Counterfactual Demand Predictions with Deep Learning

Mingyu (Max) Joo

UCR School of Business

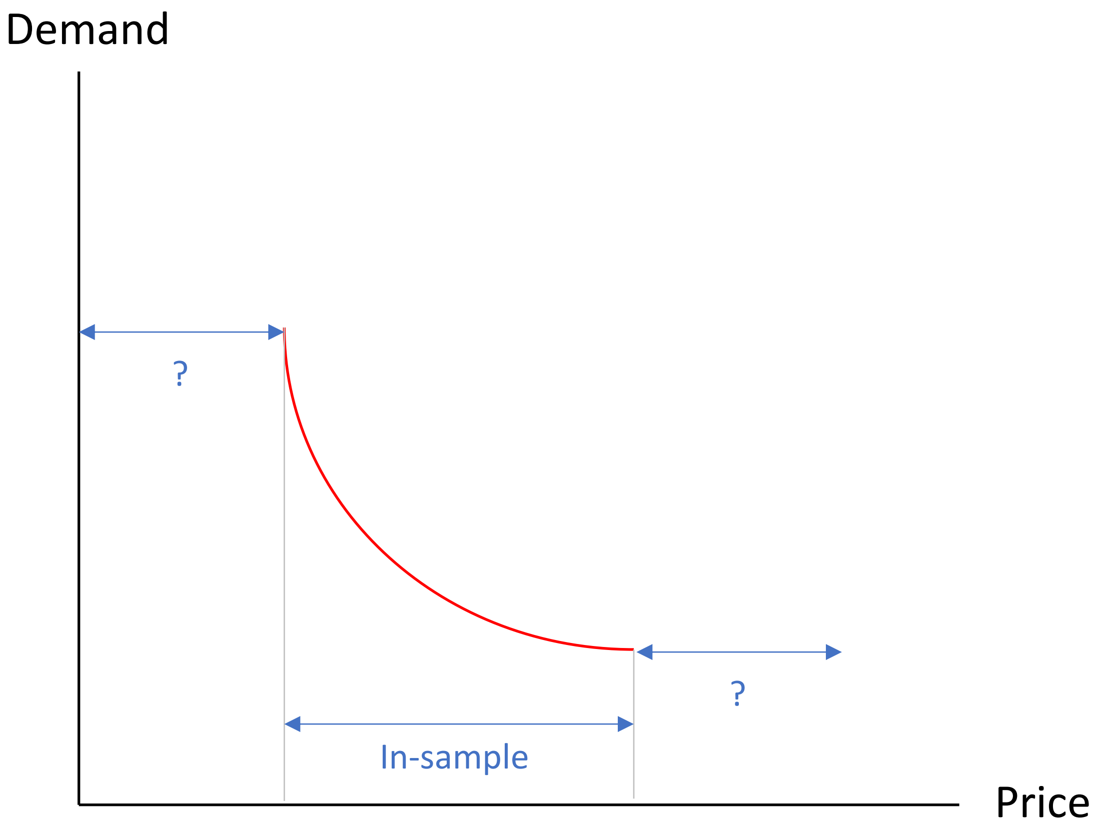
June 3, 2020

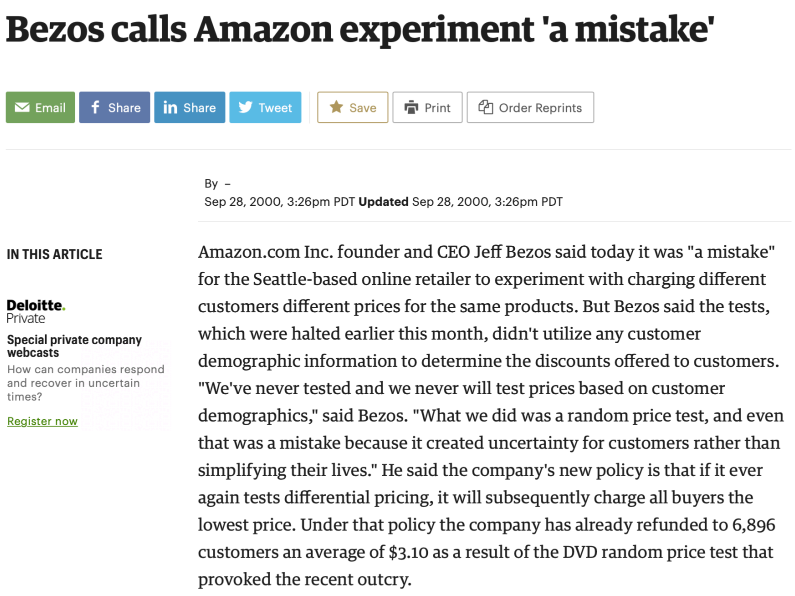
Symposium on Data Science & Statistics

Joint work with **Dong Soo Kim** (OSU), **Chul Kim** (CUNY Baruch), and **Hai Che** (UCR)

Counterfactual policy evaluation

- Counterfactual, what-if analyses guide policy-related questions





Bezos calls Amazon experiment 'a mistake'

Email | Share | in Share | Tweet | Save | Print | Order Reprints

By –
Sep 28, 2000, 3:26pm PDT **Updated** Sep 28, 2000, 3:26pm PDT

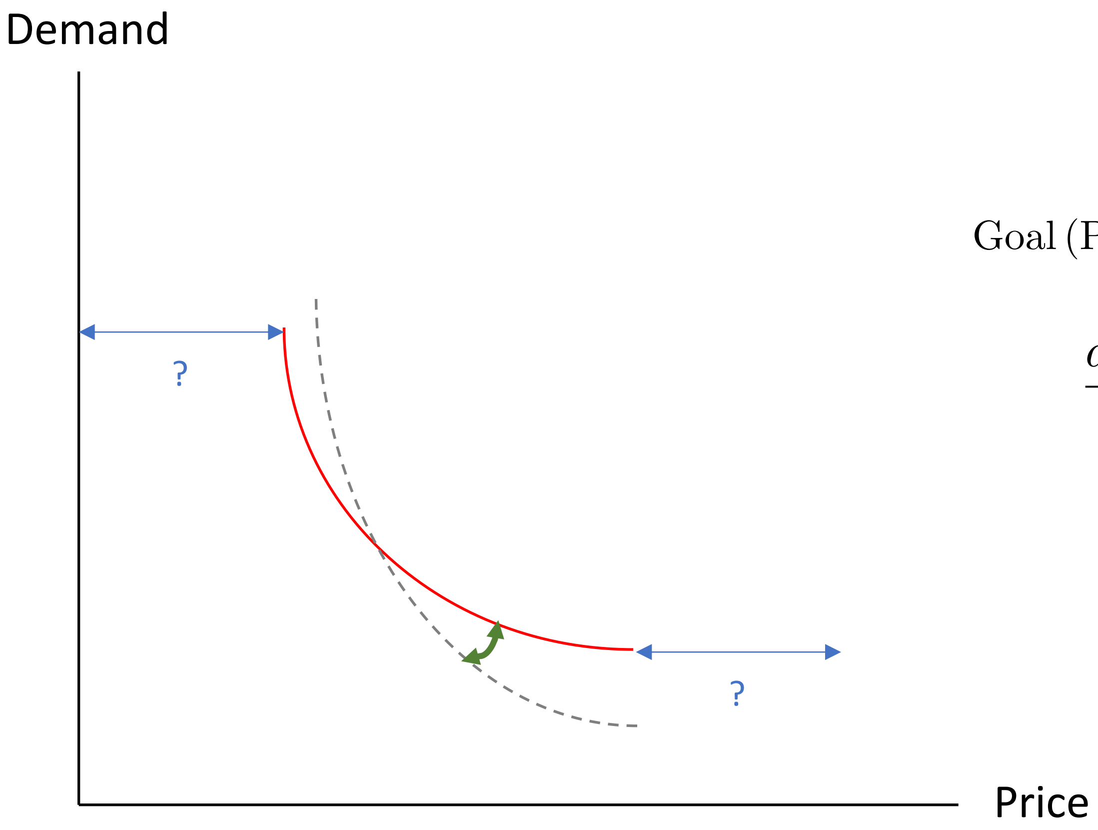
IN THIS ARTICLE

Deloitte. Private
Special private company webcasts
How can companies respond and recover in uncertain times?
Register now

Amazon.com Inc. founder and CEO Jeff Bezos said today it was "a mistake" for the Seattle-based online retailer to experiment with charging different customers different prices for the same products. But Bezos said the tests, which were halted earlier this month, didn't utilize any customer demographic information to determine the discounts offered to customers. "We've never tested and we never will test prices based on customer demographics," said Bezos. "What we did was a random price test, and even that was a mistake because it created uncertainty for customers rather than simplifying their lives." He said the company's new policy is that if it ever again tests differential pricing, it will subsequently charge all buyers the lowest price. Under that policy the company has already refunded to 6,896 customers an average of $3.10 as a result of the DVD random price test that provoked the recent outcry.

