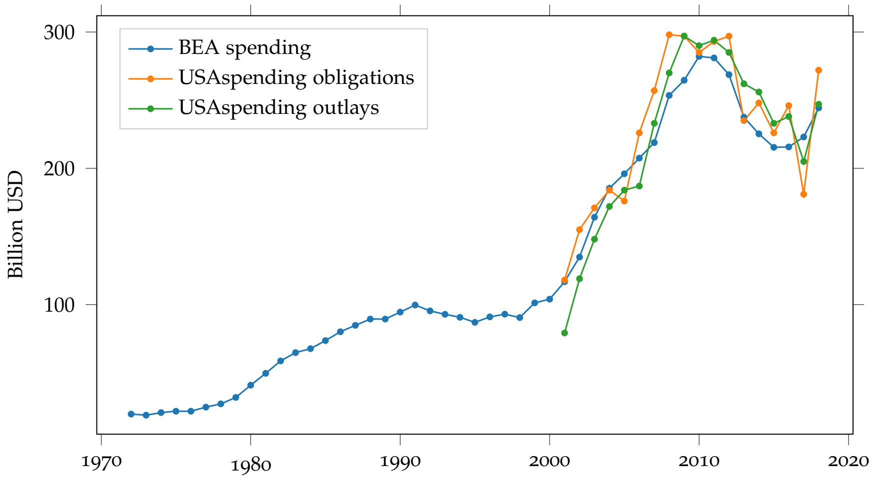
Notes: The figure shows the effects of a shock to government spending (normalized to 1 percent) identified using a Cholesky decomposition with government spending ordered first. The black line represents the estimated response, while the grey areas indicate the 68 and 90 percent confidence bands.

removed from the data set. Second, we remove long-term contracts that terminate after our sample period by removing all contracts that terminate after 2023.

Our baseline estimates use variation in obligations rather than actual outlays. This assigns the entire obligated amount to the first year of the contract. As a robustness check, we construct a proxy for outlays per contract by dividing the dollars obligated in each contract evenly among the months of the contract's duration. We then sum these amounts annually by CBSA to get a proxy for total annual outlays to the CBSAs.

Our data tracks official data on national military spending from the BEA well in terms of both magnitude and movements. This is seen in figure A.2, which plots national obligations and our proxy for outlays according to the data from USAspending together with intermediate goods and services purchased for national defense from the BEA's NIPA tables.

Figure A.2: Military spending according to USAspending and BEA data



Notes: The blue line is "Intermediate goods and services purchased" in the BEA's NIPA Table 3.11.5, "National Defense Consumption Expenditures and Gross Investment by Type.". Orange and green lines are annual obligations and outlays constructed using USAspending.gov data.

A.2.2 First-stage estimates

Figure A.3 shows the period-by-period Kleibergen-Papp F-statistics from the first-stage regression

$$\frac{\sum_{k=1}^{h}\left(G_{i,t+h}-G_{i,t}\right)}{Y_{i,t}} = \tilde{\alpha}_{i,h} + \tilde{\eta}_{t+h} + \tilde{\beta}_h s_i \times \frac{\sum_{k=1}^{h}\left(G_{t+h}^{agg}-G_t^{agg}\right)}{Y_{i,t}} + \epsilon_{i,t+h} \qquad \text{(A.1)}$$

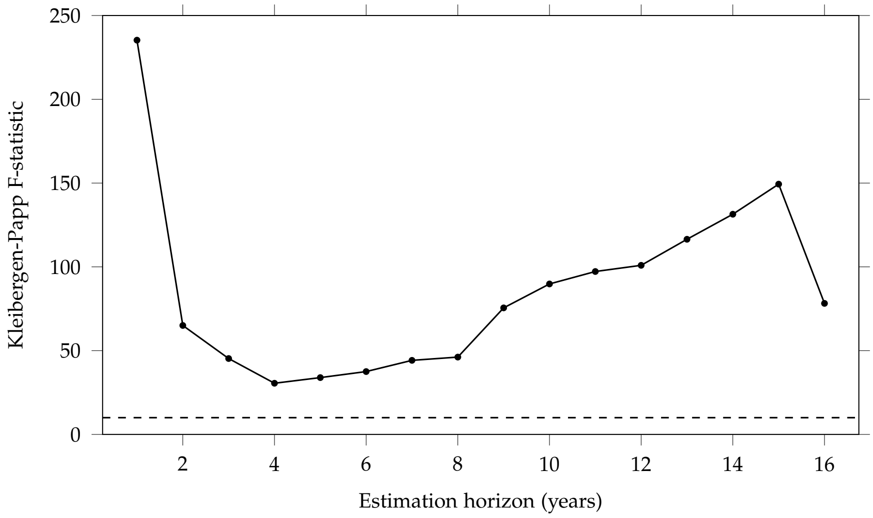
where s_i is CBSA i's average share of annual national DoD obligations over the sample period 2001-2017 and G_t^{agg} is national DoD obligations in period t.

A.2.3 Robustness of the regional estimates

This section analyzes the robustness of our regional estimates in section 2.2 to alternative specifications and potential outliers.

