

The Price of Everything Everywhere

Jeff K.W. Cheung (UC Davis)¹

This version: October 19, 2025

[Preliminary draft. Please do not circulate or cite]

Abstract:

This research project develops a novel historical dataset on *prices* of tradable goods and non-tradable services (occupation-level wages), returns on financial assets (bank deposit rates across maturities), as well as firms' balance sheets, across U.S. counties and international borders from 1850 to 2000. Using advanced machine learning methods, I examine printed *advertisements* from millions of digitized historical newspapers in order to shed light on (i) price stickiness and firms' markup structures at both domestic and international scales, (ii) the heterogeneous and distributional effects of macroeconomic policies—such as tariffs and monetary shocks—on realized inflation experienced by households from different income groups and demographic compositions across geographies, and (iii) the differential factor prices for capital and labor (e.g., interest/mortgage rates across maturities, and offered wages across service occupations) in different regions. The resulting “Global Price Initiative” database provides researchers with a highly detailed, product-level dataset that supports transparency, replicability, and versatility in macroeconomic and international research on *prices*.

¹ I am grateful for funding opportunities from Google Cloud Research Credits and the UC Davis Research Fellows Program. I thank Raphael Zhang, Inbar Schwartz, Andrew Huang for their excellent research assistance. This project builds on the Library of Congress' [Newspaper Navigator](#). All errors are mine.

1. INTRODUCTION

What exactly is “price”? In everyday life, price is what we pay to acquire a product or service. However, from a macroeconomic perspective, “price” takes on a more sophisticated role, particularly when we talk about the “price index.” Policymakers often look at price as an aggregate measure, using a single index to track the price of *all* goods and services in the economy. This is why we have terms like the Consumer Price Index (CPI), the GDP deflator, and the Federal Reserve’s preferred Personal Consumption Expenditures (PCE) index. These price indexes provide a snapshot of economic health, guiding important decisions in monetary policy, such as maintaining an explicit “inflation target”—the Federal Reserve’s goal of keeping the inflation rate at around 2%.

Stepping beyond our daily life, the concept of “price” becomes even more crucial, yet often overlooked, in assessing firms’ market power and a country’s economic capacity. At the granular level, product- and firm-specific prices reveal markup behavior and market power; interest rate quotes at the same maturities offered by different commercial banks, or wage offers for the same occupation in different cities, capture local financial and labor conditions. These prices can be further linked to regional characteristics, such as income distribution or industrial composition, to track the distributional effects of aggregate macroeconomic policies (e.g., tariffs, monetary, or fiscal policies). Despite its fundamental role, the true importance of price in these contexts is often underappreciated, even though it shapes economic analysis on both a local and global scale.

In this research proposal, I introduce a new historical dataset on price information for a wide range of tradable goods, non-tradable services, and financial assets across many countries from 1850 to 2024. Methodologically, I leverage advanced machine learning algorithms to identify, classify, and extract price data from millions of historical newspaper *advertisements*—an invaluable source of micro-level product and firm data that has been underused due to past technological constraints. The resulting “Global Price Initiative” database offers researchers an exceptionally rich resource for constructing price indexes on the cost of living, bank interest rates and mortgage rates at various maturities, housing rents and returns, and wages of different occupations across both space and time. To the best of my knowledge, this is the first effort to provide micro-level price data at time-geography-firm level sufficient to track the price dynamics across different regions and times.

1.1 Related Literature

This research project builds on the global collaborative efforts to digitize and preserve historical newspapers. Notably, the Library of Congress initiated the “[Chronicling America](#)” project and the “Beyond Words Database,” which make millions of historical U.S. newspapers publicly accessible and leverage volunteer annotations (e.g., headlines, maps, advertisements). Building on this ambitious initiative, the “[Newspaper Navigator](#)” developed by [Lee et al. \(2023\)](#), which trains a machine learning model to recognize the content categories present on each newspaper page and identify their broad location, is the foundation of my work. On the other hand, [Dell et al. \(2024\)](#) transcribe more detailed article-level information under each headline. In general, existing work focuses on textual articles (e.g., sentiment, topic) rather than visual advertisements. While manual collection of printed advertisement is labor-intensive, printed *advertisements*—designed to be eye-catching—are well-suited to automated detection.

Although economists have likewise expanded long-run asset-return datasets: [Jorda et al. \(2019\)](#) developed an extensive macro-finance database that, for the first time, includes data on housing rent returns. [Lyons et al. \(2024\)](#) create a comprehensive database of housing prices across U.S. cities from 1890 to 2006 with substantial RA effort. The Global Price Initiative offers a scalable, automated alternative by focusing on advertisement-level signals and machine-assisted extraction.

