Counterfactual Demand Predictions: Deep Learning with Microeconomic Structure

Dongsoo Kim
Ohio State University
Columbus, OH 43210
kim.5042@osu.edu

Chul Kim
CUNY Baruch
New York, NY 10010
chul.kim@baruch.cuny.edu

Mingyu Joo
UC Riverside
Riverside, CA 92521
mingyu.joo@ucr.edu

Hai Che
UC Riverside
Riverside, CA 92521
chehai@ucr.edu

Abstract

This paper proposes a framework to estimate consumer demand with rich observational data and to predict counterfactual outcomes using deep neural networks regularized by microeconomic theory. Two microeconomic assumptions, weak-separability and quasi-homotheticity, decompose price responses and demand shifters into separable functions in a linear expenditure share curve. The functional decomposition of price responses and demand shifters allow empirical specifications that are scalable yet still interpretable, and the microeconomic assumptions independent of the observed data patterns prevent overfitting in sample. The linear expenditure share curve identifies demand shifters as minimum living costs instead of average, which alleviates endogeneity concerns from strategic pricing. Synthetic data analyses show that the proposed method provides with stable extrapolation of demand curves, and is robust to correlated demand and price shocks.

1 Introduction

Counterfactual policy evaluation is one of the most important goals of empirical methods in economics. For example, the market consequences of price reduction or increase can be evaluated by extrapolating a demand curve outside of the current price range. However, such counterfactual prediction is challenging, because the extrapolations by empirical models are largely dependent upon the observed data patterns and the model's predictive performance. The extant machine learning methods accurately predict market outcomes under no regime change, which are flexible and scalable with a large number of predictors [Dzyabura and Yoganarasimhan, 2019], but they do not provide with much guidance on the causal price effects $\hat{\beta}$ required for extrapolation [Kleinberg et al., 2015]. On the other hand, the theory-driven econometric methods aim to obtain the best unbiased causal estimates with the focus of counterfactual experiment [Chintagunta and Nair, 2011], yet their predictive accuracy of the market outcome $\hat{y}$ in "big data" settings with many predictors are relatively low [Athey and Imbens, 2019].

The main goal of this research is to provide a flexible and scalable demand estimation framework using deep learning combined with theoretical regularizations to offer stable yet accurate counterfactual demand predictions. To do so, we derive a linear expenditure share curve from a rational consumer's cost function under a weakly-separable, quasi-homothetic demand system [Deaton and Muellbauer, 1980].[1] The weak-separability assumes that consumers' purchase quantities in a certain product category are independent from those in another category, allowing for a microeconomic utility maximization model in a subcategory. The quasi-homotheticity assumes that higher category budget increases purchase quantities, but does not shift product choices, leading to a linear form of budget allocation to different products.

[1] The expenditure share curve is also known as an Engel curve, where it is used to analyze income elasticities of consumption categories.

ML for Economic Policy, NeurIPS 2020.

The linear expenditure share curve theoretically decomposes budget-irrelevant demand shifters for baseline predictions, from budget-relevant price responses for causal effects into separable functions. Therefore, each component can take a flexible functional form with many predictors, where deep learning methods become applicable, and the function-valued demand shifters and price responses remain interpretable regardless of empirical specifications. The *data-independent* theoretical assumptions prevent potential overfitting and stabilize extrapolations, and the *data-dependent* functions of demand shifters and price responses enhance predictive accuracy.

This paper also proposes a modification of neural networks for estimation, where three neural net components of demand shifter, price response, and category budget are merged under theoretical regularizations of the linear expenditure share curve. The baseline fluctuations of demand shifters are identified by the minimum expenditure required for living within each category/season. As the minimum cost of living is irrelevant to consumers' income or category budget by definition, it does not confound with demand responses to seasonal price fluctuations, unlike the typical mean fixed effects that may capture both pure seasonality and aggregate response to seasonal pricing. Consumers' price responses are identified by the departure from the minimum expenditure, which depend on consumers' remaining budget beyond the minimum living cost.

The synthetic data analyses show that the proposed model offers a reasonable extrapolation of demand curves, close to the price elasticity recovered from theoretical models. Deep neural networks improve the predictive performance of causal component by allowing for flexibility, relative to parametric, likelihood-based methods. The theoretical regularization prevents overfitting problems observed in the application of "off-the-shelf" neural nets to a flexible log demand model. The proposed identification of demand shifters with minimum living costs offers robust estimates of price responsiveness and demand shifter in the presence of strategic seasonal pricing, which alleviates endogeneity concerns in the absence of ideal instruments.

