Cataloging Growth: A Re-Evaluation of 1900–1990*

Verónica Bäcker-Peral MIT
Benjamin Wittenbrink MIT

August 11, 2026

Abstract

Measuring real GDP growth requires distinguishing changes in prices from changes in product quality and composition, yet systematic quality adjustment of price indexes is unavailable for much of the twentieth century. We construct a new quality-adjusted price index for U.S. consumer goods using 5.1 million product listings from Sears catalogs, 1900–1990. We use large language models to extract product information and estimate hedonic price schedules from high-dimensional text embeddings, allowing us to infer annual changes in the cost of living. The resulting cost-of-living index implies substantially lower goods inflation than conventional deflators, and consequently implies much faster real economic growth: between 1900 and 1990, real goods consumption grew by a factor of 39 using our index, compared with a factor of 10.3 using standard goods deflators. The gap between our index and canonical ones is largest before World War II, reversing the conventional view that goods consumption growth was slower before 1945 than in the post-war decades.

JEL Classification: E01, N11, N12, O47, C43.

*We would like to thank Abhijit Banerjee and Dave Donaldson for invaluable guidance and mentorship. We are very grateful to Daron Acemoglu, Ricardo Caballero, Amy Finkelstein, Joe Hazell, Jack Kelly, Russell Legate-Yang, Atif Mian, Emi Nakamura, Christina Romer, David Romer, David Weinstein, Iván Werning, Chris Wolf and workshop participants at MIT for comments and suggestions. We thank the George and Obie Shultz Fund at MIT for financial support.

1 Introduction

The period of most transformative growth in modern US history is also amongst the least understood. Economic historians have emphasized the century from roughly 1870 to 1970 as an exceptional episode in American economic growth (Gordon, 1999; Abramovitz and David, 2000; Gordon, 2016). There is less agreement about how or when that growth took place. Conventional estimates place the peak of real output and measured productivity growth in the two decades after World War II (Jordà et al., 2017; Gordon, 1999), while qualitative accounts emphasize substantial technological progress before the war (Gordon, 2016). Progress on this question has been slow in large part because high-quality price data are scarce for most of this period. As Gordon (1999) puts it, "the history of price changes from 1914 to 1947 [is] the black hole where little is known." This paper sheds light on that black hole.

Economists have long recognized that accurate price measurement is essential for welfare comparisons. A 1915 Bureau of Labor Statistics (BLS) report lamented that "men did not attempt to measure changes in the level of prices until after they had learned to measure such subtle magnitudes as the weight of the atmosphere, the velocity of sound, fluctuations of temperature, and the precession of the equinoxes" (United States Bureau of Labor Statistics, 1915). Rapid inflation during World War I made price measurement especially salient, and the BLS responded by introducing the Consumer Price Index (CPI), which aims to be a nationally representative measure of the cost of living. How prices are measured shapes much of what we know about growth. Nominal U.S. Gross Domestic Product grew by a factor of nearly 500 over the twentieth century, but the CPI attributes 44% of average annual nominal growth to changes in the cost of living alone.[1]

There is a long-standing concern, however, that the CPI has historically overstated inflation by failing to fully account for changes in the quality of consumer goods and services (Boskin et al., 1998). The quality of goods available to consumers has improved dramatically in the last century. A refrigerator in 1900 was a large wooden box, much of it taken up by a compartment that held blocks of ice to keep the rest cold. By the 1930s, households could purchase electric refrigerators, which continued to improve over the century as they added features such as ice makers and water dispensers. Mattresses, often filled with cotton felt, sea moss, or horsehair in the early 1900s, improved markedly over the century through better spring systems and the introduction of foam and latex for greater support and comfort. Innovations in textiles led to softer fabrics, machine-washable clothing, and elastic materials. More broadly, almost every aspect of the American consumer landscape was transformed (Gordon, 2016). In the 1990s, the BLS revised the CPI methodology to account for quality change in a subset of goods and extended some of these revisions back to 1978, but for much of the earlier period there was effectively no explicit quality adjustment.[2]

This paper constructs a new quality-adjusted price index for U.S. consumer goods over the twentieth century. We digitize a nearly complete run of Sears general merchandise catalogs from 1900 to 1990, containing approximately 5.1 million product listings. For each product category and year, we estimate hedonic price schedules from high-dimensional text embeddings of the product descriptions, then aggregate these estimates into a quality-adjusted price index. We present two main results: first, the overall rise in the standard of living over the twentieth century was much larger than conventional estimates suggest. Using existing price indexes for consumer goods, real goods consumption grew by a factor of 10.3 between 1900 and 1990; using our quality-adjusted goods index, it grew by a factor of 39. This revision is driven by changes in the quality of retail goods. Unadjusted prices, which track average price changes within each category of goods, rise much more than quality-adjusted prices: quality improvements account for roughly half of the cumulative rise in unadjusted prices between 1900 and 1990. Existing BLS-based goods price indexes track our unadjusted series much more closely than our

[1] We obtain data on long-run GDP from the Macrohistory Database (Jordà et al., 2017). We obtain historical CPI data from the Minneapolis Federal Reserve Bank.

[2] We describe the BLS methodology in detail in Section A.3.

quality-adjusted series, suggesting that conventional deflators miss a quantitatively important component of quality growth. Carried through to real growth, the difference implies that goods consumption rose about 1.6 percentage points faster per year than existing deflators record; for the period of 1900-1939, the difference is much larger with an annual average of 3.1 percentage points. By 1990, real goods consumption is therefore roughly four times higher relative to 1900 under our quality-adjusted index than under existing deflators.

Our second main result is that a substantial share of 20th century output and productivity growth took place before World War II. According to existing price indexes, real goods consumption growth peaked in the postwar decades: average annual growth was 2.3% from 1900 to 1939 and 3.1% from 1946 to 1980. Replacing the conventional goods deflator with our quality-adjusted Sears index reverses this pattern. We estimate average annual real goods consumption growth of 5.4% for 1900-1939 and 4.2% for 1946-1980. We find similar patterns for multi-factor productivity. Our estimates therefore suggest that the era of fastest growth was in the prewar, not postwar, contrary to conventional estimates.

The Sears catalogs provide an unusually rich source for historical price measurement. Sears was the largest retailer in the US for most of the 20th century and published two general merchandise catalogs per year, in the spring and fall, as well as a Christmas catalog. These catalogs listed nationally posted prices for a broad range of consumer goods and described products in considerable detail. We digitize a nearly complete collection of these catalogs from 1900 to 1990 and use large language models to extract structured product records from the catalog pages.

Building on a long tradition of empirical work, we use hedonic methods to construct a cost-of-living index for US consumer goods. Hedonic methods are especially useful in settings like ours for which we have rich data on product prices and characteristics but very limited data on quantities of goods sold. For each product listing, we map the full description into a high-dimensional vector representation and estimate price schedules separately by year and product class using ridge regressions. This approach retains the economic logic of the hedonic tradition while relaxing the need to hand-code a small number of product attributes. It is especially useful in historical settings, where quality-relevant information is often contained in unstructured text and where the relevant attributes differ sharply across categories. The method is close in spirit to recent work using machine learning to estimate hedonic price indexes (Bajari et al., 2025; Ehrlich et al., 2026). We then aggregate estimates from hedonic regressions into a consumer goods price index using nationally representative weights. Under general conditions outlined in Pakes (2003), the resulting quality-adjusted index approximates the true change in the cost of living to first order. If anything, this interpretation understates the welfare gains from quality improvement, since the second-order bias goes in the direction of overstating the cost of living.

Interpreting our estimates as changes in the cost of living for consumer goods rests on three empirical requirements: the embeddings must capture price-relevant product characteristics, those characteristics must remain sufficiently comparable across adjacent years, and Sears must be informative about broader movements in retail prices for consumer goods. Across alternative specifications, samples, and text representations, the quality-adjusted index remains close to our baseline.

We first assess whether the embeddings recover price-relevant product characteristics. Our approach replaces traditional hand-coded characteristics with high-dimensional text embeddings, allowing the hedonic regression to capture a much richer description of product attributes. We estimate separate hedonic models by year and product category, using ridge regularization to discipline the high-dimensional specification. These models fit prices well: the median out-of-sample R^2 across year-category regressions is 0.94, suggesting that we capture most price-relevant characteristics. Moreover, our results are robust to alternative specifications, such as using principal components of the embeddings instead of ridge regularization. On a subsample of products, we hand-code price-relevant attributes and estimate a traditional linear hedonic regression, which yields similar estimates. We also provide evidence that our hedonic regressions capture welfare-relevant characteristics rather than marketing language. In particular, we construct alternative embeddings after removing promotional phrases and subjective

adjectives, such as "beautiful" or "superb." The resulting price index is very similar to the baseline and, if anything, implies slightly lower inflation. This suggests that our results are not driven by advertising language embedded in the product descriptions.

We next assess whether changing product characteristics undermine comparisons across adjacent years. Hedonic methods assume that even if the mix of products changes over time, the set of underlying characteristics available in the base period continue to exist in the next year. This assumption does not hold exactly: certainly some characteristics disappeared from the catalog as they became obsolete. If, however, characteristics are rare by the time they disappear, their loss should not affect our estimates. Using the text descriptions, we identify products with no close counterpart (i.e., exiting goods) in the following year and recompute the index excluding them. The results are effectively unchanged.

We finally assess whether Sears is informative about broader movements in retail prices for consumer goods. Sears was the largest retailer in the US for most of the twentieth century, accounting for about 6% of retail sales, and it sold a broad range of consumer durables and household goods, serving consumers nationwide. Its share of retail output in the post-World War II period was comparable to that of Amazon today. As an additional check, we construct a parallel quality-adjusted price index using J.C. Penney catalogs for 1969–1982. J.C. Penney was also a major national retailer, but served a higher-end segment of the market. The two catalog-based indexes are very similar over their period of overlap. This comparison suggests that the main results are not specific to Sears and are informative about broader movements in retail prices for consumer goods.

Beyond estimating aggregate quality change, the richness of the catalog descriptions allows us to isolate specific margins along which quality improved. Materials account for about one-third of total measured quality improvement, through both new synthetic fibers such as nylon and spandex and the application of existing materials to new uses, such as porcelain enamel on stoves. Electrification was another major margin: by 1939, more than half of the refrigerators and washing machines in the Sears catalog were electric, and after the war the share quickly reached 100%. Although quality changes are typically positive, economic theory predicts that during severe recessions, quality should fall as households substitute away from luxury goods (Fajgelbaum et al., 2011; Jaimovich et al., 2019). This is exactly what we find. At the onset of the Great Depression, measured quality fell by seven percent over three years, as Sears shifted toward smaller and lower-quality versions of products it had previously offered. This episode provides further evidence that the quality upgrading we measure using hedonic regressions is capturing real shifts in the economy.

We also compare our index to price indexes for other advanced economies and find that measured convergence between Europe and the United States after World War II depends crucially on price measurement. Using catalog data from La Redoute, a large French mail-order retailer, we provide suggestive evidence that quality upgrading in French consumer durables was much faster than in the US in the 1970s and 80s, so French CPI bias may have been high as well. As an extreme test, we deflate both French and US consumption with our quality-adjusted Sears index, which assumes true prices evolved similarly in the two countries; the France–US gap in average annual real consumption growth over 1950–1990 widens from 1.7 to 7.5 percentage points.

Our Sears-based estimates cover non-perishable consumer goods but not housing, food, or services. To assess the aggregate implications, we combine our index with existing price indexes for these other sectors and compare aggregate real consumption growth under two alternative consumer-goods deflators: the existing BLS index and our quality-adjusted Sears index.[3] Holding housing, food, and services fixed, our adjustment still is consequential for changes in total consumption. Using the standard BLS index, U.S. real total consumption grew by a factor of 13.3 between 1900 and 1990; using ours it grew by a

[3] We do not include investment goods, which Gordon (1990) shows saw much more dramatic quality improvement than retail goods, so price index bias for total GDP is likely higher than for total consumption. For the pre-1935 period, when to our knowledge no service-price series is available, we estimate service inflation using a two-sector goods-services model disciplined by the relationship between relative prices and relative employment.

factor of 19.8 — an average of 0.51 percentage points per year.

Related Literature This paper contributes to the literature on measurement of long-run growth in the U.S. and other advanced economies. Gordon characterizes the twentieth-century transformation of the US economy as "one big wave" that peaked in the decades after World War II (Gordon, 1999). This finding stands in contrast to evidence that key technologies had diffused well before the war. For example, by 1925, 89.5% of U.S. manufacturing establishments reported using electric power equipment (Du Boff, 1967; Devine, 1983).[4] David (1990) reconciles slow measured growth with rapid technological change by appealing to diffusion lags, which can keep a new technology from registering in measured productivity for decades. We reach a different conclusion: once we account for quality change in consumer goods, measured productivity growth was higher across much of the pre-World War II period than in 1946–1966. This revision reflects substantial quality improvements concentrated in the first half of the century and spanning a wide range of industries.

Catalog data has long been recognized as a consistent source for historical price measurement. Rees and Jacobs (1961) used Sears and Montgomery Ward catalogs to construct price indexes for clothing and home furnishings for 1890–1914, following the conventional BLS methodology of tracking repeated products and substituting the most similar alternative when a product disappears. This approach, still standard today, is well established to be biased relative to the true change in the cost of living (Pakes, 2003; Bils, 2009; Nakamura and Steinsson, 2012; Ehrlich et al., 2026). In an influential study, Gordon (1990) hand-collected data on several categories of durable goods for 1947–1983 and estimated hedonic regressions to construct a quality-adjusted index, finding that real growth in Producer's Durable Equipment was about three times higher than the official index implies. Relative to Gordon, we expand the scope of time and products and apply embedding-based methods that capture quality dimensions linear specifications may miss. Our implied revision is more modest: for 1947–1983, we find real consumption growth 1.3 times higher than the official index implies. The difference reflects that consumer goods saw less quality improvement over this period than investment durables like machinery and vehicles, so the official consumer-goods index carries less quality bias.

The hedonic approach treats goods as bundles of characteristics and estimates how market prices vary with them. Early work adjusted automobile prices for quality change (Court, 1939; Griliches, 1961, 1971), Rosen (1974) provided the canonical equilibrium interpretation, and subsequent work clarified when hedonic coefficients can be used to construct price indexes (Triplett, 1986, 2006; Pakes, 2003). Nordhaus (1996) studied the quality-adjusted price of lighting over a very long horizon, showing that conventional price indexes can miss large gains from technological change. Our contribution is to extend the hedonic tradition to a far broader historical setting, using millions of Sears product descriptions to estimate quality-adjusted indexes across many categories of consumer goods over nearly the entire twentieth century, and to measure those characteristics at scale with modern text-based methods.

The main alternatives to the hedonic approach rely on quantity or expenditure data. One strand imposes structure on preferences to construct exact price indexes (Diewert, 1976; Sato, 1976; Vartia, 1976), extended by Feenstra (1994) to allow entry and exit and widely applied to measure the gains from new varieties (Hausman, 1997; Broda and Weinstein, 2006; Redding and Weinstein, 2020) and the growth that imputation misses when entrants displace incumbents (Aghion et al., 2019). The other uses Engel curves: Hamilton (2001) and Costa (2001) infer CPI bias from shifts in food expenditure shares, while Bils and Klenow (2001) recover unmeasured quality growth in durables from the tendency of richer households to buy higher-quality goods. Both require expenditure shares or quantities, which catalog data lacks, so neither is feasible in our setting.

Our paper also relates to recent work using modern computational methods to construct price indexes from high-dimensional product data. Bajari et al. (2025) use text and image data from Amazon to

[4] Field (2003) challenges this interpretation by highlighting substantial productivity growth in the early 1940s, before the US entered the war. We find higher productivity for several decades before World War II.

estimate hedonic price indexes with deep learning methods. Ehrlich et al. (2026) compare hedonic and exact demand-based price indexes using large-scale item-level data with prices, quantities, and product attributes. Cafarella et al. (2023) use machine learning to construct hedonic price indexes in scanner data with substantial product turnover. We contribute by constructing price indexes from high-dimensional embeddings within a simpler ridge framework, which is better suited to the medium-sized datasets typical of historical price data, where richer models like neural networks are prone to overfitting.

Outline Section 2 describes how we construct our primary dataset of Sears catalogs. Section 3 presents a simple framework for constructing a cost-of-living index using hedonic methods and develops our empirical methodology. Section 4 presents our estimated price index for consumer goods and several robustness checks. Section 5 discusses the implications of our findings for measuring real GDP growth and productivity, and Section 6 concludes.

2 Data

Our primary data consist of 180 Sears general merchandise catalogs, published bi-annually, spanning 1900 to 1990. We also include the Sears Christmas books for 1945 to 1990, as well as the Sears Home Annual catalogs for 1986 to 1990. For these years, many appliances were included only in the Home Annual catalog, and so it is important to add these to our dataset.[5] We augment the Sears series with J.C. Penney catalogs (1969–1982) to broaden coverage and to assess robustness to retailer composition.