Table A.1 shows the IV estimates from alternative specifications of regression (2.1). Column (1) shows the estimates from a regression in which house prices, DoD spending and GDP have been deflated by the CBSA-level GDP deflator. Column (2) reports estimates from a regresion in which we use the proxy for outlays described in Appendix A.2.1 to measure DoD spending. Columns (3) and (4) present estimates with alternative normalizations of DoD spending (personal income and population in thousand persons). Column (5) controls for house price movements associated with industry composition by

Figure A.3: F-statistics from first-stage regression



Notes: The figure shows the period-by-period Kleibergen-Papp F-statistics from the first-stage regression (A.1). Heteroskedasticity-robust standard errors are clustered by CBSA. The dashed line indicates the rule-of-thumb value of 10 above which the instrument is strong.

adding 2-digit industry employment shares multiplied with year dummies to the regression. Column (6) controls for differential exposure to aggregate house price movements by adding three time-invariant controls multiplied by year dummies to the regression: the Wharton Regulation Index, the Saiz (2010) instrument and the Bartik-like instrument for sensitivity to regional house price movements by Guren et al. (2018). Lastly, column (7) adds state $\times$ year fixed effects to control for state-specific house price fluctuations.

The only alternative estimates that differ from the baseline estimates are those from the specification using the proxy for outlays and the specification with state $\times$ year fixed effects. The latter only uses within-state variation and reduces estimates by around a half but the estimates are still significant and display a hump-shaped pattern. The estimates from the specification using the proxy for outlays instead of obligations are larger. Additionally, they do not display the same hump-shape as the estimates based on obligations.

Table A.2 presents estimates from a sensitivity analysis to outliers. In column (1), we remove all CBSAs in the bottom and top 5th percentiles of the distribution of DoD spending shares used to construct the instrument. Column (2) reports estimates from a regression in which we use the non-winsorized change in local spending. Finally, column (3) shows the estimates when we remove all winsorized observations. The estimates from all three specifications display the same hump-shape as the baseline estimates and are

Table A.1: Robustness of regional estimates (alternative specifications)

	(1) Real variables	(2) Outlays	(3) Normalize by income	(4) Normalize by population	(5) Control for industry comp.	(6) Control for housing exposure	(7) Control for state
1-year estimate	0.31***	1.12***	0.26***	0.0037**	0.36***	0.31***	0.16**
	(0.092)	(0.212)	(0.072)	(0.002)	(0.135)	(0.081)	(0.070)
2-year estimate	0.52***	0.93***	0.41***	0.0076***	0.55***	0.46***	0.24***
	(0.097)	(0.175)	(0.087)	(0.002)	(0.155)	(0.129)	(0.077)
4-year estimate	0.49***	0.64***	0.38***	0.0093***	0.46***	0.36***	0.23***
	(0.093)	(0.116)	(0.090)	(0.002)	(0.114)	(0.103)	(0.061)
10-year estimate	0.17***	0.21***	0.12***	0.0034***	0.17***	0.10***	0.079***
	(0.034)	(0.048)	(0.034)	(0.001)	(0.044)	(0.037)	(0.024)
16-year estimate	-0.0049	-0.035	-0.011	-0.00059*	-0.015	-0.0063	0.00018
	(0.010)	(0.023)	(0.008)	(0.000)	(0.012)	(0.008)	(0.007)
CBSAs	380	380	380	380	380	255	373

Notes: The table presents the IV estimates from alternative specifications of regression (2.1). Column (1) shows the estimates when house prices, DoD spending and GDP are deflated by the CBSA-level GDP deflator. Column (2) uses DoD spending measured by the outlay proxy described in Appendix A.2.1. Column (3) normalizes DoD spending by the BEA's measure of personal income. Column (4) normalizes DoD spending by BEA's measure of population (in thousand persons). Column (5) adds year dummies multiplied the average two-digit industry employments shares over the sample period. The employment shares are calculated using data from the Census' County Business Patterns. Column (6) adds year dummies interacted with three time-invariant measures of exposure to aggregate house price flucuations (the Wharton Regulation Index, the Saiz (2010) instrument and the Guren et al. (2018) instrument). This reduces the sample size since the Wharton Regulation Index and the Saiz (2010) instrument is not available for all CBSAs. Column (7) adds state $\times$ year fixed effects. Heteroskedasticity-robust standard errors clustered by CBSA are shown in parentheses. ***, ** and * denote significance at the 0.01, 0.05 and 0.1 level respectively.

of similar magnitude. However, the peak of the hump is shifted to the fourth year in columns (2) and (3).