2 Flexible demand estimation framework with theoretical regularization

The proposed model is derived from a consumer's cost function (or expenditure function) under two common assumptions in microeconomics, weakly-separable and quasi-homothetic demand. The cost function yields the minimum outlay of consumption across products to achieve a certain level of utility given price information, where the budget allocation solved at the level of indirect utility function corresponds to Marshallian demand [Barten, 2001]. Imposing weak-separability and quasi-homotheticity on the cost function derives a linear expenditure share curve within a category, where the category-level demand can be modeled as an additively separable form of demand shifters and price responses.

2.1 Linear expenditure share curve

The first assumption, weak or functional separability, imposes that preferences from one category of goods (e.g., soft drinks) are independent from the quantities purchased in other categories (e.g., clothings). As described above, utility from demand vector $\mathbf{y}$, $u(\mathbf{y})$, can be expressed as subutilities of different categories, $v_{C_1}(\cdot), ..., v_{C_n}(\cdot)$. Then, the conditional preference ordering of goods in a category C_1 (i.e., $y_1^{C_1}, ..., y_n^{C_1}$) is not influenced by consumption of goods in another category C_n (i.e., $y_1^{C_n}, ..., y_n^{C_n}$). This assumption allows to set up a feasible subutility maximization problem in one category (e.g., soft drinks) subject to category budget allocation, which reasonably holds in many cases [Deaton and Muellbauer, 1980].

$$u(\mathbf{y}) = u\left(v_{C_1}\left(y_1^{C_1}, ..., y_n^{C_1}\right), ..., v_{C_n}\left(y_1^{C_n}, ..., y_n^{C_n}\right)\right)$$

The second one, quasi homotheticity, assumes that budget increases lead to proportional increases in expenditures beyond the fixed cost of living [Gorman, 1976]. The category expenditure $E(\cdot)$ with price vector $\mathbf{p}$ of all goods in the category and indirect utility function u is described by the minimum cost of living $a(\mathbf{p})$ (e.g., minimum amount of eggs to feed family) and the optimal allocation of remaining budget $b(\mathbf{p})\,u$ (e.g., additional eggs for baking) as a consumer's cost function

$$E(\mathbf{p}, u) = a(\mathbf{p}) + b(\mathbf{p})\,u,$$

where $a(\mathbf{p})$ and $b(\mathbf{p})$ are linearly homogeneous and concave with respect to $\mathbf{p}$. The indirect utility function with category budget m is given by $u = \frac{m-a(\mathbf{p})}{b(\mathbf{p})}$.

Taking a partial derivative of the cost function with respect to the price of i-th good in the category, p_i, and substituting the indirect utility function above, we obtain the following expenditure curve for the i-th good [Deaton and Muellbauer, 1980]:

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

where E_i is a consumer's spending on the i-th good among n goods in the category ($m = \sum_{i=0}^{n} E_i$), $a_i(\mathbf{p}) = \frac{\partial a(\mathbf{p})}{\partial p_i}$ is the minimum cost of living for the i-th good, and $b_i(\mathbf{p}) = \frac{\partial b(\mathbf{p})}{\partial p_i}$ is further budget allocation to the i-th good that varies by price responsiveness.

To obtain the i-th good's expenditure at time t, E_{it}, we normalize the composite outside good (denoted as i=0) to not require minimum quantity for living and to have a unit price across all t's (i.e., $a_{0t} \equiv 0$ and $p_{0t} \equiv 1$). We also introduce a normalized price responsiveness function by the composite outside good ($\tilde{b}_i(\mathbf{p}_{-0,t}) \equiv b_i(\mathbf{p}_{-0,t}) / b_0(\mathbf{p}_{-0,t})$ for all i's). Hence, the expenditure function for product $i \in \{0, 1, ..., n\}$ at time t is given by

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

where p_{it} is price for product i at time t, $\mathbf{p}_{-0,t}$ is price vector for all n goods except for the outside good, m_t is category budget, $\mathbf{r}_t$ is a set of environmental variables that may influence minimum expenditures (e.g., seasonality), and $\tilde{b}_0(\mathbf{p}_{-0,t})$ is normalized to be one. We estimate the function values of $a_i(\mathbf{p}_{-0,t})$, $\tilde{b}_i(\mathbf{p}_{-0,t})$, and m_t, each of which can take flexible specifications and functional forms with many predictor variables.