Counterfactual policy evaluation

- Of course, causal inference is the key requirement for counterfactual



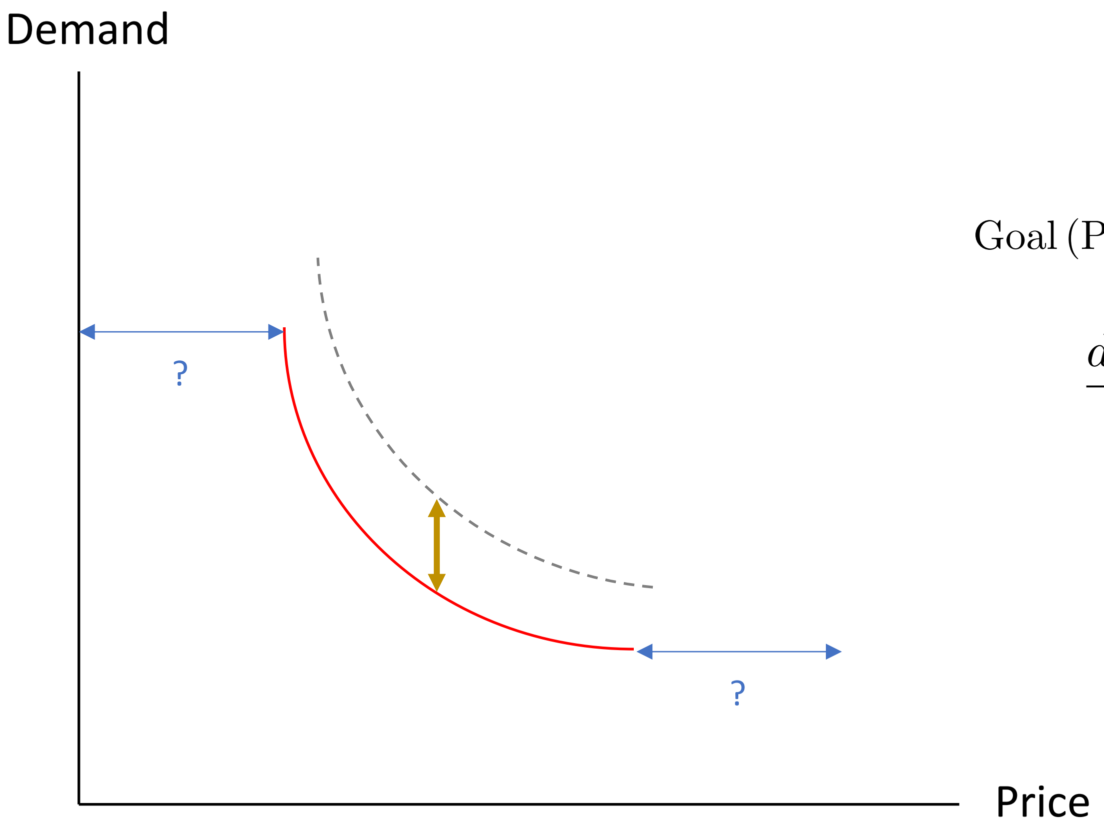
$$\text{Goal}\,(\text{Price}, \text{Demand} \mid \text{Environment}) \equiv \pi\,(p, y \mid \mathbf{x})$$

$$\frac{d\pi\,(p, y \mid \mathbf{x})}{dp} = \boxed{\frac{\partial \pi}{\partial y}\frac{\partial y}{\partial p}} + \frac{\partial \pi}{\partial p}\,(y \mid \mathbf{x})$$

Causal price effects require inference (i.e., price elasticity – so called $\hat{\beta}$)

Counterfactual policy evaluation

- However, causal inference is NOT the only requirement for counterfactual



$$\text{Goal}\,(\text{Price}, \text{Demand} \mid \text{Environment}) \equiv \pi\,(p, y \mid \mathbf{x})$$

$$\frac{d\pi\,(p, y \mid \mathbf{x})}{dp} = \frac{\partial \pi}{\partial y}\frac{\partial y}{\partial p} + \boxed{\frac{\partial \pi}{\partial p}(y \mid \mathbf{x})}$$

Demand shifters require prediction (e.g., seasonal effects – so called $\hat{\mathbf{y}}$)

Empirical models for policy evaluation

- Scalable, flexible machine learning has solved prediction policy problems
  - Who will *need* to be recommended for hip replacement surgery
  - Who will need to be *categorized* as our target customers
  - Flexible predictive methods are tuned for $\hat{y}$, but do not give useful guidance for $\hat{\beta}$
    - Athey (2017), Mullainathan and Spiess (2017)

Empirical models for policy evaluation

- Scalable, flexible machine learning has solved prediction policy problems
  - Who will *need* to be recommended for hip replacement surgery
  - Who will need to be *categorized* as our target customers
  - Flexible predictive methods are tuned for $\hat{y}$, but do not give useful guidance for $\hat{\beta}$
    - Athey (2017), Mullainathan and Spiess (2017)
- Causal inference is still required to fully resolve resource allocation problems
  - Who will *first need* to receive hip replacement surgery, under medical resource limitation
  - Who will need to be *prioritized* as our target customers, under limited couponing budget
  - Parsimonious structural models recover policy-invariant $\hat{\beta}$ at the expenses of low predictive accuracy of $\hat{y}$
    - Bajari et al. (2015), Athey and Imbens (2019)

Empirical models for policy evaluation

- Scalable, flexible machine learning has solved prediction policy problems
  - Who will *need* to be recommended for hip replacement surgery
  - Who will need to be *categorized* as our target customers
  - Flexible predictive methods are tuned for $\hat{y}$, but do not give useful guidance for $\hat{\beta}$
    - Athey (2017), Mullainathan and Spiess (2017)
- Causal inference is still required to fully resolve resource allocation problems
  - Who will *first need* to receive hip replacement surgery, under medical resource limitation
  - Who will need to be *prioritized* as our target customers, under limited couponing budget
  - Parsimonious structural models recover policy-invariant $\hat{\beta}$ at the expenses of low predictive accuracy of $\hat{y}$
    - Bajari et al. (2015), Athey and Imbens (2019)
- Focus on either prediction or inference

A hybrid approach

- Our proposal theoretically decomposes causal components and predictive components into *separable* functions

$$y = f\left(p, \mathbf{x}\right) = f\left(g\left(p\right), h\left(\mathbf{x}\right)\right)$$

Causal effect of price on demand

Demand shifter by environment

  - Each component can take flexible functional forms
  - Price responsiveness and demand shifters are interpreted as function values, so they are robust to specifications

- Flexible deep learning methods can be used for estimation and prediction of causal and predictive functions

Linear expenditure share curve

- Two standard microeconomic assumptions – functional separability and quasi-homotheticity – derives the linear cost function of good *i* given p_i , *m*

$$E_i = p_i y_i = p_i a_i(\mathbf{p}) + p_i \frac{b_i(\mathbf{p})}{b(\mathbf{p})}[m - a(\mathbf{p})]$$

where

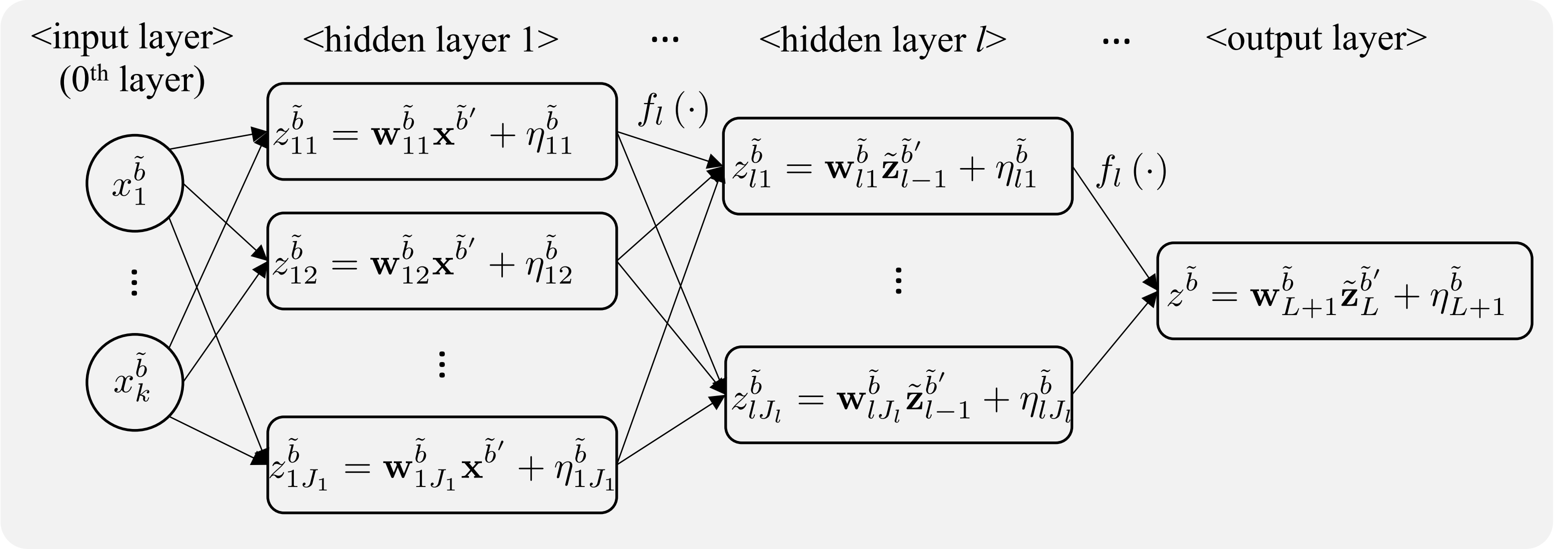
$a(\mathbf{p})$ is *budget-irrelevant* cost of living in the category and $a_i(\mathbf{p}) = \frac{\partial a(\mathbf{p})}{\partial p_i}$

