

Looking into the Consumption Black Box: Evidence from Scanner Data

Krzysztof Pytka ⓘ Daniel Runge
University of Mannheim

Summer Workshop on Macro & Finance, 02.07.2024



Dall-E prompt:
create a painting "Looking into the Consumption Black Box" painted by
Salvador Dali

Disclaimer

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

Big Picture: Short Story of Non-Durable Consumption in Macroeconomics

Motto of the Project

“There is no accounting for tastes”

Big Picture: Short Story of Non-Durable Consumption in Macroeconomics

Motto of the Project

“There is **no** accounting for tastes”

Big Picture: Short Story of Non-Durable Consumption in Macroeconomics

Motto of the Project

“There is **no** accounting for tastes”

◊ Neoclassical model with representative agent

(Ramsey, EJ 1928; Cass, REStud 1965; Koopmans, 1965; Brock&Mirman, JET 1972; Kydland&Prescott, Ecfa 1982)

- Early macroeconomic models considered a **single representative agent**.
- Consumption was aggregated into a **single product** for tractability.
- This approach ignored individual differences in consumption patterns.

◊ Models of Inequalities w/Heterogeneous Households

(Aiyagari, QJE 1994, JPE 1995; Huggett, JEDC 1993)

- Aimed to capture state-based consumption disparities.
- Generated **differences in quantities** consumed across households.
- Still focused on a single type of product.

◊ Models of Heterogeneous Goods (+Heterogeneity in Choices)

(Handbury, Ecfa 2021; Michelacci-Paciello-Pozzi, REStud 2022; Faber&Fally, REStud 2022; Neiman&Vavra, AEJ:Macro 2023)

- Recognize that households may consume different varieties of products due to:
either (i) latent heterogeneity or (ii) (intratemporal) non-homothetic preferences.
- **Differences in preferences** lead to diverse consumption baskets.
- Aim to provide more realistic representation of actual consumption behavior.

Motivation and Why is It Important?

Research Questions ⓘ

1. Do different groups of consumers buy different products?
2. How persistent are the individual choices over time?

Motivation and Why is It Important?

Research Questions ⓘ

1. Do different groups of consumers buy different products?
2. How persistent are the individual choices over time?
3. What are implications for structural models of the answers to Questions 1 and 2?

Motivation and Why is It Important?

Research Questions ⓘ

1. Do different groups of consumers buy different products?
2. How persistent are the individual choices over time?
3. What are implications for structural models of the answers to Questions 1 and 2?

◊ Importance for Macroeconomic/Industrial-Organization Perspectives:

- Direct implications for the ease or difficulty of price discrimination by retailers.
- Influences the market power of retailers.
- The level of heterogeneity in consumption bundles affects the magnitude of consumption externalities.

Motivation and Why is It Important?

Research Questions ⓘ

1. Do different groups of consumers buy different products?
2. How persistent are the individual choices over time?
3. What are implications for structural models of the answers to Questions 1 and 2?

- ◊ **Importance for Macroeconomic/Industrial-Organization Perspectives:**
 - Direct implications for the ease or difficulty of price discrimination by retailers.
 - Influences the market power of retailers.
 - The level of heterogeneity in consumption bundles affects the magnitude of consumption externalities.
- ◊ While there is a rich empirical literature documenting inequality in consumption expenditures, there are still not many studies systematically analyzing **the composition of consumption bundles across different groups of consumers**

Motivation and Why is It Important?

Research Questions ⓘ

1. Do different groups of consumers buy different products?
2. How persistent are the individual choices over time?
3. What are implications for structural models of the answers to Questions 1 and 2?

- ◊ **Importance for Macroeconomic/Industrial-Organization Perspectives:**
 - Direct implications for the ease or difficulty of price discrimination by retailers.
 - Influences the market power of retailers.
 - The level of heterogeneity in consumption bundles affects the magnitude of consumption externalities.
- ◊ While there is a rich empirical literature documenting inequality in consumption expenditures, there are still not many studies systematically analyzing **the composition of consumption bundles across different groups of consumers**
- 👉 **Challenges:**
 - *How to measure group differences in (overwhelmingly) high-dimensional choices?*
 - *Potentially severe small-sample bias (yet to be discussed).*



- ◊ **Heterogeneous preferences:**
Neiman&Vavra (AEJ:Macro, 2023), Handbury (Ecta, 2021),
Michelacci, Paciello, and Pozzi (REStud, 2022), Faber&Fally (REStud, 2022)
- ◊ **Polarization in Economics:**
Bertrand&Kamenica (AEJ: Applied, 2023), Alesina, Tabellini, and Trebbi (BP:EA 2017),
Desmet&Wacziarg (EJ, 2021), Boar&Giannone (2023)
- ◊ **Big data with small sample bias:**
Gentzkow, Shapiro, and Taddy (Ecta, 2019), Armenter&Koren (AER, 2014),
Conlon, Mortimer, and Sarkis (2023)
- ◊ **Confronting the mainstream models:**
Armenter & Koren (AER, 2014), Menzio (202*)
Defending the mainstream models: Becker (JPE, 1962)

◊ **Indistinguishable Consumption Choices:**

- Consumption choices are practically indistinguishable between rich and poor households.
- On average, the way one dollar is spent does not allow identification of whether the consumer is rich or poor.
- This suggests that non-durable consumption is not polarized at all, contrary to common preconceptions.

◊ **Unstable Individual Consumption Choices:**

- Individual consumption choices are very unstable.
- On average, only 40% of products purchased in one year are bought again in the next year.

◊ **Parsimonious Model of Consumption:**

- All products are perfect substitutes (per price); the composition of baskets is the result of random sampling.

Data: NielsenIQ Consumer Panel

- ◊ Data on Consumer Expenditures are taken from the Kilts-NielsenIQ Consumer Panel
- ◊ 40,000-60,000 households per year from 2004-2018 with socio-economic characteristics
- ◊ 1.5 million unique products including groceries, drug products, small appliances, and electronics
- ◊ Scanner data, so measurement error is expected to be rather low
 - Data is collected either with a separate scanner or an app by scanning the bar codes of the bought items
- ◊ Projection weights to be representative for the US economy

Research Questions ⓘ

1. Do different groups of consumers buy different products?
2. How persistent are the individual choices over time?
3. What are implications for structural models of the answers to Questions 1 and 2?

Consumption Polarization

Polarization in Consumption

Challenge 1: *How to measure group differences in high-dimensional choices?*

- ◊ Measuring inequality is dramatically simpler: Gini coefficient, Lorenz curve, etc.
- ◊ Measuring polarization is challenging due to high-dimensional data.
(800,000 unique products per year)

Polarization in Consumption

Challenge 1: How to measure group differences in high-dimensional choices?