2. MINIMUM-SCALABLE PROJECT

[Figure 1](#) lists advertisement categories that the Global Price Initiative examines: (i) daily grocery prices, (ii) bank promotions (interest/mortgage rates by maturity), (iii) housing prices/rents, (iv) balance sheets of commercial banks/firms, and (v) occupation-level wages from the “classifieds” column, in particular the service sectors (e.g., house cleaning). The search pipeline prioritizes advertisements containing numerical values or price-related symbols like “\$” and/or “%” and then tags each record with date, location, brand name, and retailer name. This enables markup comparisons for the same good across different regions. All digitized newspapers used for this analysis are publicly accessible from reputable institutions (e.g., the Library of Congress). Our focus on advertisements—small sub-regions of each page—supports fair, non-commercial scholarly use.

[Figure 2](#) shows the overall workflow of the Global Price Initiative: since many advertisement snippets lack price information or are too blurry for analysis, I train a binary machine learning classifier to pre-screen and select advertisement images that are most likely to contain usable price information. Then, the OCR tools (e.g., AWS textract or Google Cloud Vision) transcribe printed advertisement snippets into machine-readable text. Last but not least, generative AI tools, such as ChatGPT, summarize prices and the names

of products, brands, retailers, and locations, together with unique IDs, page URLs, and OCR-confidence scores. The resulting database is fully reproducible and scalable on cloud platforms (AWS EC2 GPU instances and S3 buckets).

Table 1 shows the template: each advertisement image provides a self-contained record with the publication date, location, product description, quoted price or wage, seller/retailer/bank name, and contextual cues such as quality or size. A pilot sample from *The Rock Island Argus* (Aug 16, 1918, IL) demonstrates the structure: seven furniture items ranging from \$19.95 to \$150, each tagged with brand and retailer. Similar records are collected across time and space to form a time–geography–product panel. The resulting microdata can be aggregated to construct price indices, wage series, and return measures at multiple spatial levels (city, county, country).

3. PROJECT FEASIBILITY AND LONG-TERM GOAL

This research project proposes a scalable methodology for measuring price changes across time and space from primary resources—printed advertisements from historical newspapers. The resulting Global Price Initiative Database will provide granular time–geography–product data that track consumer prices, bank interest rates by maturity, local wages by occupation, and asset returns across space and time. Coverage spans regions within the United States and internationally. To my knowledge, no study has yet systematically harvested advertisement-level prices on a global scale, and the closest substitutes (e.g., Orbis and Compustat) require expensive subscriptions and lack comparable historical coverage. Although font heterogeneity and irregular layouts in printed newspaper advertisements remain non-trivial challenges, the public release of this ambitious project will provide researchers with an unprecedented product-, occupation-, and firm-level granularity across geography and time, accompanied by open data and code to support transparency, replication, and policy analysis.

Reference:

- Bergin, Paul R., Reuven Glick, and Alan M. Taylor. 2006. “Productivity, Tradability, and the Long-Run Price Puzzle.” *Journal of Monetary Economics* 53 (8): 2041–66. <https://doi.org/10.1016/j.jmoneco.2006.05.012>.
- Carlson, Jacob, Tom Bryan, and Melissa Dell. 2023. “Efficient OCR for Building a Diverse Digital History.” arXiv. <https://doi.org/10.48550/ARXIV.2304.02737>.
- Dell, Melissa, Jacob Carlson, Tom Bryan, Emily Silcock, Abhishek Arora, Zejiang Shen, Luca D’Amico-Wong, Quan Le, Pablo Querubin, and Leander Heldring. 2023. “American Stories: A Large-Scale Structured Text Dataset of Historical U.S. Newspapers.” arXiv. <https://doi.org/10.48550/ARXIV.2308.12477>.
- Feenstra, Robert C., Robert Inklaar, and Marcel P. Timmer. 2015. “The Next Generation of the Penn World Table.” *American Economic Review* 105 (10): 3150–82. <https://doi.org/10.1257/aer.20130954>.
- Gerding, Felix, Espen Henriksen, and Ina Simonovska. 2024. “The Risky Capital of Emerging Markets.” *NBER Working Paper No. 20769*, April. <https://doi.org/10.3386/w20769>.
- Jordà, Oscar, Katharina Knoll, Dmitry Kuvshinov, Moritz Schularick, and Alan M. Taylor. 2019. “The Rate of Return on Everything, 1870–2015*.” *The Quarterly Journal of Economics* 134 (3): 1225–98. <https://doi.org/10.1093/qje/qjz012>.
- Lee, Benjamin Charles Germain, Jaime Mears, Eileen Jakeway, Meghan Ferriter, Chris Adams, Nathan Yarasavage, Deborah Thomas, Kate Zwaard, and Daniel S. Weld. 2020. “The Newspaper Navigator Dataset: Extracting Headlines and Visual Content from 16 Million Historic Newspaper Pages in Chronicling America.” In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 3055–62. Virtual Event Ireland: ACM. <https://doi.org/10.1145/3340531.3412767>.