$b(\mathbf{p})$ is additional expenditure of *remaining category budget* and $b_i(\mathbf{p}) = \frac{\partial b(\mathbf{p})}{\partial p_i}$

Feed-forward neural nets

- b, a, and m_t are trained as separate neural nets

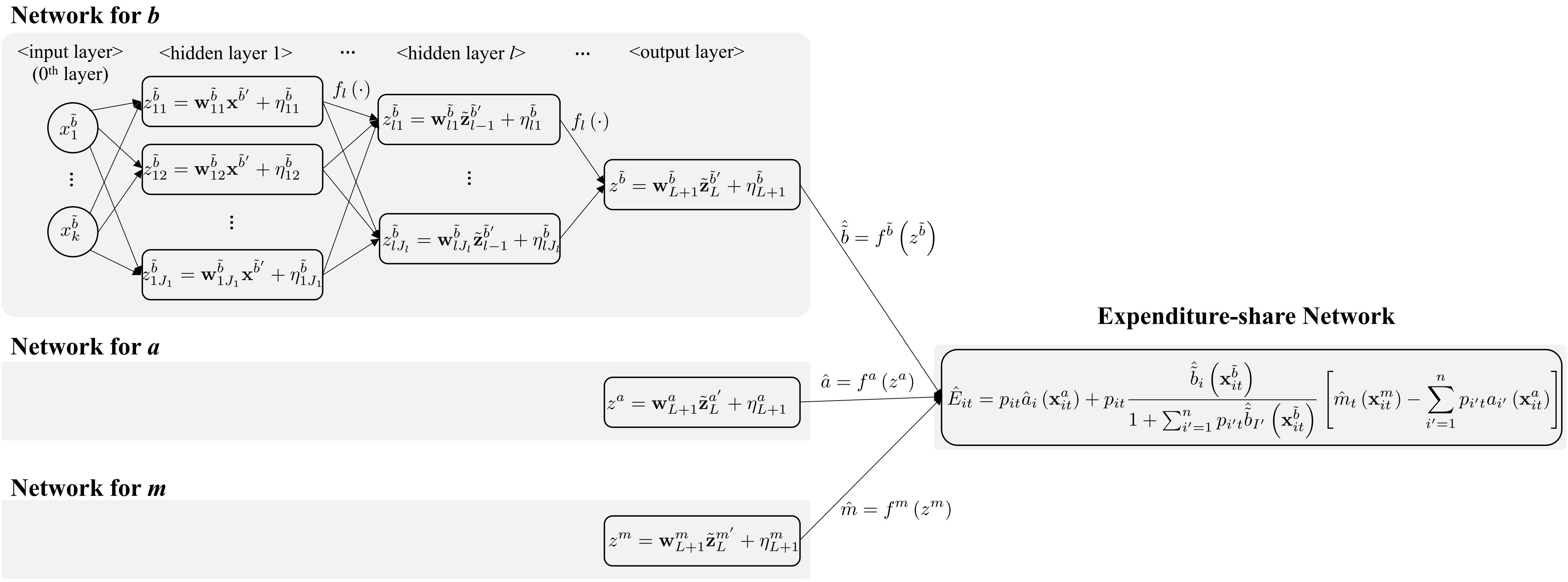
Network for *b*



- Linear combinations of input variables create output vectors
- One hidden layer with 10 nodes is used for empirical application

Feed-forward neural nets

- b, a, and m_t are trained as separate neural nets, then combined into expenditures



Loss function

Output values are evaluated by the following loss function:

$$L = \underbrace{\frac{\sum_{i,t}\left(\log \bar{E}_{it} - \log \hat{E}_{it}\right)^2}{nT}}_{\text{Sum of Squared Residuals}} + \underbrace{\theta_a \frac{\sum_{i,t}\log p_{it}}{nT}\frac{\sum_{i,t}(\log \bar{a}_{it} - \log \hat{a}_{it})^2}{nT}}_{L\text{-2 regularization for identification of the minimum demand quantity}} + \underbrace{\theta_m \frac{\sum_{i,t}(\log \bar{m}_t - \log \hat{m}_t)^2}{nT}}_{L\text{-2 regularization for identification of the category budget}}$$

Sum of Squared Residuals

L-2 regularization for identification of the minimum demand quantity

L-2 regularization for identification of the category budget

where

- *n* and *T* — # of goods and # of time periods in data
- $\bar{E}$, $\bar{a}$, and $\bar{m}$ — Expenditure, minimum quantity, and maximum expenditure observed in data
- $\hat{E}$, $\hat{a}$, and $\hat{m}$ — Fitted expenditure, minimum quantity, and maximum expenditure
- θ_a and θ_m — Tuning parameters for the regularization

Two counterfactual simulation studies

- *Extrapolation*
  - Does the proposed model predict a reasonable counterfactual demand curve outside of the observed price ranges?
  - No endogeneity

- *Endogeneity*
  - Is the proposed identification strategy robust to strategic pricing unobserved to researchers?
  - Demand fluctuations are perfectly correlated with price fluctuations
  - Seasonal price shocks with and without actual demand shocks

Extrapolation

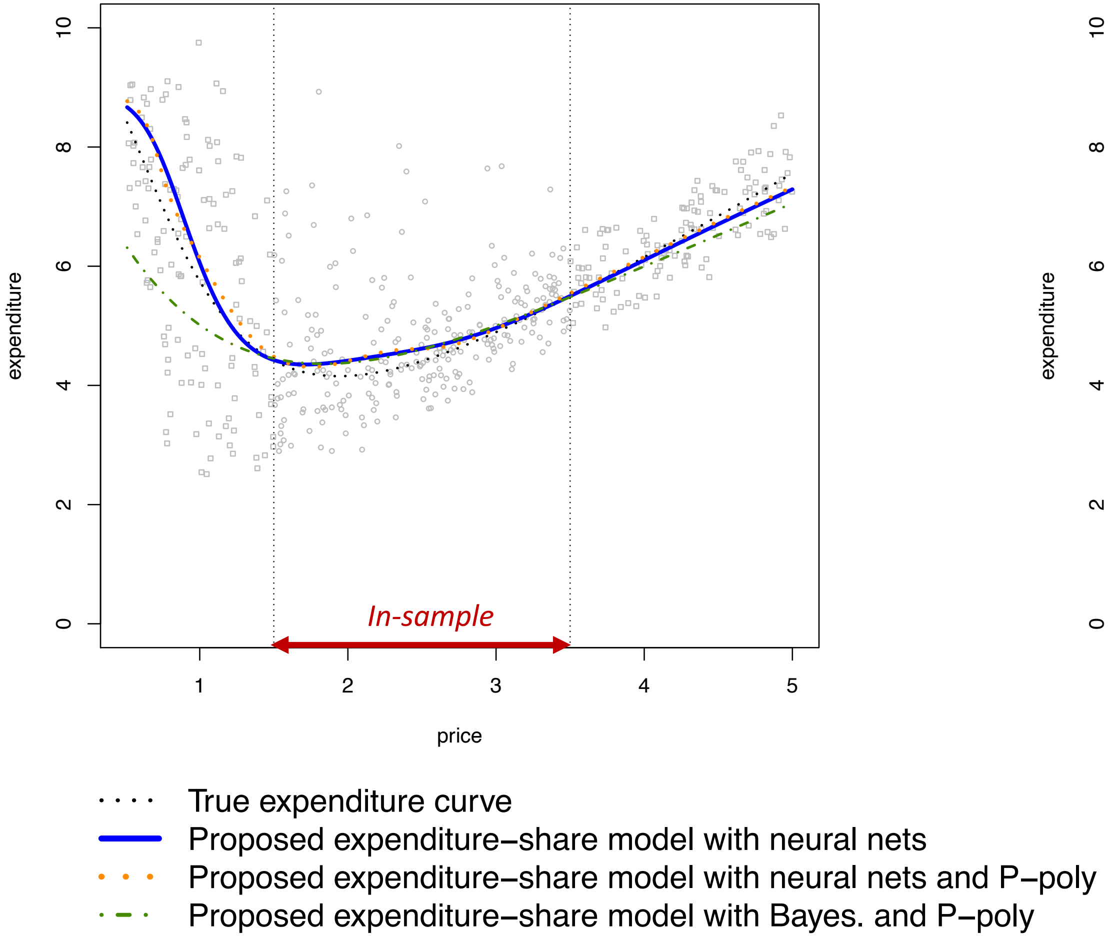
- Data generated by a translated CES function

