Adventures in Demand Analysis Using AI

Martin Spindler

Presentation at the 13th ECB Conference on Forecasting Techniques, Frankfurt am Main, Germany

based on joint work with
Philipp Bach (FU Berlin), Victor Chernozhukov (MIT), Sven Klaassen (UHH), Jan Teichert-Kluge (UHH), Suhas Vijaykumar (UCSD)

Motivation

- **Early JASA tradition:** Almost a century ago, *The Journal of the American Statistical Association* published foundational empirical studies of demand (Wright, Working, Schultz, Mills, Stigler), moving economics toward quantitative measurement and establishing econometrics.
- **Hedonic modeling roots:** Classical work on hedonics (Lancaster, 50s; Griliches, 1971) showed how product attributes drive prices and demand. Pat Bajari, Pakes, and others set the stage for modern empirical work in early 2000s.
- **Today's AI/ML opportunities:** We'll use AI-generated product representations from text and images to better represent the products than the standard tabular covariates.

- **Amazon application:** Using toy car sales ranks and prices on *Amazon.com*, the paper demonstrates how multimodal transformer-based models (text, images, tabular data) generate demand-relevant and price-relevant embeddings.
- **Findings & implications:** These embeddings dramatically improve predictions of price and quantity, capturing subtle features like branding, quality, and visual appeal, thereby strengthening the hedonic approach to demand.
- **Finding for Price Elasticity**. The AI embeddings are key determinants of price sensitivity/elasticity of demand. While the embeddings are strong effect modifiers, they are very weak confounders – once the past visibility signals and prices are accounted for.

Background Papers and Collaborators

- P. Bajari, Z. Cen, V. Chernozhukov, M. Manukonda, J. Wang, R. Huerta, J. Li, L. Leng, G. Monokroussos, S. Vijaykumar, S. Wan (2021): *Hedonic Prices and Quality Adjusted Price Indices Powered by AI*. CEMMAP Working Paper CWP04/21.
- P. Bach, V. Chernozhukov, S. Klaassen, J. Teichert-Kluge, M. Spindler, S. Vijaykumar (2024): *Adventures in Demand Analysis Using AI*. arXiv:2501.00382v1.
- P. Bach, V. Chernozhukov, S. Klaassen, J. Teichert-Kluge, M. Spindler, S. Vijaykumar (2024): *Estimation of Causal E ects with Multimodal Data*. arXiv:2402.01785.
- V. Chernozhukov, C. Hansen, N. Kallus, M. Spindler, V. Syrgkanis (2024): *Applied Causal Inference Powered by ML and AI*. CausalML-Book.org.

ELSEVIER

Journal of Econometrics

Volume 251, September 2025, 106052

Hedonic prices and quality adjusted price indices powered by AI ☆

P. Bajari [a 1], Z. Cen [b 1], V. Chernozhukov [c 1], M. Manukonda [d 1], S. Vijaykumar [e 1], J. Wang [f 1], R. Huerta [d 2], J. Li [d], L. Leng [g], G. Monokroussos [h 2], S. Wang

Show more

Add to Mendeley Share Cite

https://doi.org/10.1016/j.jeconom.2025.106052

Get rights and content

arXiv > econ > arXiv:2501.00382

Economics > General Economics

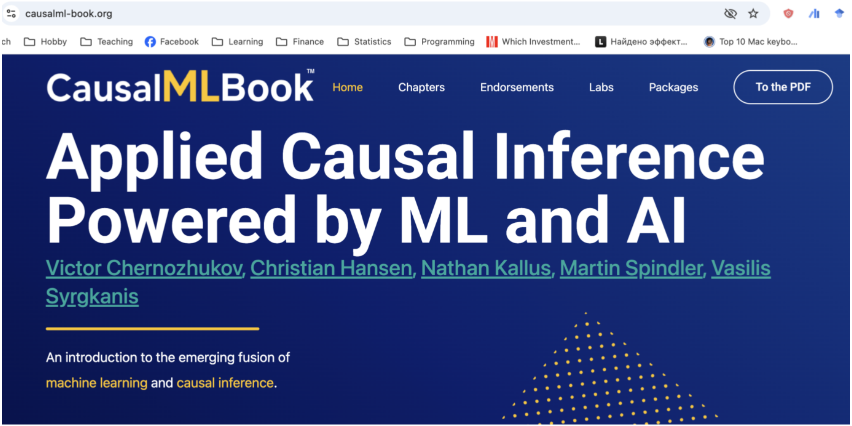
[Submitted on 31 Dec 2024]