2.2 Identification strategy

The economic interpretation of the linear expenditure share curve allows for a theory-grounded identification strategy of price responsiveness from demand shifter. First, the minimum quantity within each season identifies time-varying demand shifters. The cost of living varies over time, as consumption environment changes depending on various conditions such as seasonality, inventory level, economic conditions, etc. In the expenditure share curve, $a_i(\mathbf{p}_{-0,t} | \mathbf{r}_t)$ captures such upward or downward shifts of demand curves caused from environmental changes by *minimum* observed costs within each season shown in variation (A) in Figure 1, as opposed to *mean* expenditures. While the mean controls in the data capture both seasonality and demand fluctuation caused by seasonal pricing, minimum expenditures for living are independent of price responses, leading $a_i(\mathbf{p}_{-0,t} | \mathbf{r}_t)$ to be a relatively robust estimate of demand shifter to unobserved confounds.

The second part $\tilde{b}_i(\mathbf{p}_{-0,t})$ identifies the allocation of remaining budget after minimum expenditures $(m_t - \sum_{i'=1}^{n} p_{i't} a_{i'}(\mathbf{p}_{-0,t} | \mathbf{r}_t))$ into n goods given prices $\mathbf{p}_{-0,t}$, where the ratio $\frac{\tilde{b}_i(\mathbf{p}_{-0,t})}{1 + \sum_{i'=1}^{n} p_{i't} \tilde{b}_{i'}(\mathbf{p}_{-0,t})}$ represents substitution and price sensitivity. This component (shape of demand curve) is captured by

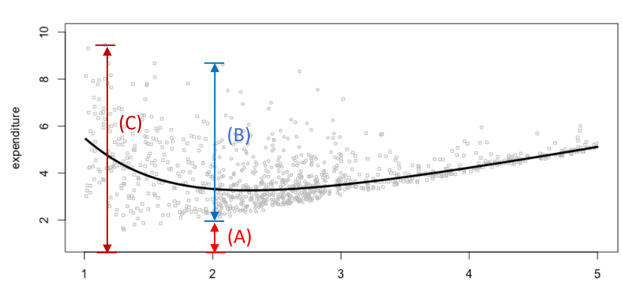


Figure 1: Identifying variation of the expenditure share curve model

observed expenditures for the excess quantity beyond cost of living shown in variation (B) in Figure 1. If $a_i(\mathbf{p}_{-0,t}|\mathbf{r}_t)$ identified by minimum costs is closer to true baseline than typical mean controls, the departure from it in response to price changes is less likely to be inflated or deflated than the departure from the mean fixed effects. Lastly, the given category budget is identified by the observed *maximum* expenditure across the entire sample periods shown in variation (C) in Figure 1.

The proposed model offers a tractable structure of three functional components, $a_i(\mathbf{p}_{-0,t}|\mathbf{r}_t)$, $\tilde{b}_i(\mathbf{p}_{-0,t})$, and m_t, with a straightforward identification strategy based on theoretical assumptions. Each of the functional components is flexible enough to take any functional form or many predictor variables to enhance predictive performance, as the structure preserves microeconomic relationships between them.

2.3 Deep neural networks

Each of $a_i(\cdot)$, $\tilde{b}_i(\cdot)$, and m_t are estimated by feed-forward neural networks (see e.g., LeCun et al. [2012, 2015], Hinton et al. [2012] for details), then the outputs are assembled into an expenditure share curve network. Figure 2 shows the network for $\tilde{b}_i(\cdot)$ with $L\ (\ni l)$ hidden layers and $J_1\ (\ni j)$ nodes in the first layer. Some input variables ($\mathbf{x}^{\tilde{b}} = \left(x_1^{\tilde{b}}, ..., x_k^{\tilde{b}}\right)$) enter to the first layer, then J_1 different outputs ($z_{lj}^{\tilde{b}}$) are produced by J_1 linear weight vectors of k input variables $\mathbf{w}_{lj}^{\tilde{b}} = \left(w_{lj1}^{\tilde{b}}, ..., w_{ljk}^{\tilde{b}}\right)$, and biases $\eta_{lj}^{\tilde{b}}$. The transformed output vector ($\tilde{\mathbf{z}}_l^{\tilde{b}}$) by an activation function of layer l ($f_l(\cdot)$), $\tilde{\mathbf{z}}_l^{\tilde{b}} = \left(f_l\left(z_{l1}^{\tilde{b}}\right), ..., f_l\left(z_{lJ_l}^{\tilde{b}}\right)\right)$, enters to the next layer as an input, and this procedure iterates until the pre-specified final layer. We use a sigmoid activation function to allow for *continuous* mapping required for our expenditure curve model. The networks for $a_i(\cdot)$ and m_t follow the same procedure with network-specific input vectors, $\mathbf{x}^a$ and $\mathbf{x}^m$.