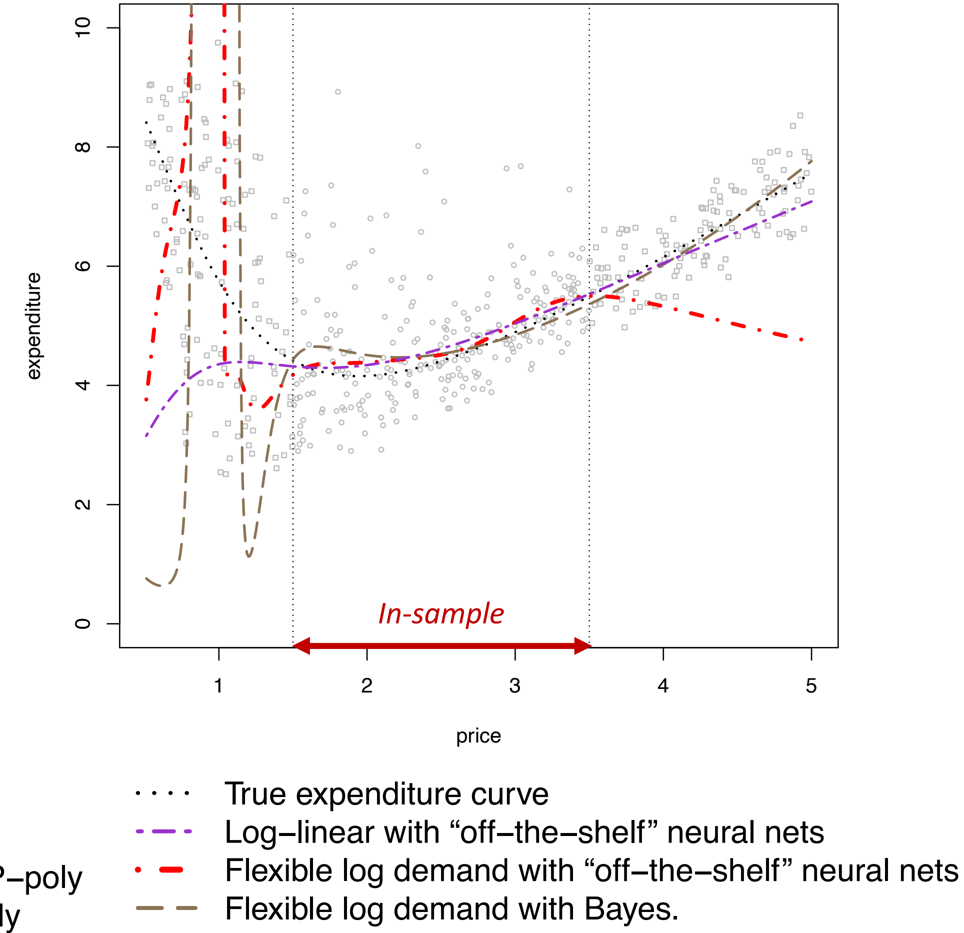
- Six empirical strategies
  - Proposed model & neural nets with **price term** as a predictor
  - Proposed model & neural nets with **price polynomials** as predictors
  - Proposed model & Bayesian estimation with **price polynomials** as predictors

  - Log demand model & “off-the-shelf” neural nets with **price term** as a predictor
  - Log demand model & “off-the-shelf” neural nets with **price polynomials** as predictors
  - Log demand model & Bayesian estimation with **price polynomials** as predictors

