Hedonic prices and quality adjusted price indices powered by AI

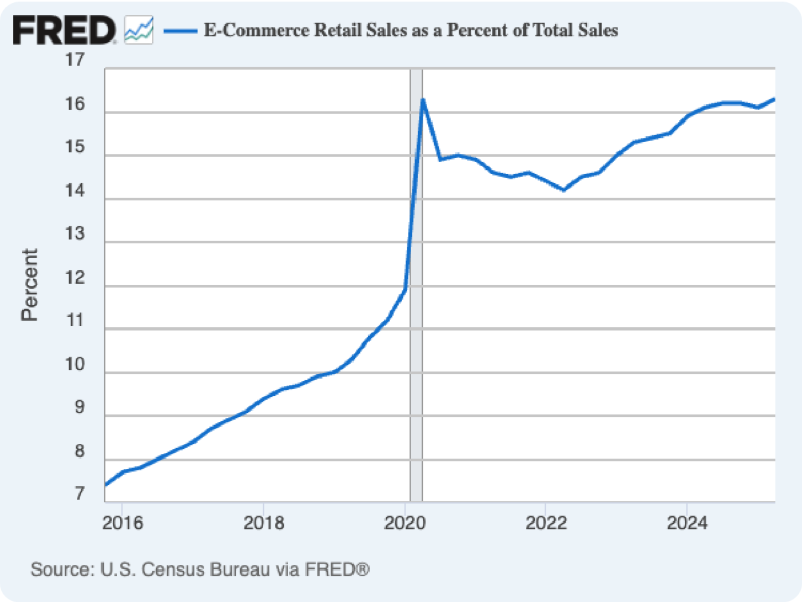
P. Bajari[1], Z. Cen[2], V. Chernozhukov[3], M. Manukonda[4], **S. V.**[5], J. Wang[6]

[1]Harvard, [2]Meta, [3]MIT, [4]Amazon, [5]UCSD, [6]Netflix

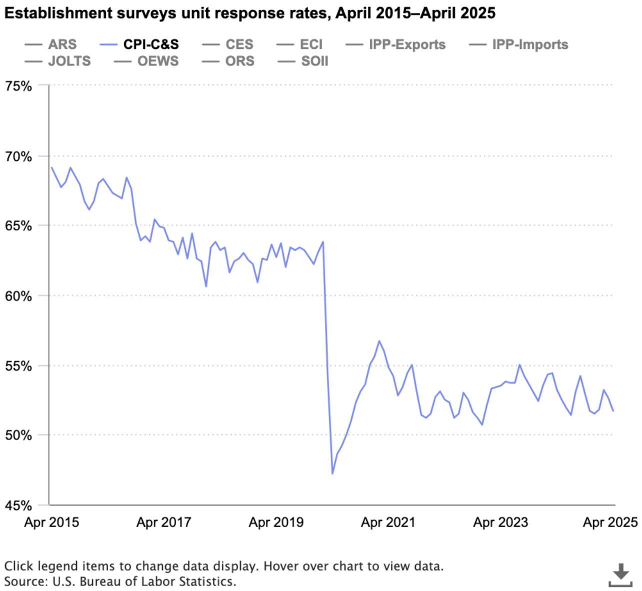
Outline

- Motivation: modernizing economic statistics and challenges in price measurement
- Methodology: AI-based hedonic models
- Application: hedonic indices for Amazon apparel
- Why this works: engineering insights
- Conclusion: implications and future work
- Appendix

Setting the stage



E-commerce share of total retail spending



CPI establishment-level response rates

- Amazon's share of U.S. e-commerce retail in 2025: **40%**
- E-commerce share of total U.S. retail in 2025: **16%** (up from 3% in 2005; 7% in 2015, *BLS 2025*)
- Broad decline in survey responses rates *(Jarmin, JEP 2019)*

What are experts saying?

> A combination of **new data sources, improved data processing and analytics,** and increased interest in economic measurement present statistical agencies with a number of opportunities to innovate and fundamentally improve and transform economic measurement.

Jarmin, Bostic and Moyer (AER, 2016)