The output layer combines transformed outputs from the last hidden layer $\tilde{\mathbf{z}}_L^{\tilde{b}}$ with final linear weight vector $\mathbf{w}_{L+1}^{\tilde{b}}$ and bias $\eta_{L+1}^{\tilde{b}}$. The final output $z^{\tilde{b}}$ is transformed by network-specific activation functions to meet theoretical requirements of the expenditure share network. We specify $f^{\tilde{b}}(\cdot) = \exp(\cdot)$ to keep non-negativity of $\tilde{b}_i(\cdot)$, $f^a(\cdot) = \text{sigmoid}(\cdot)\frac{\bar{m}_t}{\max(p_{it})}$ for cost of living to not exceed observed maximum expenditure ($\bar{m}_t$), and $f^m(\cdot) = \exp(\cdot)$ for non-negative budget. Additional empirical conditions are imposed to meet $\frac{\partial \tilde{b}_i(\mathbf{p}_{-0,t})}{\partial p_{it}} < 0$ for all i's and t's.

Evaluating the final transformed output functions at each observation, we obtain the estimated values of $\hat{\tilde{b}}_{it}$, $\hat{a}_{it}$, and $\hat{m}_t$. We also obtain $\hat{E}_{it}$ by equating them into the expenditure share network. Then we minimize the following loss function for n goods and T time periods in the observations:

$$L = \frac{\sum_{i,t}\left(\log \bar{E}_{it} - \log \hat{E}_{it}\right)^2}{nT} + \theta_a \frac{\sum_{i,t}\log p_{it}}{nT}\frac{\sum_{i,t}\left(\log \bar{a}_{it} - \log \hat{a}_{it}\right)^2}{nT} + \theta_m \frac{\sum_{i,t}\left(\log \bar{m}_t - \log \hat{m}_t\right)^2}{nT},$$

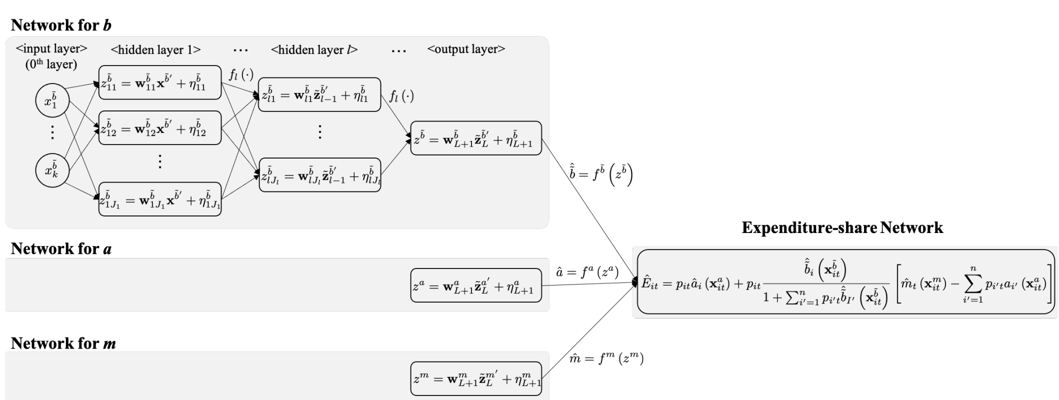


Figure 2: Overview of proposed neural networks

where $\bar{E}_{it}$ is observed expenditure from the data, $\bar{a}_{it}$ is minimum quantity of a season that t belongs to, and θ_a and θ_m are tuning parameters.

The first term in the loss function matches the estimated and observed expenditures, as typical sum of squared errors. The second term matches demand shifter estimates with observed costs of living and the third term matches category budget estimates with observed maximum expenditures, both of which are L-2 regularization terms to impose the theoretical assumptions. The average of log prices is multiplied to the demand quantities in the second term to match the scale of loss function as expenditures, not quantities. After a series of sensitivity tests, both θ_a and θ_m are set to be one. In the subsequent empirical analyses, we specify $\mathbf{x}^{\tilde{b}}$ as prices and $\mathbf{x}^a$ as seasonality factors, and do not specify $\mathbf{x}^m$, where m_t is estimated as a parameter.

3 Synthetic data analysis

The simulation study is designed to answer the following two questions: (*i*) does the proposed model predict a reasonable counterfactual demand curve outside of the observed price ranges in the presence of seasonal shifters? (*ii*) is the proposed identification strategy robust to strategic pricing unobserved to researchers in recovering demand shifters and price sensitivities in the absence of good instruments?