Extrapolation

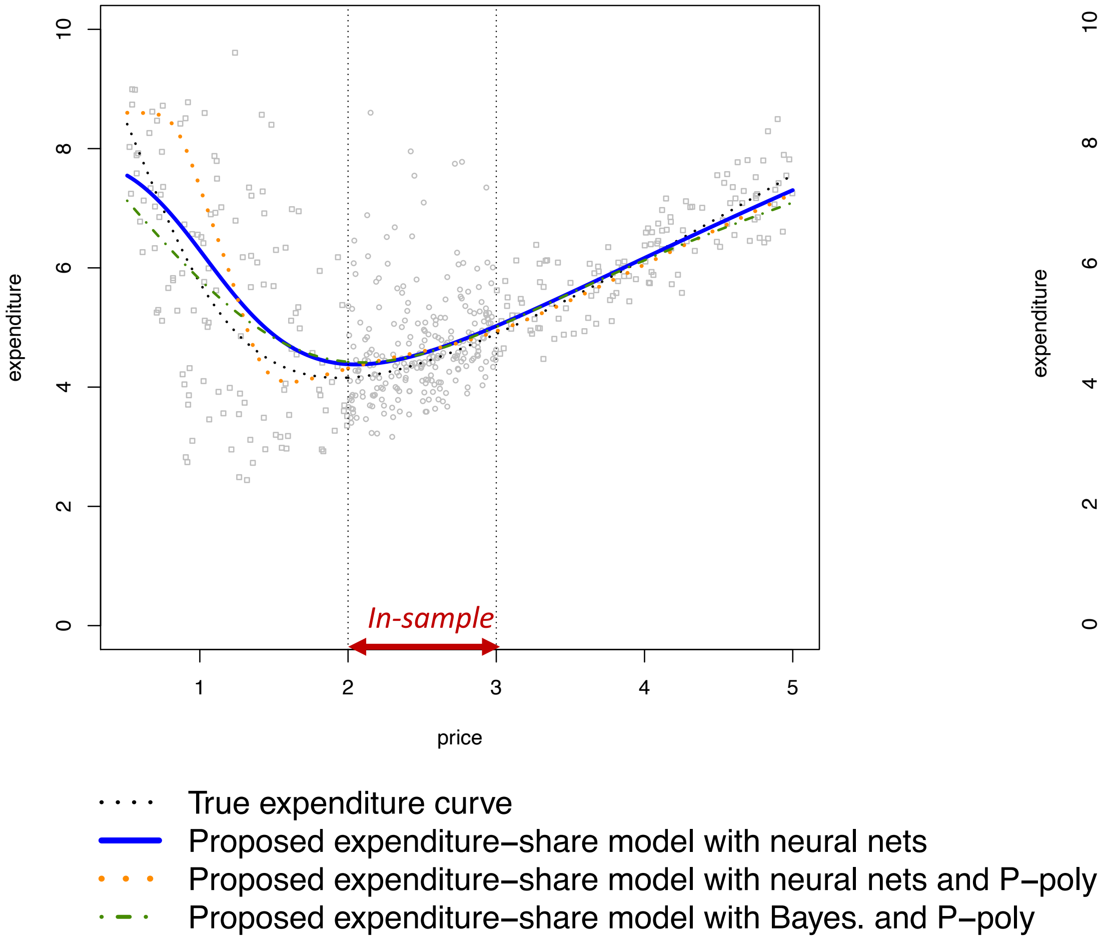
- Predicted demand curves

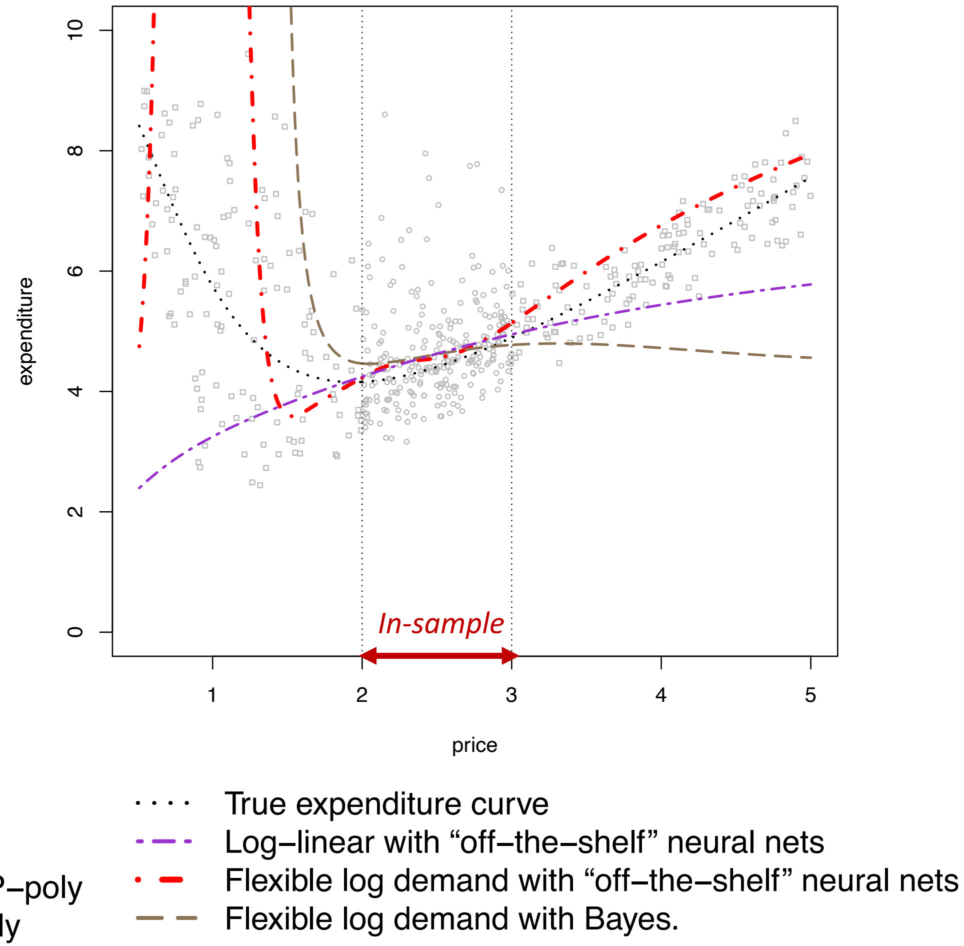




Extrapolation

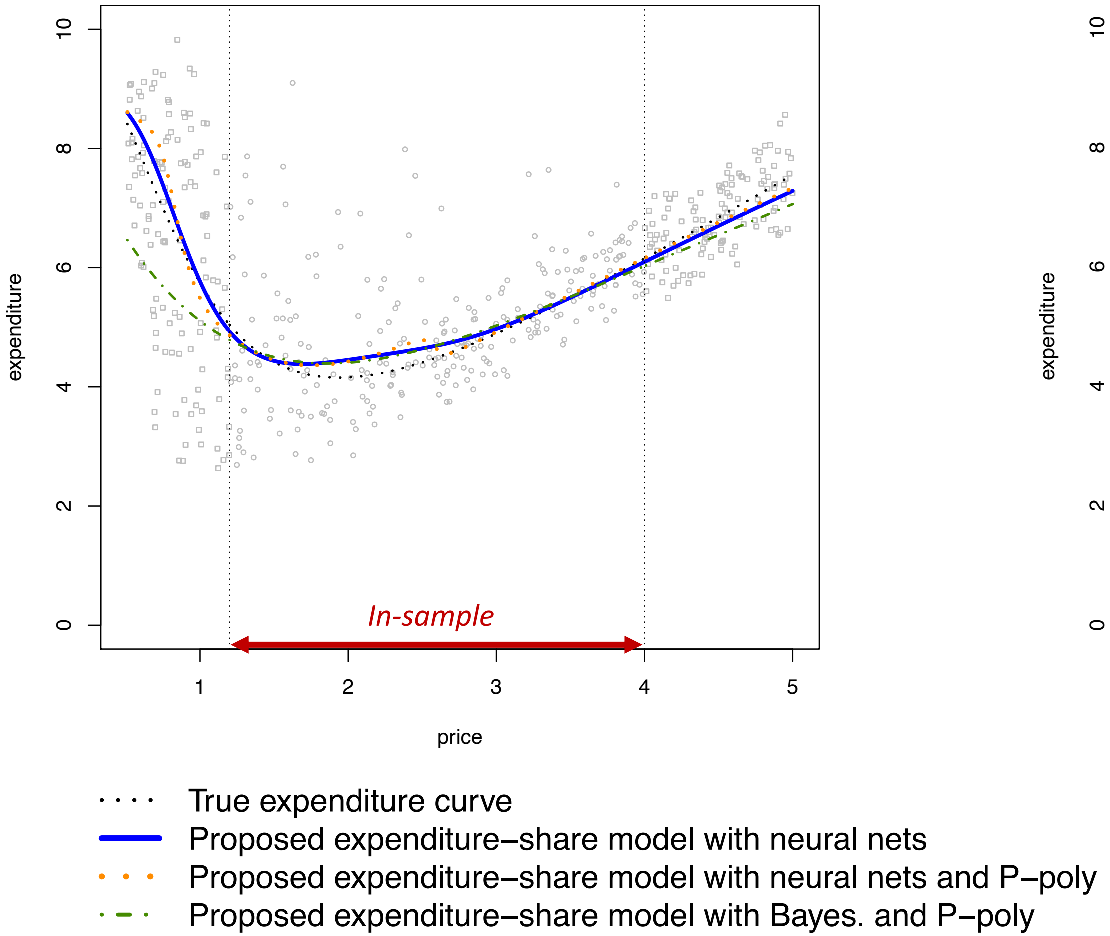
- Predicted demand curves

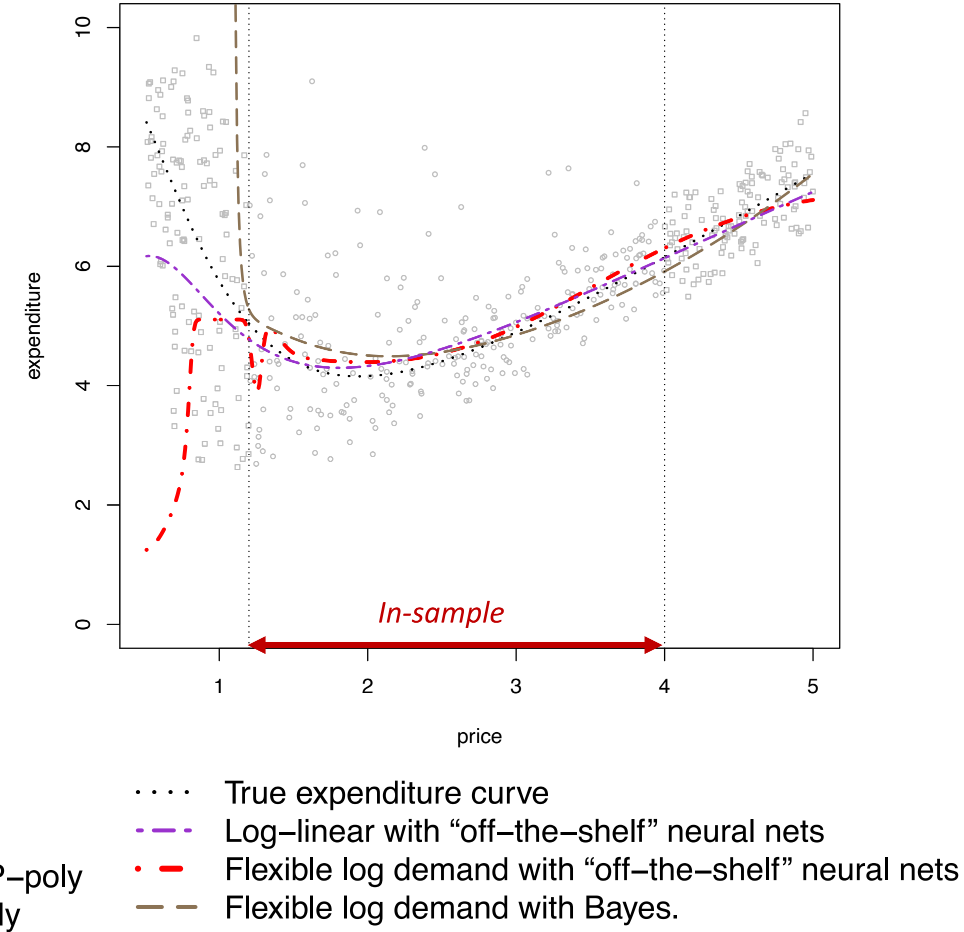




Extrapolation

- Predicted demand curves





Extrapolation

- Flexible models with "off-the-shelf" neural nets present superior in-sample MSE's

			Proposed models			Log demand models		
Price ranges	Seasons	True curve	P only neural nets	P-poly. neural nets	P-poly. Bayes.	P only neural nets	P-poly. neural nets	P-poly. Bayes.
A.Mid	Off	1.432	1.306	1.305	1.308	1.302	**1.297**	1.359
	Peak	0.916	0.873	0.871	0.872	0.875	**0.867**	0.889
B. Wide	Off	1.511	1.407	1.408	1.412	1.402	**1.368**	1.492
	Peak	0.994	0.962	0.946	0.960	0.964	**0.925**	0.992
C. Narrow	Off	1.205	1.083	1.074	1.079	1.074	**1.071**	1.088
	Peak	0.775	0.725	0.726	0.725	0.726	**0.722**	0.735

Note: Numbers in bold indicate the lowest MSE values among six competing methods.

Extrapolation

- Out-of-sample MSE's *within observed price ranges* are still better
  - Researchers may choose flexible models over proposed models

			Proposed models			Log demand models		
Price ranges	Seasons	True curve	P only neural nets	P-poly. neural nets	P-poly. Bayes.	P only neural nets	P-poly. neural nets	P-poly. Bayes.
A.Mid	Off	1.738	1.101	2.162	1.887	**1.100**	1.108	1.104
	Peak	0.934	0.903	2.409	1.437	**0.897**	0.909	0.947
B. Wide	Off	2.246	**1.210**	2.948	3.225	1.216	1.216	1.268
	Peak	1.011	0.974	3.446	2.019	**0.969**	0.983	0.997
C. Narrow	Off	1.635	0.977	1.215	1.092	0.976	0.977	**0.974**
	Peak	0.764	0.715	1.094	0.803	**0.713**	0.721	0.728

Note: Numbers in bold indicate the lowest MSE values among six competing methods.