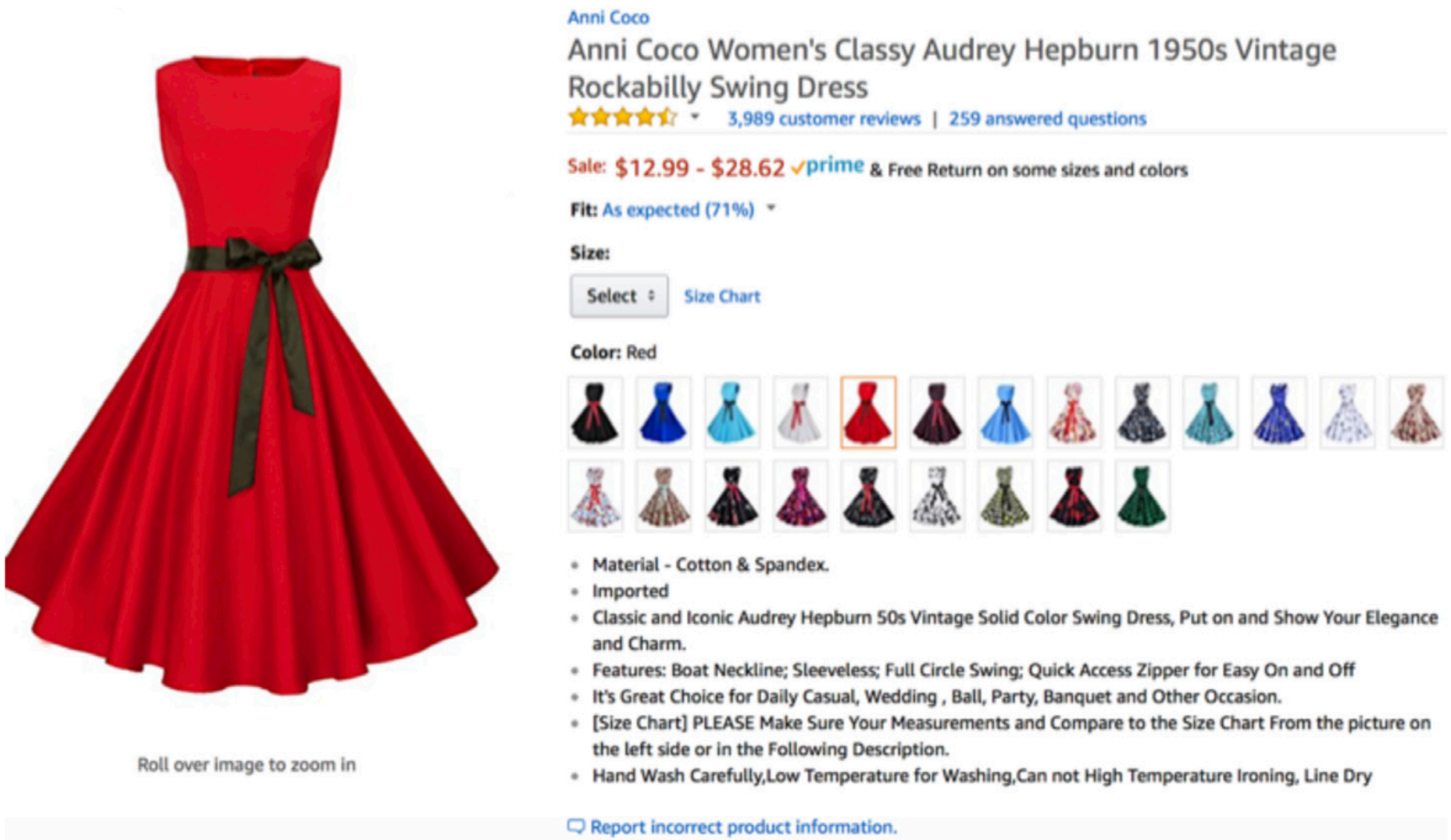
> We are at a point where measurement of key economic concepts like GDP will begin to change in critical ways … **digitized data** offer the opportunity to improve the measurement of key national economic indicators while also drastically reducing the respondent burden on households and businesses.

Ehrlich, Haltiwanger et al. (AER P&P, 2019)

> The United States needs a 21st-century national data infrastructure that blends data from multiple sources to improve the quality, timeliness, granularity, and usefulness of national statistics … The **private sector plays a critical role,** both as an important data user and as a holder of key data assets.

Groves et. al. (National Academy of Sciences, 2023)