3.1 Demand extrapolation with hypothetical prices

The synthetic data are generated by a translated CES (constant elasticity of substitution) utility function with demand for one inside good y_t, and composite outside good q_t:[2]

$$\begin{aligned} \max \quad & u\left(y_t, q_t\right) = e^{\psi+\nu_t}\{y_t - (\gamma_0 + \gamma_1 s_t + \xi_t)\}^\rho + q_t^\rho \\ \text{s.t.} \quad & p_t \cdot y_t + q_t = m_t \end{aligned}$$

where ψ is marginal utility with stochastic component ν_t, γ_0 is the baseline demand, γ_1 is a seasonal demand shifter, s_t is an indicator of peak (=1) and off-peak (=0) seasons, ξ_t is a stochastic demand shock, and ρ is a satiation parameter. Prices are randomly generated by p_t=p_0+ω_t with baseline price of p_0 and stochastic term ω_t. We assume that researchers are aware of s_t but not ξ_t, and that ξ_t is not correlated with ω_t, i.e., no endogeneity, to focus on the first question of demand curve extrapolation as an approximation of causal price effects.

The proposed model is simplified for simulation study, where $a_i\left(\cdot\right)$ is specified as a function of s_t, $\tilde{b}_i\left(\cdot\right)$ as a function of an arbitrary price polynomial vector $\mathbf{p}_t^\lambda$, and m_t as a parameter to be estimated. A benchmark model is a simple log demand function

$$\log y_t = \phi\left(s_t, \mathbf{p}_t^\lambda\right).$$

We estimate the following six cases:[3]

- Proposed model with modified deep neural nets and $\mathbf{p}_t^\lambda = (p_t)$
- Proposed model with modified deep neural nets and $\mathbf{p}_t^\lambda = \left(p_t^{-2}, p_t^{-1}, p_t, p_t^2\right)$
- Proposed model with Bayesian estimation and $\mathbf{p}_t^\lambda = \left(p_t^{-2}, p_t^{-1}, p_t, p_t^2\right)$
- Log-linear demand with “off-the-shelf" deep neural nets and $\mathbf{p}_t^\lambda = (p_t)$
- Flexible log demand with “off-the-shelf" deep neural nets and $\mathbf{p}_t^\lambda = \left(p_t^{-2}, p_t^{-1}, p_t, p_t^2\right)$
- Flexible log demand with Bayesian estimation and $\mathbf{p}_t^\lambda = \left(p_t^{-2}, p_t^{-1}, p_t, p_t^2\right)$

The focal context of demand prediction requires an approximation of a continuous mapping from one finite space (price) to another (demand), so we specify one hidden layer of neural networks for stability [Heaton, 2008]. To allow for flexible approximation, we specify 10 nodes in the hidden layer, where 10 variants of weights on predictor variables are evaluated. The large number of nodes allow to approximate a rich set of smooth functions even under the single hidden layer [Athey and

[2]Other data generating processes (e.g., quasi-linear and Stone-Geary utility functions) are also tested, and the proposed expenditure share model recovers the true expenditure curves well in all cases.

[3]The proposed model with neural nets and price polynomials was also tested and gave nearly identical, stable results as the same model with single price, so is omitted for expositional brevity.

Imbens, 2019], yet these are still small enough to not overfit. The recommended ratio of sample size to the number of parameters is larger than five to avoid overfitting [Hagen et al., 2014], and our ratio in this simulation is 14.

Simulated prices in training samples are generated in three ranges: mid range with $p_t \in (1.5, 3.5)$, narrow range with $p_t \in (2, 3)$, and wide range with $p_t \in (1.2, 4)$. In test samples, prices are regenerated in a wider range with $p_t \in (0.5, 5)$, and expenditure curves are extrapolated using the prices outside of observed ranges in the training samples. Table 1 presents out-of-sample MSE's for the *same* price ranges in the training sample, replicating typical holdout prediction tasks. In most cases, the flexible log demand model with a single price term as predictor variable estimated by the "off-the-shelf" neural nets presents the highest predictive fits except for one case in off-peak season with wide observed price range. If a researcher selects a demand forecasting method based on the best holdout prediction without regime shifts in the data (e.g., random holdout or k-fold cross validation), the log demand model with a single price term using "off-the-shelf" neural nets can be used for future decision-making.