- ◊ Measuring inequality is dramatically simpler: Gini coefficient, Lorenz curve, etc.
- ◊ Measuring polarization is challenging due to high-dimensional data.
(800,000 unique products per year)

↳ Using the methodology by **Gentzkow, Shapiro, and Taddy (Ecta, 2019)** Their application

General Idea (adapted to Consumption Choices)

1. Randomly select \$1 spent in the Nielsen universe.
2. Learn how this \$1 is spent.
3. Predict the household's group membership based on purchased products.
4. Polarization = expected predictive power.

Polarization in Consumption

Challenge 1: How to measure group differences in high-dimensional choices?

- ◊ Measuring inequality is dramatically simpler: Gini coefficient, Lorenz curve, etc.
- ◊ Measuring polarization is challenging due to high-dimensional data.
(800,000 unique products per year)

👉 Using the methodology by **Gentzkow, Shapiro, and Taddy (Ecta, 2019)** Their application

⚙️ General Idea (adapted to Consumption Choices)

1. Randomly select \$1 spent in the Nielsen universe.
2. Learn how this \$1 is spent.
3. Predict the household's group membership based on purchased products.
4. Polarization = expected predictive power.

Interpretation:

- ◊ Values between 0.5 and 1.0.
- ◊ 1.0: Perfect prediction of group membership from products (*Full polarization*).
- ◊ 0.5: No predictive power from products (*No polarization*).

Big Data with Small Sample Bias

Challenge 2: *Potentially severe small-sample bias.*

- ◊ We have $\approx 800,000$ unique products each year and only $\approx 50,000$ households.
- ◊ Many products may be bought by only a few households, leading to severe small-sample bias in estimating the probability of buying a product.
- ◊ Why? In extreme cases, if a product is bought by only one household, the predictive power of this purchase is 100%.

Big Data with Small Sample Bias

Challenge 2: Potentially severe small-sample bias.

- ◊ We have $\approx 800,000$ unique products each year and only $\approx 50,000$ households.
- ◊ Many products may be bought by only a few households, leading to severe small-sample bias in estimating the probability of buying a product.
- ◊ Why? In extreme cases, if a product is bought by only one household, the predictive power of this purchase is 100%.

Product definition	Area aggregation	No. of transactions			
		≥ 1	≥ 2	≥ 10	≥ 20
UPC	Nationwide	1	0.963	0.814	0.723
UPC	Scantrack market	1	0.713	0.287	0.209

Big Data with Small Sample Bias

Challenge 2: Potentially severe small-sample bias.

- ◊ We have $\approx 800,000$ unique products each year and only $\approx 50,000$ households.
- ◊ Many products may be bought by only a few households, leading to severe small-sample bias in estimating the probability of buying a product.
- ◊ Why? In extreme cases, if a product is bought by only one household, the predictive power of this purchase is 100%.

Product definition	Area aggregation	No. of transactions			
		≥ 1	≥ 2	≥ 10	≥ 20
UPC	Nationwide	1	0.963	0.814	0.723
UPC	Scantrack market	1	0.713	0.287	0.209

- ◊ We do not know whether purchases of products that are bought by only a few households are systematically purchased only by those households or if it is a random event.

Consumption Polarization (*cont'd*)

- ◊ Gentzkow et al. (Ecta, 2019) face data sparsity in congressional speech data.
To overcome small sample bias, they propose studying polarization of the data generating processes estimated using a penalized estimator.

Consumption Polarization (*cont'd*)

- ◊ Gentzkow et al. (Ecta, 2019) face data sparsity in congressional speech data.
To overcome small sample bias, they propose studying polarization of the data generating processes estimated using a penalized estimator.
- ◊ In our application, let $\mathbf{c}_{i,t}$ be a vector of all goods consumed by household i in year t :

$$\mathbf{c}_{i,t} \sim MN(m_{i,t}, \mathbf{q}_t^{l_{i,t}}(\mathbf{x}_{i,t}))$$

- ◊ $m_{i,t}$ is household i 's budget at time t , $l_{i,t}$ is the consumption group, $\mathbf{x}_{i,t}$ is a vector of household characteristics, and $\mathbf{q}_t^{l_{i,t}}(\mathbf{x}_{i,t})$ is a vector of choice probabilities:

$$q_{t,j}^{l_{i,t}}(\mathbf{x}_{i,t}) = \frac{e^{u_{i,j,t}}}{\sum_k e^{u_{i,k,t}}}$$

$$u_{i,j,t} = \alpha_{j,t} + \mathbf{x}'_{i,t} \boldsymbol{\gamma}_{j,t} + \varphi_{j,t} \cdot \mathbb{1}_{\{l_{i,t}=H\}}$$

Consumption Polarization (*cont'd*)

- ◊ Gentzkow et al. (Ecta, 2019) face data sparsity in congressional speech data.
To overcome small sample bias, they propose studying polarization of the data generating processes estimated using a penalized estimator.
- ◊ In our application, let $\mathbf{c}_{i,t}$ be a vector of all goods consumed by household i in year t :