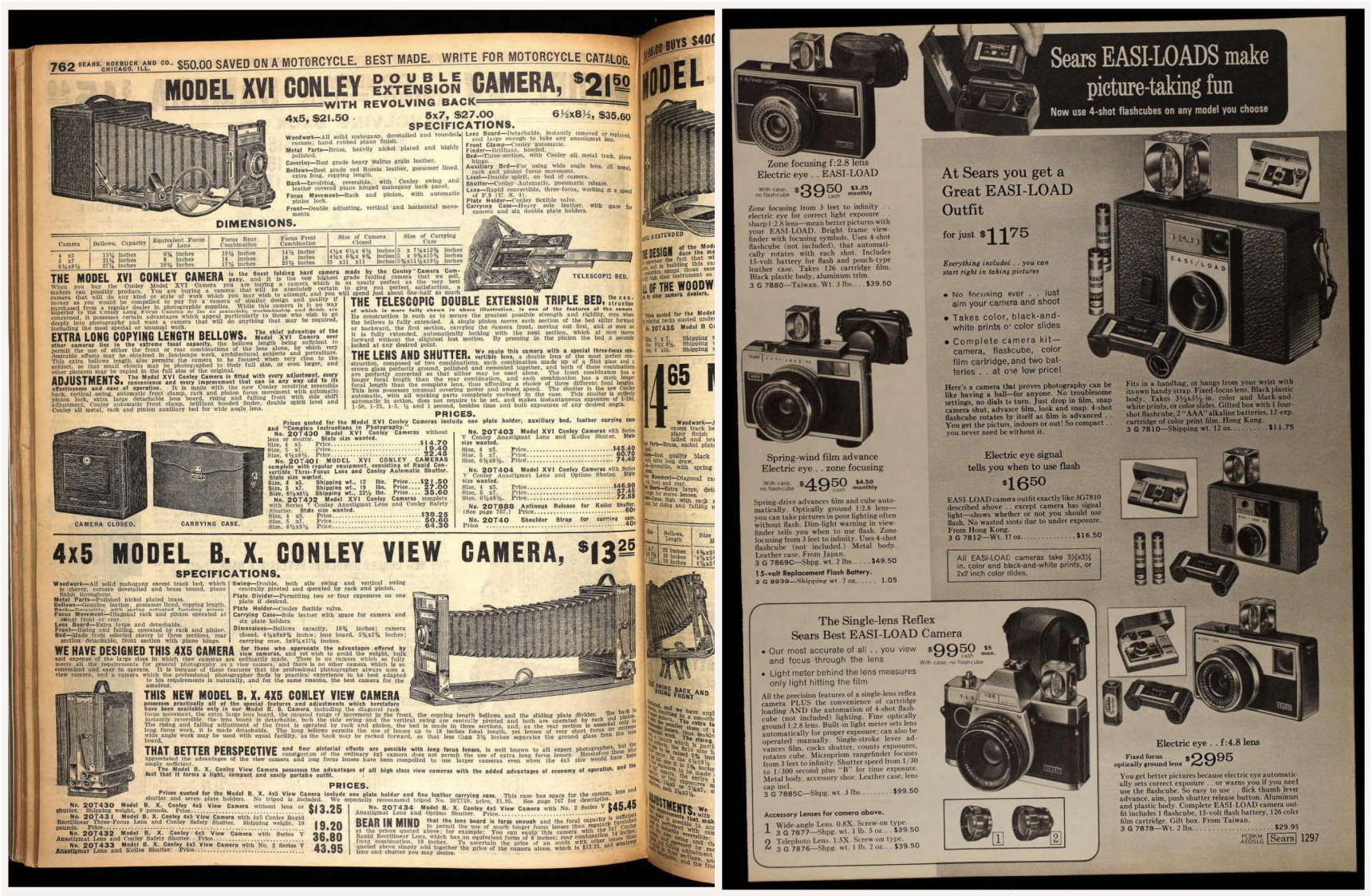
We process catalog data at the page level and extract structured product listings using a large language model (LLM) with a standardized schema, as in Bäcker-Peral et al. (2025); Asirvatham et al. (2026). Figure 1 shows an example of a page from the Sears catalog in 1910 and another from 1969, both featuring cameras. In both years, the listings are detailed, providing prices, model-level descriptions, and information on size, accessories, and other features of each distinct camera. To extract information from catalog images, we use Gemini 3 Pro, which is a state-of-the-art LLM that is highly effective at processing text from images. Each catalog contains roughly 1,000-1,500 pages, yielding 5,121,232 Sears product records.[6] For each listing we record: posted price, product title, textual description, shipping weight (when reported), brand (when reported), and a product image when present. We also ask the LLM to classify each product into a 1992 6-digit Harmonized System (HS) code, which is the classification used by the World Customs Organization.[7] The text field is particularly valuable: it contains detailed claims about materials, performance, dimensions, included accessories, and warranties, which are often precisely the characteristics that evolve as quality changes. In a second round of processing, we classify each product into a United Nations Standard Products and Services Code (UNSPSC) product category using OpenAI's GPT 5-mini. Relative to the HS classification, UNSPSC groups products by their purpose rather than material and is more detailed in its classification schema of consumer goods. In the remainder of the paper, if not otherwise specified, when we classify products into "categories" we use their UNSPSC classification. Finally, we ask the LLM to classify products which are accessories to the primary item being listed. For example, Sears often sold items like batteries or replacement parts on the same page as the item they were paired with. For our main analysis, we exclude these accessories. In Section A.4, we provide additional details on the digitization methodology.

[5] From the semiannual catalogs, we are missing the 1906 Fall and 1907 Spring catalogs, which we could not find in archival collections.

[6] For reference, the BLS collects about 80,000 quotes for goods and services per month. About 23% of these are for goods, excluding food and energy. Therefore, if the BLS had collected at this rate for the period of 1900-1990, we expect that the BLS collected about 20 million prices — about four times larger than our dataset. In practice, the BLS hand-collected a much smaller number of quotes in the pre-World War II period, so it is likely that our dataset is of approximately the same size or larger for these periods.

[7] There are 5,022 HS codes for 1992 recognized by the World Customs Organization. In our sample, we observe products spanning 3,430 HS codes.

Figure 1: Example Catalog Pages from 1910 and 1969



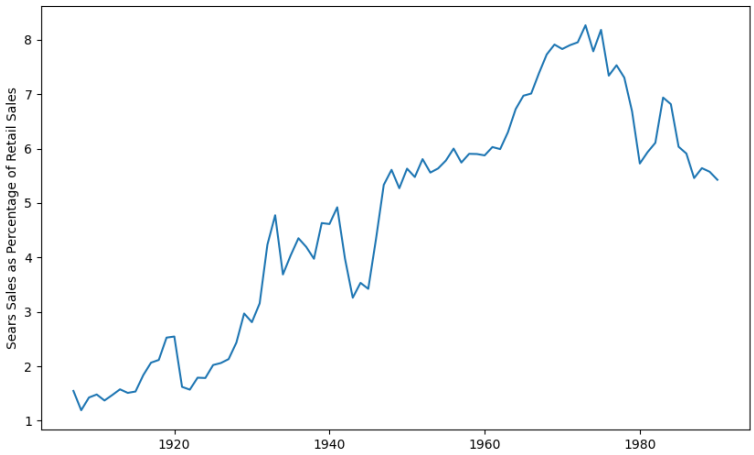
(a) 1910 **(b)** 1969

The figure shows two pages from the Sears catalogs. Panel (a) shows a page from 1910 and Panel (b) shows a page from 1969.

2.1 Historical importance of Sears

Sears was a major retail institution for much of the twentieth century, but it began as a mail-order business aimed at customers far from large urban retail centers. Richard W. Sears started by selling watches by mail in the mid-1880s and, after partnering with Alvah C. Roebuck and relocating the business to Chicago, expanded into a general merchandise catalog that offered an unusually broad variety of standardized, mass-produced goods at posted prices. In an era when many households—especially in rural areas—faced limited local selection, mail-order catalogs functioned as a central purchasing channel: consumers could shop from home and have goods delivered through the rapidly expanding national logistics system.[8] By the late 1970s, Sears reportedly printed on the order of hundreds of millions of catalog copies annually and served tens of millions of customers, with catalog sales in the billions of dollars. Figure 2 plots Sears retail sales as a share of U.S. retail sector gross output over our sample period. Sears' footprint rises from roughly 1.5 percent of retail sales in the early twentieth century to more than 6 percent for several decades following World War II, before declining in the 1980s. For reference, Amazon, Inc. accounted for 6.6% of US retail sales in 2024, so Sears can be thought of as "the Amazon of its day."

Figure 2: Sears Sales as a Percent of Retail Sector Output



The figure is a line plot of total Sears retail sales as a percent of US retail sector output. For the post-1947 period, we use retail sector gross output data from the BEA. Before this, we use retail sector value added from historical census data, and re-normalize it to match the BEA level for 1947-1957. This effectively assumes that the input-output share of retail was equal to the 1947-1957 level for the first half of the twentieth century. We exclude sales from Sears subsidiaries like Allstate Insurance.

The Sears catalog data do not span the full CPI consumption basket. In particular, Sears excludes housing, services, and food consumption, all of which are central CPI components.[9] For that reason, we interpret our results as evidence on inflation and quality change for *consumer goods*, rather than as a direct measure of the aggregate cost of living.

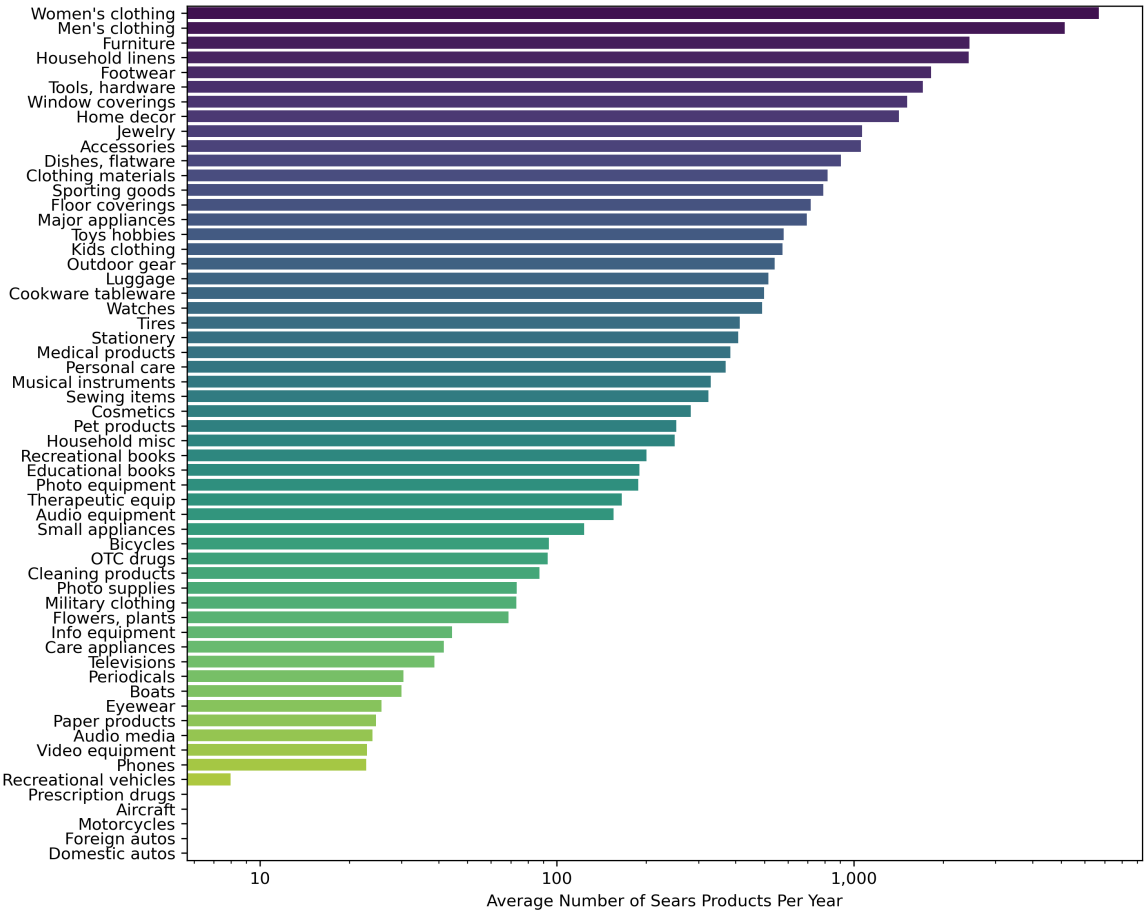
Within consumer goods, however, Sears covers many of the major categories that appear in official price indexes—apparel, household furnishings, recreational goods, appliances, and a wide range of durable and semi-durable items. Figure 3 shows the wide coverage of products in the Sears catalogs. We map products from the Sears catalogs to the 58 consumer goods categories from the BLS Personal

[8] Two infrastructure developments were particularly important for making this model scalable: the maturation of the railroad network (which lowered the cost of long-distance freight) and the introduction of Rural Free Delivery in the late 1890s (which made home delivery feasible for rural addresses).

[9] Sears did, however, sell mail-order houses from 1908-1942.

Consumption Expenditures (PCE) series and plot the average number of products that appear each year in the Sears catalog for each PCE category. All PCE categories are covered, with the exception of aircraft, motorcycles, automobiles, and prescription drugs. The figure indicates that the Sears catalog featured an especially large collection of apparel, furniture and house furnishings, but also covered a variety of distinct goods from aspirin to bicycles.

Figure 3: Sears Coverage of Consumer Goods



The figure shows a bar plot indicating the average number of products listed annually by Sears within each of the Personal Consumption Expenditure (PCE) categories for consumer goods. The x-axis uses a logarithmic scale.

How does Sears data compare to historical data used for price indexes? Modern price indexes use substantially more diverse and representative data than what can be obtained just from Sears catalogs, but historical data collection efforts were more limited. The official CPI was first published in 1921 and covered food, clothing, rent, fuel, house furnishings and a miscellaneous category, but had no coverage of services. According to the BLS, the official CPI was extended back to 1913 "using food prices only." Before 1913, there is no official CPI data, but the Minneapolis Federal Reserve Bank publishes a historical series that extends the CPI back to 1800 using a variety of data sources. For the period of 1890-1912, they use cost-of-living estimates from Rees and Jacobs (1961), which are based on catalog data, like our paper. However, rather than using the full sample of data, Rees and Jacobs (1961) select only one item for each of the ninety-three categories of consumer goods from a 1918 consumer goods survey. Then they track the price of each of these categories over time by selecting the most similar item in adjacent years. As a result, this index uses only ninety-three price quotes per year, which is a fraction of the full Sears catalog data. In contrast, we have thousands of price quotes in each year. Therefore, for at least

about a fifth of the century, official index data was also subject to important limitations.

3 Empirical framework

We are interested in how household welfare changed over the twentieth century. Let household h's utility in period t be

$$u_t^h = U^h(\mathbf{q}_t^h), \tag{3.1}$$

where $\mathbf{q}_t^h$ denotes its vector of consumption. Rather than compare utility levels directly, we can express welfare in dollar units using the expenditure function

$$e^h(\mathbf{p}, u) \equiv \min_{\mathbf{q}} \left\{ \mathbf{p} \cdot \mathbf{q} : U^h(\mathbf{q}) \geq u \right\}. \tag{3.2}$$

Let $x_t^h \equiv e^h(\mathbf{p}_t, u_t^h)$ be household h expenditures at time t. Expanding $\log e^h(\mathbf{p}_t, u)$ around u_{t-1}^h and re-arranging,

$$\log \frac{u_t^h}{u_{t-1}^h} \approx \frac{1}{\eta(\mathbf{p}_t, u_{t-1}^h)} \log \frac{e^h(\mathbf{p}_t, u_t^h)}{e^h(\mathbf{p}_t, u_{t-1}^h)} \tag{3.3}$$

where $\eta(\mathbf{p}_t, u) \equiv \frac{\partial \log e^h(\mathbf{p}_t, u)}{\partial \log u}$ is the elasticity of expenditure with respect to utility. Therefore, changes in utility are identified up to a normalization term η. We assume that preferences are homothetic, in which case we can set $\eta = 1$ and Equation (3.3) holds with equality, as shown in Section A.5.

Then, using the identity $x_t^h = e^h(\mathbf{p}_t, u_t^h)$, we can see that,

$$\log \frac{u_t^h}{u_{t-1}^h} = \log \frac{x_t^h}{x_{t-1}^h} - \log \frac{e^h(\mathbf{p}_t, u_{t-1}^h)}{e^h(\mathbf{p}_{t-1}, u_{t-1}^h)} \tag{3.4}$$

Aggregating over all households,

$$\sum_h \lambda_h \log \frac{u_t^h}{u_{t-1}^h} = \sum_h \lambda_h \log \frac{x_t^h}{x_{t-1}^h} - \sum_h \lambda_h \log \frac{e^h(\mathbf{p}_t, u_{t-1}^h)}{e^h(\mathbf{p}_{t-1}, u_{t-1}^h)} \tag{3.5}$$

where λ_h denotes the welfare weight assigned to household h.[10] Equation (3.5) clarifies the two objects that we need to estimate log changes in utility. The first term is a (geometric) average of changes in nominal consumption across all households, which, to a first order, can be approximated by aggregate changes in nominal consumption. The second is a cost-of-living index, in the sense of Konüs (1939), which captures the change in the cost of living between $t-1$ and t. We denote the cost-of-living index by,

$$COLI_t \equiv \sum_h \lambda_h \log \frac{e^h(\mathbf{p}_t, u_{t-1}^h)}{e^h(\mathbf{p}_{t-1}, u_{t-1}^h)}, \tag{3.6}$$

Intuitively, the $COLI_t$ describes how much more expensive it is today to be as well off as yesterday. This is the welfare object that we would ideally like to estimate.

In practice, the cost-of-living index is not directly observed because it depends on household utility functions, which are unknown. A standard approximation evaluates the cost of the base-period consumption bundle at current-period prices. Let $\mathbf{q}_{t-1}^h$ denote the vector of household h's base-period quantity of each good purchased, and let $\mathbf{p}_t$ denote the price vector of each good period t. The corresponding Laspeyres index is

$$L_t \equiv \sum_h \lambda_h \log \frac{\mathbf{q}_{t-1}^h \cdot \mathbf{p}_t}{\mathbf{q}_{t-1}^h \cdot \mathbf{p}_{t-1}} \tag{3.7}$$

[10] We use a log specification, as in Ehrlich et al. (2026), because it is far more empirically robust. However, log changes are to a first order equal to a more standard implementation using level changes.

By revealed preference, $COLI_t \leq L_t$ because the Laspeyres index holds the base-period bundle fixed and therefore ignores the possibility that households substitute away from goods whose relative prices rise. For small price changes, the same expression can also be interpreted as a first-order approximation to the expenditure function, since the envelope theorem implies that induced changes in optimal quantities are second order.

The empirical difficulty is that Equation (3.7) requires prices for the same goods in adjacent periods. This requirement is difficult to satisfy in settings with substantial product turnover. Products enter and exit the Sears catalog frequently, and continuing product categories often change in quality, style, materials, and other observable characteristics. As a result, the period-t price of a base-period product is often unobserved.

The solution used by the Bureau of Labor Statistics is called a matched-model methodology, which restricts attention to products observed in consecutive periods and computes price changes only for this matched set. Let Ω_t be the set of all products sold at time t and let $\mathcal{M}_t = \Omega_t \cap \Omega_{t-1}$ denote the set of products observed in both $t-1$ and t. A matched-model Laspeyres index takes the form

$$L_t^{MM} \equiv \sum_{i \in \mathcal{M}_t} s_{i,t-1} \log \frac{p_{it}}{p_{i,t-1}}, \tag{3.8}$$

where $s_{i,t-1}$ are the base-period expenditure weights for product i, given by,

$$s_{i,t-1} \equiv \sum_h \lambda_h s_{i,t-1}^h = \sum_h \lambda_h \frac{p_{i,t-1} q_{i,t-1}^h}{\sum_{\ell \in \Omega_{t-1}} p_{\ell,t-1} q_{\ell,t-1}^h}. \tag{3.9}$$

This procedure is transparent, but if products that appear in successive years are not representative of the true change in the cost of living, Equation (3.8) will be biased. Pakes (2003) emphasizes this problem in the context of personal computers, showing that matched-model indexes can sharply misstate quality-adjusted price change when product entry and exit are related to changes in quality-adjusted prices. Related work has similarly emphasized that matched-model methods can perform poorly in markets with rapid quality change, entry, and exit (Aizcorbe and Pho, 2005; Ehrlich et al., 2026). In Section A.3, we show that the matched-model methodology is biased in our sample as well.