Private sector, digitized data



A detail page describing product characteristics, from Amazon.com

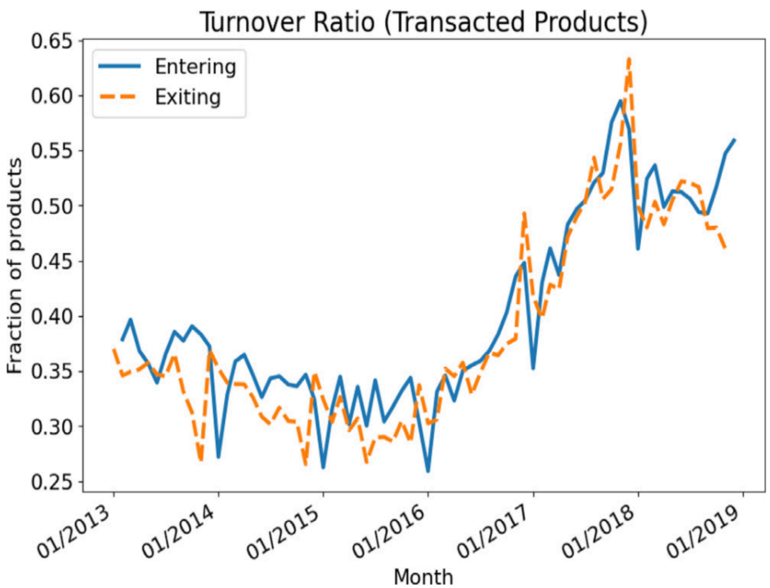
This paper: price indices

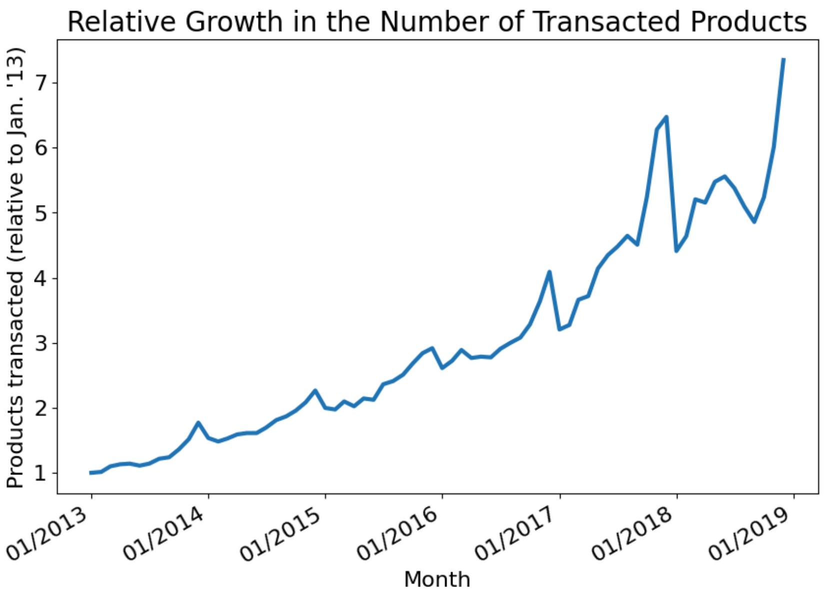
- Price indices like the CPI are essential for tracking inflation, consumer welfare and the cost of living.
- However, traditional indices face significant challenges:
  - rapid product turnover;
  - increasing variety of products;
  - extremely low market shares.
- Quality adjustment can help, but requires:
  - manual tabulation of product characteristics;
  - extensive surveys and interviews;
  - human expertise.
- **This paper:** use AI to perform quality adjustment using electronic images and descriptions.

The paper in a nutshell

- Traditionally, quality adjustment requires:
  - extensive human expertise,
  - field surveys,
  - manual construction of product characteristics.
- This is a major challenge for agencies like the Bureau of Labor Statistics
  - increasing product variety, declining response rates
- Our approach: use unstructured data and AI to learn and generate characteristics X_j.
- We process large-scale unstructured and tabular data:
  - text (using BERT)
  - images (using ResNET)
  - prices, quantities (from Amazon first-party apparel sales)
- Resulting models:
  - are highly scalable and cost-effective,
  - achieve very high out-of-sample predictive accuracy for prices ($R^2 \approx 80\%$–90%).

Preview: Amazon apparel data (2013-2018)





- First-party sales: Amazon buys from seller, sells to consumer
- Tens of millions of products (20 terabytes)
- Key features:
  - **Entry and exit**-very high product turnover (40-60% monthly)
  - Rapid growth in product variety
  - Many active products with no sales in a given month

Constructing price indices

- **Basic problem:** track prices for a representative basket of products over time.
  - How to handle product entry and exit?
- Standard (matched) indices restrict to products sold in both time periods s and t, e.g.,

$$R_{s,t}^{\text{Laspeyres}} = \frac{\sum_{i \in \mathcal{C}_t \cap \mathcal{C}_s} P_{it} Q_{is} / \sum_{i \in \mathcal{C}_t \cap \mathcal{C}_s} Q_{is}}{\sum_{i \in \mathcal{C}_s} P_{is} Q_{is} / \sum_{i \in \mathcal{C}_s} Q_{is}}, \qquad R_{s,t}^{\text{Paasche}} = \frac{\sum_{i \in \mathcal{C}_t} P_{it} Q_{it} / \sum_{i \in \mathcal{C}_t} Q_{it}}{\sum_{i \in \mathcal{C}_t \cap \mathcal{C}_s} P_{is} Q_{it} / \sum_{i \in \mathcal{C}_t \cap \mathcal{C}_s} Q_{it}},$$

$$R_{s,t}^{\text{Fisher}} = \sqrt{R_{s,t}^{\text{Laspeyres}} R_{s,t}^{\text{Paasche}}}.$$

- Here P_{it} and Q_{it} are price and aggregate demand for product i in period t; $\mathcal{C}_t$ contains all products transacted in period t.
- Selection bias: entering and exiting products likely do not resemble stable products, so blue terms give a skewed picture of aggregate prices
- Chain drift bias: using very short periods t can improve selection bias, but introduces a new source of bias from rapidly compounding errors

Hednoic models and quality adjustment

- **Hedonic models** represent each product i as a bundle of valuable characteristics, X_i *(Court '39, Lancaster '66, Griliches '61, Rosen '74)*
  - key assumption: supply, demand, and ultimately prices are determined by X_i
- Many important uses in economics, including (but not limited to) quality adjustment *(Griliches '61, Rosen '74, Pakes '02, Bajari and Benkard '05)*
  - learn mapping from characteristics X_i to prices P_{it} in period t (hedonic regression):

$$P_{it} \approx \hat{h}_t(X_i)$$

  - use $\hat{h}_t$ to impute prices of missing products at time t;
  - measure inflation by tracking changes in prices, holding characteristics fixed
- Key advantages:
  - hedonic prices can be computed for a fixed bundle of characteristics, X_i,
  - even as individual products i enter and exit the market;
  - allows long-run comparisons (e.g. year-over-year); limits chain drift.