Extrapolation

- However, proposed models fit better *outside of observed price ranges*
  - Proposed models offer more accurate counterfactual predictions for optimal pricing

			Proposed models			Log demand models		
Price ranges	Seasons	True curve	P only neural nets	P-poly. neural nets	P-poly. Bayes.	P only neural nets	P-poly. neural nets	P-poly. Bayes.
A.Mid	Off	3.873	4.133	**3.964**	5.204	10.906	93.326	*Inf.*
	Peak	3.205	3.641	**3.497**	4.183	8.534	91.067	*Inf.*
B. Wide	Off	3.877	**4.015**	4.403	5.459	6.472	19.112	*Inf.*
	Peak	3.328	**3.565**	3.917	4.420	4.236	17.515	*Inf.*
C. Narrow	Off	3.523	**3.423**	4.653	3.462	10.755	475.564	*Inf.*
	Peak	2.812	2.869	4.304	**2.818**	10.493	523.131	*Inf.*

Note: Numbers in bold indicate the lowest MSE values among six competing methods.

Endogeneity

- Data generated under cyclical pricing
  - Firm is aware of the season, but not aware of the amount of seasonal demand shock
  - Firm raises prices during the peak season
  - Firm raises prices in reaction to random positive shocks

- Two empirical strategies based on the out-of-sample fit (within range)
  - Proposed model & neural nets with **price term** as a predictor
  - Log demand model & “off-the-shelf” neural nets with **price term** as a predictor

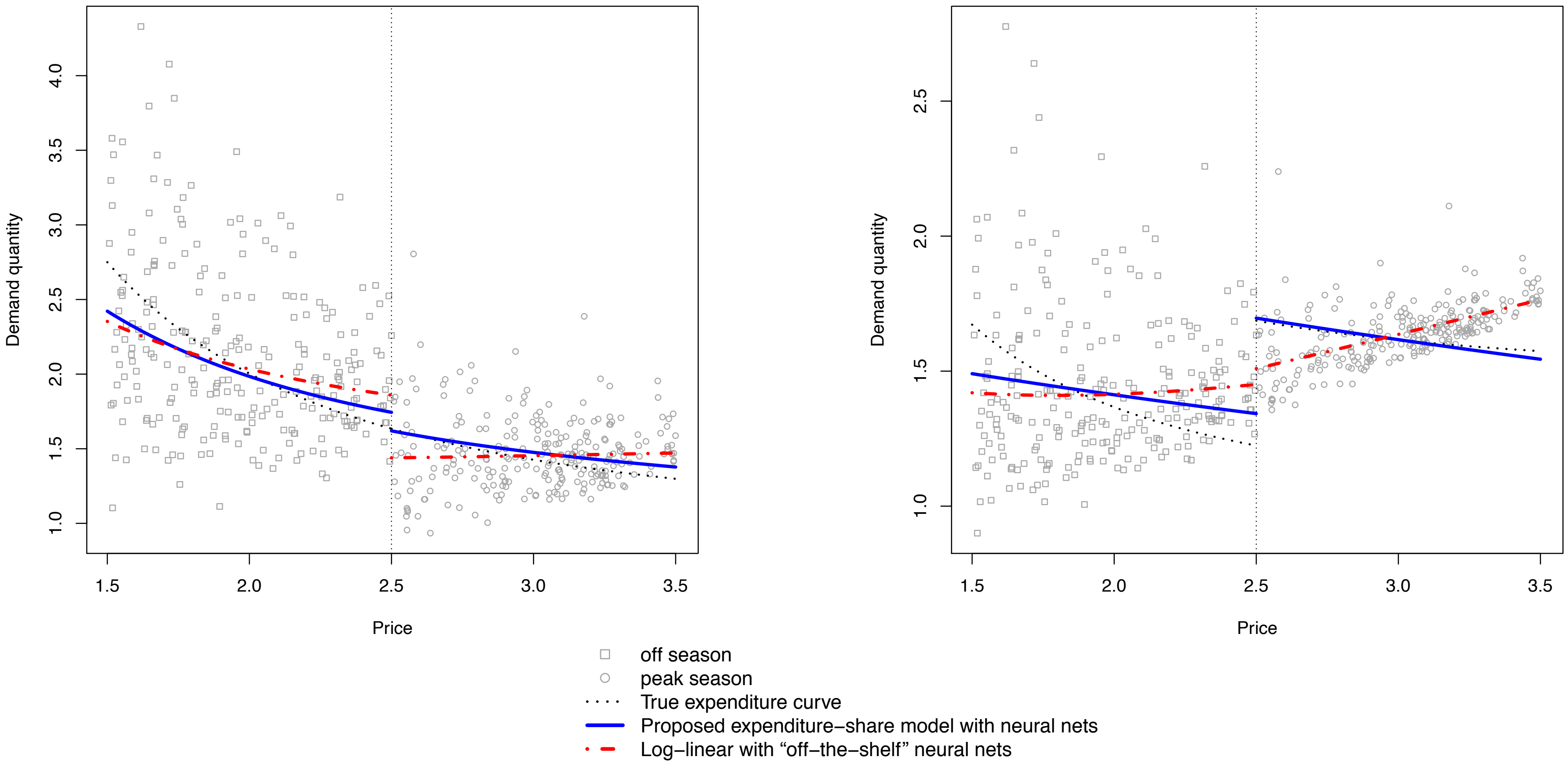
- Two different situations
  - There IS true demand shock during peak season
  - There IS NOT true demand shock during peak season

Endogeneity

- Predictive demand curves robust to endogeneity bias

w/o true demand shock

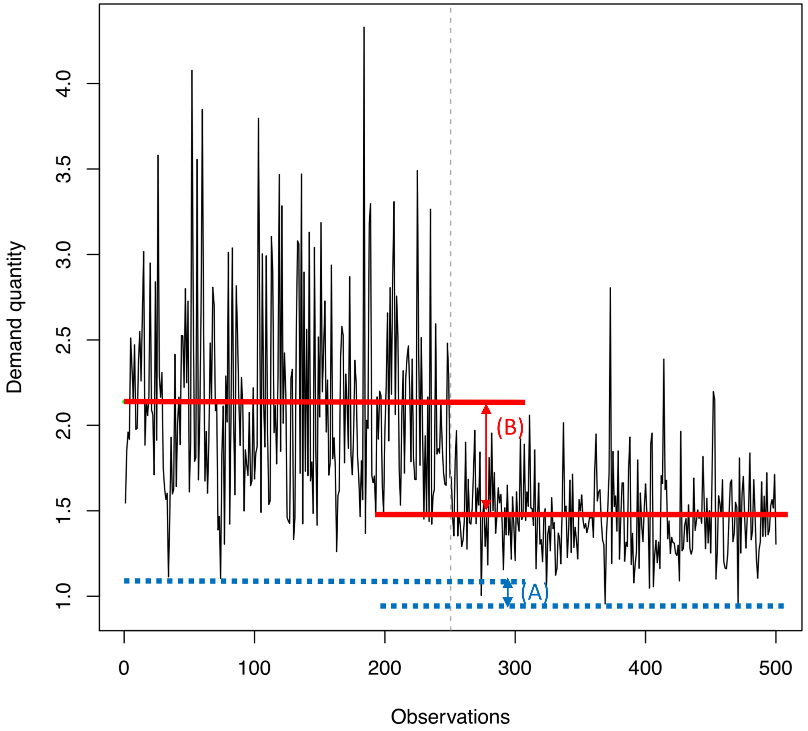
w/ true demand shock



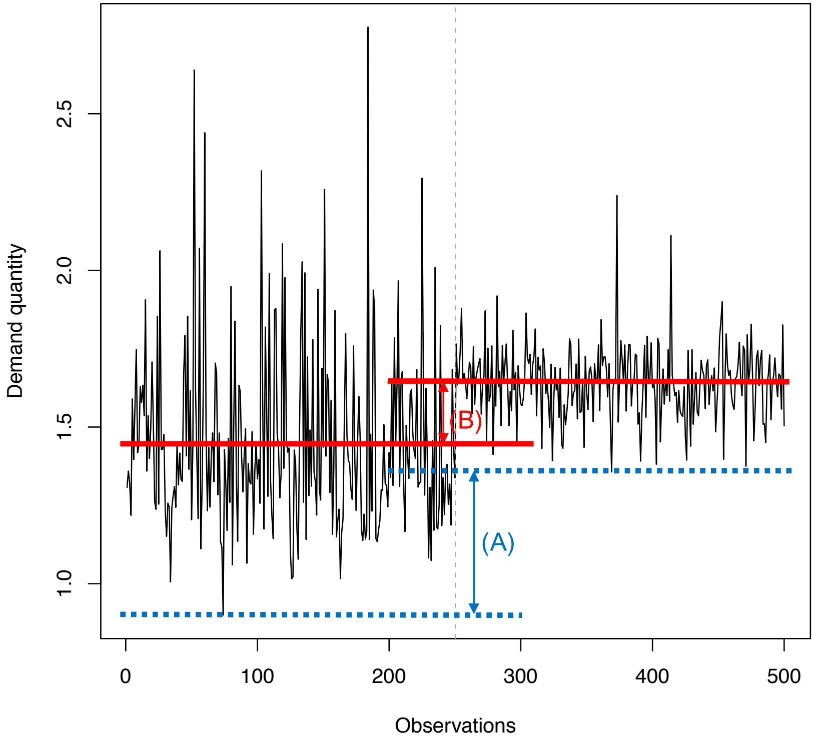
Endogeneity

- Minimum quantity captures seasonality more robust to strategic pricing
  - Mean fixed effects include confounds – i.e., aggregate demand shift due to strategic pricing
  - Price responses are also more robust in the proposed framework w/o good instruments