Adventures in Demand Analysis Using AI

Philipp Bach, Victor Chernozhukov, Sven Klaassen, Martin Spindler, Jan Teichert-Kluge, Suhas Vijaykumar

This paper advances empirical demand analysis by integrating multimodal product representations derived from artificial intelligence (AI). Us
\textit{this http URL}, we combine text descriptions, images, and tabular covariates to represent each product using transformer-based emb
capture nuanced attributes, such as quality, branding, and visual characteristics, that traditional methods often struggle to summarize. More
embeddings for causal inference tasks. We show that the resulting embeddings substantially improve the predictive accuracy of sales ranks a
more credible causal estimates of price elasticity. Notably, we uncover strong heterogeneity in price elasticity driven by these product-specif
that AI-driven representations can enrich and modernize empirical demand analysis. The insights generated may also prove valuable for app

Comments: 42 pages, 9 figures
Subjects: **General Economics (econ.GN)**; Artificial Intelligence (cs.AI); Applications (stat.AP); Machine Learning (stat.ML)
Cite as: arXiv:2501.00382 **[econ.GN]**
(or arXiv:2501.00382v1 **[econ.GN]** for this version)
https://doi.org/10.48550/arXiv.2501.00382


causalml-book.org
ch
Hobby
Teaching
Facebook
Learning
Finance
Statistics
Programming
Which Investment...
Найдено эффект...
Top 10 Mac keybo...
CausalMLBook™
Home
Chapters
Endorsements
Labs
Packages
To the PDF
Applied Causal Inference Powered by ML and AI
Victor Chernozhukov, Christian Hansen, Nathan Kallus, Martin Spindler, Vasilis Syrgkanis
An introduction to the emerging fusion of machine learning and causal inference.

A Toy Example

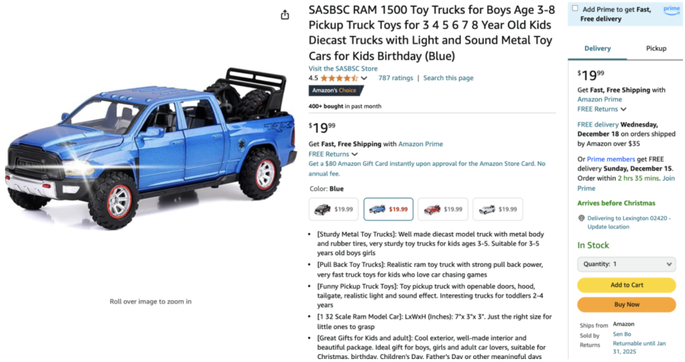


Figure: A toy example with image and text description

Toy Data for the Toy Example

- Scraped toy truck data from Keepa.com
- Sales rank of each item i at time point t
- List price of each item at each time point
- Text description of the product
- Image of the product
- Other tabular features (e.g., ratings, browse node data)

Quantity and Price Signals

- **Quantity signal:**

$$Q_{it} = \log(1/\text{Time-Averaged Sales Rank of } i \text{ in period } t)\,.$$

- **Price signal:**

$$P_{it} = \log(\text{Time-Averaged Price of } i \text{ in period } t)\,.$$

- **Data structure:** Each period $t = 1, \ldots, T$ spans 4 weeks. We set $T = 12$ periods in total, directly adjacent to each other.
- **Temporal changes:** We also examine changes in signals across periods:

$$Q_{it} := Q_{it} - Q_{i(t-1)}, \qquad P_{it} := P_{it} - P_{i(t-1)}.$$

We call the log inverse rank the "quantity signal" and log price the "price signal". Under the power law, log actual sales is **proportional** to log inverse rank.