$$\mathbf{c}_{i,t} \sim MN(m_{i,t}, \mathbf{q}_t^{l_{i,t}}(\mathbf{x}_{i,t}))$$

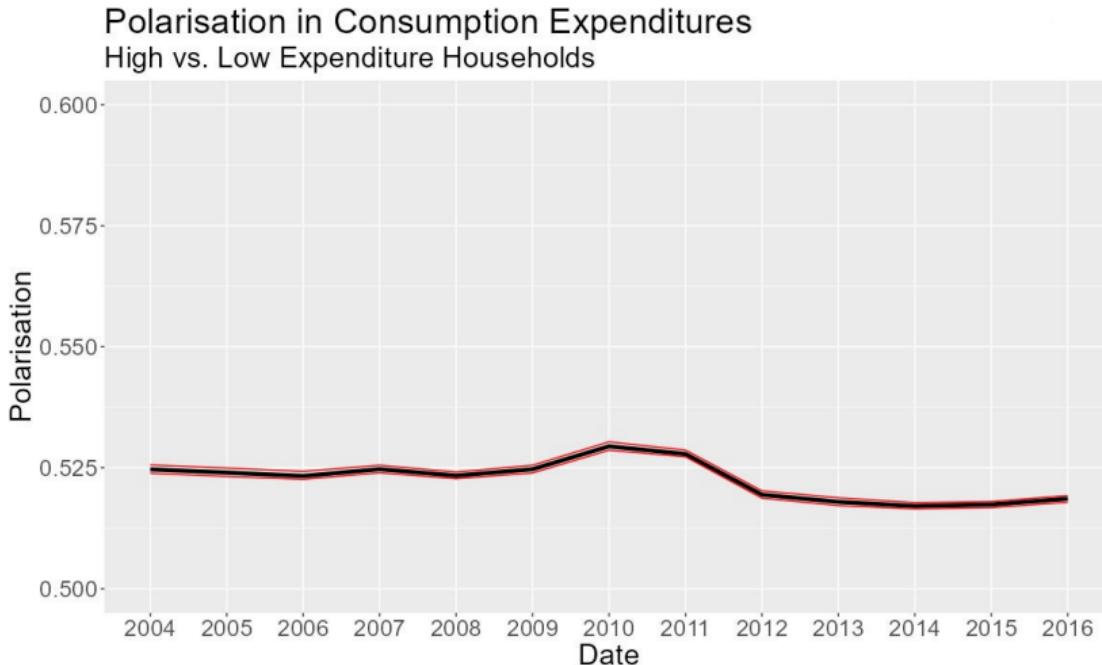
- ◊ $m_{i,t}$ is household i 's budget at time t , $l_{i,t}$ is the consumption group, $\mathbf{x}_{i,t}$ is a vector of household characteristics, and $\mathbf{q}_t^{l_{i,t}}(\mathbf{x}_{i,t})$ is a vector of choice probabilities:

$$q_{t,j}^{l_{i,t}}(\mathbf{x}_{i,t}) = \frac{e^{u_{i,j,t}}}{\sum_k e^{u_{i,k,t}}}$$

$$u_{i,j,t} = \alpha_{j,t} + \mathbf{x}'_{i,t} \boldsymbol{\gamma}_{j,t} + \varphi_{j,t} \cdot \mathbb{1}_{\{l_{i,t}=H\}}$$

- ◊ Parameter $\varphi_{j,t}$ is subject to regularization using the LASSO penalty.
 Between-group differences caused by noise are shrunk to zero.
- ◊ How is it possible to estimate a multinomial logit with 800,000 (sic!) categories?
 Distributed Poisson approximation by Taddy (AoAS, 2013).

Polarization in Nielsen consumption



Result 1

The polarization is extremely low. The choices made by high- and low-consumption households have almost no predictive power.

Research Questions ⓘ

1. Do different groups of consumers buy different products?
2. How persistent are the individual choices over time?
3. What are implications for structural models of the answers to Questions 1 and 2?

Persistence in Composition of Baskets

Defining the Persistence Measure

- ◊ We measure the persistence within the consumption basket of a household by computing the share of expenditures within a year spent on products already bought in the previous year
- ◊ Formally, we compute:

$$O_i^E = \frac{\sum_{j \in (\mathcal{U}_{i,t-1} \cap \mathcal{U}_{i,t})} e_{i,t}(j)}{\sum_{j \in \mathcal{U}_{i,t}} e_{i,t}(j)}$$

where $\mathcal{U}_{i,t}$ is the set of products consumed by household i in year t and $e_{i,t}(j)$ the expenditures on product j by household i in year t .

Defining the Persistence Measure

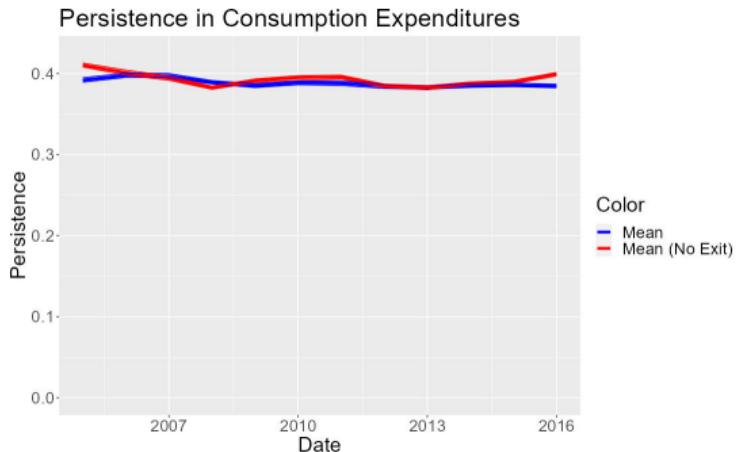
- ◊ We measure the persistence within the consumption basket of a household by computing the share of expenditures within a year spent on products already bought in the previous year
- ◊ Formally, we compute:

$$O_i^E = \frac{\sum_{j \in (\mathcal{U}_{i,t-1} \cap \mathcal{U}_{i,t})} e_{i,t}(j)}{\sum_{j \in \mathcal{U}_{i,t}} e_{i,t}(j)}$$

where $\mathcal{U}_{i,t}$ is the set of products consumed by household i in year t and $e_{i,t}(j)$ the expenditures on product j by household i in year t .

- ◊ As a first step, we compute average persistence for each year
- ◊ To ensure that results are not driven by product exits and entries, we recompute persistence by using only expenditures on products that are available within the US in both years as a robustness check

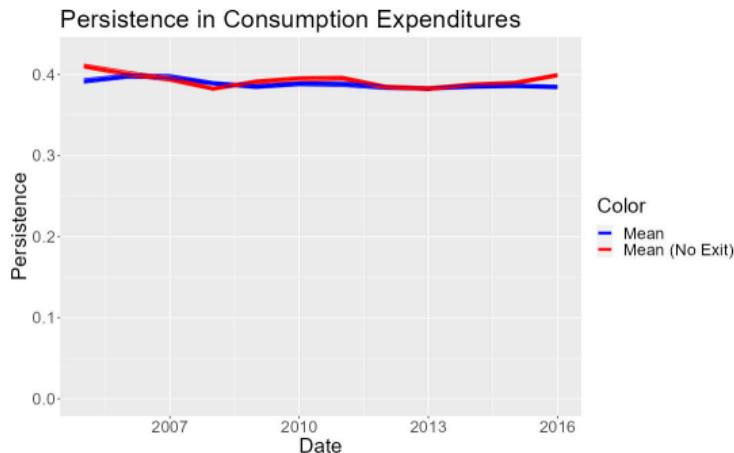
Results: Average Persistence [UPC level]



Result 2a

At the barcode level average persistence is quite low within the US. This result is robust to controlling for product entry and exit.