A common solution to this problem is to use a hedonic regression to impute missing prices of entering or exiting goods. Formally, suppose that product i belongs to a product category $j(i) \in \mathcal{J}$ and is described by a vector of characteristics $X_{it} \in \mathcal{X}^{j(i)}$. In a competitive equilibrium, Rosen (1974) shows that prices will be a function of product characteristics,

$$p_{it} = P_{j(i),t}(X_{it}).$$

Suppose that we can estimate these price schedules, $\hat{P}_{jt}(\cdot)$, which we can use to infer the price of any product, given a vector of characteristics X. Then, we can construct a hedonic Laspeyres index,

$$\widehat{COLI}_t = \sum_i s_{i,t-1} \log \frac{\hat{P}_{j(i),t}(X_{i,t-1})}{P_{j(i),t-1}(X_{i,t-1})}, \tag{3.10}$$

which evaluates base period products at next period prices using the estimated hedonic price schedule. In Section A.5 we show that the hedonic Laspeyres index approximates the true COLI to a first order. Moreover, Pakes (2003) shows that under general conditions $\widehat{COLI}_t$ will be an upper bound of $COLI_t$, and therefore will underestimate total welfare changes over time. Intuitively, Equation (3.10) is a way to impute price changes for all goods $i \in \Omega_{t-1}$ even if they cannot be matched exactly to a good in the subsequent period. Therefore, the hedonic Laspeyres index avoids the selection problems of the matched-model approach.

To reiterate, the index in Equation (3.10) is a first order approximation of the object of interest. With richer quantity data, one could construct closer approximations to the cost of living. For example,

if product-level prices and quantities were observed in adjacent periods, one could use Törnqvist-style indexes that average expenditure shares across periods. Such indexes allow expenditure shares to respond to relative price changes and therefore provide a better approximation to substitution than a base-weighted Laspeyres index (Ehrlich et al., 2026).[11]

Our setting does not permit that approach. Sears catalogs provide rich information on product descriptions and posted prices, but not transaction quantities. The same limitation prevents us from implementing exact demand-based price indexes of the kind used in Broda and Weinstein (2006) and Redding and Weinstein (2020), which require stronger information on quantities and expenditure shares, and allow researchers to estimate gains from access to new product varieties. Our empirical strategy instead uses the comparative advantage of the Sears data: detailed product descriptions over a long historical period. Hedonic regressions allow us to decompose these long-run price changes into a quality change component and a cost-of-living change component. To the degree that households benefited from expansion in varieties over this period, in addition to changes in quality, our approach will therefore underestimate the rise in welfare.

3.1 Methodology

Equation (3.10) makes clear that constructing an approximation to the cost-of-living index requires two objects for each category of goods j and year t. The first is a hedonic price schedule, $\widehat{P}_{jt}(\cdot)$, that maps the observable characteristics of a good into a predicted price. This delivers the counterfactual numerator $\widehat{P}_{jt}(X_{ij,t-1})$, the price that a base-period good would command under year-t pricing. The second is a set of nationally representative base-period expenditure shares, $s_{i,t-1}$. This section shows how we construct each in turn.

3.1.1 Hedonic price schedules

To adjust for quality change, we must price a fixed bundle of product characteristics at two adjacent dates. This requires a flexible mapping $\widehat{P}_{jt}(\cdot)$ from characteristics to prices, estimated separately for each category and year.

The classic approach to this problem involves encoding a finite, manually chosen list of characteristics for each product category, which can be included on the right hand side of a hedonic regression (Gordon, 1990; Pakes, 2003). For refrigerators, for instance, Gordon (1990) includes total capacity, freezer capacity, defrost capability, and the presence of a crisper. We report results from this approach in Section 4.1, but it has two limitations that motivate our preferred method. First, hand-coding forces the researcher to pre-specify which attributes are price relevant. This introduces a researcher degree of freedom and risks omitting important characteristics, a concern that is especially acute for goods such as apparel and furniture, where dozens of styles and materials may bear on price. Second, the approach does not scale. Gordon (1990) reports that assembling his Sears durable-goods dataset for 1948–1988 took decades; replicating such an effort across the millions of listings in our sample would be prohibitive even with modern tools.

We instead construct flexible representations of product characteristics that better take advantage of our rich text data, as in Bajari et al. (2025) and Ehrlich et al. (2026). For each product i in category j and year t, we observe a textual description T_{ijt} which lists the product's features, materials, and dimensions. On average, descriptions are about 580 characters long. We map each description into a $d = 1{,}536$-dimensional vector,

$$\boldsymbol{\phi}_{ijt} = \phi(T_{ijt}) \in \mathbb{R}^d,$$

[11] Ehrlich et al. (2026) also implement a two-step hedonic regression as in Erickson and Pakes (2011) which combines hedonic methods with the matched-model approach. In our data, this is not feasible because we do not have barcode level identifiers and as such cannot track exact products over time.

using OpenAI's text-embedding-3-small model.[12]

The embedding provides a coordinate system for product characteristic space. In a conventional hedonic regression, the econometrician selects a small set of characteristics—capacity, horsepower, material—codes them by hand, and prices them. The embedding replaces the coding step: the language model maps descriptions with similar content to nearby points in $\mathbb{R}^d$, so any feature the catalog records—a porcelain tub, a five-year guarantee, a larger oven—shifts the description in a particular direction in embedding space. Individual coordinates of $\boldsymbol{\phi}_{ijt}$ carry no economic interpretation, but directions in the space do: products described identically receive identical vectors, and products that differ in a feature differ in a systematic way. Because Sears sold by mail, the catalog description is also the natural object to include in a hedonic regression — it contains precisely the information that buyers used to make consumption decisions.

For each category j and year t, we estimate the hedonic regression

$$\log P_{ijt} = \alpha_{jt} + \boldsymbol{\phi}'_{ijt}\beta_{jt} + Z'_{ijt}\beta^Z_{jt} + \varepsilon_{ijt}, \tag{3.11}$$

where α_{jt} is a category-year intercept, β_{jt} maps the embedding coordinates into log price, and β^Z_{jt} loads on the structured covariates Z_{ijt}. The first covariate we include is shipping weight because it is observed for most of the observations in our sample and is a powerful determinant of price. Shipping weight is reported for 80% of all goods, and is overall the most common attribute reported in Sears catalogs, since it was used to calculate shipping costs. For the 20% of goods where it is not reported, it is interpolated using a procedure described in Section A.7. We also include season dummies indicating whether the product appeared in the Spring, Fall, or Christmas catalog.

The estimated schedule lets us price any base-period product under year-t pricing. For a product i observed at $t-1$ with characteristics $X_{ij,t-1} \equiv (T_{ij,t-1}, Z_{ij,t-1})$, the counterfactual log price is

$$\log \widehat{P}_{jt}(X_{ij,t-1}) = \widehat{\alpha}_{jt} + \phi(T_{ij,t-1})'\widehat{\beta}_{jt} + Z'_{ij,t-1}\widehat{\beta}^Z_{jt}. \tag{3.12}$$

This object is exactly the numerator in (3.10): it holds the good's characteristics fixed at their base-period values and revalues them at the year-t schedule. The corresponding denominator, $P_{j,t-1}(X_{ij,t-1})$, is the good's price in the base period.

A practical obstacle is that the embedding is high dimensional and its coordinates are dense and strongly correlated. We therefore estimate (3.11) by partially penalized Ridge regression, penalizing only the embedding coefficients β_{jt} and leaving the intercept α_{jt} and the covariate coefficients β^Z_{jt} unpenalized. We describe this procedure in detail in Section A.8. Our methodology is similar to Bajari et al. (2025), but in their paper, the authors use a much more flexible neural network model with big data from Amazon. In our dataset, the median product category has around 50 products per year, so even if we pooled across multiple years, a neural network will be prone to overfitting; Ridge is much more stable. Our estimates are robust to instead extracting the principal components of the embedding matrix and estimating (3.11) by OLS, as shown in Figure A.2.

3.1.2 Expenditure weights

The index in Equation (3.10) weights each product-level price change by its base-period expenditure share, $s_{i,t-1}$. In practice, these shares are unobserved in our data. We observe catalog prices and characteristics, but, with rare exceptions, we observe neither the quantities Sears sold nor—what the index ultimately requires—national expenditure at the level of individual products.

Our strategy is to recover $s_{i,t-1}$ in two steps, relying on nationally representative survey data wherever it exists and on the catalog-based weights only where it does not. National consumer surveys report representative expenditure weights, but only at a moderate level of aggregation. The Bureau of

[12] We remove prices from the text description, so that embeddings do not include price information.

Labor Statistics (BLS) organizes spending into four major retail categories, indexed by k—household goods, apparel, entertainment, and other—and roughly 60 subcategories, indexed by ℓ. We map each subcategory ℓ to one or more UNSPSC product codes and assign every Sears listing to a subcategory through its code. We take the expenditure weight on each subcategory, $w^{\text{BLS}}_{\ell,t-1}$, directly from the national source. These weights are nationally representative by construction and require no assumption about Sears.

The national source varies over the century. For the postwar period we use historical CPI weights published by the BLS for 1947-1959, and Personal Consumption Expenditure (PCE) weights starting in 1959, which are available on an annual basis thereafter. Before World War II, we rely on two surviving national consumer-spending surveys: the *Cost of Living in the United States* survey (fielded 1917–1919) and the *Study of Consumer Purchases in the United States* (fielded 1935–1936). We apply the resulting weights to four periods: the *Cost of Living* weights for 1900–1926, the *Study of Consumer Purchases* weights for 1927–1943, the 1947 BLS weights for 1943–1947, and contemporaneous BLS CPI weights thereafter.[13]

National sources do not resolve spending below the subcategory level, so we cannot read product-level shares from any representative survey. We instead use the catalog itself, under the assumption that within a subcategory, expenditure is allocated proportionally across all catalog products in that subcategory. Formally,

$$s_{i,t-1} = \frac{1}{N_{\ell(i),t-1}} \times w^{\text{BLS}}_{\ell(i),t-1} \tag{3.13}$$

where $\ell(i)$ denotes the subcategory to which product i is mapped, $N_{\ell(i),t-1}$ is the number of catalog products in subcategory ℓ at $t-1$.

This construction rests on the assumption that within a subcategory, the Sears product mix is representative of what households actually purchased. This assumption is not directly testable at the product level, but we can provide suggestive evidence in its favor.

Evidence on representativeness. To assess the assumption that *within* BLS categories Sears is representative of national consumption, we check whether it aligns with the BLS expenditure weights *across* categories. We define the Sears share of subcategory ℓ within major category k as its share of distinct listings,

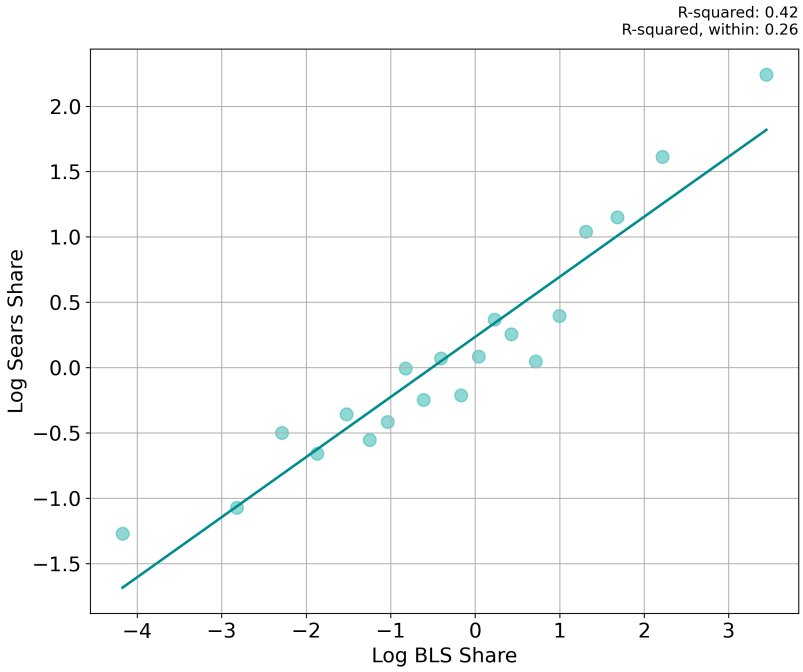
$$\text{Sears Share}_{\ell,k,t} = \frac{\text{Num Products}_{\ell,k,t}}{\sum_{\ell'} \text{Num Products}_{\ell',k,t}}. \tag{3.14}$$

Analogously, we define Sears $\text{Share}_{k,t}$ for each major category. At the level of major categories, the catalog is *not* representative: Figure A.1 shows that Sears over-weights apparel and household goods relative to the BLS. This is exactly why we anchor cross-category weights to the BLS rather than to the catalog. Within major categories, however, the catalog tracks national spending well. Figure 4 plots log Sears $\text{Share}_{\ell,k,t}$ against log BLS $\text{Share}_{\ell,k,t}$, controlling for category k and year fixed effects; the strong positive relationship indicates that, conditional on a major category, the internal composition of the catalog mirrors national expenditure patterns. This supports the assumption that within BLS subcategory, the products that are listed more frequently in Sears will also be those that had higher national expenditure shares.[14]

[13] Because nationally representative weights exist only at scattered benchmark years, we hold the weights fixed within each period, which is consistent with the methodology used by the BLS.

[14] Ideally, we would want the slope in Figure 4 to be one, such that products with higher expenditures have proportional weights in the Sears catalog. Because the slope is less than one, these results suggest that we will under-weight more popular goods.

Figure 4: Binscatter of Sears Product Shares vs. CPI Subcategory Weights



Binscatter with 100 bins. The horizontal axis is $\log \text{Sears Share}_{\ell,k,t}$, the share of subcategory ℓ within major category k in the Sears catalog in year t, defined in (3.14). The vertical axis is the corresponding $\log \text{BLS Share}_{\ell,k,t}$ reported by the BLS. Both axes residualize on category k and year t fixed effects. We restrict to years in which BLS surveys were fielded.

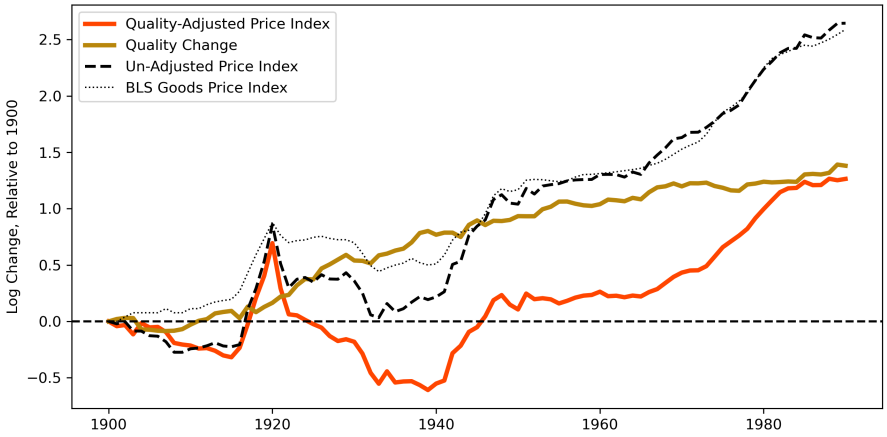
4 The cost of living: 1900 to 1990

Using this procedure, we estimate a hedonic price schedule for each UNSPSC category j in each period t. We then use these schedules to construct the consumer-goods cost-of-living index in Equation (3.10) for each year in the data.[15] We aggregate over consumer goods categories using nationally representative consumer expenditure weights to produce a single cost-of-living index for consumer goods. The resulting index is shown by the solid red line in Figure 5. The index declines weakly during the first two decades of the sample. It rises sharply during World War I, falls back rapidly afterward, declines substantially during the Great Depression, and then rises again during World War II. Inflation is effectively zero in the postwar decades before increasing again in the late 1960s and 1970s.

The dashed black line shows average unadjusted price changes for the same sample of consumer goods. The unadjusted price index is constructed using average price changes within product categories, and therefore does not control for changes in quality. These unadjusted price increases are much larger than the quality-adjusted increases. The gap between the unadjusted and quality-adjusted series captures the cumulative value of quality improvements. When unadjusted prices rise faster than quality-adjusted prices, the implied average quality of goods is increasing. Over the full 90-year period, these quality

[15] In the aggregation, we exclude the top and bottom 1% of observations based on predicted quality-adjusted inflation, $\pi_{ijt} \equiv \log \widehat{P}_{jt}(X_{ij,t-1}) - \log P_{ij,t-1}$. We restrict attention to consumer-goods categories with at least 20 items in the year of estimation. We exclude food because Sears sold food only in the pre-WWII period and primarily sold packaged foods, which are not representative of the broader food category. We also drop 0.75% of observations for which the hedonic regression in Equation (3.11) has an R^2 below 0.5. We also drop 1907, because we only have a fall catalog for 1907 and a spring catalog for 1906, and some goods like apparel vary a lot between these two seasons and therefore cannot be compared.

Figure 5: Cost-Of-Living Index For Consumer Goods, 1900-1990



The figure shows our estimates of the cost-of-living index for consumer goods in the solid red line. The dashed black line shows average cumulative price changes for the same sample of consumer goods. The difference between the black and red line is plotted in gold and represents cumulative quality change of consumer goods. The dotted black line shows estimates of the price index for consumer goods published by the BLS. Before 1935 there is no consumer goods index so we use the national CPI.

adjustments are quantitatively large, accounting for about half of the total increase in unadjusted prices.

Finally, the dotted black line shows our best available estimate of the aggregate consumer goods price level from existing sources. For 1900–1935, we use the Minneapolis Fed historical CPI series, which draws on several sources but does not purport to be comprehensive.[16] For the post-1935 period, we use BLS price indexes for durable and nondurable goods, aggregated using Laspeyres expenditure shares to form a consumer-goods price index.[17] This BLS-based series tracks the unadjusted price index closely and lies well above the quality-adjusted index. As a result, deflating nominal output using the existing aggregate price index would understate real GDP growth.

4.1 Robustness

Our consumer goods cost-of-living index relies on several key assumptions: (1) we can effectively estimate the hedonic price schedule, $P_{jt}(\cdot)$, (2) the set of underlying hedonic characteristics X_j is approximately constant year-over-year for each product category, and (3) the Sears sample is representative of national goods consumption. We will now provide some evidence to validate these assumptions in our setting.