Table A.2: Robustness of regional estimates (outliers)

	(1) Remove extreme DoD shares	(2) Non-winsorized	(3) Remove winsorized
1-year estimate	0.35***	0.22***	0.41**
	(0.112)	(0.075)	(0.173)
2-year estimate	0.55***	0.34***	0.48***
	(0.122)	(0.078)	(0.180)
4-year estimate	0.49***	0.40***	0.55***
	(0.113)	(0.111)	(0.112)
10-year estimate	0.16***	0.14***	0.22***
	(0.037)	(0.037)	(0.059)
16-year estimate	-0.0083	-0.012	-0.034
	(0.009)	(0.008)	(0.025)
CBSAs	342	380	379

Notes: The table presents the IV estimates from regression (2.1). Column (1) shows the estimates when removing CBSAs in the bottom and top 5th percentiles of the distribution of average DoD spending shares used to construct the instrument. Column (2) presents estimates when the cumulative change in DoD spending is not winsorized. Column (3) removes all winsorized observations. Heteroskedasticity-robust standard errors clustered by CBSA are shown in parentheses. ***, ** and * denote significance at the 0.01, 0.05 and 0.1 level respectively.

B Model appendix

We now turn to presenting the additional details of our general equilibrium model.

B.1 Final good firms

The representative final good firm maximizes profits:

$$P_t Y_t - \int_0^1 Q_t(j) p_t(j) dj$$

subject to the production technologies

$$Y_t = \left[\int_0^1 Q_t(j)^{\omega} dj \right]^{\frac{1}{\omega}}$$

$$Q_t(j) = F_t(j)^{\tau + \frac{\rho - 1}{\rho}} \left[\sum_{i=1}^{F_t(j)} m_t(j,i)^{\rho} \right]^{\frac{1}{\rho}}$$

The problem is solved in two steps. First, the input of aggregate sectoral goods is found by solving

$$\min_{\{Q_t(j)\}_{j=0}^1} \int_0^1 Q_t(j) p_t(j) dj \qquad \text{subject to } Y_t = \left[\int_0^1 Q_t(j)^{\omega} dj \right]^{\frac{1}{\omega}}$$

This leads to the standard demand function and price index:

$$Q_t(j) = \left(\frac{p_t(j)}{P_t}\right)^{\frac{1}{\omega-1}} Y_t$$
$$P_t = \left[\int_0^1 p_t(j)^{\frac{\omega}{\omega-1}} dj\right]^{\frac{\omega-1}{\omega}}$$

Second, the firm decides the mix of inputs within each sector by solving the following:

$$\min_{\{m_t(j,i)\}_{i=1}^{F_t(j)}} \sum_{i=1}^{F_t(j)} p_t(j,i)m_t(j,i) \qquad \text{s.t. } Q_t(j) = F_t(j)^{\tau+\frac{\rho-1}{\rho}} \left[\sum_{i=1}^{F_t(j)} m_t(j,i)^{\rho}\right]^{\frac{1}{\rho}}$$

which has the first order condition

$$p_t(j,i) - p_t(j)\frac{1}{\rho}F_t(j)^{\tau+\frac{\rho-1}{\rho}} \left[\sum_{i=1}^{F_t(j)} m_t(j,i)^{\rho}\right]^{\frac{1}{\rho}-1} \rho m_t(j,i)^{\rho-1} = 0$$

Rewriting the first order condition and inserting the expression for $Q_t(j)$ results in the following demand function:

$$m_t(j,i) = \left(\frac{p_t(j,i)}{p_t(j)}\right)^{\frac{1}{\rho-1}} \frac{Q_t(j)}{\left(F_t(j)^{\tau+\frac{\rho-1}{\rho}}\right)^{\frac{\rho}{\rho-1}}} = \left(\frac{p_t(j,i)}{p_t(j)}\right)^{\frac{1}{\rho-1}} \left(\frac{p_t(j)}{P_t}\right)^{\frac{1}{\omega-1}} \frac{Y_t}{\left(F_t(j)^{\tau+\frac{\rho-1}{\rho}}\right)^{\frac{\rho}{\rho-1}}}$$

Lastly, we derive the consumption-based price index for sector j by inserting the demand function into the cost function $Q_t(j)p_t(j) = \sum_{i=1}^{F_t(j)} p_t(j,i)m_t(j,i)$:

$$p_t(j) = \frac{1}{F_t(j)^{\tau+\frac{\rho-1}{\rho}}} \left[\sum_{i=1}^{F_t(j)} p_t(j,i)^{\frac{\rho}{\rho-1}}\right]^{\frac{\rho-1}{\rho}}$$

B.2 Intermediate firms

The intermediate firm i in sector j maximizes real profits:

$$\frac{p_t(j,i)}{P_t} m_t(j,i) - w_t^l n_t^l(j,i) - -w_t^b n_t^b(j,i) - r_t^k k_{t-1}(j,i)$$

subject to the production function, the demand for its good and the sectoral price index:

$$
\begin{aligned}
m_t(j,i) &= z_t k_{t-1}(j,i)^{\mu} \left[n_t^b(j,i)^{\alpha} n_t^l(j,i)^{1-\alpha} \right]^{1-\mu} - \varphi \\
m_t(j,i) &= \left(\frac{p_t(j,i)}{p_t(j)} \right)^{\frac{1}{\rho-1}} \left(\frac{p_t(j)}{P_t} \right)^{\frac{1}{\omega-1}} \frac{Y_t}{\left(F_t(j)^{\tau+\frac{\rho-1}{\rho}} \right)^{\frac{\rho}{\rho-1}}} \\
p_t(j) &= \frac{1}{F_t(j)^{\tau+\frac{\rho-1}{\rho}}} \left[\sum_{i=1}^{F_t(j)} p_t(j,i)^{\frac{\rho}{\rho-1}} \right]^{\frac{\rho-1}{\rho}}
\end{aligned}
$$