Results: Average Persistence [UPC level]

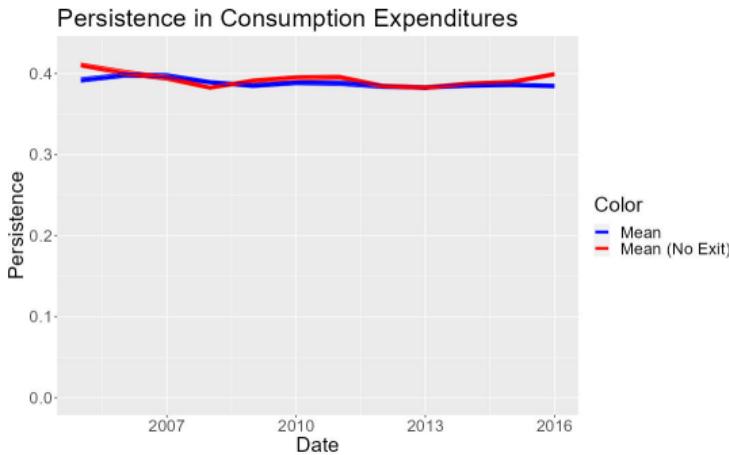


Result 2a

At the barcode level average persistence is quite low within the US. This result is robust to controlling for product entry and exit.

- ◇ There also is on average no difference of economic significance between households that experience income changes and households that do not [▶ By income changes](#)

Results: Average Persistence [UPC level]

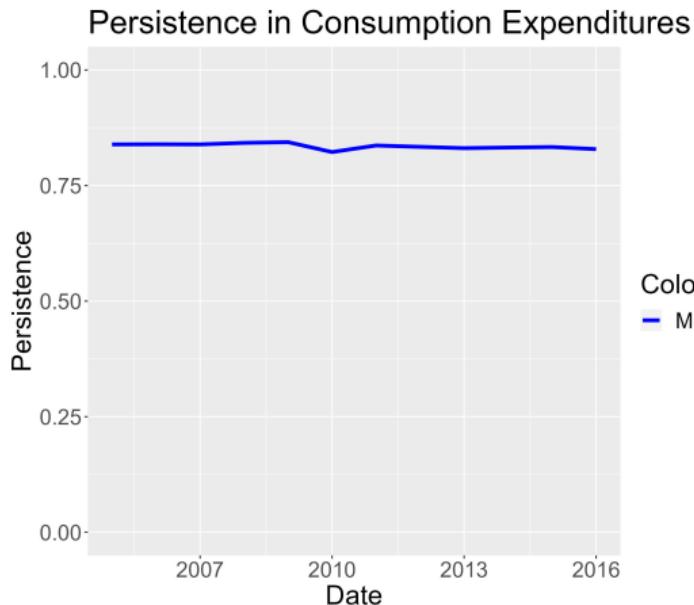


Result 2a

At the barcode level average persistence is quite low within the US. This result is robust to controlling for product entry and exit.

- ◊ There also is on average no difference of economic significance between households that experience income changes and households that do not [By income changes](#)
- ◊ Importantly, products that leave the basket are not more expensive than those that stay in the basket (therefore, cannot be rationalized by price search models).

Results: Average Persistence [Product-module level]



Result 2b

Persistence is much higher on the module level.

- ◊ We can interpret the difference between the barcode and the module level as households having relatively stable preferences for the kinds of products they like to consume. But when it comes to the specific product itself choices are much less persistent.

Heterogeneity in Persistence

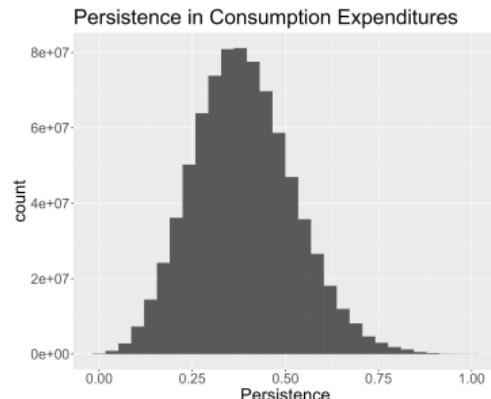
Question

We have seen that persistence is quite low on **average**. But is there some **heterogeneity** between households?

Heterogeneity in Persistence

Question

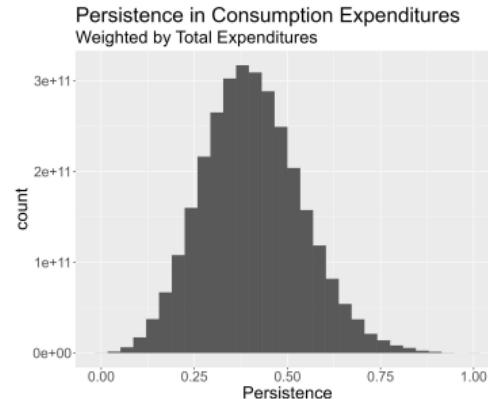
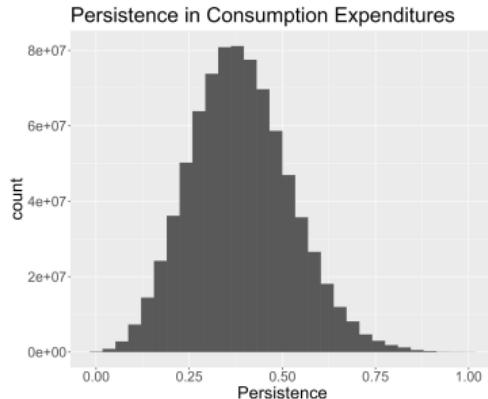
We have seen that persistence is quite low on **average**. But is there some **heterogeneity** between households?



Heterogeneity in Persistence

Question

We have seen that persistence is quite low on **average**. But is there some **heterogeneity** between households?



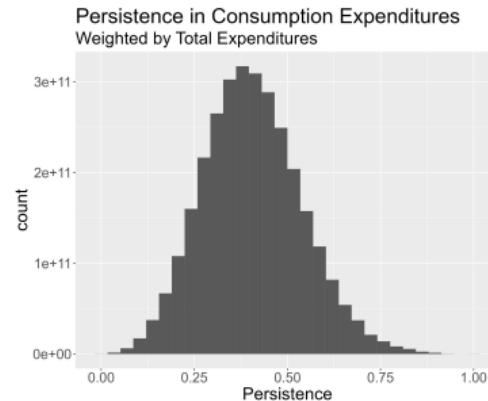
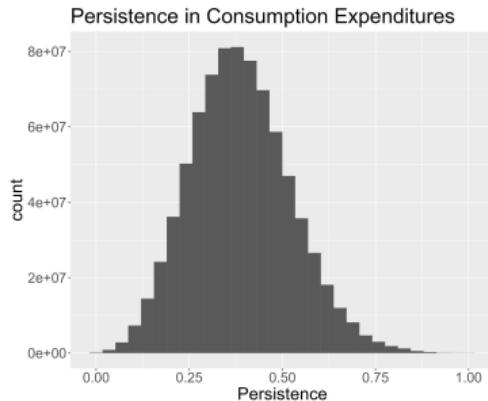
Result 3

- ◊ There is substantial heterogeneity in persistence. More than 50% of aggregate non-durable consumption are accounted for by households with a persistence of less than 50%.