4.1.1 Robustness to Functional Form

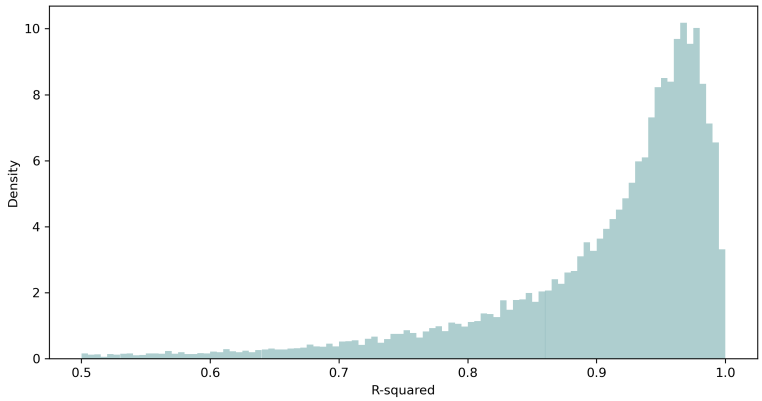
Our estimating equation, Equation (3.11), imposes some structure on the form of the hedonic price schedule in the sense that embedding characteristics enter linearly, which we may worry is misspecified. To test this, we calculate an out-of-sample R^2 for each hedonic price schedule using five-fold cross-validation: we split the sample into five folds, repeatedly estimate the hedonic regression on four folds, predict prices for the omitted fold, and compute R^2 from the pooled out-of-sample prediction errors across the five omitted folds. The distribution of R^2 is shown in Figure 6 and shows that out-of-sample

[16] The 1900-1913 CPI is based on Albert Rees' work. The 1913-1935 CPI is estimated by the BLS and primarily collected price quotes for consumer goods categories such as food, clothing, rent, fuel and house furnishings.

[17] The BLS methodology differs from ours since it tracks matched models over time. Systematic quality adjustment was only introduced in the 1980s for a subset of products, and it is used only when exact models cannot be matched. This methodology is explained further in Section A.3.

predictive power of the flexible hedonic model is quite high with a median R^2 of 0.94. Therefore, it is likely that the embeddings are picking up on most price-relevant attributes.

Figure 6: Distribution of R^2 For Hedonic Schedule



The figure shows a histogram of the distribution of R^2 across all of our estimated hedonic regressions from Equation (3.11). The median R^2 is 0.94.

As a baseline, we compare our methodology to traditional hedonic methods that rely on hand-coded attributes. For 28 categories of technology-intensive goods—including refrigerators, washing machines, cameras, and televisions—we construct a detailed list of candidate attributes (about 50 per category) and extract them from the catalog text. The attributes include numeric measures (e.g., capacity, horsepower, screen size, number of vacuum tubes) and binary indicators (e.g., color vs. black-and-white television, automatic defrost, flash included). The goal is to recover a structured representation of quality that is comparable within categories over moderate time spans. In Section A.6, we explain in detail our methodology to extract structured product attributes from text data.

We choose a standard log-linear parameterization for the hedonic price schedule,

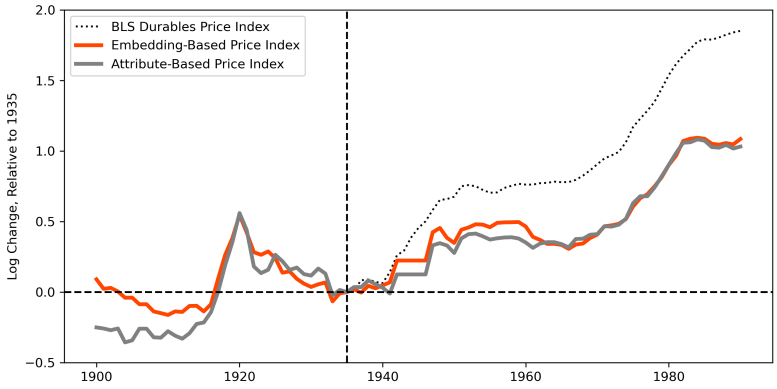
$$\log P_{ijt} = \alpha_{jt} + X'_{ijt}\beta_{jt} + \varepsilon_{ijt} \tag{4.1}$$

where X_{ijt} is the vector of attributes and β_{jt} are the attribute coefficients for category j in year t. A practical issue is that the relevant attribute set changes over time so not all of our attributes will be relevant in every year. To choose a parsimonious attribute set we use LASSO (Least Absolute Shrinkage and Selection Operator).

We can now compare the implied quality-adjusted inflation estimates produced by this method with our embedding-based estimates. In Figure 7, the red line shows estimates of the technological goods cost-of-living index constructed using the embedding-based regression, Equation (3.11). The gray line shows estimates using the attribute-based regression, Equation (4.1). Both series are normalized relative to 1935. Remarkably, the two show very similar patterns, suggesting that the high-dimensional embedding attributes are picking up on the same features that a human may have selected as attributes. Indeed, if we plot a binscatter of the embedding-based price index for each product j in each year against the attribute-based price index, the two exhibit a very strong correlation, shown in Figure A.5. Moreover, both methods have very high out-of-sample R^2 for the subset of technological goods, with a median of 0.94 for both the attribute-based method and the embedding-based method.

One concern about the embedding-based method, relative to the attribute-based method, is that it may pick up on advertisement strategies in addition to welfare-relevant characteristics. To test for this, we produce a new set of embeddings, excluding from the product description a wide selection of advertising terms (such as “best offer” or “brand new”) as well as subjective attributes such as “beautiful”,

Figure 7: Cost-Of-Living Index for Technological Goods, Embeddings vs. Attribute Method



The figure shows our quality-adjusted price indexes for a subset of 28 technology-intensive durable goods. The red line shows the quality-adjusted price index using our preferred embedding-based methodology described in Section 3.1. The grey line shows the quality-adjusted price index using the attribute-based methodology. Both series are normalized to zero in 1935. The dotted black line shows the BLS official price index for durable goods.

"wonderful" or "luxurious". While this term-based approach may not strip all advertisement from the embedding, it will help inform as to the direction and size of bias caused by including marketing language. The new series, constructed with the updated embeddings, is shown in the red line in Figure A.6. Reassuringly, the two series are very close, and the series without marketing language actually implies even lower quality-adjusted inflation than the baseline series. Therefore, this suggests that implied quality improvement is driven by true product upgrading as opposed to changes in language and marketing strategies.

4.1.2 Robustness to Changing Characteristics

A key advantage of a hedonic regression, relative to looking only at price changes of the subset of goods which appear in consecutive years, is that we can study implicit price dynamics of *all goods*, not just surviving goods. However, the framework we presented in Section 3 implicitly assumes that even if the set of goods available changes from one year to the next, the underlying set of characteristics of the goods available at each time $t-1$ continue to exist at time t, such that the time t hedonic schedule can be used to predict prices for goods in the previous year. This assumption is, of course, imperfect. There are many characteristics that disappeared at discrete points in our sample to make way for new product characteristics, from wringers on washing machines to girdles on dresses.

As a robustness exercise, we compute the cost-of-living index in Equation (3.10), excluding products that are very different from all of the products in the following year (i.e. exiting goods). Empirically, we implement this using our text embeddings, ϕ_i. For any pair of goods in category j, denoted i and i', the cosine similarity of the product is given by,

$$sim(i,i') = \frac{\phi_i \cdot \phi_{i'}}{||\phi_i||\,||\phi_{i'}||}$$

For each product i at time t, we construct a measure of its novelty relative to time $t+1$ as,

$$\sigma_{i,t+1} = \max_{i' \in \mathcal{I}_{j,t+1}} sim(i,i')$$

The novelty variable $\sigma_{i,t+1}$ captures how similar product i is to its most similar alternative at time $t+1$. Similarly, we can compute $\sigma_{i,t-1}$ as similarity to products in the previous period. We then compute

the cost-of-living index, excluding all products for which $\sigma_{i,t+1} < 0.7$. This excludes 11.1% of products which were significantly different from any product in the next year. We also compute a version in which we exclude all products for which $\sigma_{i,t-1} < 0.7$ or $\sigma_{i,t+1} < 0.7$, such that only "stable" goods are included. The new cost-of-living index using each of these samples is shown in the solid grey and red lines, respectively, in Figure A.7. Our baseline estimates are shown in the dashed line. The three series are very similar over the entire period, with slightly lower inflation implied when we exclude exiting goods, suggesting that our main conclusions are not driven by new or dying goods.

4.1.3 Robustness to Sears Sample

To construct a cost-of-living index, we need a representative sample of all goods and services purchased by Americans in each year. Sears was clearly not representative of many sectors of the economy, including services, food and automobiles, which we discuss further in Section 5.2, but our analysis thus far has assumed that it was broadly representative of a large selection of consumer goods including apparel, furniture, appliances, toys and tools. According to an internal document released in 1978, Sears described itself as follows: "We are not a fashion store; we are not a store for the whimsical nor the affluent. Sears is a family store for middleclass, homeowning Americans." To assess the representativeness of Sears, we supplement our data collection with J.C. Penney catalogs for the period of 1969–1982. J.C. Penney targeted a somewhat higher end of the market, and therefore will give us a sense of whether different tiers of the retail sectors saw very different movements in prices. We construct flexible quality-adjusted price indexes and un-adjusted price indexes using J.C. Penney data and plot both indexes in Figure A.8. We normalize the J.C. Penney index to the Sears index level in 1969. For the period of overlap, the indexes are consistent, indicating high inflation in the 70s. This is suggestive that the trends in Sears were comparable to the broader US retail sector.

4.2 What accounts for quality growth?

A limitation of our high-dimensional embedding approach is that it is difficult to interpret the drivers of quality improvement. In this section, we aim to partially open this black box by isolating a few important components of quality change, and providing examples of cases where quality change is subtle and difficult to recover fully using traditional methods.

One component of quality that we can measure over time is improvements in materials. In particular, for a large selection of materials, including textiles like rayon and cotton, as well as metals like copper or steel, we identify whether a product contains the material via keyword search. We then estimate a hedonic regression like Equation (3.11) but controlling only for materials and shipping weight.[18] The implied quality improvement driven only by changes in materials and physical size, respectively, are shown in Figure A.3. We can see that about a third of quality improvements are accounted for by changes in materials.[19]

One example in which materials innovation was very important was in domestic ranges and cooktops. The highest quality ranges sold by Sears in the early 1900s were made of cast iron designed for coal or wood fuel. Over the next decades, porcelain enamel was popularized, since it was easier to clean and more resistant to rust. In the early 1920s, ovens also introduced asbestos lining which provided thermal insulation, although this was phased out in post-World War II period, as the dangers of asbestos were discovered. After the war, the most significant materials changes included the use of chromium accents instead of nickel, and the use of glass for cooktops and oven doors, which was previously not possible.

[18] We select the relevant materials for each product-year pair via LASSO.

[19] We must be careful about interpreting these estimates causally; materials may be correlated with other features, which will lead us to overstate the impact of materials changes on total quality change. This is a benefit of our embedding based approach, which captures many aspects of quality in a single specification.

Our materials dummy variables account for about 60% quality improvement from 1900–1930 in domestic ranges, excluding movements driven by pure changes in size, and even more in the postwar period.

Figure 8: Women's Dresses, 1922–1928



(a) 1922 **(b)** 1925 **(c)** 1928

There are clear cases, however, where materials do not account for the full story. Consider, for instance, the example of women's dresses, which experienced huge quality improvements in the first half of the century. Average prices of women's dresses were approximately constant from 1922 to the onset of the Great Depression. Figure 8 shows images of three dresses targeted toward "stout" women, listed for $5.98 in the spring edition of the 1922, 1925, and 1928 Sears catalogs. In 1922, $5.98 could buy a plaid gingham dress that emphasized its good quality and durability. By 1925, the same amount could garner a more intricate silk and wool dress with embroidered sleeves. The equivalently priced dress in 1928 was made entirely of silk and had an even more ornate cut and pattern. Materials improvements, for instance with the proliferation of silk, explain a portion of quality improvement in women's dresses in the 1920s, but a much larger share is driven by changes in style and detail, which are captured by our embedding approach. Therefore, our quality-adjusted price index finds that quality-adjusted prices fell

substantially over this period, even if the average sticker price remained constant.

Technological goods offer another case in which quality improvements can be decomposed into interpretable components. As discussed in Section 4.1.1, for a subset of 28 technology-intensive goods we extract detailed product attributes that capture most dimensions of quality change. One key dimension during this period was the arrival of electricity. In the 1910s, domestic refrigerators were cooled by hand-delivered blocks of ice and washing machines were turned by a hand crank (see Figure A.9); electrification transformed both. Figure A.10 traces how electricity diffused into each. Electricity was adopted more quickly for washing machines and by the 1930s, about half of Sears listed washing machines had an electric motor. It is not until after the war, however, that features common to washing machines today, like water level and temperature control, were introduced. Electric refrigerators appeared later than washing machines but diffused rapidly; by 1939, nearly three-quarters of listed models were electric and included a freezer. Here too, the postwar period brought major refinements, including automatic defrosting, water dispensers, and other features. The figure thus highlights substantial quality improvement both before and after World War II, much of which official price statistics may have failed to capture. These quality improvements had important economic consequences; Greenwood et al. (2005) argue that the diffusion of appliances like these, by sharply reducing the time required for laundry, cooking, and cleaning, helped drive the rise in married women's labor force participation over the twentieth century.

Quality improvements were almost uniformly positive for the period that we study, with one important exception: the Great Depression. Economic theory predicts that during periods of severe economic downturn, households will substitute away from high-quality luxury goods towards lower quality alternatives (Fajgelbaum et al., 2011; Jaimovich et al., 2019). This is exactly what we observe. Between 1929 and 1932, our estimates imply that the quality of consumer goods fell by 7 percent, which is the largest sustained quality decline in our sample. In Figure A.11, we show some of the margins along which these quality declines were implemented. Panel (a) shows the use of silk in apparel items over the century. Silk was common before the Great Depression, and was used in about half of women's apparel items in the late 1920s. The use of silk plummeted throughout the entirety of the depression.[20] Panel (b) shows a similar pattern with curtain lengths. Before the Great Depression, curtains were typically 80 to 100 inches long — effectively floor-to-ceiling length. During the depression, average curtain length fell to about 50 inches, just enough to cover the typical US window. Finally, Panel (c) shows the prevalence of electricity among several technological goods including ranges, heaters and refrigerators. Electricity was on the rise when the depression hit, but plateaued from 1929 to 1939. It then plummeted during the war years, when many technological goods had production quotas, before taking off again in the post-war period. All three examples show that the Great Depression appears to have had a large effect on the composition of goods that consumers purchased and the rate of quality adjustment.

5 Implications for real growth

Table 1 summarizes the average annual inflation rates implied by several price indexes. The first four rows report CPI-based estimates: the national CPI, the BLS durable-goods index, the BLS nondurable-goods index, and the implied BLS price index for all consumer goods. In the postwar period, measured inflation was lower for goods than for the overall consumption basket, and especially low for durable goods. The fifth row reports our preferred quality-adjusted inflation estimate for consumer goods, constructed using the embedding-based hedonic methodology described in Section 3. The final row reports the difference between the best estimate of consumer goods inflation based on BLS data and our estimates.

The comparison shows that our methodology implies substantially lower inflation for consumer goods than existing BLS-based measures. In the postwar period, the average annual bias is 0.96pp for

[20] This trend is not specific to Sears. Historical reports on foreign imports show that silk imports fell substantially during the Great Depression and stayed low after World War II

1945–1990. These gaps suggest that measured goods inflation may have overstated true quality-adjusted price growth in periods of rapid product improvement. The differences are even larger in the prewar period, when the official CPI was less comprehensive and historical price measurement is necessarily more limited; average annual bias for this period is 3.1. These differences have large implications for measured real growth in consumer goods. Using the best available BLS estimates for consumer goods, real goods consumption grew by a factor of 10.3 between 1900 and 1990. Using our quality-adjusted goods price index instead, real goods consumption grew by a factor of 39.

Table 1: Price Inflation by Decade

	Series	1900s	1910s	1920s	1930s	1940s	1950s	1960s	1970s	1980s
1	CPI	0.80	7.00	0.11	-1.97	5.62	2.07	2.33	7.10	5.55
2	Durables				1.86	6.13	1.12	0.97	6.01	4.10
3	Nondurables				-0.35	7.12	1.73	2.13	7.14	4.42
4	Goods				0.06	6.90	1.57	1.80	6.76	4.27
5	Our Method	-1.89	6.87	-4.14	-4.14	8.23	1.00	1.66	5.40	3.49
6	BLS vs. Our Method	2.69	0.13	4.25	2.04	-1.33	0.57	0.15	1.36	0.78

The table reports average annual inflation by decade from the 1900s through the 1980s. The first four rows report BLS-based price indexes: the national CPI, durable goods, nondurable goods, and all consumer goods. The fifth row reports our quality-adjusted goods inflation estimate, constructed using the embedding-based hedonic methodology in Equation (3.11). The final row reports the average annual difference between our estimates and the BLS, where we use the CPI goods index when available and the national CPI otherwise.

It is worth observing that our estimates imply higher inflation and thus lower welfare gains than the BLS data in one decade: the 1940s. The 1940s price changes are dominated by wartime inflation, which accelerated quickly in the first years after the US entered WWII. We find steeper inflation in goods than existing BLS accounts, which we believe is at least partially explained by the fact that the BLS uses a matched-model methodology, which means that they attempt to track the *exact same* product over time. In Section A.3 we show that this approach can substantially underestimate inflation during periods of rapid price increases. During inflationary periods, firms are much more likely to shift to new products which face less stickiness than older product versions, and thus matched-model price indexes are particularly downward biased.

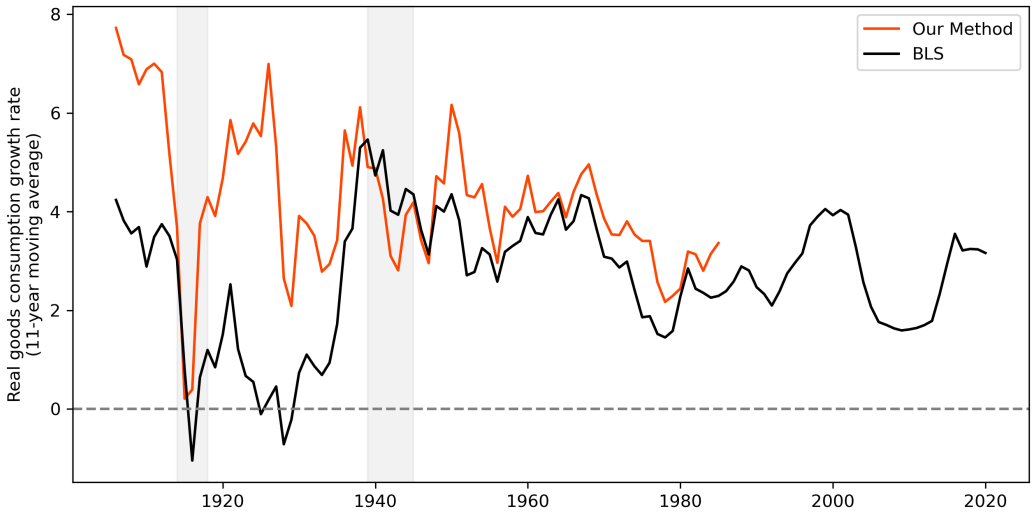
5.1 Re-evaluating trends in consumption growth

Our cost-of-living estimates reshape not only the estimated level of welfare gains over the twentieth century but also the timing of consumption growth. A central theme in the long-run growth literature is that the decades between 1870 and 1970 were exceptional relative to both earlier and subsequent periods (Gordon, 1999; Abramovitz and David, 2000; Gordon, 2016). Qualitative accounts emphasize the breadth of technological change over this entire interval. Gordon (2016), for example, highlights the diffusion of electricity, indoor plumbing, central heating, refrigeration, automobiles, telephones, radio, motion pictures, antibiotics, and modern household appliances as innovations that transformed everyday life. Many of these changes took place well before World War II; for example, by 1940, nearly 80 percent of housing units in the U.S. had electric lighting.