Figure 1: Advertisement categories from historical newspapers

Panel (a): Bank interest rates

Call us quick,
before our
better judgment
wins out over
our generosity.

1 Month CD	33 Month CD
4.03%	5.50%
Annual Percentage Yield	Annual Percentage Yield

A lot of people will be second-guessing us on this one. After all, it's not every day you see a bank being this generous.

But hey, if we wanted to be another me-too bank then we'd have just offered another me-too rate on our CDs. But that's not how we work.

We believe our customers deserve a little more. A little more attention. A little more respect. And, of course, a little more money.

Compass Bank
Our Prime Interest Is You.

Panel (b): Balance sheets:

**Report of The
Scandinavian American Bank**
At Close of Business, June 30, 1916.

RESOURCES	
Loans and Discounts	\$ 2,104,714.55
Real Estate	410,100.48
Furniture and Fixtures	33,750.00
Ware, Stocks and Bonds	1,052,570.22
Cash and Exchange	1,866,343.74
Customers' Liability under Letters of Credit	4,000.00
Credits granted on Domestic and Foreign Acceptances	250,000.00
	\$12,731,448.99

LIABILITIES	
Capital	\$ 500,000.00
Surplus	500,000.00
Undivided Profits	11,016.08
Domestic Deposits	\$4,165,279.61
Time and Savings Deposits	7,506,554.28
Domestic and Foreign Acceptances	250,000.00
	\$12,731,448.99

New Accounts Invited
On the basis of strength, conservatism and real banking service, we are pleased to invite additional accounts from banks and business, firms, corporations and individuals, to each of whom will be accorded courteous treatment and every accommodation that is consistent with good, sound banking.

**The Puget Sound Mortgage Co.'s
Collateral Gold Bonds**
In Denominations of \$100 and \$500
The 1-year Bonds bear 5 per cent.
The 2-year Bonds 6 per cent.
3 and 5-year Bonds 6 per cent.

The security behind these bonds consists of First Mortgages on Improved Seattle Real Estate and Seattle Local Improvement Bonds—all deposited with The Scandinavian American Bank as trustee. On top of that and as a matter of course there is the full security of The Puget Sound Mortgage Co., amounting to \$100,000, as a further security. Interest and principal collected without charge by The Scandinavian American Bank and either forwarded or placed to credit of customer's account as directed.

Scandinavian American Bank
USE OUR BALLARD BRANCH IF MORE CONVENIENT.
RESERVE OVER \$11,000,000.

Panel (c): Grocery prices

YOU SAVE MORE! LOWEST FOOD PRICES IN THE SOUTH!

Ground Beef 97¢	Top Sirloin Steak 2.97¢	Sweet Oranges 97¢
World Classics Drinks 97¢	Pork & Beans 41¢	Food Club Milk 1.97¢
Hunt's Ketchup 97¢	Potato Chips 87¢	Fryer Breast 1.17¢
Puritan or Citric Oil 1.17¢	Coca-Cola 1.17¢	Boneless Cube Steak 2.37¢
Agave Fresh Tomatoes 47¢	Green Cabbage 5.97¢	Zeigler's Beans 1.37¢
Model Tablets 2.97¢	Snow White 1.97¢	Bologna or Wieners 97¢
Black & Lamb Sausages 67¢	Mushrooms 1.97¢	Smoked Sausage 1.97¢
Turkey Breast 3.97¢	Headless Shrimp 3.97¢	Ice Cream 1.97¢
		Supersoft Bread 2.97¢

Sack and Save Foods
MUSCLE SHOALS - FLORENCE
1100 Woodward Ave., South Plaza • 300 One Creek Parkway

Panel (d): Wages and rents

899 Mobile Homes For Rent
PETERSVILLE area, 2 bedroom on large private lot, close to shopping center, \$400 + security deposit. References: 110-1470, call between 8 A.M. - 4 P.M. only.