The first order conditions with respect to $k_{t-1}(j,i)$, $n_t^b(j,i)$ and $n_t^l(j,i)$ are

$$
\begin{aligned}
r_t^k &= \mu \frac{p_t(j,i)}{P_t} \frac{z_t k_{t-1}(j,i)^{\mu} \left[n_t^b(j,i)^{\alpha} n_t^l(j,i)^{1-\alpha} \right]^{1-\mu}}{x_t(j,i) k_{t-1}(j,i)} \\
w_t^b &= (1-\mu)\alpha \frac{p_t(j,i)}{P_t} \frac{z_t k_{t-1}(j,i)^{\mu} \left[n_t^b(j,i)^{\alpha} n_t^l(j,i)^{1-\alpha} \right]^{1-\mu}}{x_t(j,i) n_t^b(j,i)} \\
w_t^l &= (1-\mu)(1-\alpha) \frac{p_t(j,i)}{P_t} \frac{z_t k_{t-1}(j,i)^{\mu} \left[n_t^b(j,i)^{\alpha} n_t^l(j,i)^{1-\alpha} \right]^{1-\mu}}{x_t(j,i) n_t^l(j,i)}
\end{aligned}
$$

The elasticity of demand according to the demand curve and the sectoral price index is given by

$$
\begin{aligned}
\varepsilon_{m_t(j,i)} = &\left(\frac{m_t(j,i)}{p_t(j,i)} \frac{1}{\rho-1} + \left(\frac{1}{\omega-1} - \frac{1}{\rho-1} \right) \frac{m_t(j,i)}{p_t(j)} \frac{\rho-1}{\rho} \frac{p_t(j)}{\sum_{i=1}^{F_t} p_t(j,i)^{\frac{\rho}{\rho-1}}} \frac{\rho}{\rho-1} p_t(j,i)^{\frac{\rho}{\rho-1}-1} \right) \\
&\times \frac{p_t(j,i)}{m_t(j,i)}
\end{aligned}
$$

Reducing this and substituting out $\sum_{i=1}^{F_t} p_t(j,i)^{\frac{\rho}{\rho-1}}$ results in the following expression:

$$
\varepsilon_{m_t(j,i)} = \frac{1}{\rho-1} + \left(\frac{1}{\omega-1} - \frac{1}{\rho-1} \right) \left(\frac{p_t(j,i)}{p_t(j) F_t(j)^{\tau}} \right)^{\frac{\rho}{\rho-1}} \frac{1}{F_t}
$$

Since the firm sells the good in a monopolistic competitive market, it will set its price at a markup over marginal costs. The markup follows from inserting the elasticity into the standard markup rule:

$$
x_t(j,i) = \frac{1}{1 + \frac{1}{\varepsilon_{m_t(j,i)}}} = \frac{\varepsilon_{m_t(j,i)}}{1 + \varepsilon_{m_t(j,i)}}
$$

Marginal costs are derived by minimizing the following:

$$w_t^l n_t^l(j,i) + w_t^b n_t^b(j,i) + r_t^k k_{t-1}(j,i) + \lambda_t(j,i)\left(m_t(j,i) - z_t k_{t-1}(j,i)^{\mu}\left[n_t^b(j,i)^{\alpha} n_t^l(j,i)^{1-\alpha}\right]^{1-\mu} + \varphi\right)$$

where $\lambda_t(j,i)$ is the multiplier on the production function.

The first order conditions with respect to $k_{t-1}(j,i)$, $n_{i,t}^b$ and $n_{i,t}^l$ are

$$r_t^k - \lambda_t(j,i)\mu \frac{z_t k_{t-1}(j,i)^{\mu}\left[n_t^b(j,i)^{\alpha} n_t^l(j,i)^{1-\alpha}\right]^{1-\mu}}{k_t(j,i)} = 0$$

$$w_t^b - \lambda_t(j,i)(1-\mu)\alpha \frac{z_t k_{t-1}(j,i)^{\mu}\left[n_t^b(j,i)^{\alpha} n_t^l(j,i)^{1-\alpha}\right]^{1-\mu}}{n_t^b(j,i)} = 0$$

$$w_t^l - \lambda_t(j,i)(1-\mu)(1-\alpha) \frac{z_t k_{t-1}(j,i)^{\mu}\left[n_t^b(j,i)^{\alpha} n_t^l(j,i)^{1-\alpha}\right]^{1-\mu}}{n_t^l(j,i)} = 0$$