Quantitative accounts based on productivity data typically place the peak of U.S. growth in the post-World War II decades. For example, Gordon (1999) finds that the "'winner and still champion' interval" was the period of 1950–1964, which saw annual productivity growth of 2.35 percent. National accounts data on real consumer goods growth, shown in the black line Figure 9, align with this result, showing slower growth rates before World War II compared to the decades after. However, Gordon

(1999) caveats that "consideration of data gaps points to the history of price changes from 1914 to 1947 as the black hole where little is known," so it is therefore difficult to draw conclusions about this period based on official data. Our paper aims to shed light on this black hole. We find that consumer goods growth was vastly understated for most of the first half of the twentieth century. Existing estimates imply that consumption grew more slowly before the war than after; our estimates reverse this ranking, with growth in 1900–1939 roughly 2.5 percentage points higher than in 1946–1990. Our estimates also imply faster growth than previously measured in the late 1940s and 1950s, a period of rapid quality improvement in consumer durables. Thereafter, the gains from quality adjustment remain positive but shrink. Qualitatively, our estimates support existing accounts of the postwar growth slowdown, but extend the pattern to the full century: rather than peaking in the postwar decades, consumption growth has been declining since the start of the century.

Figure 9: Real Goods Consumption Growth



The figure shows an 11-year moving average of real goods consumption growth since 1900 deflated using our consumer goods price index in red, and using the BLS consumer goods price index in black. We use BEA data on nominal goods consumption starting in 1929 and extend it backwards using data from Carter et al. (2006).

5.2 Impact on total expenditures

The estimates above imply large revisions to real growth in consumer goods. To assess their implications for aggregate real expenditure, however, we must also account for services, food and housing. For food prices, we use the CPI food index (U.S. city average), which begins in 1913, and extend it backwards using food wholesale prices from the U.S. Historical Statistics bulletin published by the Census Bureau. For housing rental prices, we use a novel historical dataset from Lyons et al. (2026), which uses historical newspapers to assess rental prices from 1890 to 2006. The remaining challenge is to construct a price index for services over the twentieth century. For the post-1935 period, we use the available BLS service price estimates. In doing so, we assume that these BLS service indexes correctly account for quality change in services. Under this assumption, nominal service expenditures deflated by the BLS service price index provide an estimate of real service expenditure. Thus, the exercise isolates the implications of our revised goods price index while treating measured service inflation as correctly quality-adjusted.

The main limitation is that the BLS collected essentially no data on the price of services before 1935. We therefore extrapolate pre-1935 service inflation using the relationship between relative prices and relative employment implied by a simple two-sector model. This procedure is necessarily imperfect: it does not replace direct measurement of service prices. Rather, it provides a disciplined way to extend the

available service price series backward using observable employment data and the postwar relationship between relative employment and relative prices. Given the absence of direct historical service price measurement, this extrapolation is the best feasible approach for constructing an aggregate expenditure measure over the full period.

Formally, suppose households have CES preferences over goods consumption c_g and services consumption c_s:

$$U(c_g, c_s) = \left[\alpha c_g^{\frac{\sigma-1}{\sigma}} + (1-\alpha) c_s^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}},$$

where σ is the elasticity of substitution between goods and services. Production in the two sectors is linear in homogeneous labor:

$$y_g = A_g \ell_g, \qquad y_s = A_s \ell_s,$$

with total labor $\ell_g + \ell_s = L$. The terms A_g and A_s capture sectoral productivity, broadly interpreted to include quality-adjusted productivity. With competitive firms, prices equal marginal costs:

$$p_g = \frac{w}{A_g}, \qquad p_s = \frac{w}{A_s}.$$

Hence relative prices are inversely related to relative productivity:

$$\frac{p_g}{p_s} = \frac{A_s}{A_g}.$$

This is the standard mechanism from Baumol (1967): if productivity grows faster in goods than in services, then services become relatively more expensive.

CES demand implies

$$\frac{c_g}{c_s} = \frac{\alpha}{1-\alpha}\left(\frac{p_g}{p_s}\right)^{-\sigma}.$$

Using market clearing, $c_g = y_g = A_g \ell_g$ and $c_s = y_s = A_s \ell_s$. Combining the demand equation with the relative-price condition yields

$$\log \frac{\ell_s}{\ell_g} = \kappa + (1-\sigma) \log \frac{p_s}{p_g}, \tag{5.1}$$

where κ is a constant determined by preference parameters. Equation (5.1) links the relative employment share of services to the relative price of services. If goods and services are close substitutes, then a rise in goods-sector productivity lowers goods prices and induces households to substitute toward goods, limiting the growth of service employment. If instead goods and services are complements, or poor substitutes, then the same relative productivity growth raises the relative price of services while also increasing the service share of employment. The latter pattern is the classic Baumol prediction.

Figure A.12 shows this relationship in the data. The red line plots the log relative price of services, using BLS service price estimates and our quality-adjusted goods price index. The black line plots the log relative employment ratio, $\log(\ell_s/\ell_g)$. Outside the war period, both series rise over the twentieth century: services become relatively more expensive, and employment shifts toward services. This positive relationship implies an elasticity of substitution below one.

We estimate Equation (5.1) using the period 1946–1965. We exclude World War II because rationing and price controls severely distorted relative prices. We also exclude the post-1965 period because the rise of imports weakened the connection between domestic sectoral employment and domestic consumption prices. Over 1946–1965, imports were relatively stable as a share of GDP, making the closed-economy approximation more plausible. The resulting estimate is $\hat{\sigma} = 0.55$ with Newey-West standard error 0.07.

We then use this estimate to extrapolate the relative price of services before 1935. Specifically, for years before the BLS service price index is available, we use observed relative employment and the estimated relationship in Equation (5.1) to infer the implied relative price $\log(p_s/p_g)$. Combining this

predicted relative price with our quality-adjusted goods price index yields an estimated service price index for the pre-1935 period, normalized to connect smoothly to the observed BLS service index once it becomes available. The dashed red line in Figure A.12 shows the extrapolated relative price series.

Finally, we combine goods, services, food and housing using Laspeyres expenditure weights to estimate aggregate real expenditure, using appropriate expenditure weights for each sector. The key assumption is that food and service-sector quality growth is already properly incorporated into the BLS food and service price index. Under that assumption, the difference between our aggregate real expenditure estimate and the BLS-based estimate is driven by the revised treatment of goods quality.

Under these assumptions, real expenditure grew by a factor of 19.8 from 1900 to 1990. By contrast, using existing BLS-based deflators for consumer goods implies growth by a factor of 13.3. Thus, even holding measured service inflation fixed, quality adjustment for consumer goods meaningfully increases estimated real expenditure growth over the twentieth century. This estimate is likely conservative if food and service quality also improved in ways not fully captured by official price indexes. The twentieth century saw large improvements in medical care, education, transportation, communication, and other services. Even for food, Cafarella et al. (2023) show that half of measured food price inflation from 2007 to 2015 is driven by quality improvements; it is likely that over the twentieth century–which saw tremendous changes in food regulation, globalization, and technologies like refrigeration—quality improvements in food were much larger. To the extent that these quality improvements are recorded as higher prices rather than higher real service output, our estimates understate total real growth.

5.3 Implications for productivity

The long-run growth literature often evaluates aggregate performance through multi-factor productivity (MFP): the component of real output growth that remains after accounting for growth in capital and labor inputs (Solow, 1957). According to conventional data, real output growth was fairly high in the early twentieth century, but existing work generally attributes much of this growth to rapid accumulation of labor and capital rather than to unusually rapid productivity growth (Gordon, 1999). As a result, this literature typically places the peak of MFP growth in the postwar period (Gordon, 1999). Our estimates imply a different chronology: once output growth is adjusted for price-index bias, MFP growth appears substantially higher in the first three decades of the twentieth century.

Formally, MFP growth is defined as

$$\Delta \log MFP_t = \Delta \log Y_t - \left(\alpha_{K,t} \Delta \log K_t + \alpha_{L,t} \Delta \log L_t\right), \tag{5.2}$$

where Y_t is real output, K_t is capital input, L_t is labor input, and $\alpha_{K,t}$ and $\alpha_{L,t}$ are the corresponding factor cost shares. Let measured real output growth be

$$\Delta \log \tilde{Y}_t = \Delta \log Y_t + b_t,$$

where b_t denotes the measurement error in real output growth induced by price-index bias. Measured MFP growth is therefore

$$\Delta \log \widetilde{MFP}_t = \Delta \log MFP_t + b_t.$$

Thus, given an estimate of b_t, corrected MFP growth is obtained by subtracting this output-growth bias from measured MFP growth:

$$\Delta \log MFP_t = \Delta \log \widetilde{MFP}_t - b_t.$$

For our sample, we have shown that the official price deflator overstates inflation, and therefore understates real growth, so b_t is negative. As a result, our correction implies that true MFP growth was higher than existing estimates suggest.

Table A.1 applies this correction to historical estimates of MFP. We first report estimates based on Kendrick (1961), along with the revised series from Gordon (1999). We then show true MFP growth,

under the assumption that the price-index bias for total output is the same as the bias we measure for consumer goods in Table 1. This adjustment is large: we find peak MFP growth of 4.31pp annually for the period of 1913-1928, compared to MFP growth of 2.64pp in the period of 1950-1964, where Gordon's MFP series peaks.

As we have discussed, the Sears index covers only consumer goods, so we may worry that it is not representative of the full economy. We therefore also report a more conservative adjustment using CPI bias implied by the aggregate expenditure price index constructed in Section 5.2. This index assumes that price measurement in other sectors is correct and applies our adjustment only where consumer-goods prices enter the aggregate deflator. As in Gordon (1999), we exclude housing from these calculations. To the extent that services, food, or other sectors also experienced unmeasured quality improvements, these conservative estimates will still understate true productivity growth. Even under this restricted correction, MFP growth remains highest in 1913–1928, exceeding the 1950–1964 rate by 0.75 percentage points per year.

5.4 Cross country comparisons

An important branch of empirical macroeconomics uses cross-country evidence to study long-run growth, convergence, productivity, and income differences across countries (e.g. Baumol, 1986; Barro, 1991; Mankiw et al., 1992). Accurate comparisons of output growth, however, depend crucially on price measurement. For instance, an important period of convergence in modern history is post-World War II Europe, in which many European countries experienced sustained higher growth rates than the US, which has been cited as an example of global convergence (Abramovitz, 1986; Crafts, 1995). We plot real consumption for several of these countries in Figure A.13a. Notably, these countries also experienced much faster recorded inflation in the postwar period relative to the US, as shown in Figure A.13b. To the degree that existing price indexes are biased by quality upgrading, we will underestimate growth in Europe and therefore underestimate the strength of convergence.

Table 2: Technology Diffusion, Sears vs. La Redoute

(a) Sears

Year	Radios: Stereo	Washing Machines: Spin Dry	Ironing Machines: Steam	Cameras: Built-in Light Meter
1940s	0.00	0.00	0.05	0.06
1950s	0.03	0.53	0.22	0.04
1960s	0.19	0.75	0.62	0.62
1970s	0.40	0.97	0.95	0.64
1980s	0.67	0.98	0.95	0.80

(b) La Redoute

Year	Radios: Stereo	Washing Machines: Spin Dry	Ironing Machines: Steam	Cameras: Built-in Light Meter
1967	0.00	0.57	0.00	0.20
1989	0.74	1.00	0.92	0.50

The table shows the share of products that contain specific features in Sears and La Redoute catalogs. Panel (a) plots the share of (1) radios with stereo sound, (2) washing machines with spin dry, (3) ironing machines with a steam feature, and (4) cameras with a built-in light meter in each decade from the 1940s to the 1980s.

Estimating a quality-adjusted index of the kind we construct here for multiple countries is beyond the scope of this paper, but we present suggestive evidence that Europe experienced rapid quality upgrading

in consumer goods after the war. We obtain catalogs for La Redoute, a large French mail-order retailer, for 1967 and 1989, and compare the quality of consumer durables that appear in both Sears and La Redoute. Panel (a) of Table 2 reports the prevalence in Sears of four durable-good features: stereo sound on radios, spin dry on washing machines, steam on electric irons, and a light meter on cameras. Each was rare in Sears in the 1940s and was widely adopted over the 1950s and 1960s. Panel (b) reports the corresponding shares in La Redoute. These features were common in Sears by the 1960s but remained rare in La Redoute in 1967: no radio offered stereo sound, no iron offered steam, and only half of washing machines offered spin dry. By 1989, adoption in La Redoute matched Sears. This suggests that French retailers saw much faster quality upgrading in the seventies and eighties than the US. As a result, we may worry that price indexes for this period were more biased in France, and therefore that existing data understates the degree of convergence in the postwar period. These revisions could plausibly be very large. As an upper bound, we deflate both French and US consumption by our quality-adjusted Sears index. Conventional deflators put France's average annual real consumption growth at 1.7pp above the US over 1950–1990; ours put the difference at 7.5pp.

6 Conclusion

This paper constructs a quality-adjusted price index for U.S. consumer goods spanning the twentieth century. Using newly digitized catalog data, we recover millions of product-level prices and descriptions, estimate hedonic price schedules from high-dimensional text representations, and use those schedules to price fixed bundles of product characteristics over time. The resulting index offers a new view of long-run inflation, quality change, and real growth in the retail goods sector.

Our central finding is that quality improvement in consumer goods over this period was large. Posted prices rose substantially, but quality-adjusted prices rose far less. Existing goods deflators track unadjusted prices more closely than our quality-adjusted index, implying that conventional measures understate real growth in the sector. The magnitudes are large: replacing standard goods deflators with our index raises measured growth in real goods consumption from a factor of 10.3 to 39 between 1900 and 1990, and raises aggregate real expenditure growth from 13.3 to 19.8 once other sectors are held fixed.

The timing of this revision is also important. Conventional estimates imply that goods consumption and productivity growth accelerated after World War II. Our quality-adjusted index reverses this pattern. Real goods consumption growth was faster from 1900 to 1939 than from 1946 to 1980, and the same pattern appears in our implied estimates of multi-factor productivity. This suggests that a substantial share of twentieth-century growth occurred before the war, during precisely the period for which historical price measurement is weakest.

Our methodological contribution is to show that historical product text can measure quality change at scale. Traditional hedonic methods require hand-coding product attributes, which confines them to a small number of well-defined categories. Text embeddings instead let us estimate flexible, category-specific hedonic schedules across a broad range of goods and over long horizons. This is especially valuable in historical settings, where structured attributes are rarely available but product descriptions are rich.

Our index covers only consumer goods and omits major sectors such as services, food, and housing. Even holding inflation in those sectors fixed, quality adjustment of consumer goods alone materially raises measured real expenditure growth over the century. More broadly, scalable methods for digitizing and analyzing historical text open new opportunities to measure economic change over periods for which structured data have never existed. Extending this approach to other retailers, countries, and historical records is a natural next step toward a fuller account of one of the most transformative periods in economic history.

References

Abramovitz, M. (1986). Catching up, forging ahead, and falling behind. *The Journal of Economic History*, 46(2):385–406.

Abramovitz, M. and David, P. A. (2000). American macroeconomic growth in the era of knowledge-based progress: The long-run perspective. In Engerman, S. L. and Gallman, R. E., editors, *The Cambridge Economic History of the United States*, volume 3, pages 1–92. Cambridge University Press, Cambridge.

Aghion, P., Bergeaud, A., Boppart, T., Klenow, P. J., and Li, H. (2019). Missing growth from creative destruction. *American Economic Review*, 109(8):2795–2822.

Aizcorbe, A. M. and Pho, Y. H. (2005). Differences in hedonic and matched-model price indexes: Do the weights matter? Working Paper WP2005-06, Bureau of Economic Analysis.

Asirvatham, H., Mokski, E., and Shleifer, A. (2026). GPT as a measurement tool. NBER Working Paper 34834, National Bureau of Economic Research.

Bäcker-Peral, V., Meursault, V., and Severen, C. (2025). Can LLMs credibly transform the creation of panel data from diverse historical tables? Working Paper 25-28, Federal Reserve Bank of Philadelphia.

Bajari, P., Cen, Z., Chernozhukov, V., Manukonda, M., Vijaykumar, S., Wang, J., Huerta, R., Li, J., Leng, L., Monokroussos, G., and Wang, S. (2025). Hedonic prices and quality adjusted price indices powered by AI. *Journal of Econometrics*, 251:106052.

Barro, R. J. (1991). Economic growth in a cross section of countries. *The Quarterly Journal of Economics*, 106(2):407–443.

Baumol, W. J. (1967). Macroeconomics of unbalanced growth: The anatomy of urban crisis. *American Economic Review*, 57(3):415–426.

Baumol, W. J. (1986). Productivity growth, convergence, and welfare: What the long-run data show. *American Economic Review*, 76(5):1072–1085.

Bils, M. (2009). Do higher prices for new goods reflect quality growth or inflation? *Quarterly Journal of Economics*, 124(2):637–675.

Bils, M. and Klenow, P. J. (2001). Quantifying quality growth. *American Economic Review*, 91(4):1006–1030.

Boskin, M. J., Dulberger, E. R., Gordon, R. J., Griliches, Z., and Jorgenson, D. W. (1998). Consumer prices, the Consumer Price Index, and the cost of living. *Journal of Economic Perspectives*, 12(1):3–26.

Broda, C. and Weinstein, D. E. (2006). Globalization and the gains from variety. *The Quarterly Journal of Economics*, 121(2):541–585.

Cafarella, M., Ehrlich, G., Gao, T., Haltiwanger, J. C., Shapiro, M. D., and Zhao, L. (2023). Using machine learning to construct hedonic price indices. Working Paper 31315, National Bureau of Economic Research.