... "powered by artificial intelligence"

Two observations

1. quality adjustment works by accurately imputing prices for missing products
    - out-of-sample prediction with known reference distribution
    - great application for AI *(Mullainathan and Spiess, JPE '17)*
2. AI works by exploiting latent, low-dimensional structure (or structured sparsity) behind images and text

- precisely what is assumed by the hedonic model
- latent representation in characteristic space, X_i, explains prices *(Rosen '74)*

AI pipeline overview

1) Feature extraction using foundation models

- Models pre-trained on vast amounts of unlabeled data
  - Text embeddings (BERT): W_i (768 dimensions)
  - Image embeddings (ResNet): I_i (2048 dimensions)
  - Combined representation: $X_i = (W_i, I_i)$.

2) Price prediction using learned representation

- Multi-task neural network trained on prices and quantities
  - single network predicts full time-series of prices for each product
- Learns dense "value embedding" V_i (256 dimensions)
- Estimates hedonic prices as $\hat{H}_{it} = \hat{\theta}_t' V_i$.

3) Index construction

- Fisher Hedonic Price Index using predicted prices $\hat{H}_{it}$
- Yearly chaining to minimize chain drift

Many design choices were made for practical reasons and after significant R&D. We will revisit these after looking at the empirical results

- Key takeaways: transfer learning, multi-task network, long chaining.

Hedonic price model

Empirical hedonic model:

$$P_{it} = H_{it} + \epsilon_{it} = h_t(X_i) + \epsilon_{it}, \quad \mathbb{E}[\epsilon_{it}|X_i] = 0,$$

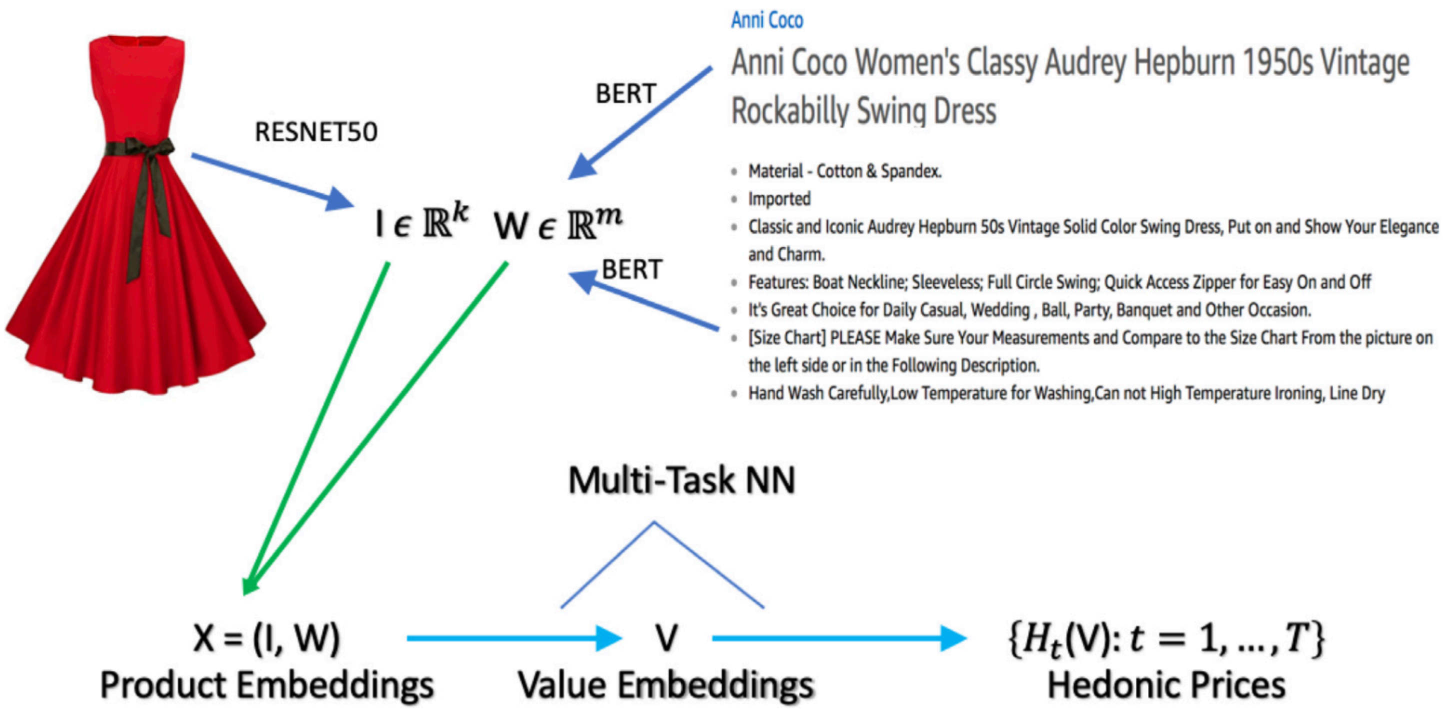
where:

- P_{it}: observed price of product i at time t
- X_i: AI-generated features (embeddings)
- $h_t(\cdot)$: hedonic price function (estimated via neural network)

Multi-task architecture:

