The Price of Everything Everywhere

Non-technical Summary

Jeff K.W. Cheung (UC Davis)

This version: January 22, 2026

What is "price"? In everyday life, it is simply what we pay for goods and services. In economics, however, prices play a broader and more fundamental role tracking the health of an economy, especially when aggregated into indices such as the Consumer Price Index (CPI), the GDP deflator, and the Federal Reserve's preferred Personal Consumption Expenditures (PCE) index. These measures guide monetary policy, including the well-known 2% inflation-target goal.

Yet prices matter far beyond aggregate inflation. At a granular level, prices reveal how markets function: firm- and product-level prices shed light on markups and market power. Wage offers for the same occupation across cities capture local labor market conditions. Interest rates quoted by different banks reflect variation in credit conditions. Linking prices to geography and demographics makes it possible to study how macroeconomic policies—such as tariffs, monetary shocks, or fiscal stimulus—affect different regions and populations. When observed over time, such data also allow researchers to track changes in the cost of living and the evolution of income inequality across regiones. Despite their importance, systematic micro-level price data over long historical horizons have largely been unavailable.

The "Global Price Initiative" database addresses this gap by constructing a new historical dataset from millions of digitized historical newspaper *advertisements* published from 1850 to recent years. These advertisements contain rich local economic information, including grocery prices, housing rents, mortgage and deposit rates, banks' balance sheets, and occupation-level wages drawn from classified listings. The textual descriptions in advertisements also provide information about product characteristics, which may allow adjustment for over-time quality improvement.

Recent advances in machine learning, cloud computing/storage, and modern AI tools make it possible to recover this granular price information at an unprecedented scale for the first time. Figure 1 shows the data construction pipeline: first, advertisement regions are identified using the Newspaper Navigator Database, supplemented with a custom binary
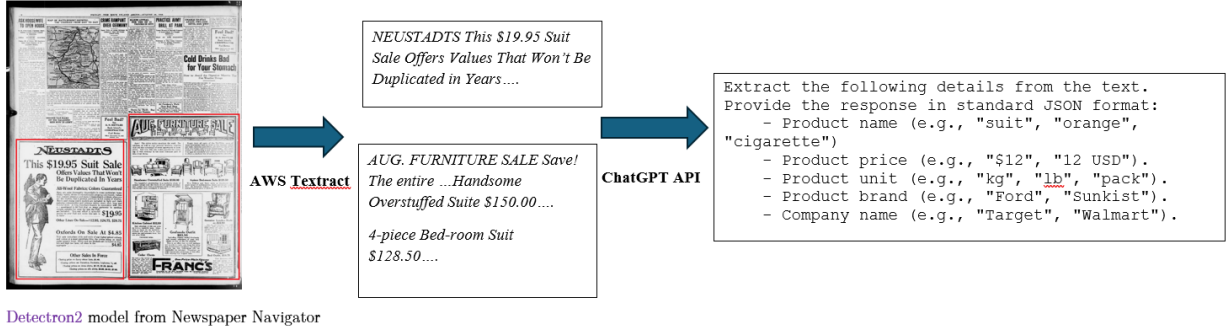
classification model trained to detect advertisements likely to contain price information.[1] Optical character recognition tools then convert scanned images into text. Large language models—API access to Google Gemini or ChatGPT—are subsequently used to extract structured fields—including prices, product descriptions, retailer/firm names—and link them to publication dates and locations. The output is a standardized time–geography–firm–product dataset generated in an automated and reproducible way.

Table 1 shows a pilot sample from *The Rock Island Argus* (Illinois, August 16, 1918) and illustrates the richness of the data. On a single date, seven furniture items are observed with prices ranging from $19.95 to $150, each tagged with brand and retailer. As such observations accumulate across time and space, they form a high-frequency panel of local economic conditions.

The resulting database enables the construction of detailed price indices, wage series, and return measures across counties, states, and countries. To my knowledge, no existing project has systematically collected advertisement-level price data at comparable scale, and commercial alternatives are costly and lack similar historical coverage. Although historical sources pose challenges—including inconsistent layouts and imperfect print quality—the public release of this dataset, together with open code and documentation, offers an unprecedented empirical foundation for research on *prices*, inflation, market power, and the distributional effects of macroeconomic policies.

---

[1] Because of resource constraints, the project initially focuses on the most promising advertisements—such as those containing symbols like "$" or "%"—and temporarily excludes images that are too blurry to process reliably. Additional financial resources would allow the project to scale rapidly and expand coverage.

Figure 1: The workflow of the Global Price Initiative

NEUSTADTS This $19.95 Suit Sale Offers Values That Won't Be Duplicated in Years….

AUG. FURNITURE SALE Save! The entire …Handsome Overstuffed Suite $150.00….

4-piece Bed-room Suit $128.50….

AWS **Textract**

ChatGPT API

```
Extract the following details from the text.
Provide the response in standard JSON format:
    - Product name (e.g., "suit", "orange",
"cigarette")
    - Product price (e.g., "$12", "12 USD").
    - Product unit (e.g., "kg", "lb", "pack").
    - Product brand (e.g., "Ford", "Sunkist").
    - Company name (e.g., "Target", "Walmart").
```

Detectron2 model from Newspaper Navigator

Notes: This figure shows the workflow of the Global Price Initiative. In particular, this is a newspaper page from "The Rock Island Argus" (August 16, 1918; IL).

Table 1: A snapshot of the dataset (pilot extract)

Date	Location	Product Name	Price	Company Name
1918-08-16	Rock Island cty, IL	Suit	19.95	NEUSTADTS
1918-08-16	Rock Island cty, IL	Handsome Overstuffed Suite	$150	Franc's
1918-08-16	Rock Island cty, IL	4-piece Bedroom Suite	$128.59	Franc's
1918-08-16	Rock Island cty, IL	Kitchen Cabinet	$32.50	Franc's
1918-08-16	Rock Island cty, IL	Grafonola Outfit	$83.50	Franc's
1918-08-16	Rock Island cty, IL	Genuine Leather Rocker	$22.50	Franc's
1918-08-16	Rock Island cty, IL	Bed Outfit	$18.75	Franc's

Notes: This table presents a snapshot of advertisement data extracted from Figure 1. It highlights the richness of information available even in a single advertisement snippet. For example, we know the publication date ("1918-08-16"), the location (Rock Island County, Illinois) that can be mapped to income and demographic profiles, the product name (e.g., suit, kitchen cabinet), the price ($18.75–$128.59 for different products), and the retailer's name (e.g., Franc's). Together, these fields enable analysis of how firms price the same products across markets and over time.