Carter, S. B., Gartner, S. S., Haines, M. R., Olmstead, A. L., Sutch, R., and Wright, G., editors (2006). *Historical Statistics of the United States: Earliest Times to the Present, Millennial Edition*. Cambridge University Press, New York. 5 volumes.

Costa, D. L. (2001). Estimating real income in the United States from 1888 to 1994: Correcting CPI bias using Engel curves. *Journal of Political Economy*, 109(6):1288–1310.

Court, A. T. (1939). Hedonic price indexes with automotive examples. In *The Dynamics of Automobile Demand*, pages 98–119. General Motors Corporation, New York.

Crafts, N. F. R. (1995). The golden age of economic growth in western europe, 1950–1973. *The Economic History Review*, 48(3):429–447.

David, P. A. (1990). The dynamo and the computer: An historical perspective on the modern productivity paradox. *American Economic Review*, 80(2):355–361.

Devine, Warren D., J. (1983). From shafts to wires: Historical perspective on electrification. *The Journal of Economic History*, 43(2):347–372.

Diewert, W. E. (1976). Exact and superlative index numbers. *Journal of Econometrics*, 4(2):115–145.

Du Boff, R. B. (1967). The introduction of electric power in american manufacturing. *The Economic History Review*, 20(3):509–518.

Ehrlich, G., Haltiwanger, J., Jarmin, R., Johnson, D., Olivares, E., Pardue, L., Shapiro, M. D., and Zhao, L. Y. (2026). Quality adjustment at scale: Hedonic versus exact demand-based price indices. *American Economic Review*, 116(6):1955–1995.

Erickson, T. and Pakes, A. (2011). An experimental component index for the CPI: From annual computer data to monthly data on other goods. *American Economic Review*, 101(5):1707–1738.

Fajgelbaum, P. D., Grossman, G. M., and Helpman, E. (2011). Income distribution, product quality, and international trade. *Journal of Political Economy*, 119(4):721–765.

Feenstra, R. C. (1994). New product varieties and the measurement of international prices. *American Economic Review*, 84(1):157–177.

Field, A. J. (2003). The most technologically progressive decade of the century. *American Economic Review*, 93(4):1399–1413.

Gordon, R. J. (1990). *The Measurement of Durable Goods Prices*. National Bureau of Economic Research Monograph. University of Chicago Press, Chicago.

Gordon, R. J. (1999). U.S. economic growth since 1870: One big wave? *American Economic Review*, 89(2):123–128.

Gordon, R. J. (2016). *The Rise and Fall of American Growth: The U.S. Standard of Living since the Civil War*. Princeton University Press, Princeton, NJ.

Greenwood, J., Seshadri, A., and Yorukoglu, M. (2005). Engines of liberation. *The Review of Economic Studies*, 72(1):109–133.

Griliches, Z. (1961). Hedonic price indexes for automobiles: An econometric analysis of quality change. In *The Price Statistics of the Federal Government*, pages 173–196. National Bureau of Economic Research, New York. Report of the Price Statistics Review Committee.

Griliches, Z., editor (1971). *Price Indexes and Quality Change: Studies in New Methods of Measurement*. Harvard University Press, Cambridge, MA.

Hamilton, B. W. (2001). Using Engel's Law to estimate CPI bias. *American Economic Review*, 91(3):619–630.

Hausman, J. A. (1997). Cellular telephone, new products and the CPI. SSRN Scholarly Paper.

Jaimovich, N., Rebelo, S., and Wong, A. (2019). Trading down and the business cycle. *Journal of Monetary Economics*, 102:96–121.

Jordà, Ò., Schularick, M., and Taylor, A. M. (2017). Macrofinancial history and the new business cycle facts. *NBER Macroeconomics Annual*, 31(1):213–263.

Kendrick, J. W. (1961). *Productivity Trends in the United States*. Number 71 in General Series. Princeton University Press, Princeton, NJ. Published for the National Bureau of Economic Research.

Konüs, A. A. (1939). The problem of the true index of the cost of living. *Econometrica*, 7(1):10–29. Translated from the 1924 Russian original.

Lyons, R. C., Shertzer, A., Gray, R., and Agorastos, D. (2026). The price of housing in the United States, 1890–2006. *The Quarterly Journal of Economics*, 141(1):559–603.

Mankiw, N. G., Romer, D., and Weil, D. N. (1992). A contribution to the empirics of economic growth. *The Quarterly Journal of Economics*, 107(2):407–437.

Nakamura, E. and Steinsson, J. (2012). Lost in transit: Product replacement bias and pricing to market. *American Economic Review*, 102(7):3277–3316.

Nordhaus, W. D. (1996). Do real-output and real-wage measures capture reality? The history of lighting suggests not. In Bresnahan, T. F. and Gordon, R. J., editors, *The Economics of New Goods*, pages 27–70. University of Chicago Press, Chicago.

Pakes, A. (2003). A reconsideration of hedonic price indexes with an application to PC's. *American Economic Review*, 93(5):1578–1596.

Redding, S. J. and Weinstein, D. E. (2020). Measuring aggregate price indices with taste shocks: Theory and evidence for CES preferences. *The Quarterly Journal of Economics*, 135(1):503–560.

Rees, A. and Jacobs, D. P. (1961). The cost-of-living index. In *Real Wages in Manufacturing, 1890–1914*, pages 74–119. Princeton University Press.

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, 82(1):34–55.

Sato, K. (1976). The ideal log-change index number. *The Review of Economics and Statistics*, 58(2):223–228.

Solow, R. M. (1957). Technical change and the aggregate production function. *The Review of Economics and Statistics*, 39(3):312–320.

Triplett, J. E. (1986). The economic interpretation of hedonic methods. *Survey of Current Business*, 66(1):36–40.

Triplett, J. E. (2006). *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes: Special Application to Information Technology Products*. OECD Publishing, Paris.

United States Bureau of Labor Statistics (1915). Index numbers of wholesale prices in the United States and foreign countries. Technical Report 173, United States Bureau of Labor Statistics, Washington, DC. Wholesale Prices Series, No. 3.

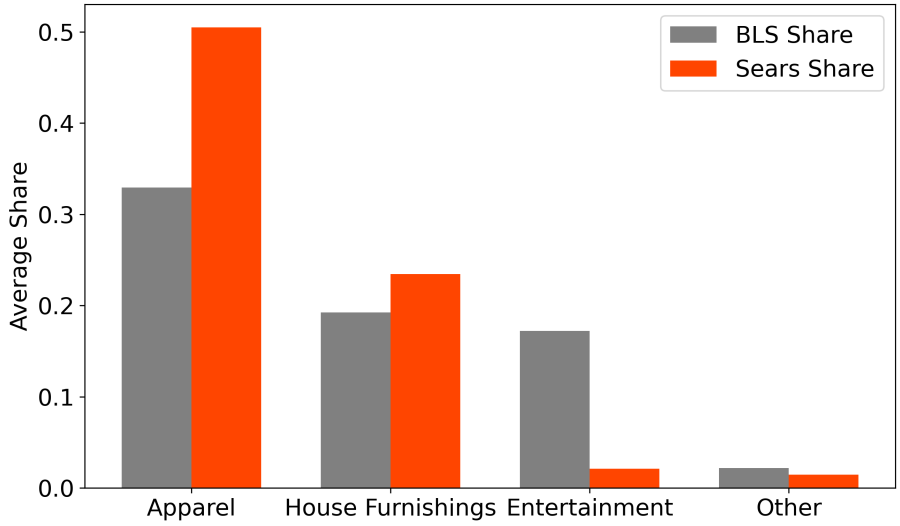
U.S. Bureau of Labor Statistics (2025). Consumer Price Index: Calculation. BLS Handbook of Methods. Last modified January 30, 2025. Accessed June 8, 2026.

Vartia, Y. O. (1976). *Relative Changes and Index Numbers*. Number 4 in ETLA Series A. The Research Institute of the Finnish Economy, Helsinki.

A Appendix

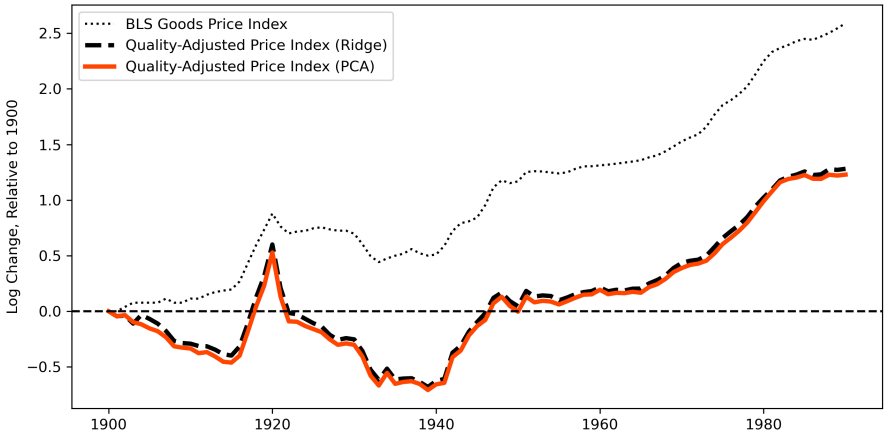
A.1 Figures

Figure A.1: Sears Product Shares vs. CPI Weights



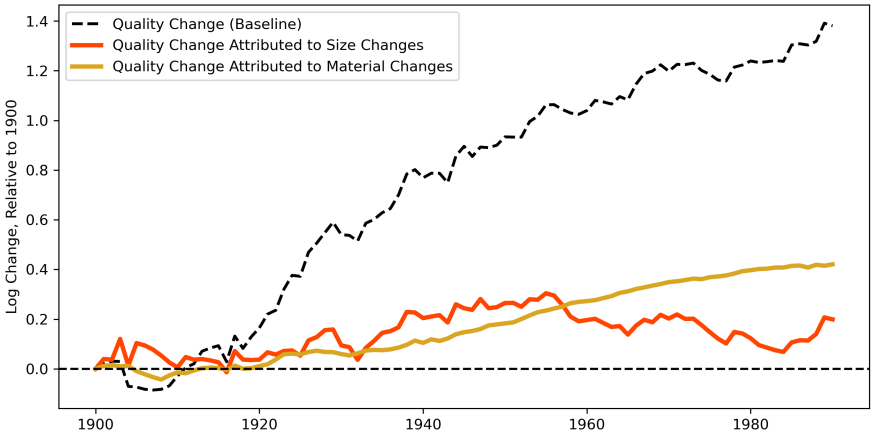
The figure is a bar plot showing the relative weights assigned to each major BLS retail goods category based on the frequencies of products in Sears catalogs vs. the BLS reported weights.

Figure A.2: Ridge vs. PCA Regression



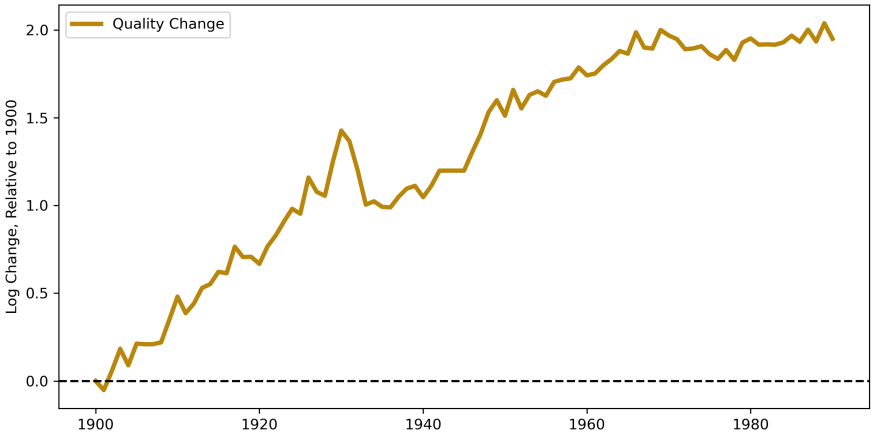
The figure presents in red a quality-adjusted price index constructed using a hedonic regression as in Equation (3.11), where the hedonic controls are selected via Principal Component Analysis. In particular, our matrix of high-dimensional embeddings, ϕ_{ijt}, is reduced to a smaller n-dimensional matrix of n principal components, ϕ^n_{ijt}. Then, Equation (3.11) can be estimated via OLS. The number of principal components, n, is chosen to minimize out-of-sample error via cross-validation. The black line shows our baseline estimates. The dotted line shows the BLS price index for consumer goods.

Figure A.3: Total Quality Improvement and Materials Quality Improvement



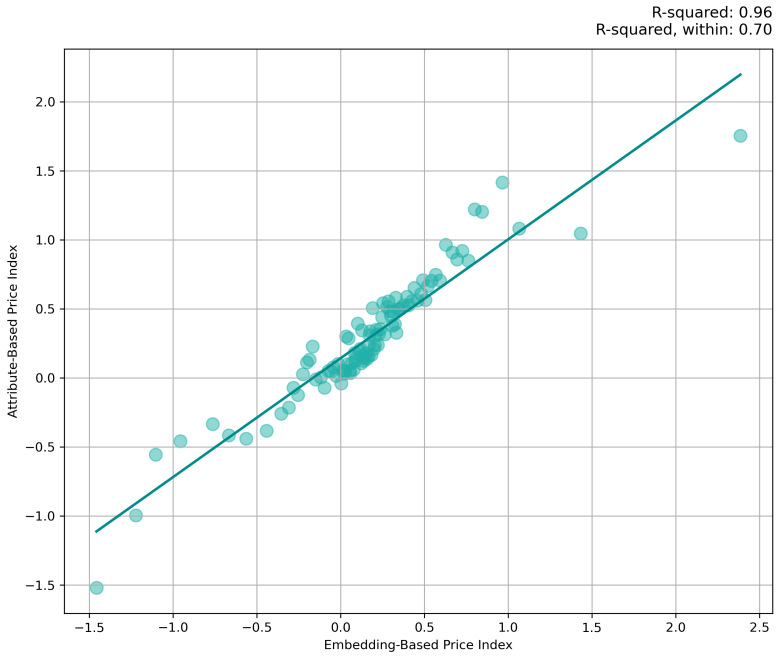
The figure shows the share of quality improvement that can be attributed to changes in materials and changes in product size. Total cumulative quality improvement is shown in the dashed black line. Quality changes attributed to changes in product size are shown in red. Quality changes attributed to materials changes is shown in gold.

Figure A.4: Quality Improvements for Technological Goods



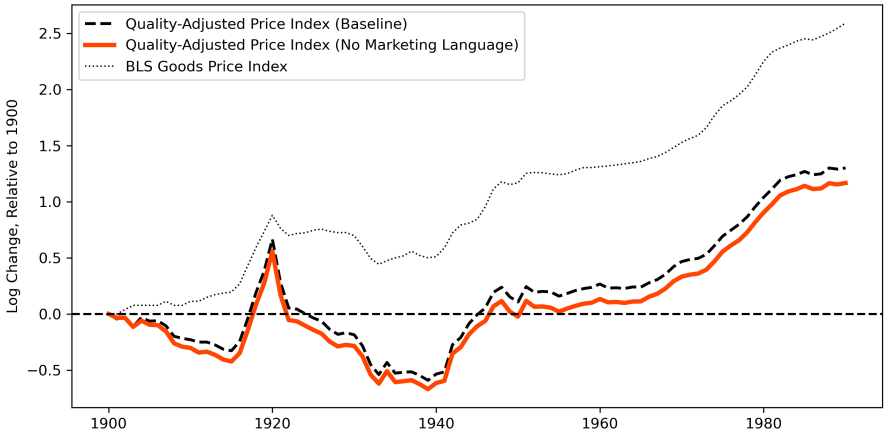
The figure shows cumulative quality improvements for the subset of 28 technology-intensive durable goods described in Section 4.1.1.

Figure A.5: Embedding vs. Attribute-Based Indexes for Technological Products



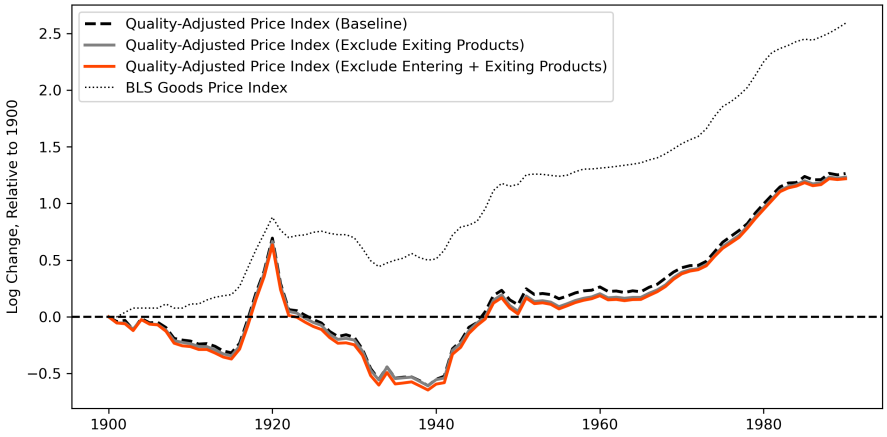
The figure shows a binscatter of the price index for each product category j at each year t constructed using our preferred embedding-based methodology (x-axis) vs. our attribute-based methodology (y-axis). Each price index is constructed via the methodology described in Section 3.1, but aggregating for each product category separately. The sample includes 28 technology-intensive durable goods categories for which we run attribute-based hedonic regressions. We control for year t and product category j fixed effects. The relationship is positive with a slope of 0.98 and very statistically significant with a within-R^2 of 0.70. The binscatter has 100 bins.

Figure A.6: Price Indexes, Excluding Marketing Language



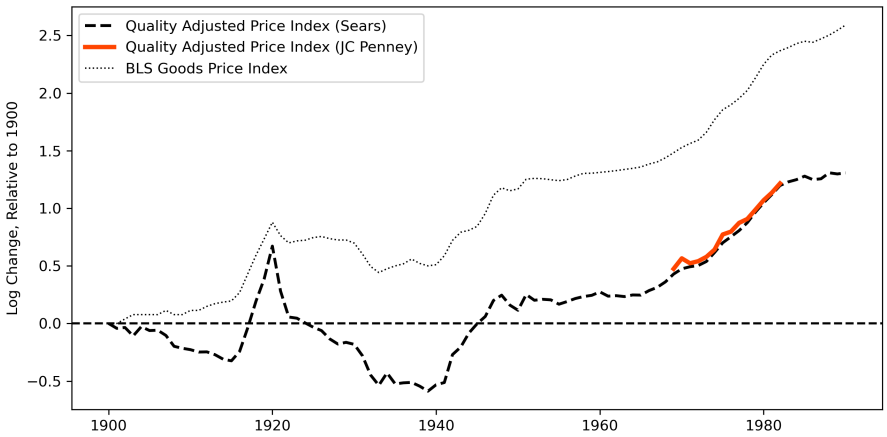
The figure presents in the solid red line a quality-adjusted price index following the methodology described in Section 3.1 with the exception that we strip embedding text from marketing terms and very subjective adjectives. The goal is to construct an embedding representative of product descriptions that emphasizes tangible product characteristics rather than subjective language for advertisement purposes. The dashed black line shows our baseline estimates. The dotted black line shows the BLS price index for consumer goods.