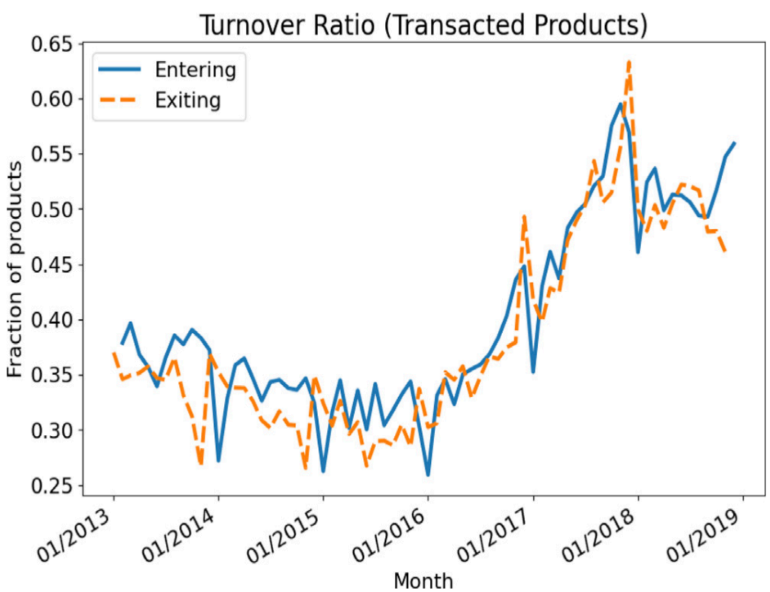
$$Z_i = \begin{bmatrix} \text{Text}_i \\ \text{Image}_i \end{bmatrix} \xrightarrow{e} X_i \xrightarrow{g_1} E_i^{(1)} \cdots \xrightarrow{g_m} V_i \xrightarrow{\theta'} \{H_{it}\}_{t=1}^{T},$$

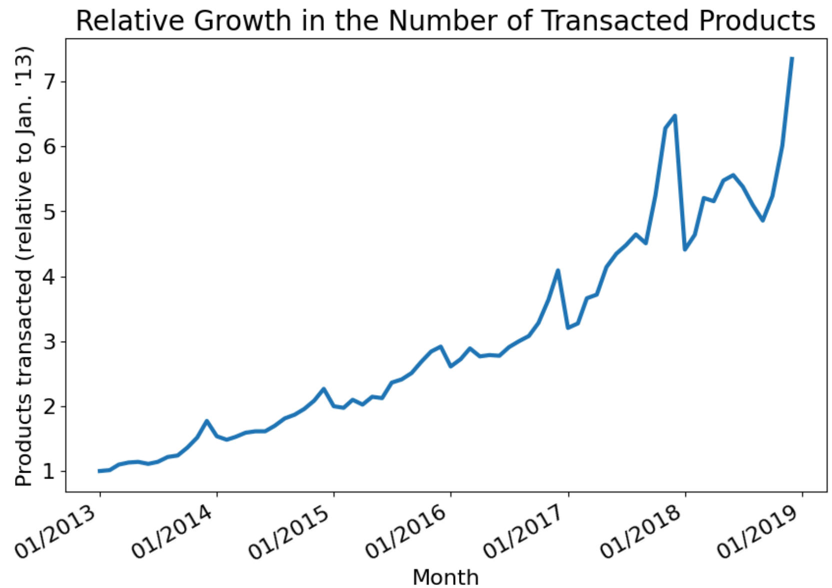
where V_i is the value embedding (learned representation). Predicts time series of prices.



Conceptual overview of AI-based hedonic price construction.

Application: hedonic indices for Amazon apparel



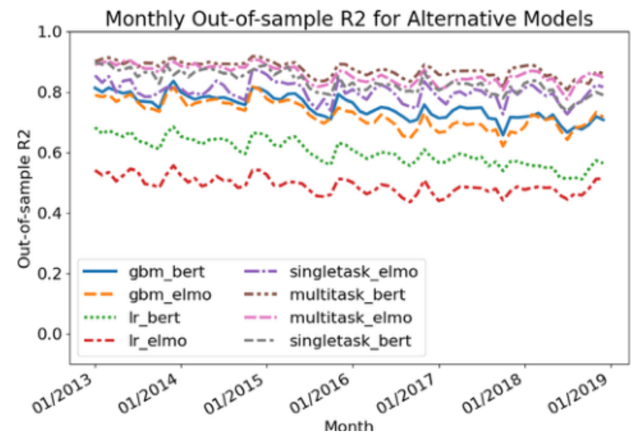


- First-party sales: Amazon buys from seller, sells to consumer
- Tens of millions of products (20 terabytes)
- Key features:
  - **Entry and exit**-very high product turnover (40-60% monthly)
  - Rapid growth in product variety
  - Many active products with no sales in a given month

Predictive performance (out-of-sample)

Method	Out-of-sample R^2
Linear (catalog features)	30–45%
Linear (AI embeddings)	55–65%
Random forest (AI embeddings)	70–80%
Single-task NN (AI embeddings)	75–85%
Multi-task NN (AI embeddings)	**80–90%**

Out-of-sample R^2 in holdout dataset.



Predictive performance by month.