w/o true demand shock



w/ true demand shock

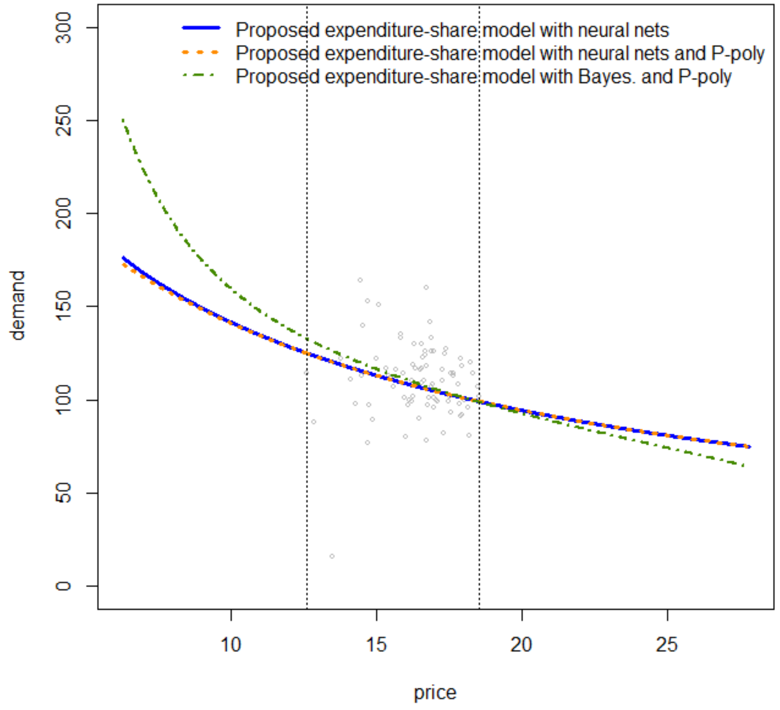


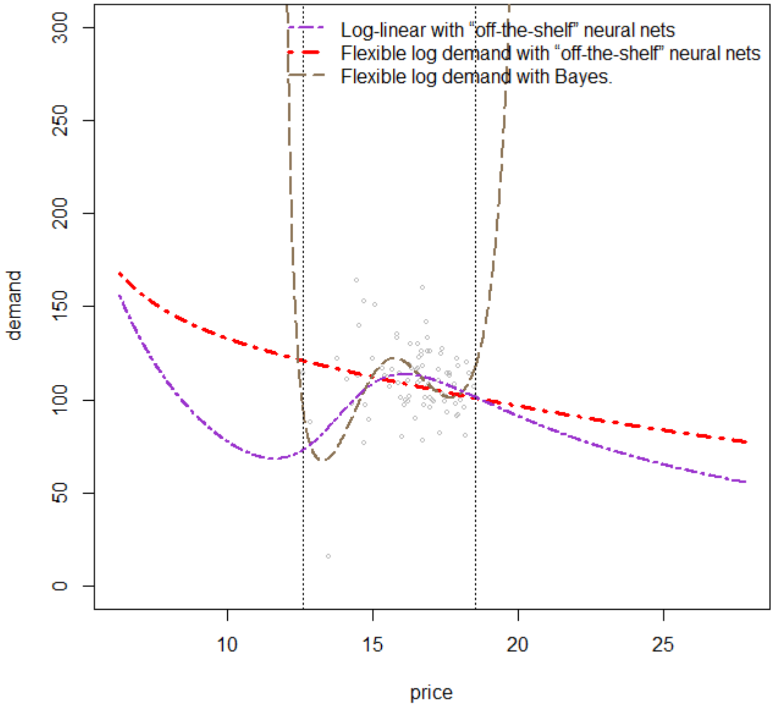
Preliminary application

- Complete transaction data in the diaper category of a grocery store in the Bay area
  - Nationwide chain
  - Information available at the individual level
  - Accurate price/sales information observed
- We analyze weekly aggregation of the sales quantity
  - Estimation sample: May 2005 ~ Dec 2006
  - Holdout sample: Jan 2007 ~ May 2007
  - Small number of data points
    - 84 observations in sample
    - 22 observations out of sample
- This is a preliminary test
  - Further analysis in progress

All diapers

- Proposed models predict more stable demand curves, despite small # of obs.
  - Flexible predictive models potentially overfit the sample





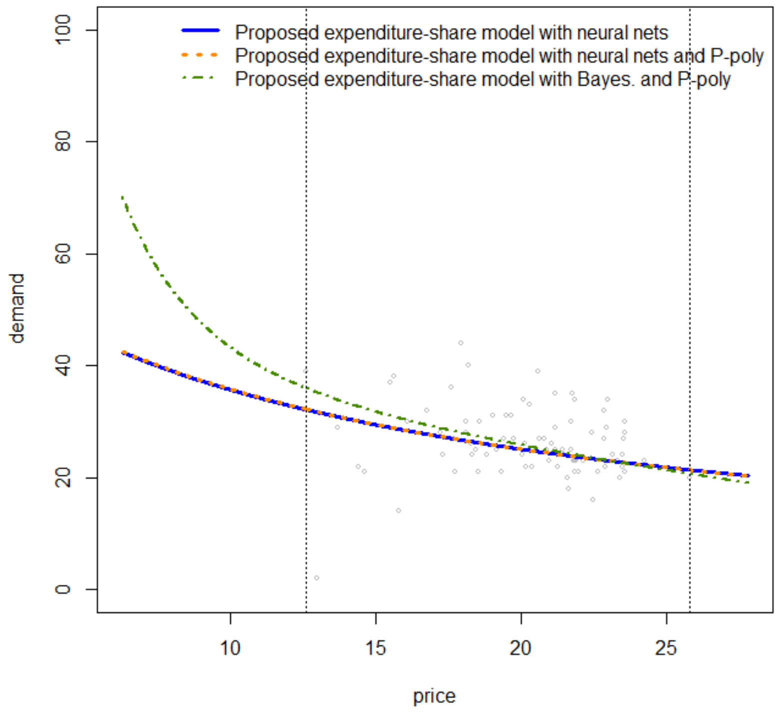
All diapers

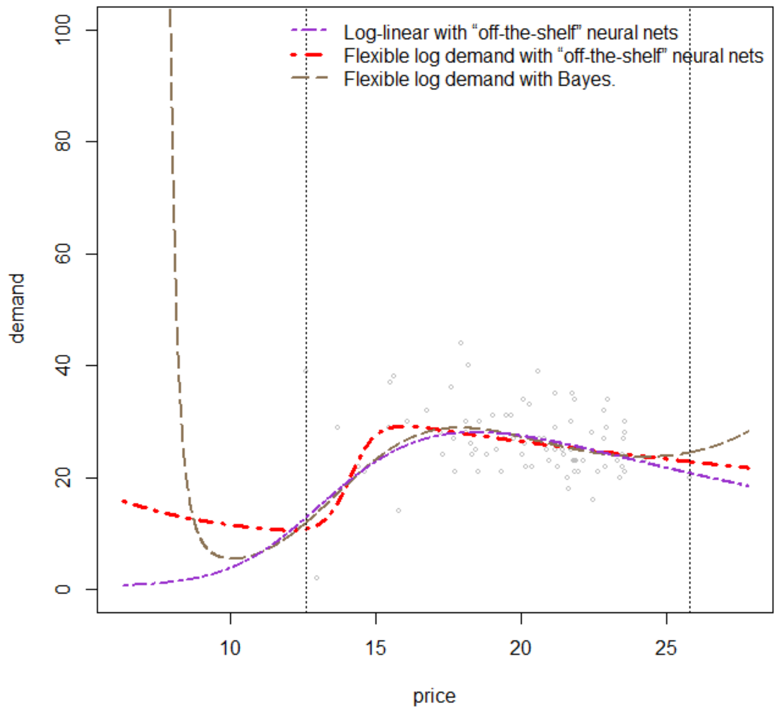
- Proposed models improve out of sample MSE's by about 9~44%
  - This is without regime shifts observed