Table 1: Out-of-sample MSE's of expenditure curves (*within* observed price range)

		Expenditure share curve models			Flexible log demand models		
Price ranges	Seasons	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	1.101	2.162	1.887	**1.100**	1.108	1.104
	Peak	0.903	2.409	1.437	**0.897**	0.909	0.947
B. Wide	Off	**1.210**	2.948	3.225	1.216	1.216	1.268
	Peak	0.974	3.446	2.019	**0.969**	0.983	0.997
C. Narrow	Off	0.977	1.215	1.092	0.976	0.977	**0.974**
	Peak	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.

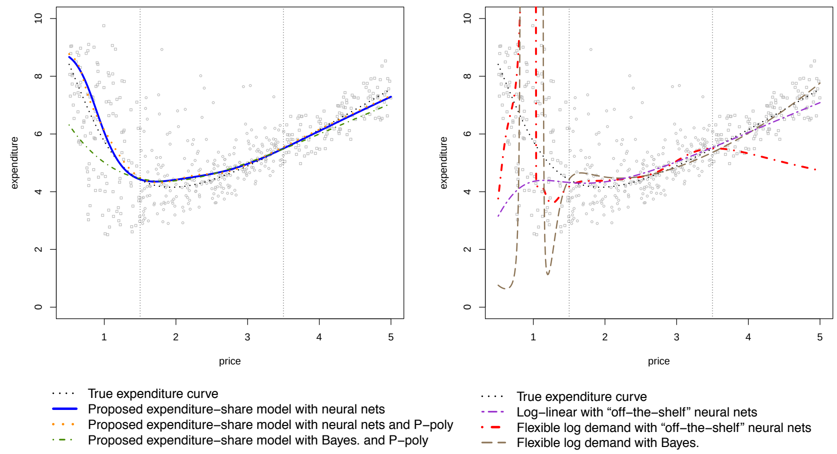
However, Table 2 shows that such model selection based on a random holdout validation may not select the best model for a researcher's policy evaluation. The out-of-sample MSE's measured *outside* of the observed price ranges are much lower for the proposed models relative to the log demand models in all cases. When extrapolating the expenditure curves with respect to prices that are not previously implemented, the proposed models' MSE's are very close to the true expenditure curve's, indicating that causal price effects are well approximated. The modified deep neural networks with expenditure share function improve the counterfactual prediction relative to Bayesian estimation with price polynomials in most cases. Although the predictive models (i.e., flexible log-demand models with "off-the-shelf" neural nets) present higher predictive fits within the observed price range, the proposed model significantly outperforms in extrapolations.

Table 2: Out-of-sample MSE's of expenditure curves (*outside* of observed price range)

		Expenditure share curve models			Flexible log demand models		
Price ranges	Seasons	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	4.133	**3.964**	5.204	10.906	93.326	*Inf.*
	Peak	3.641	**3.497**	4.183	8.534	91.067	*Inf.*
B. Wide	Off	**4.015**	4.403	5.459	6.472	19.112	*Inf.*
	Peak	**3.565**	3.917	4.420	4.236	17.515	*Inf.*
C. Narrow	Off	**3.423**	4.653	3.462	10.755	475.564	*Inf.*
	Peak	2.869	4.304	**2.818**	10.493	523.131	*Inf.*

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

These findings are confirmed by the visualization of extrapolated expenditure curves. Figure 3 illustrates the expenditure curves predicted by the proposed models for peak seasons with mid-range of observed prices. Two vertical dotted lines indicate lower and upper bounds of observed price ranges, so the tails outside of those lines are counterfactual predictions. The proposed model with modified deep neural nets well approximates counterfactual price effects on both tails of extrapolations in the left panel. However, the right panel shows that flexible log demand models with both "off-the-shelf"



Proposed Models   Benchmark Models

Figure 3: Extrapolated expenditure curves

deep neural nets and Bayesian estimation with price polynomials tend to overfit in sample, presenting poor predictive expenditures in the extrapolated regions. The simulation results confirm that the proposed models outperform flexible predictive methods in approximating causal effects. They also indicate that the proposed method in analyzing observational data can offer a stable counterfactual prediction in response to firms' price changes without experimental data.

3.2 Demand prediction in the presence of price endogeneity

The proposed and benchmark models are estimated under two extreme cases of endogenous pricing – perfect seasonal pricing and perfect counter-seasonal pricing by the firm. The data are generated by the same translated CES function in the previous subsection with the same seasonal and random demand shifters, where γ_t=γ_0+$\gamma_1 s_t$+ξ_t. The endogenous prices are generated by

$$p_t = p_0 + p_1 s_t + \omega_t,$$

where p_1 represents an endogenous price shock. For perfect seasonal pricing, the firm raises their price upon peak season (p_1>0 for s_t=1) and positive stochastic demand shocks ($cor\,(\xi_t, \omega_t)$ =1). For perfect counter-seasonal pricing, the firm reduces their price upon peak season (p_1<0 for s_t=1) and positive stochastic demand shocks ($cor\,(\xi_t, \omega_t)$ =−1).