Substituting these into the production function leads to the following:

$$m_t(j,i) = z_t k_{t-1}(j,i)\left(\frac{r_t^k}{w_t^b}\frac{\alpha(1-\mu)}{\mu}\left(\frac{w_t^b}{w_t^l}\frac{1-\alpha}{\alpha}\right)^{1-\alpha}\right)^{1-\mu} - \varphi$$

$$m_t(j,i) = z_t n_t^b(j,i)\left(\frac{\mu}{(1-\mu)\alpha}\frac{w^b}{r^k}\right)^{\mu}\left(\frac{w_t^b}{w_t^l}\frac{1-\alpha}{\alpha}\right)^{(1-\alpha)(1-\mu)} - \varphi$$

$$m_t(j,i) = z_t n_t^l(j,i)\left(\frac{\mu}{(1-\mu)(1-\alpha)}\frac{w^l}{r^k}\right)^{\mu}\left(\frac{w_t^l}{w_t^b}\frac{\alpha}{1-\alpha}\right)^{\alpha(1-\mu)} - \varphi$$

which can be inserting into the cost function $C_t(j,i) = w_t^l n_t^l(j,i) + w_t^b n_t^b(j,i) + r_t^k k_{t-1}(j,i)$ to get an expression for costs:

$$C_t(j,i) = A\left(r_t^k\right)^{\mu}\left(w_t^b\right)^{\alpha(1-\mu)}\left(w_t^l\right)^{(1-\alpha)(1-\mu)}\left(\frac{m_t(j,i)+\varphi}{z_t}\right)$$

where $A = \frac{1}{(1-\mu)^{1-\mu}(1-\alpha)^{(1-\alpha)(1-\mu)}\mu^{\mu}\alpha^{\alpha(1-\mu)}}$.

B.3 Steady state

The non-stochastic steady state of the economy is described in the following section, where variables without time subscripts denote steady state values.

We derive the interest rate and the capital rental rate in steady state from (3.13), (3.14) and (3.15):

$$R = \frac{1}{\beta^l} \tag{B.1}$$

$$r^k = \frac{1}{\beta^l} - (1-\delta) \tag{B.2}$$

The capital-to-output ratio is derived by combining equations (3.36) and (3.38):

$$\frac{K}{Y} = \frac{\mu}{\frac{1}{\beta^l} - (1-\delta)}$$

while steady state government spending as a share of output is determined by the parameter θ:

$$\frac{G}{Y} = \theta$$

Combining the two ratios above with the capital accumulation schedule (3.9) and the aggregate resource constraint (3.45) gives us the consumption-to-output ratio:

$$\frac{C}{Y} = 1 - \bar{G} - \frac{\delta\mu}{\frac{1}{\beta^l} - (1-\delta)} \tag{B.3}$$

Next, we derive the income shares by using (3.36), (3.38), (3.39) and (3.40):

$$\frac{r^k K}{Y} = \mu \tag{B.4}$$

$$\frac{w^b N^b}{Y} = (1-\mu)\alpha \tag{B.5}$$

$$\frac{w^l N^b}{Y} = (1-\mu)(1-\alpha) \tag{B.6}$$

The steady state markup is simply given by the (3.34):

$$x = \frac{(1-\omega)\,F - (\rho - \omega)}{\rho\,(1-\omega)\,F - (\rho-\omega)}$$

Lastly, we compute the consumption and housing shares of the two households. The housing demand equation (3.5) and the Euler equation (3.7) in combination with (B.1) are given by

$$q\lambda^b = \mathrm{Y}^b \left(H^b\right)^{-\sigma_h} + \beta^b\lambda^b q + \mu^b m \beta^l q(1-\gamma) \tag{B.7}$$

$$\mu^b = \lambda^b \frac{1 - \frac{\beta^b}{\beta^l}}{1-\beta^b\gamma} \tag{B.8}$$

Substituting the latter equation into the former yields

$$\mathrm{Y}^b \left(H^b\right)^{-\sigma_h} = q\lambda^b \left[1 - \beta^b - \frac{\beta^l - \beta^b}{1-\beta^b\gamma} m(1-\gamma)\right]. \tag{B.9}$$

The housing demand equation for the patient households is given by

$$\mathrm{Y}^l \left(H^l\right)^{-\sigma_h} = q\lambda^l \left(1 - \beta^l\right) \tag{B.10}$$

The budget constraint multipliers follow from (3.4) and (3.10):

$$\lambda^b = \left(1 - h^b \beta^b\right) \left(\left(1 - h^b\right) C^b\right)^{-\sigma_c}$$
$$\lambda^l = \left(1 - h^l \beta^l\right) \left(\left(1 - h^l\right) C^l\right)^{-\sigma_c}$$