Heterogeneity in Persistence

Question

We have seen that persistence is quite low on **average**. But is there some **heterogeneity** between households?



Result 3

- ◊ There is substantial heterogeneity in persistence. More than 50% of aggregate non-durable consumption are accounted for by households with a persistence of less than 50%.
- ◊ Heterogeneity in persistence is latent and not associated with any observable characteristics. ▶ Random-forest analysis

Research Questions ⓘ

1. Do different groups of consumers buy different products?
2. How persistent are the individual choices over time?
3. What are implications for structural models of the answers to Questions 1 and 2?

Mapping the Data to Existing Models

- ◊ The composition of consumption bundles is not influenced by income level. The choices made by high- and low-consumption households have almost no predictive power. This suggests that in contrast to models w/intratemporal non-homothetic preferences, consumers' choices are not income-dependent.
- ◊ In models with (latent) heterogeneous preferences, the very low stability of consumption baskets over time can be generated only when households would experience significant preference shocks in each period.

Mapping the Data to Existing Models

- ◊ The composition of consumption bundles is not influenced by income level. The choices made by high- and low-consumption households have almost no predictive power. This suggests that in contrast to models w/intratemporal non-homothetic preferences, consumers' choices are not income-dependent.
- ◊ In models with (latent) heterogeneous preferences, the very low stability of consumption baskets over time can be generated only when households would experience significant preference shocks in each period.
- ◊ These findings motivate a thought experiment: a model where differences in consumption bundles are driven by factors other than heterogeneous preferences. Instead, simple information frictions are considered, where households need to find goods before consumption. Consequently, different baskets result from different realizations of randomness rather than from choices driven by different preferences.

Shopping Spree of Beckerian Consumers

Extreme example of a *shopping spree*



I Gave People \$1,000,000 But ONLY 1 Minute To Spend It!

209M views • 3 years ago

 MrBeast 

(its epic) New Merch - <https://mrbeast.store> Check out Viewstats! - <https://www.viewstat>

5:16 Je veux remercier M. Jimmy, Mr Beast et tout le monde. Ça rendra la vie bien plus

CC

Shopping Spree Model

- ◊ Each household i has its own budget $m_{i,t}$ and nothing can be saved for the future.
- ◊ No information about products: to buy a product, they need to find it.
- ◊ There is no order in preferences; households randomly draw products during each purchase.
- ◊ Goal of each household: **to spend as much as possible.**

¹ \mathbf{e}_j is a vector with 1 at the j -th position and 0 elsewhere.

Shopping Spree Model

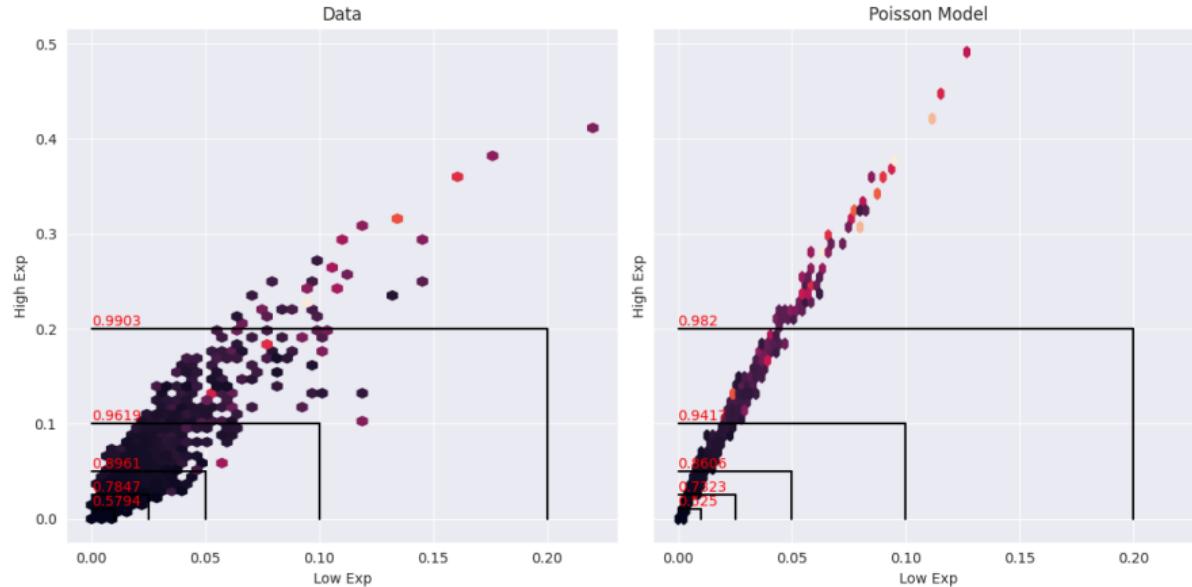
- ◊ Each household i has its own budget $m_{i,t}$ and nothing can be saved for the future.
- ◊ No information about products: to buy a product, they need to find it.
- ◊ There is no order in preferences; households randomly draw products during each purchase.
- ◊ Goal of each household: to spend as much as possible.