In both cases, we test two subcases of γ_1=0, where the firm practices seasonal or counter-seasonal pricing without knowing that there is no actual seasonal demand shock, and γ_1>0, where there exists actual seasonal demand shock. The firm is assumed to know both s_t and ξ_t, but the researcher only knows s_t. The seasonal shock γ_1 is to be estimated, and there is no time-series pattern of ξ_t. We present in-sample prediction of demand curves to show how each method recovers price elasticity and seasonal demand shock in the presence of seasonal price endogeneity. For brevity, we report the results from proposed model with modified deep neural nets and log demand model with "off-the-shelf" deep neural nets using a single price term.

Figure 4 presents the predictive demand curves. The lefthand side of each panel indicates low price season (off-peak for seasonal pricing/peak for counter-seasonal pricing), and the righthand side indicates high price season (peak for seasonal pricing/off-peak for counter-seasonal pricing), where demand shocks may occur on the vertical dotted line in the center. Panels A and B without true demand shocks present that the proposed model (solid line) well approximates near-zero demand shocks and price elasticities in the true curves (dotted line). However, the log-linear demand model (dashed line) predicts large, false demand shocks, as the seasonal dummies in the model capture demand responses to firm's seasonal pricing as demand shocks. Panels C and D in Figure 4 with true demand shocks show that the proposed model accurately predicts non-zero demand shocks as well for both seasonal and counter-seasonal pricing. The shapes of demand curves by the proposed

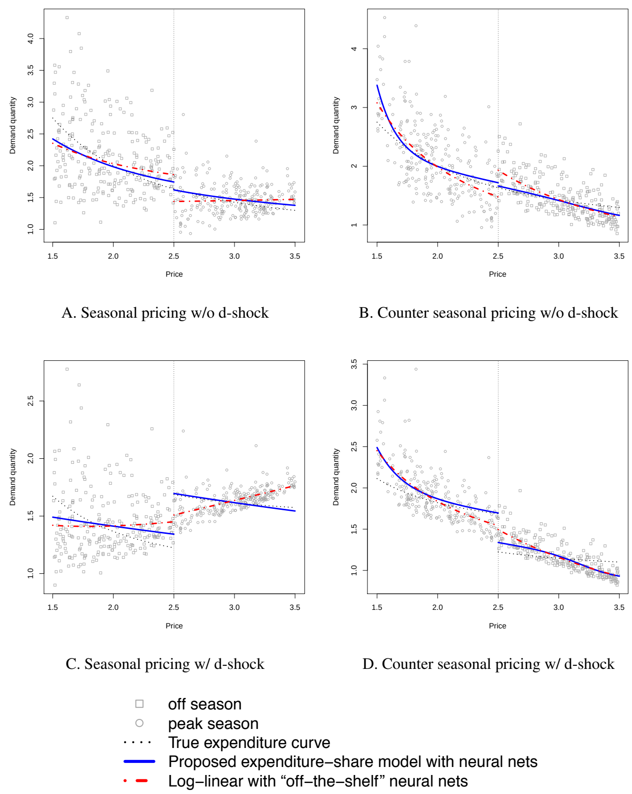


Figure 4: Predictive demand curves with price endogeneity

model are also much closer to true demand curves. The log demand model estimates price elasticities with typical downward biases in seasonal pricing and upward biases in counter-seasonal pricing. The results show that the proposed model may alleviate potential endogeneity biases.

4 Discussion and conclusion

The proposed theoretical structure offers a reasonable regularization between price responsiveness and demand shifter. It allows to take flexible forms and potentially many predictors within each functional component. Despite its flexibility, the structure still preserves microeconomic relationships between functional components, so that causal effects and counterfactual outcomes can be stably approximated. The identification strategy based on behavioral assumptions potentially alleviates endogeneity concerns mainly from seasonal pricing. The proposed framework can be easily applied to the market demand analysis using observational data for policy evaluation.

The simulation study shows how microeconomic theory can improve machine learning methods to predict counterfactual outcomes upon *drastic* policy regime changes in the absence of good instruments. The literature has focused on how policy evaluation can be enhanced under binary treatment assignment or local policy changes. Our results show that, a method that produces good out-of-sample predictions within the same policy regime may produce poor predictions upon policy regime changes. It is also shown that theoretical knowledge can help approximate causal effects, especially when the knowledge from observed data cannot be applied.