Figure A.7: Price Indexes, Restricting to Stable Products



The figure presents in the solid gray line a quality-adjusted price index which excludes products that are very different from any product in the subsequent year (i.e. exiting goods). The solid red line shows a quality-adjusted price index which excludes both exiting goods and entering goods, which are defined analogously as products that are very different from any product in the preceding year. The dashed black line shows our baseline estimates. The dotted black line shows the BLS price index for consumer goods.

Figure A.8: Price Indexes, Sears vs. J.C. Penney



The figure presents in the solid red line a quality-adjusted price index for the period of 1969–1982 based on J.C. Penney catalog data. The dashed black line shows our baseline estimates. The J.C. Penney indexed is normalized to the level of our baseline estimates in 1969. The dotted black line shows the BLS price index for consumer goods.

Figure A.9: Example of Appliances from 1910



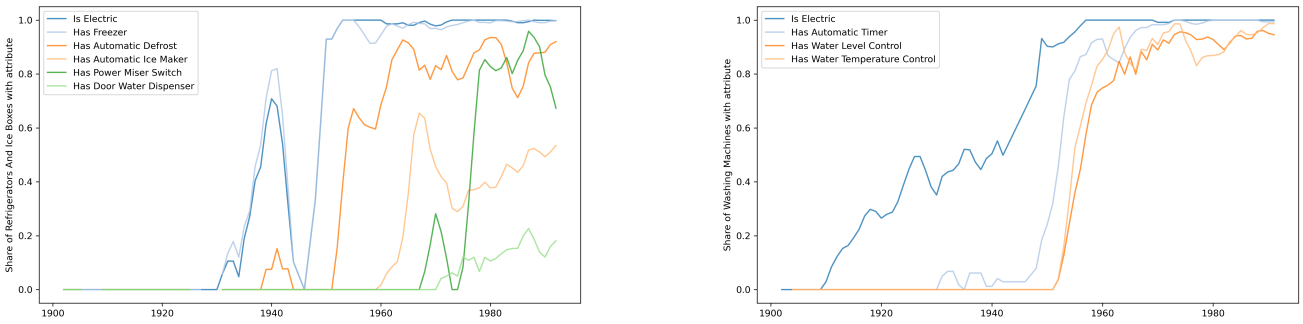
(a) Refrigerator



(b) Washing Machine

The figure presents sections from the 1910 Sears catalog featuring in Panel (a) a refrigerator and in Panel (b) a washing machine. Both products were not electric. The refrigerator used large blocks of ice which cooled the rest of the cabinet, and the washing machine was operated manually as displayed in the image.

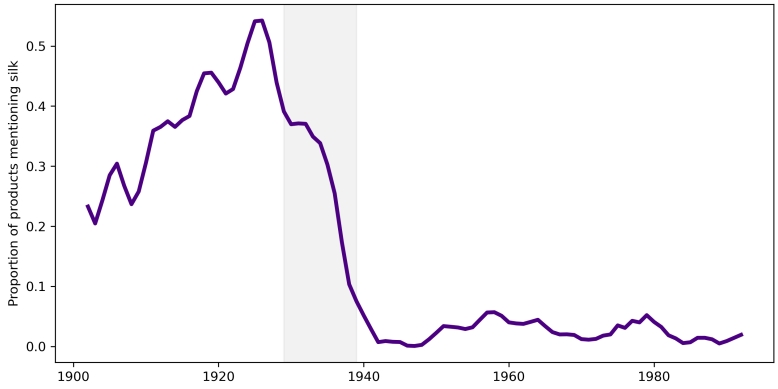
Figure A.10: Diffusion of Electricity Into Appliances



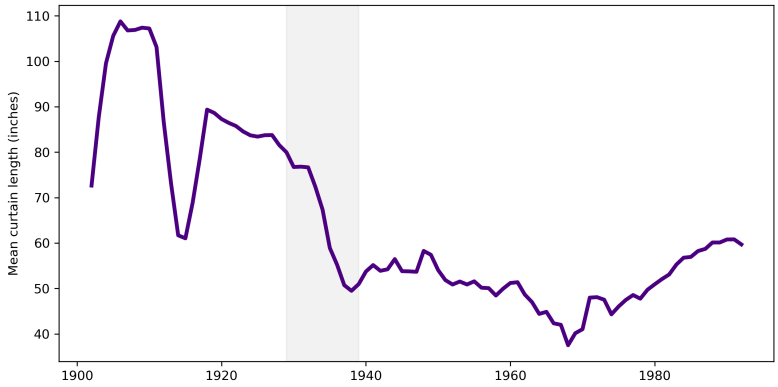
(a) Refrigerator **(b)** Washing Machine

The figure presents the diffusion of several technological advances in refrigerators and washing machines over time. Panel (a) shows the share of refrigerators listed on the Sears catalog that featured: electricity, freezers, automatic defrosting, an ice maker, a power miser switch, and a door water dispenser. Panel (b) shows the share of washing machines that featured: electricity, an automatic timer, water level control, and water temperature control. All lines are three-year moving averages.

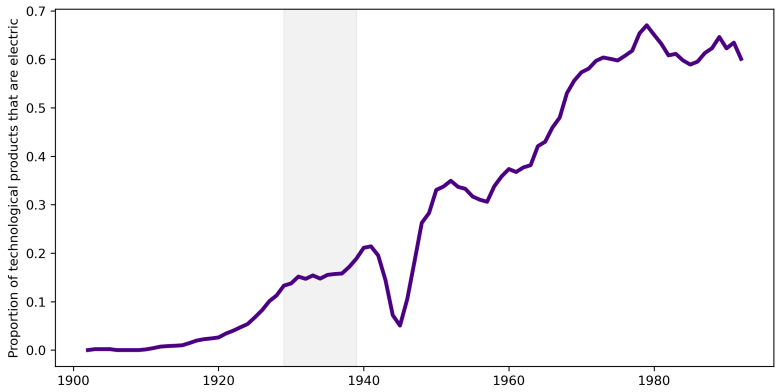
Figure A.11: Quality Declines During the Great Depression



(a) Silk Usage in Women's Apparel



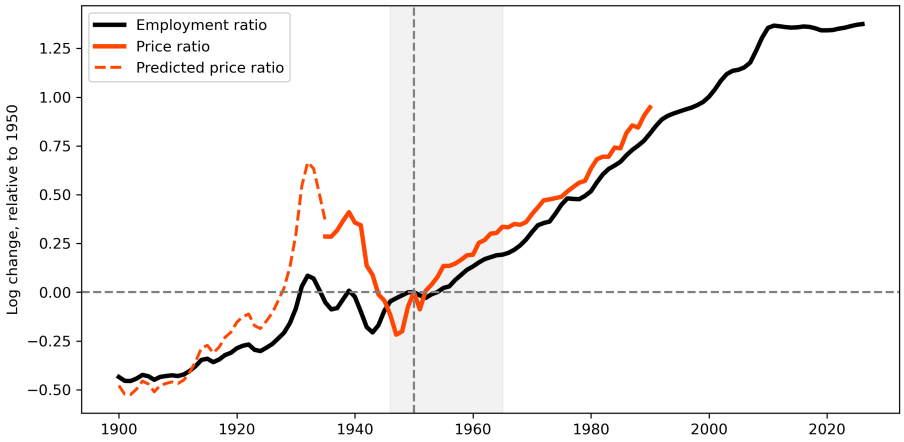
(b) Curtain Length



(c) Electrical Goods

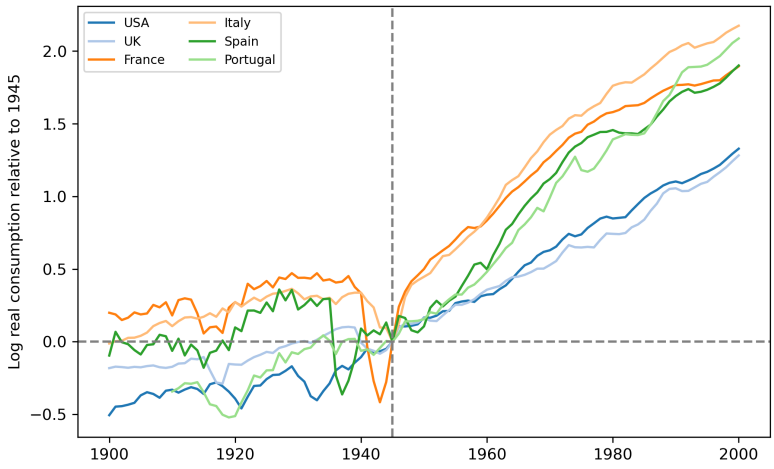
The figure shows three dimensions along which quality changed over the twentieth century. Panel (a) shows the proportion of women's apparel products (shirts, dresses, jackets, skirts, pajamas and sweaters) that used silk. Silk was a common material even at the start of the sample and its usage grew to more than 50% of women's apparel in the 1920s. Throughout this period, it was a luxury material and commanded a price premium of 0.8 log points on average. Silk usage collapsed during the Great Depression and remained low for the rest of the century. Panel (b) shows average curtain length in the Sears catalog. Before the Great Depression, curtains tended to be close to wall-length at around 80 inches (with the exception of WWI). During the Great Depression, average curtain length fell to approximately 50 inches, closer to the typical window size in the US. Panel (c) shows the proportion of goods that use electricity amongst a set of durable goods that shifted from non-electric to electric versions during this period: refrigerators, washing machines, typewriters, stoves, heaters, etc. Electricity adoption took off in the 1920s, slowed during the Great Depression, collapsed during WWII, and then took off again in the decades after the war. All three panels show 3-year moving averages.

Figure A.12: Employment and Prices in Services vs. Goods

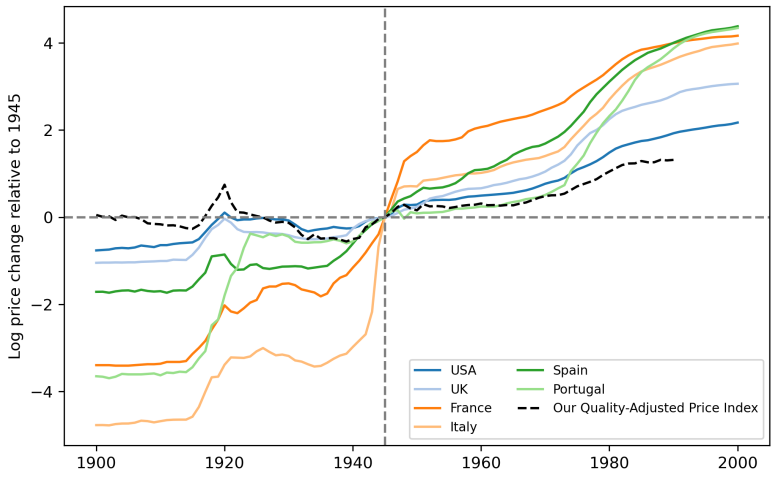


The figure shows the log change in the ratio of employment in service sectors vs. goods-producing sectors, normalized to zero in 1950. The solid red line shows the log change in the ratio of service prices, based on the BLS, and goods prices, based on our estimates. The dashed red line shows the predicted price ratio going back to 1900 based on the estimated elasticity between goods and services estimated in Section 5.2. The shaded gray region shows the period that is used to estimate the elasticity.

Figure A.13: Long-Run Changes in Price and Real Consumption, Advanced Economies



(a) Real Consumption



(b) CPI

The figure shows in Panel (a) log real consumption and in Panel (b) a log price index for several countries, normalized to zero in 1945. Panel (b) also shows our baseline quality-adjusted price index for reference. The data is obtained from the Macrohistory Database (Jordà et al., 2017).

A.2 Tables

Table A.1: Revised MFP Estimates

Period	MFP, Unadjusted			MFP, Gordon Adjustment		
	Gordon (1999)	Goods Index	Goods + Services Index	Gordon (1999)	Goods Index	Goods + Services Index
1891–1913	1.20	4.37	2.60	0.86	4.02	2.26
1913–1928	1.43	4.74	3.12	1.01	4.31	2.69
1928–1950	1.90	2.77	1.90	1.01	1.88	1.01
1950–1964	2.35	3.13	2.43	1.86	2.64	1.94
1964–1972	2.07	2.24	1.94	1.69	1.86	1.56
1972–1979	1.12	2.46	1.36	1.04	2.38	1.28
1979–1988	0.90	1.27	0.95	0.34	0.71	0.39

The table shows historical multi-factor productivity (MFP) estimates from Gordon (1999), adjusted using our quality-adjusted price index. For 1891-1913, we assume that CPI bias was the same in the period of 1891-1900, which is not covered by our sample, as in the period of 1900-1913.

A.3 Comparing with BLS Methodology

The Consumer Price Index (CPI) methodology has several differences from the methodology we implement in this paper. The CPI is constructed as a weighted average of price relatives for a representative sample of goods and services purchased by urban consumers. At the lowest level, the BLS divides the CPI into item–area cells and computes elementary indexes from individual price quotes; these are then aggregated using expenditure weights derived primarily from the Consumer Expenditure Surveys (U.S. Bureau of Labor Statistics, 2025).

To track how prices change over time, the BLS typically implements a matched-model approach: the price of the same sampled item in the same outlet is compared across periods, and, if a sampled item must be replaced, the data collector selects the closest available substitute within the same outlet. Historically, there was not a systematic process to account for differences between the original product and its closest available substitute, which resulted in systematic bias in the CPI (Boskin et al., 1998). In the 1990s, the Boskin Commission introduced major changes to the CPI methodology, including explicit hedonic adjustments for a subset of products. However, even today, the BLS has no quality adjustment for 85% of items. Even in the cases where quality adjustment exists, they consider only a small category of attributes. For instance, the hedonic model for refrigerators includes only 12 attributes, which may not be sufficient to capture subtle quality dimensions like warranty length or appearance.

Another limitation of the BLS methodology is that it is primarily a matched-model method, and uses hedonic methods only when necessary. As has been well documented in the literature, the matched-model approach introduces severe selection problems by focusing only on a small and unrepresentative sample of goods (Pakes, 2003; Ehrlich et al., 2026). Pakes (2003) emphasizes a selection problem in matched-model indexes: products that can be matched across periods are the survivors, and survival is itself partly determined by consumer demand. If surviving products are systematically more desirable than discontinued products, their price may rise faster than those of the full product universe. Other plausible sources of bias may run in the opposite direction. For instance, prices may be stickier for continuing products if sellers face higher costs to updating prices for old products relative to slightly different new versions of products. Moreover, old products may be more obsolete or unfashionable relative to new versions of products, which could result in lower inflation of matched models. Therefore, the direction of the selection bias from the matched-model approach is ambiguous.

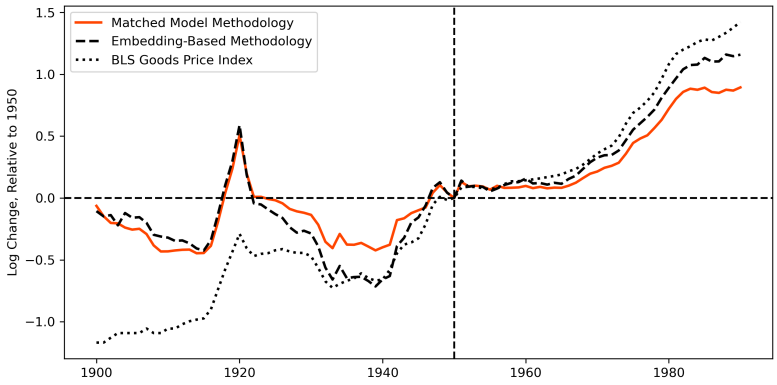
Unlike modern scanner data, which contain persistent product-level identifiers, the Sears catalogs do not allow us to directly observe whether a product continues from one catalog to the next. We therefore infer continuing products using a conservative matching algorithm. A product in catalog t and a product in catalog $t+1$ are classified as a matched model if they satisfy all of the following criteria: (1) they have the same UNSPSC code, (2) they have the same shipping weight, (3) they have the same brand,

(4) the product in catalog $t+1$ is the nearest neighbor of the product in catalog t among all products in catalog $t+1$, and the product in catalog t is likewise the nearest neighbor of the product in catalog $t+1$ among all products in catalog t, and (5) their cosine similarity in embedding space exceeds 0.80. Under these criteria, 16% of products have a matched model in the subsequent year. Based on these matched models, we can construct a matched-model Laspeyres index,

$$L_t^{MM} \equiv \sum_{i \in \mathcal{M}_t} s_{i,t-1} \log \frac{p_{it}}{p_{i,t-1}},$$

where $i \in \mathcal{M}_t$ are the set of matched models, $s_{i,t-1}$ are the BLS expenditure weights for each product in our sample and p_{it} and $p_{i,t-1}$ are the price quotes observed for product i.

Figure A.14: Cost-Of-Living Index Using Matched Models

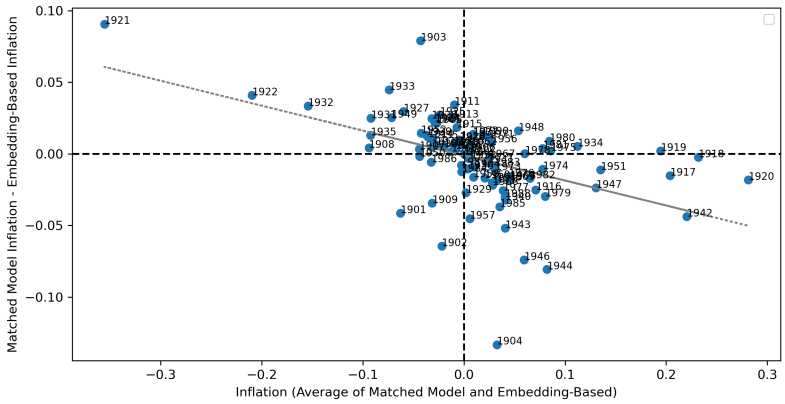


The figure shows in the solid red line a quality-adjusted price index based on a matched-model methodology. The dashed black line shows our baseline estimates. The dotted black line shows the BLS official goods price index.