Decision Problem of the Consumer

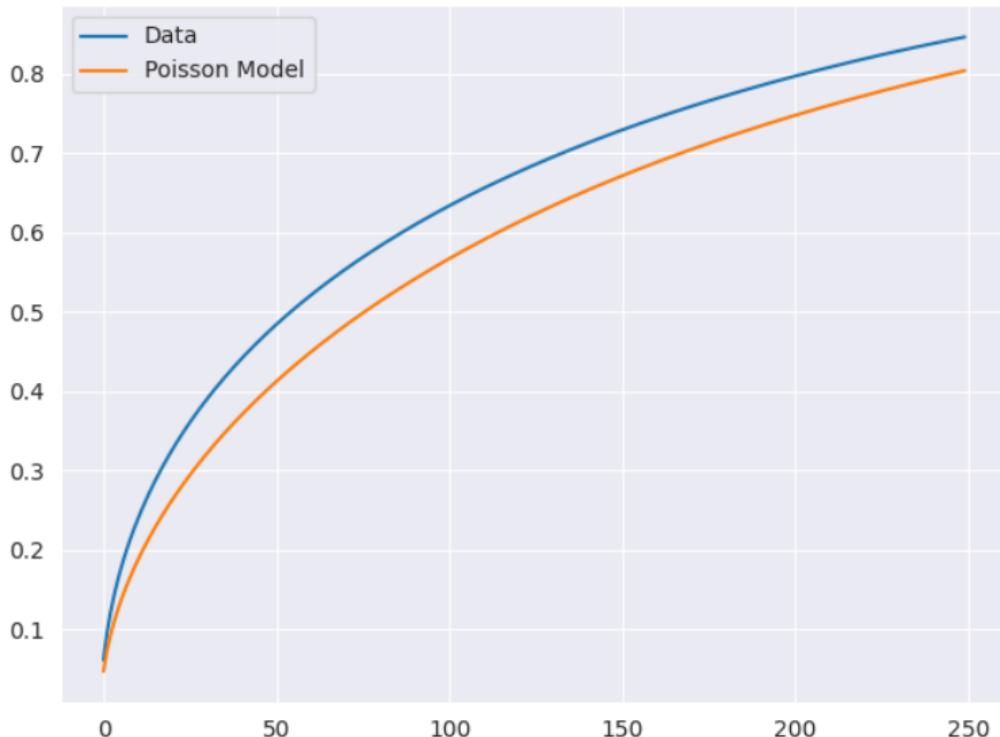
0. The consumer's consumption $\mathbf{c}_{i,t}$ is a zero vector at the beginning of the period.
1. The consumer randomly draws product j from the set of all products. The probability of drawing product j is product-specific and common across all consumers. The number of purchased products, n_j , is drawn from a product-specific Poisson distribution.
2. Update $\mathbf{c}_{i,t} \leftarrow \mathbf{c}_{i,t} + \mathbf{e}_j n_j$.¹
3. If the budget constraint for the new bundle is not violated ($\mathbf{p}'_{i,t} \mathbf{c}_{i,t} < m_{i,t}$), go to **Step 1**. Otherwise, **Stop**.

¹ \mathbf{e}_j is a vector with 1 at the j -th position and 0 elsewhere.

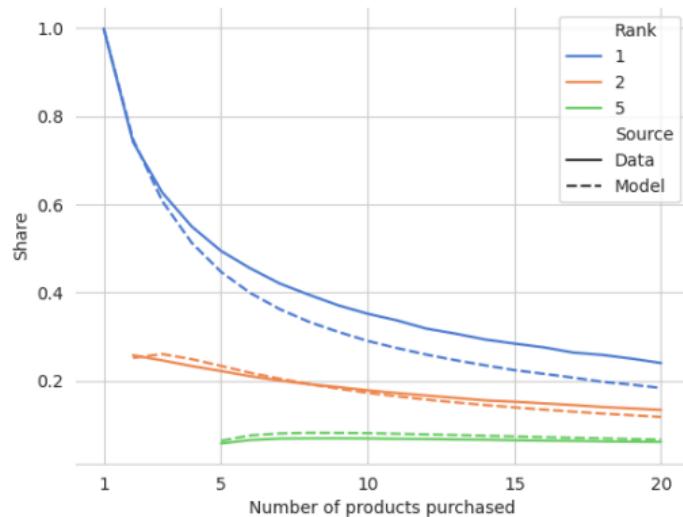
Extensive Margins in Choices: Model vs. Data



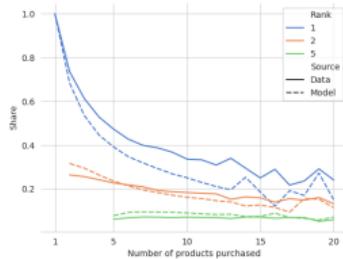
CDF of the top individual products



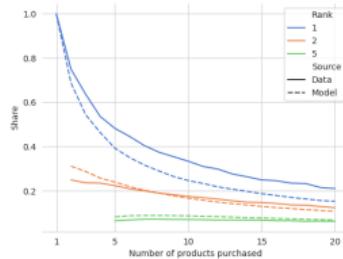
Average Spendings on Different Ranked Items



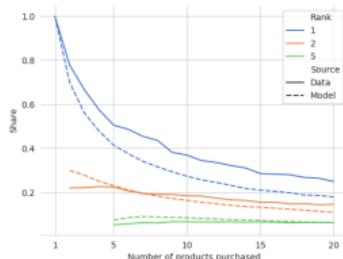
Category Spendings on Different Ranked Items



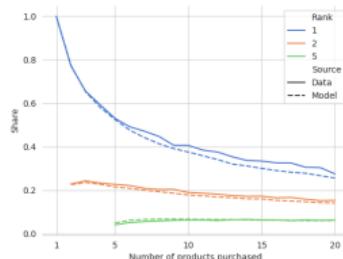
(a) Seafood Canned



(b) Cereal



(c) Yogurt



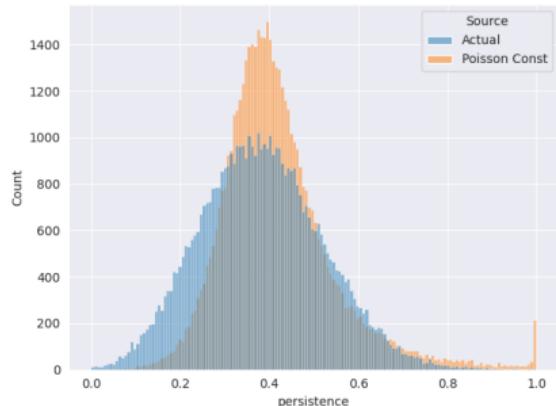
(d) Pet Food

Shopping Spree Model with Partially Persistent Choices

- ◊ Despite its extreme simplicity, the model is able to capture the main cross-sectional features of the data.
- ◊ Ad-hoc dynamics introduced by repeating the same problem over and over again would generate persistence much lower than observed in the data.
- ◊ However, if we introduce a small amount of persistence in the model, we can generate a very interesting result. Namely, we assume that each product purchased in the previous period is purchased again with probability $\rho = 0.4$. This can be interpreted as either (i) a smaller level of frictions for products purchased in the past or as (ii) a form of breadth of variety à la Menzio (Ecta, 2023).

Households Persistence: Model vs. Data

- ◊ Despite the fact that the model is calibrated to capture only the average dynamics of the data, it is able to capture the heterogeneity in persistence of consumption bundles. The reason for this is that documented extreme values of persistence are driven by households with a relatively fewer number of transactions.



☺ Ok... so what?