699-Homes For Rent
6 MILES northwest of Rogersville, brick 2 bedroom, 2 bath, no pets, \$275 month plus deposit. 209-837-4640

COUNTRY LIVING - 2800
sq. ft. 4 bedroom brick, 1/2 basement central heat air, fenced 8 acres ideal for horses or cows, Wilson High School District, Florence, \$350 month. Call 764-6033.

CREEKWOOD Forest, 3 bedroom, 2 bath town-house cathedral ceiling. Clean, \$550 + deposit. 386-6520, 764-1221.

DONNEGAN SLOUGH, Weekly rental, Cable TV, 3 bedrooms, 1 1/2 bath, pier, beachhouse. 766-1238

EXECUTIVE HOMES
In prime area. Priced from \$550 per month. Call Julia at Warner Properties 766-0069

FLORENCE - 2 bedroom, unfurnished, \$290 month. \$105 deposit. No pets. 766-7442

FLORENCE - Large 3 bedroom, 2 bath home, dec. central heat/air, on double size lot, quiet neighbor hood. \$700 - \$700 deposit. 766-4137

KILLEN AREA
2-story, 4 bedroom, large lot, \$600 month + deposit. 386-7185.

MSC - 3 bedroom, 2 bath brick, double garage, excellent location, fenced fireplace, more. \$600 month. 767-2673 after 5.

MUSCLE SHOALS - 3 bedroom, 1 1/2 bath brick, \$450 month, \$100 deposit, no pets, references, partially furnished. Call 381-9336.

710-Homes For Sale
CHEROKEE - N. Pike Rd. 2 bedroom, large lot, utility building, large garden spot. 381-1046

CONTRACTORS
Or handy person special! Small frame house in Sheffield. Free and clear deed. Final \$18,200. 383-7018

FORD ST., M.S.C.
1,250 sq. ft. 3 bedroom, 1 1/2 bath, central heat/cooling. 24x30 detached garage, ceiling fans, appliances stay. 383-7354.

GREAT NEIGHBORS
H. Sherrod Ave. 3 bedroom, 1 1/2 bath, central heat/air, large fenced backyard, extra nice house & lot. \$57,500. Call for appointment. 767-3662.

GREENHILL
3 bedroom, 2 bath, city water, gas, cable. 2 miles from school. House only 14 months old \$13,900 down and assume payments of \$301.43. Seller will finance part of down payment. Must have good credit. Call 760-1066.

LOOK for the NEESE REAL ESTATE ads this weekend! We hold the key to your next home! 767-7000.

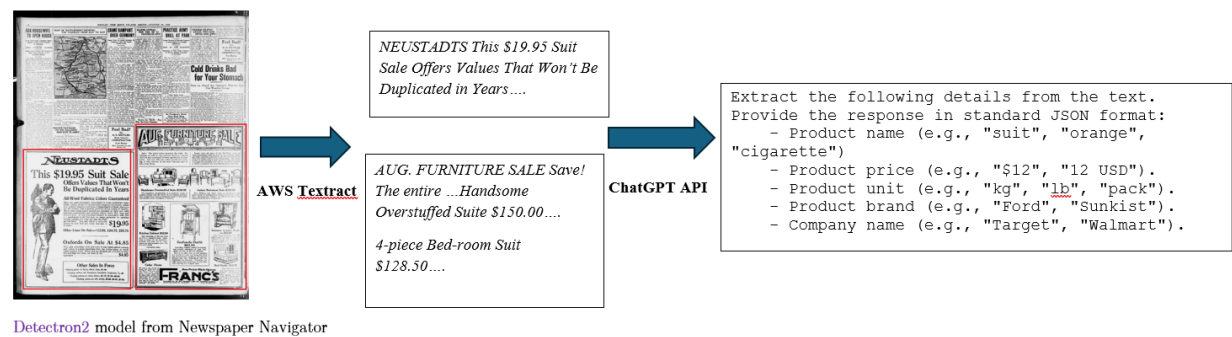
NEW 4 BEDROOM
With game room, Jackson Hts. M.S.C.
Robert Eaton 381-4663

NEW BRICK
3 bedroom, 2 1/2 bath, bonus room, Jackson Hts. MS Robert Eaton 381-4663

NEW LISTING
Margrunt, 3 bedroom, 2 1/2 bath brick with view of lake. \$650.
Alan Realty 381-8330

NEW BY BUILDER
Brick, 2 story, in Aycock Hts. Over 2400 sq. ft. on 1/2 acre corner lot, 3 or 4

Figure 2: The workflow of the Global Price Initiative



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.