Dividing (B.10) by (B.9) and inserting the steady state expressions for the budget constraint multipliers together with the consumption and housing market clearing conditions (3.46) and (3.47) gives us an expression for the housing and consumption shares of the impatient households:

$$\begin{aligned}
\frac{Y^l \left(H^l\right)^{-\sigma_h}}{Y^b \left(H^b\right)^{-\sigma_h}} &= \frac{q\lambda^l \left(1-\beta^l\right)}{q\lambda^b \left[1 - \beta^b - \frac{\beta^l - \beta^b}{1-\beta^b\gamma} m(1-\gamma)\right]} \\
\left(\frac{H}{H^b} - 1\right)^{-\sigma_h} &= \frac{Y^b}{Y^l} \frac{1-\beta^l}{1 - \beta^b - \frac{\beta^l - \beta^b}{1-\beta^b\gamma} m(1-\gamma)} \frac{\lambda^l}{\lambda^b} \\
\left(\frac{H}{H^b} - 1\right)^{-\sigma_h} &= \frac{Y^b}{Y^l} \frac{1-\beta^l}{1 - \beta^b - \frac{\beta^l - \beta^b}{1-\beta^b\gamma} m(1-\gamma)} \frac{\left(1 - h^l\beta^l\right)\left(\left(1-h^l\right)\left(C - C^b\right)\right)^{-\sigma_c}}{\left(1 - h^b\beta^b\right)\left(\left(1-h^b\right)C^b\right)^{-\sigma_c}} \\
\left(\frac{H}{H^b} - 1\right)^{-\sigma_h} &= \frac{Y^b}{Y^l} \frac{1-\beta^l}{1 - \beta^b - \frac{\beta^l - \beta^b}{1-\beta^b\gamma} m(1-\gamma)} \frac{1 - \beta^l h^l}{1 - \beta^b h^b} \left(\frac{1-h^l}{1-h^b}\right)^{-\sigma_c} \left(\frac{C}{C^b} - 1\right)^{-\sigma_c} \qquad \text{(B.11)}
\end{aligned}$$

Similarly, we derive an additional expression for the housing and consumption shares of the impatient households by inserting the borrowing constraint (3.3) into their budget constraint (3.2), multiplying both sides by $\frac{C}{Y}$ and inserting the interest rate (B.1), the labor income share (B.5), their lump-sum tax (3.43) and government spending as a share of output:

$$\frac{C^b}{C} = \frac{Y}{C}\left(\left(\beta^l - 1\right) m \frac{qH}{Y}\frac{H^b}{H} + \alpha(1-\mu) - \theta\tau^b\right) \qquad \text{(B.12)}$$

The housing wealth-to-output ratio, $\frac{qH}{Y}$, is calibrated, while the consumption share, $\frac{C}{Y}$, follows from (B.3) so (B.11) and (B.12) are solved numerically for $\frac{H^b}{H}$ and $\frac{C^b}{C}$. The steady state budget constraint of the patient households has not been used in the derivation of the steady state but will hold by Walras' law.

B.4 Log-linearized model

The model is log-linearized around the non-stochastic steady state. For any generic variable X_t, we let $\hat{X}_t = \ln X_t - \ln X$ denote the log-deviation from steady state. We replace B_t^l and B_t^b by B_t throughout the following.

B.4.1 Optimality conditions of the impatient households

Log-linearization of (3.4), (3.6) and (3.3) gives us the following:

$$\hat{\lambda}_t^b = -\frac{\sigma_c^b}{\left(1-\beta^b h^b\right)\left(1-h^b\right)}\left(\hat{C}_t^b - h^b \hat{C}_{t-1}^b - \beta^l h^b E_t\left\{\hat{C}_{t+1}^b - h^b \hat{C}_t^b\right\}\right) \tag{B.13}$$

$$\hat{w}_t^b + \lambda_t^b = \psi^b \hat{N}_t^b \tag{B.14}$$

$$\widehat{B}_t^b = \gamma \widehat{B}_{t-1}^b + (1-\gamma)\left(\mathrm{E}_t \widehat{q}_{t+1} + \widehat{H}_t^b - \widehat{R}_t\right) \tag{B.15}$$

Log-linearization of (3.5) results in

$$q\lambda^b\left(\hat{q}_t + \hat{\lambda}_t^b\right) = -\sigma_h Y^b\left(H^b\right)^{-\sigma_h}\hat{H}_t^b + \beta^b\lambda^b q E_t\left\{\hat{\lambda}_{t+1}^b + \hat{q}_{t+1}\right\} + \mu^b m(1-\gamma)\frac{q}{R}E_t\left\{\hat{\mu}_t^b + \hat{q}_{t+1} - \hat{R}_t\right\},$$

which is rewritten using (B.1), (B.9) and (B.8):

$$\begin{aligned} q\lambda^b\left(\hat{q}_t + \hat{\lambda}_t^b\right) &= -\sigma_h q\lambda^b\left[1-\beta^b - \frac{\beta^l-\beta^b}{1-\beta^b\gamma}m(1-\gamma)\right]\hat{H}_t^b + \beta^b\lambda^b q E_t\left\{\hat{\lambda}_{t+1}^b + \hat{q}_{t+1}\right\} \\ &\quad + \lambda^b\frac{1-\frac{\beta^b}{\beta^l}}{1-\beta^b\gamma}m(1-\gamma)\frac{q}{R}E_t\left\{\hat{\mu}_t^b + \hat{q}_{t+1} - \hat{R}_t\right\}, \end{aligned}$$

$$\begin{aligned} \hat{q}_t + \hat{\lambda}_t^b &= -\sigma_h\left[1-\beta^b - \frac{\beta^l-\beta^b}{1-\beta^b\gamma}m(1-\gamma)\right]\hat{H}_t^b + \beta^b E_t\left\{\hat{\lambda}_{t+1}^b + \hat{q}_{t+1}\right\} \\ &\quad + \frac{\beta^l-\beta^b}{1-\beta^b\gamma}m(1-\gamma)E_t\left\{\hat{\mu}_t^b + \hat{q}_{t+1} - \hat{R}_t\right\}. \end{aligned} \tag{B.16}$$