Takeaways:

- Great news for hedonic models: a low dimensional representation of product image and description, V_i, explains 80-90% of price variation using linear coefficients!
- AI-based embeddings (transfer learning) dramatically improve accuracy across specifications.
- Multi-task model achieves best performance by sharing information across time periods
  - single task model: separate network for each time period.
- Comparisons: *Pakes ('03)* obtains 30-50% adjusted R^2 for personal computers; *Greenlees & McClelland ('10)* obtain R^2 = 29% using scanner data for apparel sales.

Fisher Hedonic Price Index

Laspeyres and Paasche indices:

$$R_{t,\ell}^{L,H} = \frac{\sum_{i \in \mathcal{C}_{t-\ell}} H_{it} Q_{i(t-\ell)}}{\sum_{i \in \mathcal{C}_{t-\ell}} H_{i(t-\ell)} Q_{i(t-\ell)}}, \quad R_{t,\ell}^{P,H} = \frac{\sum_{i \in \mathcal{C}_t} H_{it} Q_{it}}{\sum_{i \in \mathcal{C}_t} H_{i(t-\ell)} Q_{it}}$$

Fisher index (geometric mean):

$$R_{t,\ell}^{F,H} = \sqrt{R_{t,\ell}^{L,H} \cdot R_{t,\ell}^{P,H}}$$

Full imputation:

- Use predicted prices $H_{it} = \hat{h}_t(X_i)$ for all products, even when observed
- Hybrid (using a mix of real and predicted prices) are numerically similar

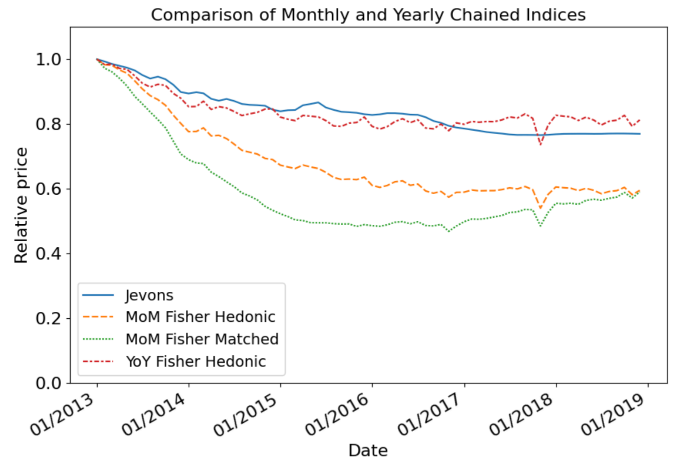
Chaining:

- Monthly chaining ($\ell = 1$): captures short-term changes
- Yearly chaining ($\ell = 12$): reduces chain drift (our preferred index)

Inflation estimates for apparel (main results)

Index, chaining frequency	Avg. annual inflation
Fisher Hedonic, yearly (FHPI)	**–0.98%**
Fisher Hedonic, monthly	–5.27%
Fisher Matched, monthly	–3.12%
Jevons posted price, daily	–3.01%
Adobe DPI, monthly	–2.02%
BLS Urban CPI (apparel)	–0.31%

Inflation estimates for FHPI and comparison indices



Yearly FHPI (preferred) vs. alternative indices

Interpretation:

- Yearly-chained FHPI shows moderate deflation (-0.98%)
- Closest to CPI (-0.31%), suggesting quality adjustment is important
- Frequent chaining suffers from severe chain drift

Why indices differ

- **Monthly FHPI (-5.27%)**:
  - Chain drift from compounding errors
- **Matched Fisher (-3.12%)**:
  - Selection bias from product entry/exit
  - Missing quality adjustment for new products
- **Jevons Index (-3.01%):**
  - No quantity weighting
  - Posted prices vs. transaction prices
- **Adobe DPI (-2.02%):**
  - Different product mix
  - Selection bias from entry/exit
- **CPI (-0.31%):**
  - Different methodology (within-subgroup measurement)
  - Different product mix

Key design takeaways

Structured Sparsity:

- Original representation: millions of pixels + thousands of words
- Effective representation V_i: hundreds of dimensions
- Deep learning exploits latent structure postulated by hedonic theory

Transfer Learning:

- BERT: pre-trained on language understanding tasks (open source)
- ResNet-50: pre-trained on image classification (open source)
- Trained to perform price prediction taking pre-trained embeddings X_i as fixed inputs
  - highly scalable and effective approach
  - produces informative "value embedding," V_i

Multi-task Learning:

- Predicts complete time series of prices simultaneously
- Shares information across time periods
- More accurate than period-by-period estimation

BERT: Understanding Text

Bi-directional Encoder Representations from Transformers

Devlin et al. '18, Vaswani et al., '18

- Transformer architecture with self-attention mechanism (open source, precursor to ChatGPT)
- Learns contextual word embeddings
- Pre-trained on masked word prediction and sentence ordering
  - first with a massive amount of publicly available text
  - then with Amazon product descriptions

Example: "Classic red wool cardigan sweater"

- Traditional: Each word treated independently
- BERT: Understands "classic" modifies style, "wool" indicates material, "cardigan" specifies garment type