In sum, we believe that the proposed framework has a potential to overcome major challenges of the existing theoretical demand models (e.g., lack of flexibility) and weaknesses of machine learning methods (e.g., lack of tractability and theoretical regularization) by imposing a microeconomic structure. We hope that the proposed framework is extended and applied to estimate and predict the market demand to enhance the economic policy-maker’s decision.

References

Susan Athey and Guido W. Imbens. Machine learning methods economists should know about. *Annual Reviews*, page Forthcoming, 2019.

Anton P. Barten. Consumer economics. In Neil J. Smelser and Paul B. Baltes, editors, *International Encyclopedia of the Social & Behavioral Sciences*, pages 2669–2674. Elsevier Ltd., 2001.

Pradeep K. Chintagunta and Harikesh S. Nair. Discrete-choice models of consumer demand in marketing. *Marketing Science*, 30(6):977–996, 2011.

Angus Deaton and John Muellbauer. *Economics and consumer behavior*. Cambridge university press, 1980.

Daria Dzyabura and Hema Yoganarasimhan. Machine learning and marketing. In Natalie Mizik and Dominique M. Hanssens, editors, *Handbook of Marketing Analytics: Methods and Applications in Marketing, Public Policy, and Litigation Support*, pages 255–279. Edward Elgar Publishing, Inc., 2019.

William M. Gorman. Tricks with utility functions. In M. Artis and A.R. Nobay, editors, *Essays in Economic Analysis*, pages 211–243. Cambridge: Cambridge University Press, 1976.

Martin Hagen, Howard B. Demuth, Mark H. Beale, and Orlando De Jesús. *Neural network design*. Martin Hagen, 2nd edition, 2014.

Jeff Heaton. *Introduction to Neural Networks with Java*. Heaton Research, 2nd edition, 2008.

Geoffrey E. Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan R. Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. http://arxiv.org/abs/1207.0580, 2012.

Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan, and Ziad Obermeyer. Prediction policy problems. *American Economic Review: Papers & Proceedings*, 105(5):491–495, 2015.

Yann A. LeCun, Léon Bottou, Genevieve B. Orr, and Klaus-Robert Müller. Efficient backprop. In G. Montavon and G.B. Orr, editors, *Neural Networks: Tricks of the Trade. Lecture Notes in Computure Science*, volume 7700. Springer, Berlin, Heidelberg, 2012.

Yann A. LeCun, Yoshua Bengio, and Geoffrey E. Hinton. Deep learning. *Nature*, 521:436–444, 2015.

Broader impact

The proposed method mainly improves the prediction accuracy of consumers' demand system combining deep learning with microeconomic theory. There has been a long tradition of demand estimation in econometrics (e.g., BLP 1995), because it is important and required in evaluating the consequences of many economic policy decisions. Therefore, we believe that the proposed model can enhance the prediction accuracy under the big data environment by using deep learning, and also enhance the generalizability of empirical findings by regularizing the model with microeconomic theory. It also contributes to an economic agent's decision-making and corresponding welfare, such as suppliers' optimal pricing decisions. Consumers' welfare may decrease, if firms optimize their prices with better demand prediction. However, a more informed economic policy can be made as well, if the market demand is more accurately predicted with policy variation.

Relevance to the workshop

This paper methodologically contributes to the ML for Economic Policy Workshop by proposing a theoretical framework to apply machine learning methods to study consumer demand system for economic decision-making. Machine learning methods in economic policy evaluation have been applied to mainly resolve *prediction* policy problems [Kleinberg et al., 2015], due to their limited guidance on the causal effects of focal policy variables. The recent research in machine learning methods has significantly improved their ability to make causal inferences by statistically debiasing focal parameters, but such methods still rely on a strong assumption of generalizability of local patterns to the global policy regions, which limit the applicability of machine learning to extrapolation. The proposed method presents a theoretical framework that a researcher can use machine learning methods for demand studies. Unlike the existing methods that flexibly capture the entire utility function in demand models, the proposed model offers a theoretical decomposition of critical components in demand studies, namely, demand shifters and price responses as separate functions. We believe that the proposed model shows how the data-independent, theoretical knowledge help generalize the local data patterns, instead of assuming a constant pattern for extrapolation. In addition, we hope to bring attention to the importance of theoretical knowledge in machine learning applications to economic decision-making.

Declaration of originality

The author(s) declare that the submitted work has not been published or first made available before January 1, 2017, and is original work.