- ◊ **Purpose:** Highlight that empirical patterns typically modeled with heterogeneous preferences can also be generated by very different models. The exercise is analogous to:
 - “Balls-and-bins” model (Armenter&Koren, AER 2014) in the context of gravity models of international trade.
 - Price search (Menzio, 202*) in the context of the Dixit-Stiglitz monopolistic competition model.
 - Irrational consumers (Becker, JPE 1962) in the context of the utility-maximization theory.
- ◊ The proposed model, though simple, captures main features of the data very well.
Unlike models of heterogeneous choices, it relies on randomness, treating consumers as “zero-intelligence” agents with perfect substitutes.

☺ Ok... so what?

- ◊ **Purpose:** Highlight that empirical patterns typically modeled with heterogeneous preferences can also be generated by very different models. The exercise is analogous to:
 - “Balls-and-bins” model (Armenter&Koren, AER 2014) in the context of gravity models of international trade.
 - Price search (Menzio, 202*) in the context of the Dixit-Stiglitz monopolistic competition model.
 - Irrational consumers (Becker, JPE 1962) in the context of the utility-maximization theory.
- ◊ The proposed model, though simple, captures main features of the data very well.
Unlike models of heterogeneous choices, it relies on randomness, treating consumers as “zero-intelligence” agents with perfect substitutes.
- ◊ Policy implications differ significantly from heterogeneous preference models. Policies beneficial through the substitution effect of individual choices would not be beneficial in this model. For example, increasing product variety would be welfare decreasing in our setup, in contrast to being welfare increasing in a heterogeneous-preference framework by Neiman&Vavra (AEJ:Macro, 2023).

Concluding Thoughts

- ◊ Using scanner data, we document that consumption choices made by rich and poor households are surprisingly similar. Information on what product was purchased does not allow us to identify the group of households.
 **New empirical challenge for** models with intratemporal non-homothetic preferences.
- ◊ Individual consumption choices are very unstable. A product purchased in one year will be purchased again with a probability of 0.4 in the next year.
 **New empirical challenge for** models with latent heterogeneous preferences.
- ◊ We propose a new model where differences in consumption baskets are a result of randomness. This model fits the data quite well.
- ◊ The paper strikes a cautious note on policy implications drawn from models with heterogeneous choices on goods.

Thank you for your attention and have a nice lunch!

Group Differences in High-dimensional Choices

- ◊ Gentzkow, Shapiro, and Taddy (Ecta, 2019) discuss a bias in their study on the polarization of US politics using congressional speech data.
- ◊ Developed a model of a speech-generating process.
- ◊ Analyzed the bias formally and introduced estimators to **overcome finite-sample bias.**
- ◊ Their methods help recover a more **accurate estimate of polarization.**
- ◊ Their methodology is well suited for studying consumption polarization because the structure of speech data is very similar to scanner data.

▶ Gentzkow on Gentzkow et al. (Ecta, 2019)

▶ Back

Gentzkow on Gentzkow et al. (Ecta, 2019)



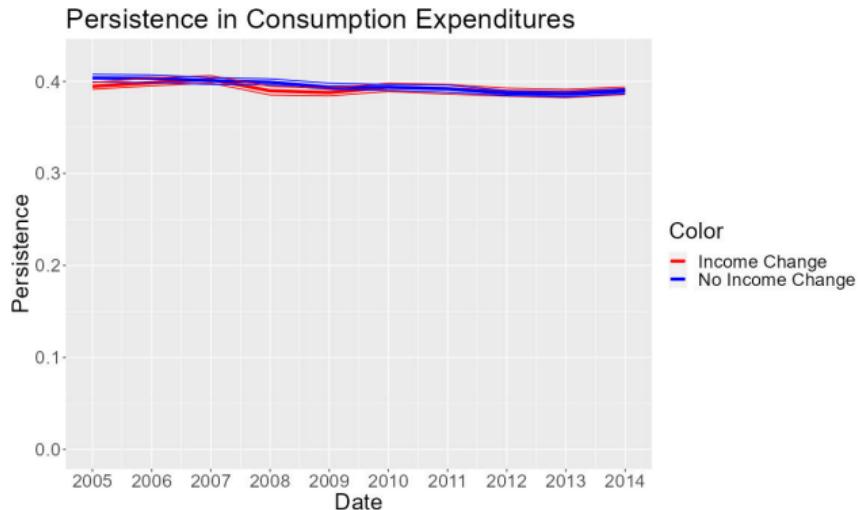
"Let's look at how different the choices of two different groups are in that high-dimensional space (...) It could be which UPCs did they buy in the supermarket."

Group Averages

Quantile	Size	Male Age	Female Age	Race	Income
1st	2.16	4.38	6.24	1.38	20
2nd	2.26	5.22	6.54	1.29	20.2
3rd	2.32	5.65	6.63	1.25	20.1
4th	2.33	5.94	6.65	1.21	20
5th	2.19	6.16	6.37	1.17	19.3

▶ Back

Persistence and Income Change



▶ Back

Zooming into the Heterogeneity

Question

Can we attribute the heterogeneity to observable characteristics of the households?

Zooming into the Heterogeneity

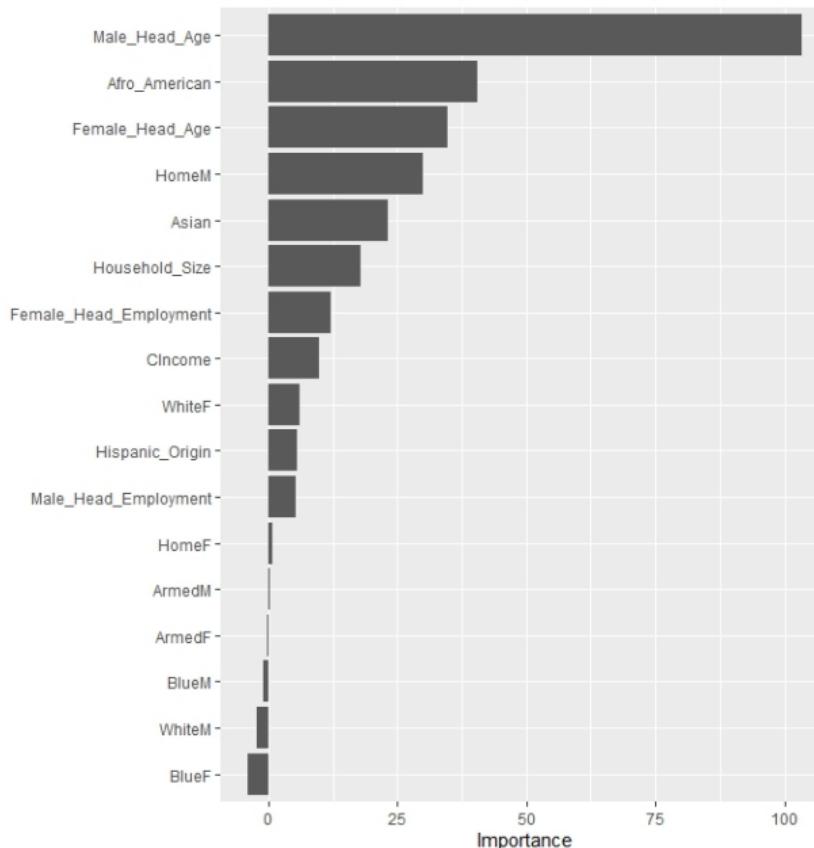
Question

Can we attribute the heterogeneity to observable characteristics of the households?