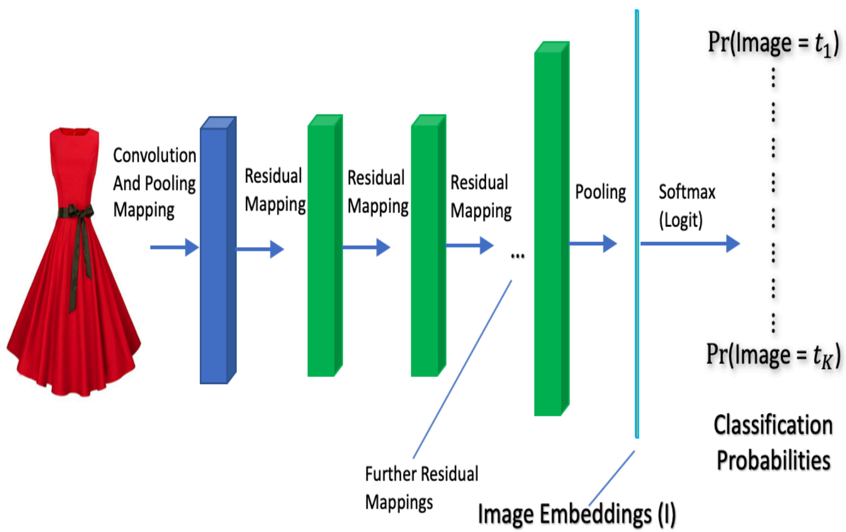
- Output: 768-dimensional embedding capturing product attributes

ResNet-50: Understanding Images

Residual Neural Network for Image Classification
He et al., '16

- Convolutional neural network with 50 layers
- Key innovation: "residual" connections improve training convergence
- Pre-trained on ImageNet data (14M images, 22K categories)

- **Extracts visual features:**
  - Shapes and patterns
  - Colors and textures
  - Style and design elements



- Output: 2048-dimensional embedding that summarizes visual information about product

Summary of contributions

1. Methodological Innovation

- Large-scale hedonic model and hedonic price indices using AI-generated features
- Provides adaptable framework for using deep learning to construct price indices
- Open-source components enable replication

2. Empirical Findings

- 80–90% R^2 validates hedonic model as good approximation
- Documents quality-adjusted deflation in apparel
- Attests to importance of chain drift

3. Policy Relevance

- Complements traditional CPI methodology
- Fast and scalable approach for statistical agencies
- Addresses challenges specific to e-commerce setting

Implications for Economic Measurement

Advantages of AI-based approach:

- **Scalability:** Handles millions of products automatically in a matter of hours
- **Cost:** Less than $500 in cloud computing + $460/month storage
  - not counting R&D expenses, which were significant
- **Timeliness:** Can produce indices quickly or in real-time
- **Coverage:** Works across product categories without domain experts

Timeliness:

- E-commerce continues rapid growth
- Electronic data increasingly available
- Traditional survey methods are becoming costlier
- Need for quality adjustment intensifies with increasing product variety

Recap

- **Problem:** Modernizing price statistics
  - more goods transacted online
  - increasing variety; high rates of entry and exit
  - traditional quality adjustment is costly and time consuming
- **Solution:** AI automatically extracts features from text and images
  - BERT for product descriptions
  - ResNet-50 for product images
  - multi-task neural network for price prediction
- **Performance:** 80–90% out-of-sample R^2
- **Finding:** Quality-adjusted apparel prices declined 0.98% annually (2014–2019)
- **Impact:** Scalable, cost-effective alternative method for price level measurement

Future Research Directions

1. **Model Explainability**
    - Attribute price changes to specific product characteristics
    - Interpret value embeddings
2. **Better Use of Images**
    - Images currently add little predictive power beyond text
    - Explore newer technology: vision transformers and multimodal models
3. **Extensions**
    - Other product categories
    - Real-time index construction
    - Integration with official statistics
4. **Welfare Analysis**
    - Link to consumer surplus measurement
    - Decompose price vs quality changes
    - Consumer characteristics: study heterogeneity in the impact of price changes

Historical remarks

Origins and purpose. Initiated by P. Bajari and V. Chernozhukov and first presented at the 2018 FESAC meeting, the project was conceived explicitly as a public-service effort: a concrete blueprint for modernizing official inflation measurement using electronic data. *FESAC agenda, December 2018*

Policy engagement. The work has been widely shared with and discussed with major statistical agencies and central banks in the U.S., U.K., Europe, and Japan, generating strong engagement among economists working on price statistics and inflation measurement.

Academic reception. While the project attracted little interest from general-interest economics journals, it received substantial enthusiasm from the econometrics community, culminating in publication interest from the Journal of Econometrics (thank you!)

Broader impact. Beyond official statistics, modern hedonic methods have clear business applications —for example, helping firms and platforms identify high-quality products and better deals for consumers.

Engineering effort. The paper's work was a major engineering effort, involving both econometricians (Bajari, Chernozhukov, Vijaykumar) and machine learning and data processing engineers at Amazon (Cen, Manukonda, Wang).

In the meantime: Following the release of BERT as an open-source project from Google in 2018, transformer-based models have become deeply embedded in modern society. The initial ChatGPT demo was released around the time of the paper's R&R decision; two versions later, it now has an estimated 700 million weekly users.

Thank You!

Questions?