In addition, (3.7) becomes

$$\lambda^b\widehat{\lambda}_t^b + \beta^b\gamma\mu^b E_t\left\{\widehat{\mu}_{t+1}^b\right\} = \mu^b\widehat{\mu}_t^b + \beta^b\lambda^b R E_t\left\{\widehat{\lambda}_{t+1}^b + \widehat{R}_t\right\}.$$

Rewriting this using (B.8) results in

$$\widehat{\lambda}_t^b + \beta^b\gamma\frac{1-\frac{\beta^b}{\beta^l}}{1-\beta^b\gamma}E_t\left\{\widehat{\mu}_{t+1}^b\right\} = \frac{1-\frac{\beta^b}{\beta^l}}{1-\beta^b\gamma}\widehat{\mu}_t^b + \frac{\beta^b}{\beta^l}E_t\left\{\widehat{\lambda}_{t+1}^b + \widehat{R}_t\right\}. \tag{B.17}$$

The log-linearized budget constraint becomes

$$\begin{aligned} &\frac{C^b}{C}\frac{C}{Y}\hat{C}_t^b + \frac{qH}{Y}\frac{H^b}{H}\left(\hat{H}_t^b - \hat{H}_{t-1}^b\right) + m\frac{qH}{Y}\frac{H^b}{H}\left(\hat{R}_{t-1} + \hat{B}_{t-1}\right) = \\ &(1-\mu)\alpha\left(\hat{w}_t^b + \hat{N}_t^b\right) + m\frac{qH}{Y}\frac{H^b}{H}\beta^l\hat{B}_t - \theta\alpha\hat{G}_t \end{aligned} \tag{B.18}$$

by dividing through by Y and using the labor income share of the impatient households state along with the borrowing constraint and the log-linearized version of (3.43).

B.4.2 Optimality conditions of the patient households

Log-linearization of (3.10), (3.12), (3.13), (3.9) results in

$$\hat{\lambda}_t^l = -\frac{\sigma_c^l}{(1-\beta^l h^l)(1-h^l)}\left(\hat{C}_t^l - h^l\hat{C}_{t-1}^l - \beta^l h^l E_t\left\{\hat{C}_{t+1}^l - h^l\hat{C}_t^l\right\}\right) \tag{B.19}$$

$$\hat{\lambda}_t^l = E_t\left\{\hat{\lambda}_{t+1}^l\right\} + \hat{R}_t \tag{B.20}$$

$$\hat{w}_t^l + \hat{\lambda}_t^l = \psi^l \hat{N}_t^l \tag{B.21}$$

$$\hat{K}_t = (1-\delta)\hat{K}_{t-1} + \delta\hat{I}_t \tag{B.22}$$

Similar to the derivation for the impatient households, log-linearization of (3.11) in combination with (B.10) becomes

$$\hat{q}_t + \hat{\lambda}_t^l = -\sigma_h\left(1-\beta^l\right)\hat{H}_t^l + \beta^l E_t\left\{\hat{q}_{t+1} + \hat{\lambda}_{t+1}^l\right\} \tag{B.23}$$

The log-linearized first-order conditions for capital and investment, (3.14) and (3.15), are

$$\hat{q}_t^k = E_t\left\{\hat{\lambda}_{t+1}^l\right\} - \hat{\lambda}_t^l + \beta^l r^k \hat{r}_{t+1}^k + \beta^l(1-\delta)E_t\left\{q_{t+1}^k\right\} + \beta^l\delta^2\phi E_t\left\{\hat{I}_{t+1} - \hat{K}_t\right\}$$

$$\hat{q}_t^k = \phi\delta\left(\hat{I}_t - \hat{K}_{t-1}\right)$$

Combining these two equations to eliminate $\hat{q}_t^k$ and inserting (B.22) results in

$$\phi\left(\hat{K}_t - \hat{K}_{t-1}\right) + \hat{\lambda}_t^l = E_t\left\{\hat{\lambda}_{t+1}^l + \beta^l r^k \hat{r}_{t+1}^k + \beta^l\phi\left(\hat{K}_{t+1} - \hat{K}_t\right)\right\} \tag{B.24}$$

Lastly, the budget constraint (3.8) becomes

$$\frac{C^l}{C}\frac{C}{Y}\hat{C}_t^l + \frac{qH}{Y}\frac{H^l}{H}\left(\hat{H}_t^l - \hat{H}_{t-1}^l\right) + \frac{\delta K}{Y}\hat{I}_t + m\frac{qH}{Y}\frac{H^b}{H}\beta^l\hat{B}_t =$$
$$(1-\mu)(1-\alpha)\left(\hat{w}_t^l + \hat{N}_t^l\right) + m\frac{qH}{Y}\frac{H^b}{H}\left(\hat{R}_{t-1} + \hat{B}_{t-1}\right) + \mu\left(\hat{r}_t^k + \hat{K}_{t-1}\right) - \theta(1-\alpha)\hat{G}_t \tag{B.25}$$