- ◊ We employ a random forest algorithm to find the household characteristics that are most important to explain differences within persistence
- ◊ We can then use accumulated local effects to see the size and direction of the influence of individual characteristics
 - Identifies the effect of singular variables while holding all else constant

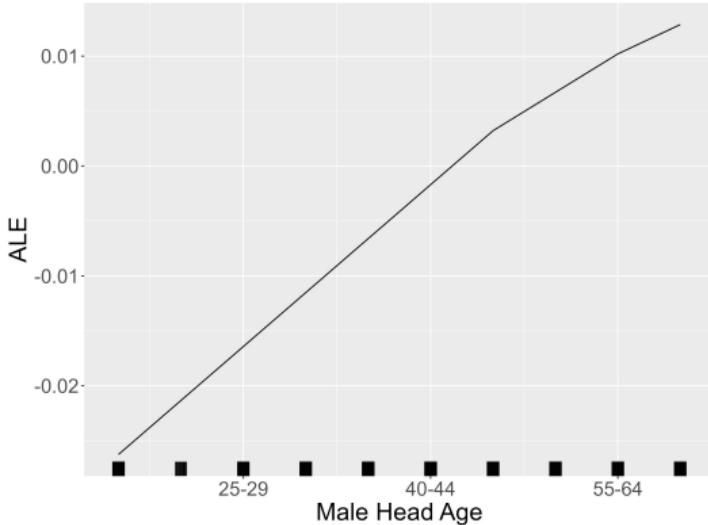
▶ Alternatively: Subgroup analysis

Variable Importance

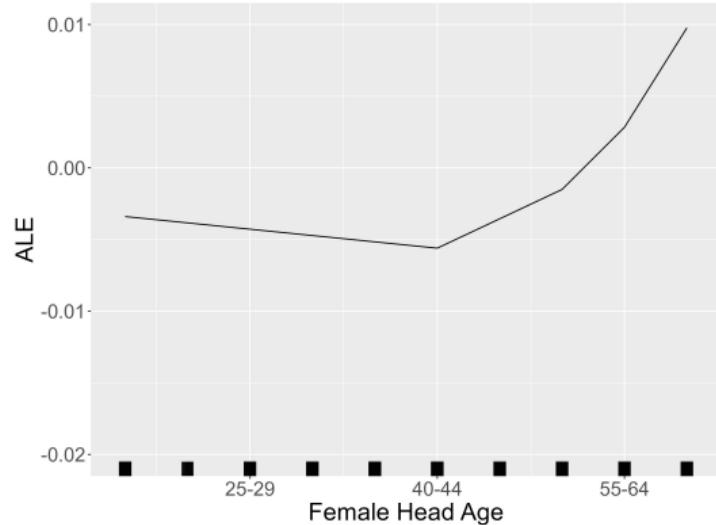


Effect Age

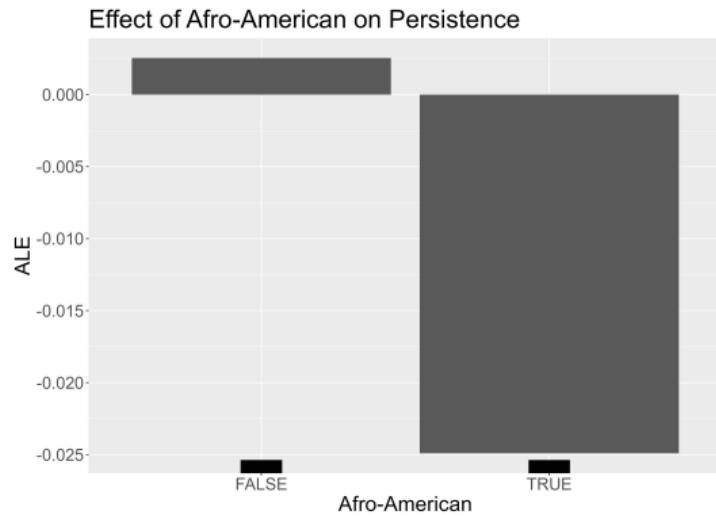
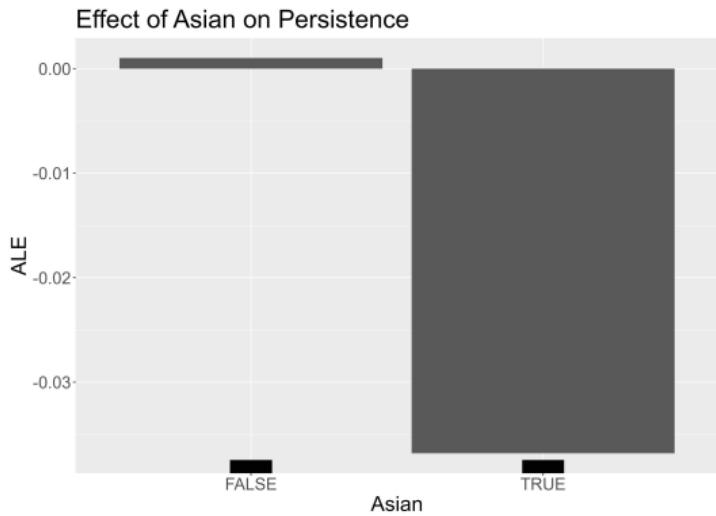
Effect of Male Head Age on Persistence



Effect of Female Head Age on Persistence



Effect Race



Results: Persistence Summary

- ◊ We have seen that households are **on average** highly impersistent with their consumption choices
- ◊ More than 50% of aggregate non-durable consumption is accounted for by households with a persistence of less than 50%
- ◊ **But** there is substantial heterogeneity within persistence
- ◊ Most of the heterogeneity is latent and only some can be explained by household characteristics.