Paper: “Hedonic prices and quality-adjusted price indices powered by AI” Journal of Econometrics 251 (2025)

Appendix: Model Design Details

Multi-Task Neural Network Architecture

$$Z_i \xrightarrow[\text{BERT/ResNet}]{e} X_i \xrightarrow{g_1} E_i^{(1)} \xrightarrow{g_2} E_i^{(2)} \xrightarrow{g_3} V_i \xrightarrow{\theta'} \{H_{it}\}_{t=1}^{T}$$

- Input dimension: 2816 (768 text + 2048 image)
- Hidden layers: 3 (with ReLU activation)
- Value embedding: 256 dimensions
- Output: 72 time periods (months)
- Loss function: MSE weighted by quantity + temporal smoothness penalty

$$\mathcal{L}(\theta) = \sum_{i=1}^{I} \sum_{t=1}^{T} Q_{it}(P_{it} - \theta_t' V_i)^2 + \lambda \sum_i \sum_t |\theta_{t+1}' V_i - \theta_t' V_i|$$

Appendix: Model Training Details

Sample split (by products, preserving time series):

- Training: 60%
- Validation: 20% (used for hyperparameter tuning)
- Testing: 20% (used to compute out-of-sample accuracy)

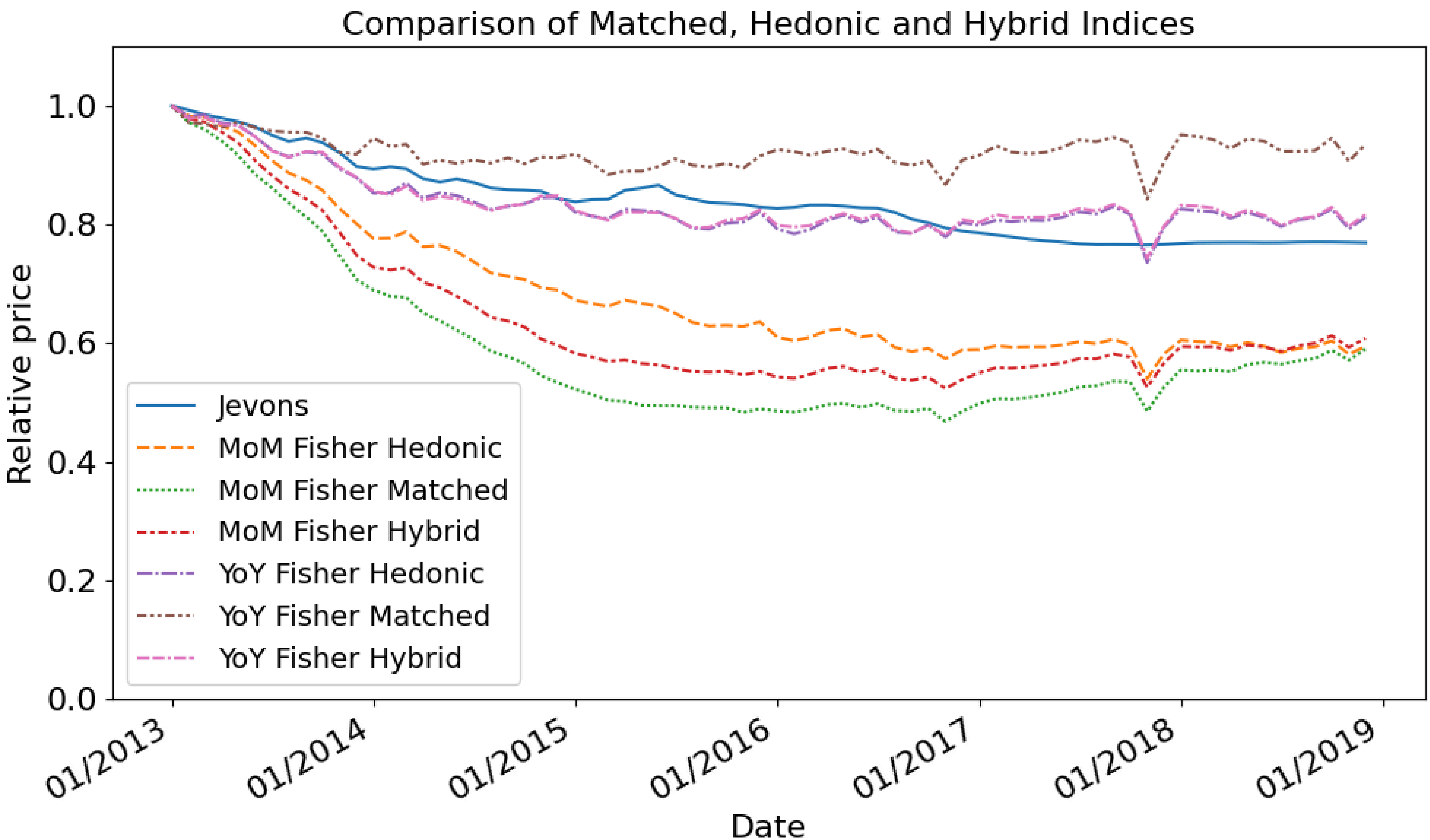
Optimization:

- Adam optimization algorithm (gradient descent with momentum)
- Regularization via temporal smoothness:

$$\lambda \sum_i \sum_t |\theta'_{t+1} V_i - \theta'_t V_i|$$

- Train model, assess fit in validation data, repeat

Appendix: chaining and imputation methods



Appendix: effect of "sibling products"