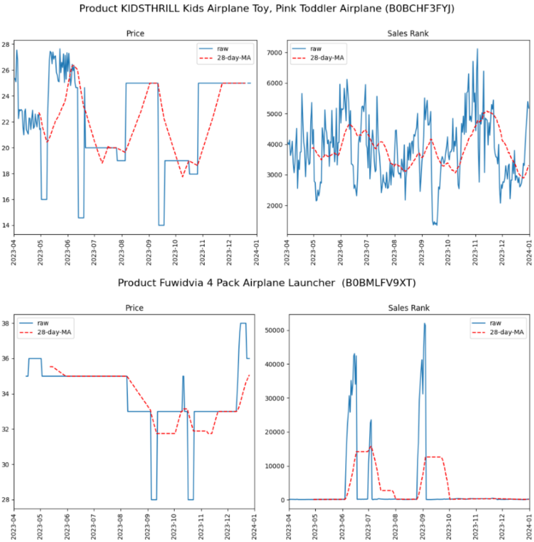


Figure 2: Price and sales rank series for two example products.

AI-Based Representation of Products

We will use various encoder models based on Transformers to convert product data into useful numerical features:

- Convert text description to dense embeddings D_{it} using Transformers (RoBERTa model; also LLama 3)
- Convert image to dense embeddings I_{it} using Transformers (BEiT model)
- Convert tabular data to embeddings T_{it} using Transformers (SAINT model)

Key Ideas: Self-Supervision, Attention, Fine-Tuning

- **Self-supervision.** Turn "unlabeled" data into massive amounts of "labeled" data by masking. For example,

$$S = (\text{Well made diecast model truck with metal body.}) =)$$

$$W = (\text{“Well made [m] model truck with [m] body”}; \quad Y = S$$

  Predict Y using W via neural networks. Use hidden layers of neural networks as representations E —"embeddings"—of S.
- **Attention Mechanisms**: Capture important interactions or co-occurrences of words and pieces – “Attention Is All You Need” by Vaswani, et al.
- **(Causal) Fine-Tuning:** Update embeddings to perform well on downstream tasks—for example, predicting quantity and price signals. Estimation of causal effects relies on residualization of price and quantity, and we fine-tune on the residualization tasks.

Diagram

$$Z_i = \begin{array}{l}\text{Text}_i\\ \text{Image}_i\end{array} \overset{e}{\longmapsto} E_i := \begin{array}{c} A_{T_i}\\ \Big\uparrow \\ W_i \\ I_i \\ \Big\downarrow \\ A_{I_i}\end{array} \overset{m}{\longmapsto} \{\hat{Q}_{it}, \hat{P}_{it}\}_{t=1}^{T}.$$

- The neural embedding map e maps text and images Z into numerical vectors E that are used to predict "auxiliary" targets—masked chunks of text and images.
- The neural map m takes the embeddings and predicts labels in downstream tasks related to causal analysis. This part can be used for fine-tuning e – updating e to perform better at the downstream task.

Embeddings: AI Representation of Products

We concatenate the embeddings: $E_{it} = (D_{it}; I_{it}; T_{it}).$ We will use centered and normalized embeddings:

$$X_{it}^{e} = \frac{E_{it} - \frac{1}{n}\sum_i E_{it}}{\|E_{it} - \frac{1}{n}\sum_i E_{it}\|}.$$

Each product is then a point on the sphere.

Does AI Understand Products?

We can evaluate this in different ways:

1. **Qualitative.** Examine clusters of products with similar embeddings, and check if AI-similarity agrees with HI-similarity.
2. **Quantitative.** Check whether AI-embeddings improve predictions of price and quantity signals.

Both tasks are very important for demand analysis, including:

- Computation of hedonic prices (prices as functions of characteristics)
- Hedonic inflation price indices
- Demand and price prediction for new products
- Predicting changes in demand due to changes in price

PCAs and Cluster Visualizations

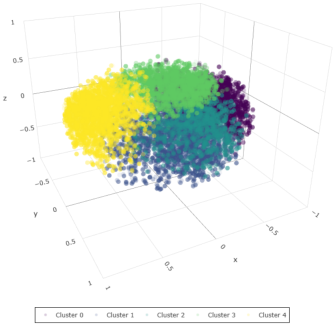


Figure 3: 3D-representation of product embeddings (with image) and five clusters

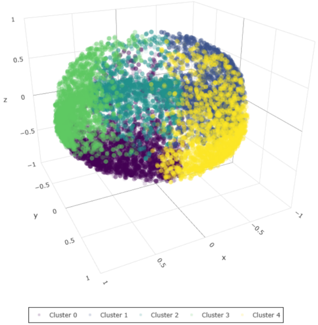


Figure 4: 3D-representation of product embeddings (no image) and five clusters