	In-sample (before 2007)	**Out-of-sample (after 2007)**
P only NN (proposed)	105,928.20	**55,404.57**
P-poly NN (proposed)	105,845.70	**55,497.91**
P-poly Bayes (proposed)	107,594.90	**62,428.08**
P only NN (log demand)	**94,440.84**	98,498.49
P-poly NN (log demand)	**102,349.40**	60,975.31
P-poly Bayes (log demand)	**101,641.00**	89,882.08

Pampers

- Proposed models predict more stable demand curves, despite small # of obs.
  - Flexible predictive models potentially overfit the sample





Pampers

- Proposed models improve out of sample MSE’s by about 23~28%
  - This is without regime shifts observed

	In-sample (before 2007)	Out-of-sample (after 2007)
P only NN (proposed)	14,149.01	**13,542.20**
P-poly NN (proposed)	14,153.27	**13,545.53**
P-poly Bayes (proposed)	14,278.67	**14,462.53**
P only NN (log demand)	**12,894.89**	18,912.17
P-poly NN (log demand)	**12,739.95**	17,711.84
P-poly Bayes (log demand)	**12,944.71**	18,673.13

Takeaways

- We suggest a theory-based identification strategy to decompose demand fluctuations into environment-driven shifters and budget-driven price responses
  - Our strategy yields a good predictive performance for counterfactual pricing
  - It also presents a superior out-of-sample prediction in real-world data

- The proposed model is robust to endogeneity concerns
  - It exploits further information in the data – i.e., minimum quantity
  - It is useful especially when there is no good instrument

- The theoretical regularization combined with neural nets offers accurate prediction of demand shifters and reasonable approximation of causal price effects
  - Scalable and flexible method, yet stable across policy regime spaces

Thank you

- Questions and comments: mingyu.joo@ucr.edu

- We can share a preliminary version of manuscript
  - Empirical application not included yet

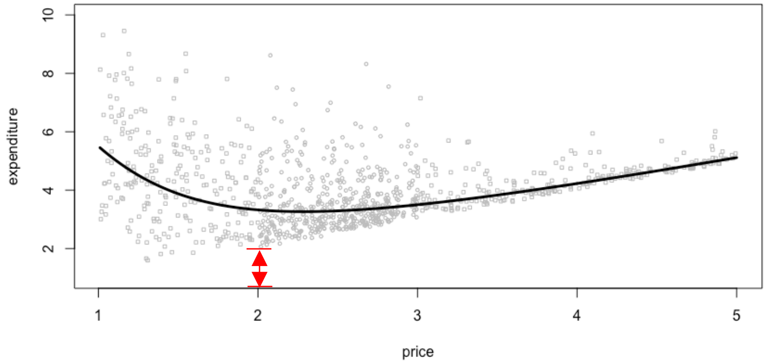
- Stay well!

Appendix

Identification strategy (within-season)

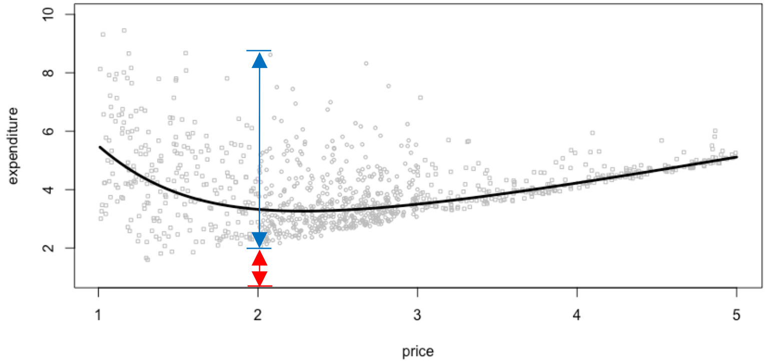
A. Minimum cost-of-living function

- Distinguished from "mean" fixed effects in most demand models, the ***minimum demand quantity*** identifies the cost-of-living function - $a_i(\cdot)$ - as the baseline demand shifter.



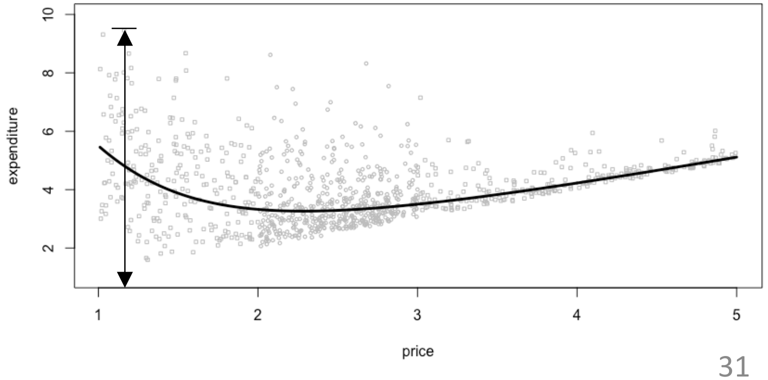
B. Price sensitivity function

- ***Expenditure for the excess quantity*** identifies the (pure) price sensitivity - $\tilde{b}_i(\cdot)$ - as the response to price fluctuations.



C. Category budget

- The maximum expenditure over time identifies the category budget - m_t - as the upper-bound of the expenditure.



Theoretical assumptions

Functional (or weak) separability (Deaton and Muellbauer 1980)

- Preferences independent from other categories (e.g., soft drinks vs. clothing)
- Allowing for a separable sub-utility maximization subject to category budget allocation

$$u(\mathbf{y}) = u\left(\underline{v_{C_1}(y_1, y_2, y_3)}, \ldots, v_{C_K}(y_{N-1}, y_N)\right)$$

Theoretical assumptions

Functional (or weak) separability (Deaton and Muellbauer 1980)

- Preferences independent from other categories (e.g., soft drinks vs. clothing)
- Allowing for a separable sub-utility maximization subject to category budget allocation

$$u(\mathbf{y}) = u\left(v_{C_1}(y_1, y_2, y_3), \ldots, v_{C_K}(y_{N-1}, y_N)\right)$$

Quasi-homotheticity (Gorman 1976, Deaton and Muellbauer 1980)

- Budget increases lead to proportional increases in expenditures, beyond the fixed cost-of-living.

$$\max_{y_1, y_2, y_3} \quad v_{C_1}(y_1, y_2, y_3) \qquad \text{s.t.} \quad p_1 y_1 + p_2 y_2 + p_3 y_3 = E_{C_1}$$

$$\Rightarrow \quad E(\mathbf{p}, v) = a(\mathbf{p}) + b(\mathbf{p})\, v$$

$a(\mathbf{p})$: The minimum cost-of-living (e.g., eggs to feed family)

$b(\mathbf{p})\, v$: Optimal allocation of remaining budget (e.g., additional eggs for baking)

Empirical model

- Log expenditure of good *i* at time period *t* is given by

$$\log E_{it} = \log \left\{ p_{it} a_i \left(\mathbf{p}_{-0,t} | \mathbf{r}_t \right) + p_{it} \frac{\tilde{b}_i \left(\mathbf{p}_{-0,t} \right)}{1 + \sum_{i'=1}^{n} p_{i't} \tilde{b}_{i'} \left(\mathbf{p}_{-0,t} \right)} \left[m_t - \sum_{i'=1}^{n} p_{i't} a_{i'} \left(\mathbf{p}_{-0,t} | \mathbf{r}_t \right) \right] \right\} + \epsilon_{it}$$

  - Outside option price is normalized to be one ($p_{0t} \equiv 1$)
  - Minimum cost-of-living for outside option is normalized to be zero ($a_{0t} \equiv 0$)
  - Budget-relevant cost for outside option is normalized to be one ($\tilde{b}_i \left(\mathbf{p}_{-0,t} \right) \equiv b_i \left(\mathbf{p}_{-0,t} \right) / b_0 \left(\mathbf{p}_{-0,t} \right)$)
  - Environmental variables control for seasonal fluctuations ($\mathbf{r}_t$)

Empirical model

- Log expenditure of good *i* at time period *t* is given by

$$\log E_{it} = \log \left\{ p_{it} a_i \left(\mathbf{p}_{-0,t} | \mathbf{r}_t\right) + p_{it} \frac{\tilde{b}_i \left(\mathbf{p}_{-0,t}\right)}{1 + \sum_{i'=1}^{n} p_{i't} \tilde{b}_{i'} \left(\mathbf{p}_{-0,t}\right)} \left[m_t - \sum_{i'=1}^{n} p_{i't} a_{i'} \left(\mathbf{p}_{-0,t} | \mathbf{r}_t\right) \right] \right\} + \epsilon_{it}$$

  - Outside option price is normalized to be one ($p_{0t} \equiv 1$)
  - Minimum cost-of-living for outside option is normalized to be zero ($a_{0t} \equiv 0$)
  - Budget-relevant cost for outside option is normalized to be one ($\tilde{b}_i \left(\mathbf{p}_{-0,t}\right) \equiv b_i \left(\mathbf{p}_{-0,t}\right) / b_0 \left(\mathbf{p}_{-0,t}\right)$)
  - Environmental variables control for seasonal fluctuations ($\mathbf{z}_t$)

- Functions to be estimated by neural nets are