B.4.3 Symmetric firm equilibrium conditions

The log-linearized factor prices (3.38), (3.39) and (3.40) are

$$\hat{r}_t^k = (1+\tau)\left(\hat{z}_t + \left(\left(\mu - \frac{1}{1+\tau}\right)\hat{K}_{t-1} + (1-\mu)\left(\alpha\hat{N}_t^b + (1-\alpha)\hat{N}_t^l\right)\right)\right) - \frac{x-(1+\tau)}{x-1}\hat{x}_t \tag{B.26}$$

$$\hat{w}_t^b = (1+\tau)\left(\hat{z}_t + \left(\mu\hat{K}_{t-1} + (1-\mu)\left(\left(\alpha - \frac{1}{(1+\tau)(1-\mu)}\right)\hat{N}_t^b + (1-\alpha)\hat{N}_t^l\right)\right)\right) - \frac{x-(1+\tau)}{x-1}\hat{x}_t \tag{B.27}$$

$$\hat{w}_t^l = (1+\tau)\left(\hat{z}_t + \left(\mu\hat{K}_{t-1} + (1-\mu)\left(\alpha\hat{N}_t^b + \left(1-\alpha - \frac{1}{(1+\tau)(1-\mu)}\right)\hat{N}_t^l\right)\right)\right) - \frac{x-(1+\tau)}{x-1}\hat{x}_t \tag{B.28}$$

while log-linearization of (3.36) and (3.37) results in

$$\hat{Y}_t = (1+\tau)\left(\hat{z}_t + \left(\mu\hat{K}_{t-1} + (1-\mu)\left(\alpha\hat{N}_t^b + (1-\alpha)\hat{N}_t^l\right)\right)\right) - \frac{x-(1+\tau)}{x-1}\hat{x}_t \tag{B.29}$$

$$\hat{F}_t = \frac{1}{1+\tau}\left(\hat{Y}_t + \frac{x}{x-1}\hat{x}_t\right) \tag{B.30}$$

We rewrite the markup (3.34) as

$$F_t\left(\rho x_t - 1\right)\left(1-\omega\right) = \left(x_t - 1\right)\left(\rho - \omega\right)$$

which is log-linearized:

$$F\left(\rho x - 1\right)\left(1-\omega\right)\hat{F}_t + \rho x F\left(1-\omega\right)\hat{x}_t = x\left(\rho-\omega\right)\hat{x}_t$$

Inserting $F = \frac{(x-1)(\rho-\omega)}{(\rho x-1)(1-\omega)}$ into the equation above and rewriting yields the following:

$$\hat{F}_t = \frac{x}{x-1}\frac{\rho-1}{\rho x-1}\hat{x}_t \tag{B.31}$$

The log-linearized expression for TFP in equation (3.41) is

$$T\hat{F}P_t = \tau\hat{F}_t - \hat{x}_t + (1-\tau)\hat{z}_t \tag{B.32}$$

B.4.4 Market clearing conditions and shock processes

The market clearing conditions (3.45), (3.46) and (3.47) become

$$\hat{Y}_t = \frac{C}{Y}\hat{C}_t + \frac{G}{Y}\hat{G}_t + \frac{I}{Y}\hat{I}_t$$

$$\hat{C}_t = \frac{C^b}{C}\hat{C}^b_t + \frac{C^l}{C}\hat{C}^l_t \tag{B.33}$$

$$0 = \frac{H^b}{H}\hat{H}^b_t + \frac{H^l}{H}\hat{H}^l_t \tag{B.34}$$

The good market clearing condition is not included in the Dynare code since it is redundant by Walras' law.

The log-linearized shock processes for government spending (3.42) and productivity are

$$\hat{G}_t = \gamma_g \hat{G}_t + \epsilon_{g,t} \tag{B.35}$$

$$\hat{z}_t = \gamma_z \hat{z}_t + \epsilon_{z,t} \tag{B.36}$$

The 23 equations (B.13) to (B.36) define the log-linearized model for the 23 endogeneous variables $\hat{\lambda}^b_t$, $\hat{\lambda}^l_t$, $\hat{C}^b_t$, $\hat{C}^l_t$, $\hat{w}^b_t$, $\hat{w}^l_t$, $\hat{N}^b_t$, $\hat{N}^l_t$, $\hat{H}^b_t$, $\hat{H}^l_t$, $\hat{B}_t$, $\hat{q}_t$, $\hat{R}_t$, $\hat{\mu}^b_t$, $\hat{G}_t$, $\hat{K}_t$, $\hat{I}_t$, $\hat{r}^k_t$, $\hat{x}_t$, $\hat{F}_t$, $\hat{Y}_t$, $\hat{C}_t$ and $\hat{z}_t$.