Figure A.14 shows in red the implied cost-of-living index using the matched-model Laspeyres index. The dashed black line shows our baseline estimates using hedonic methods. In the post-World War II period, we find that the matched-model approach vastly *underestimates* inflation relative to the hedonic method. More generally, we find that the matched-model approach underestimates the magnitude of inflation, such that in deflationary periods it overstates inflation and in inflationary periods it understates inflation. This is shown formally in Figure A.15. We show a scatter plot of the difference in annual inflation implied by the matched-model method vs. the hedonic regression, against the average inflation implied by both methods. The relationship is negative and statistically significant. In years with deflation, matched models fell less in price than the broader set of goods; in years with inflation, matched models rose less in price.

This result is consistent with the findings in Pakes (2003), who find that for the case of computers for 1995-1999, where average inflation was negative, a matched-model method overstated inflation. This can be partially attributed to stickier prices for matched models relative to new goods. 33% of matched models in our sample experience zero annual inflation, even though in general the price level was changing. This suggests that there are important price frictions for continuing goods, which companies can avoid by selling slightly different variations of products.

Figure A.15: Differences Between Matched Model and Baseline Method



The figure shows a scatter plot of annual inflation (average between our preferred methodology and the matched model methodology) on the x-axis and the annual matched-model methodology gap on the y-axis. The matched-model methodology gap is defined as the difference between the matched-model methodology and our preferred methodology.

A.4 Digitizing Data Using LLMs

Our methodology requires extracting detailed structured information from more than one hundred thousand pages. Such a project would cost hundreds of thousands of dollars using conventional professional digitization services (Bäcker-Peral et al., 2025). For that reason, we use Gemini 3 Pro — the most powerful LLM for OCR tasks at the time of writing — to extract this information in a cost efficient manner. On average, digitization uses 3,000 input tokens and 9,000 output tokens per page, which at the current Gemini 3 API rate totals a cost of about $0.06 per page – effectively two orders of magnitude lower than the cost of outsourcing to a digitization company.

For each page in the catalog, we use the following prompt to extract structured information from the text.

```
You are an expert data extraction agent specializing in historical retail catalogs.
You are processing a page from a {company} catalog from {year}.

Inputs: Images of the catalog page (which you must analyze directly).

Your task is to analyze the images and to identify every unique product. Extract its
key details, grouping all variants under a single parent product. Variants must have
the same description and brand and must be the same type of product but can have
minor differences.

Output Requirements: You must return only a valid JSON object. Do not include any
introductory text, explanations, or apologies.

The JSON should have the following structure:
{{
  "products": [
    {{
      "name": "Product Name",
      "description": "General description",
      "short_description": "Summary",
```

```
      "hs_code": "123456",
      "is_accessory": false,
      "brand_name": "Brand",
      "versions": [
        {{
          "name": "Version Name",
          "price": "$X.XX",
          "installments_price":"$X.XX",
          "downpayment_price":"$X.XX",
          "monthly_price":"$X.XX",
          "attributes": {{
            "weight": "5kg"
          }},
          "product_code": "SKU123"
        }}
      ]
    }}
  ]
}}
```

Guidelines:
products: A list containing one object for each distinct product group found on the page. If no products are found, return an empty array: []. If the page is a glossary or index page that does not list specific products, return an empty array.

products.name: The main, overarching name for the product (e.g., "Craftsman Electric Drill", "Women's Day Dress").

products.description: The FULL description of the product, using the exact language from the catalog. Do not skip or abbreviate any parts of the description. Include slogans or parts of the description that are purely for advertisement purposes.

products.short_description: A one sentence summary of the product in your own words

products.hs_code: Your best inference of the {hs_year} 6-digit Harmonized System (HS) code for this product category. Do not classify it as an antique. Use your expert knowledge to carefully pick the code that best describes this product's purpose and specific features.

products.is_accessory: true or false
-true: If the item is an accessory (e.g., "replacement belt," "drill bits," "sewing machine needles") that is used with another primary product
-false: If the item is a standalone product (e.g., "drill," "dress," "refrigerator").

products.brand_name: The name of the brand of the product (e.g. "Kodak"). Use "" if unknown.

products.versions: A list containing all specific buyable variants of the parent product.

```
- Each variant must belong to the specific product category of the parent product and
should all have the same HS code.
- Accessories for the product (like cases or replacement parts) must be listed as a
different parent product.
- If a product has a table of sizes (e.g., 6ft, 8ft, 10ft) with different prices,
each row is a separate Version.
- If a product has multiple colors but ONE price and ONE catalog number, treat it as
a single Version with a "colors" attribute.

products.versions.name: The specific name or identifier for this version (e.g.,
"1/2-HP Model").

products.versions.price: The price for this specific version, extracted exactly as a
string (e.g., "$19.95"). If a cash and installments price is listed, use the cash
price.
products.versions.installments_price: The price for this specific version if paid in
installments. This is sometimes called the 'easy payment' price. If an installment
price is not listed, use the empty string "".
products.versions.downpayment_price: If installments_price exists, the downpayment
that must be paid at the time of purchase.
products.versions.monthly_price: If installments_price exists, the monthly payment
that must be paid every month until repayment.

products.versions.attributes: A dictionary of key attributes of this specific
version. Extract these directly from the image. If any of the following attributes
are available, always include them:
- shipping_weight
- warranty_length
- material
- origin_country
- num_units (The number of identical copies of the product that come in the package.
This will typically be 1.)
- includes_accessories: a comma-delineated list with any accessories that the product
comes with
Include all attributes that are relevant to the buyer's decision, especially those
which are not included in the product description but appear in the image or in a
table.

products.versions.product_code: The specific catalog number for this version

Other rules:
- Ensure all keys and string values in the JSON are enclosed in double quotes.
- When creating a string value that has a quotation inside the string, you must not
use double quotes (") if the string is delimited with double quotes. Instead, convert
the quotes inside the string to single quotes (').
CORRECT: "description": "A 3-button 'brain' for switching modes."
INCORRECT: "description": "A 3-button "brain" for switching modes."
- Return nothing but a JSON object. The output MUST be a valid JSON object.
```

We manually audited hundreds of pages and found low error rates in product extraction and price

transcription.

A.5 Details on the Hedonic Cost-of-Living Approximation

This appendix provides additional details on the framework in Section 3.

A.5.1 Set-up and homotheticity assumption

Our goal is to capture changes in welfare between time $t-1$ and time t for each household, h,

$$\log \frac{u_t^h}{u_{t-1}^h}$$

Assume that preferences are homothetic, and without loss of generality, assume that they are homogenous of degree-one. Then,

$$U^h(\alpha \mathbf{q}) = \alpha U^h(\mathbf{q})$$

for all h. We can then write the consumption bundle as $\mathbf{q} \equiv u\mathbf{z}$, where u is the utility level achieved by household h. Then,

$$\begin{aligned} e(\mathbf{p}, u) &= \min_{\mathbf{z}} \{p \cdot (u\mathbf{z}) \ : \ U(\mathbf{z}) \geq 1\} \\ &= u \min_{\mathbf{z}} \{p \cdot \mathbf{z} \ : \ U(\mathbf{z}) \geq 1\} \\ &= u e(\mathbf{p}, 1) \end{aligned}$$

Define $\tilde{e}(\mathbf{p}) = e(\mathbf{p}, 1)$. Then,

$$\frac{e(\mathbf{p}, u_t)}{e(\mathbf{p}, u_{t-1})} = \frac{\tilde{e}(\mathbf{p}) u_t}{\tilde{e}(\mathbf{p}) u_{t-1}} = \frac{u_t}{u_{t-1}}$$

as desired.

A.5.2 Cost of living and the first-order Laspeyres approximation

For household h, let $e^h(\mathbf{p}, \bar{u})$ denote the expenditure function

$$e^h(\mathbf{p}, \bar{u}) = \min_{\mathbf{q} \geq 0} \left\{ \mathbf{p} \cdot \mathbf{q} : u^h(\mathbf{q}) \geq \bar{u} \right\}. \tag{A.1}$$

Let $\mathbf{q}_{t-1}^h$ be household h's cost-minimizing bundle at $(\mathbf{p}_{t-1}, u_{t-1}^h)$, which exists and is unique under standard assumptions. Then

$$e^h(\mathbf{p}_{t-1}, u_{t-1}^h) = \mathbf{p}_{t-1} \cdot \mathbf{q}_{t-1}^h.$$

Define the household-level base-period expenditure share on product i as

$$s_{i,t-1}^h \equiv \frac{p_{i,t-1} q_{i,t-1}^h}{\sum_{\ell \in \Omega_{t-1}} p_{\ell,t-1} q_{\ell,t-1}^h}. \tag{A.2}$$

For products with zero household expenditure, $s_{i,t-1}^h = 0$.

Suppose first that every base-period product can be priced in period t. Let

$$\Delta \log p_{it} \equiv \log p_{it} - \log p_{i,t-1}.$$

By Shephard's lemma,

$$\frac{\partial \log e^h(\mathbf{p}, u_{t-1}^h)}{\partial \log p_i} = \frac{p_i q_i^h(\mathbf{p}, u_{t-1}^h)}{e^h(\mathbf{p}, u_{t-1}^h)}.$$

Evaluating this derivative at the base-period price vector $\mathbf{p}_{t-1}$ gives $s^h_{i,t-1}$. Therefore, a first-order expansion of the log expenditure function around $\mathbf{p}_{t-1}$ implies

$$\log \frac{e^h(\mathbf{p}_t, u^h_{t-1})}{e^h(\mathbf{p}_{t-1}, u^h_{t-1})} = \sum_{i \in \Omega_{t-1}} s^h_{i,t-1} \Delta \log p_{it} + O\Big(\|\Delta \log \mathbf{p}_t\|^2\Big). \tag{A.3}$$

Thus, to a first order, the household-level log cost-of-living index is an expenditure-weighted average of log price relatives.

The exact household-level Laspeyres index is

$$L^h_t \equiv \log \frac{\mathbf{q}^h_{t-1} \cdot \mathbf{p}_t}{\mathbf{q}^h_{t-1} \cdot \mathbf{p}_{t-1}}. \tag{A.4}$$

Using the base-period expenditure shares, this can be written as

$$L^h_t = \log \left(\sum_{i \in \Omega_{t-1}} s^h_{i,t-1} \exp(\Delta \log p_{it}) \right). \tag{A.5}$$

A second-order Taylor expansion of (A.5) gives

$$L^h_t = \sum_{i \in \Omega_{t-1}} s^h_{i,t-1} \Delta \log p_{it} + \frac{1}{2} \sum_{i \in \Omega_{t-1}} s^h_{i,t-1} \left(\Delta \log p_{it} - \sum_{\ell \in \Omega_{t-1}} s^h_{\ell,t-1} \Delta \log p_{\ell t} \right)^2 + O\Big(\|\Delta \log \mathbf{p}_t\|^3\Big). \tag{A.6}$$

Hence the expenditure-weighted average of log price relatives is the first-order log approximation to the exact Laspeyres index.

Because the base-period bundle is feasible at period-t prices, revealed preference implies

$$\log \frac{e^h(\mathbf{p}_t, u^h_{t-1})}{e^h(\mathbf{p}_{t-1}, u^h_{t-1})} \leq L^h_t. \tag{A.7}$$

The Laspeyres index is therefore an upper bound on the true household-level cost-of-living change. Its first-order log approximation is the object used in the empirical index.

Aggregating across households gives

$$COLI_t = \sum_h \lambda_h \log \frac{e^h(\mathbf{p}_t, u^h_{t-1})}{e^h(\mathbf{p}_{t-1}, u^h_{t-1})} = \sum_{i \in \Omega_{t-1}} s_{i,t-1} \Delta \log p_{it} + O\Big(\|\Delta \log \mathbf{p}_t\|^2\Big), \tag{A.8}$$

where

$$s_{i,t-1} \equiv \sum_h \lambda_h s^h_{i,t-1}. \tag{A.9}$$

This is the aggregate expenditure share used in Equation (3.9).

A.5.3 Product turnover and matched-model indexes

The preceding derivation assumes that every product purchased in period $t-1$ can be priced in period t. This assumption fails when products enter and exit. Let Ω_t denote the set of products observed in period t, and let

$$\mathcal{M}_t \equiv \Omega_t \cap \Omega_{t-1}$$

denote the set of products observed in both periods.

A matched-model index computes price changes only for products in $\mathcal{M}_t$:

$$L_t^{MM} = \sum_{i \in \mathcal{M}_t} s_{i,t-1} \log \frac{p_{it}}{p_{i,t-1}}. \tag{A.10}$$

This object differs from the first-order Laspeyres approximation in (A.8) because price relatives are missing for products in $\Omega_{t-1} \setminus \mathcal{M}_t$. If matched products are not representative of all base-period products, then (A.10) does not recover the first-order approximation to the cost-of-living index. The problem is selection on product survival: continuing products may have different quality change, price rigidity, markdown behavior, or life-cycle pricing than entering and exiting products.

A.5.4 Hedonic schedules

The hedonic approach replaces the missing period-t price of a base-period product with the price of its base-period characteristics under the period-t hedonic schedule. Let product i belong to product category $j(i) \in \mathcal{J}$, and let

$$X_{it} \in \mathcal{X}^{j(i)}$$

denote its observed characteristics. Following Rosen (1974), prices are equilibrium functions of characteristics:

$$p_{it} = P_{j(i),t}(X_{it}). \tag{A.11}$$

The schedule $P_{jt}(\cdot)$ is an equilibrium price schedule, not a structural demand curve. It does not by itself recover household preferences or supply functions.

For each base-period product $i \in \Omega_{t-1}$, define the hedonic log price relative

$$\Delta^H \log p_{it} \equiv \log P_{j(i),t}(X_{i,t-1}) - \log P_{j(i),t-1}(X_{i,t-1}). \tag{A.12}$$

Since product i is observed in period $t-1$,

$$P_{j(i),t-1}(X_{i,t-1}) = p_{i,t-1}.$$

The empirical analogue replaces the period-t schedule with the estimated schedule:

$$\widehat{\Delta^H \log p_{it}} \equiv \log \widehat{P}_{j(i),t}(X_{i,t-1}) - \log P_{j(i),t-1}(X_{i,t-1}). \tag{A.13}$$

The corresponding hedonic Laspeyres approximation is

$$\widehat{COLI}_t = \sum_{i \in \Omega_{t-1}} s_{i,t-1} \left[\log \widehat{P}_{j(i),t}(X_{i,t-1}) - \log P_{j(i),t-1}(X_{i,t-1}) \right]. \tag{A.14}$$

This is the same object as Equation (3.10). It prices the base-period characteristics of every product in Ω_{t-1}, including products that cannot be exactly matched to period-t products.

A.5.5 A scalar-quality example

To see why evaluating the period-t schedule at base-period characteristics is the relevant first-order object, consider one product category with a scalar quality attribute $Q \in \mathcal{Q}$. Let $P_t(Q)$ denote the period-t hedonic price schedule for that category. Collect all other goods in a vector $\mathbf{r}$, with price vector $\mathbf{p}_{-j,t}$.

Household h's expenditure function can be written as

$$e^h(P_t, \mathbf{p}_{-j,t}, \bar{u}) = \min_{Q,\mathbf{r}} \left\{ P_t(Q) + \mathbf{p}_{-j,t} \cdot \mathbf{r} : U^h(Q, \mathbf{r}) \geq \bar{u} \right\}. \tag{A.15}$$

Let

$$(Q^h_{t-1}, \mathbf{r}^h_{t-1})$$

denote the base-period bundle attaining u^h_{t-1}, and let

$$(\widetilde{Q}^h_t, \widetilde{\mathbf{r}}^h_t)$$

denote the period-t Hicksian bundle that attains the same utility u^h_{t-1}. Define

$$\Delta Q^h \equiv Q^h_{t-1} - \widetilde{Q}^h_t, \qquad \Delta \mathbf{r}^h \equiv \mathbf{r}^h_{t-1} - \widetilde{\mathbf{r}}^h_t.$$

Both bundles attain utility u^h_{t-1}. A first-order Taylor expansion of the utility function around the compensated period-t bundle gives

$$0 \approx U^h_Q(\widetilde{Q}^h_t, \widetilde{\mathbf{r}}^h_t)\Delta Q^h + \nabla_{\mathbf{r}} U^h(\widetilde{Q}^h_t, \widetilde{\mathbf{r}}^h_t) \cdot \Delta \mathbf{r}^h. \tag{A.16}$$

The first-order conditions for the compensated problem are

$$P'_t(\widetilde{Q}^h_t) = \mu^h_t U^h_Q(\widetilde{Q}^h_t, \widetilde{\mathbf{r}}^h_t), \tag{A.17}$$

$$\mathbf{p}_{-j,t} = \mu^h_t \nabla_{\mathbf{r}} U^h(\widetilde{Q}^h_t, \widetilde{\mathbf{r}}^h_t). \tag{A.18}$$

Multiplying (A.16) by μ^h_t and using (A.17)–(A.18) yields

$$0 \approx P'_t(\widetilde{Q}^h_t)\Delta Q^h + \mathbf{p}_{-j,t} \cdot \Delta \mathbf{r}^h. \tag{A.19}$$

Now expand the period-t hedonic schedule around $\widetilde{Q}^h_t$:

$$P_t(Q^h_{t-1}) \approx P_t(\widetilde{Q}^h_t) + P'_t(\widetilde{Q}^h_t)\Delta Q^h. \tag{A.20}$$

Using

$$\mathbf{r}^h_{t-1} = \widetilde{\mathbf{r}}^h_t + \Delta \mathbf{r}^h,$$

the period-t cost of the base-period bundle is

$$\begin{aligned} P_t(Q^h_{t-1}) + \mathbf{p}_{-j,t} \cdot \mathbf{r}^h_{t-1} &\approx P_t(\widetilde{Q}^h_t) + \mathbf{p}_{-j,t} \cdot \widetilde{\mathbf{r}}^h_t + P'_t(\widetilde{Q}^h_t)\Delta Q^h + \mathbf{p}_{-j,t} \cdot \Delta \mathbf{r}^h \\ &\approx e^h(P_t, \mathbf{p}_{-j,t}, u^h_{t-1}), \end{aligned} \tag{A.21}$$

where the last line uses (A.19).

Thus, up to second-order error, the period-t expenditure required to attain period-$t-1$ utility equals the period-t cost of the base-period bundle:

$$e^h(P_t, \mathbf{p}_{-j,t}, u^h_{t-1}) \approx P_t(Q^h_{t-1}) + \mathbf{p}_{-j,t} \cdot \mathbf{r}^h_{t-1}. \tag{A.22}$$

Taking logs and applying the first-order log approximation in (A.3), the contribution of this category to the household-level log cost-of-living index is

$$s^h_{j,t-1}\left[\log P_t(Q^h_{t-1}) - \log P_{t-1}(Q^h_{t-1})\right], \tag{A.23}$$

where

$$s^h_{j,t-1} = \frac{P_{t-1}(Q^h_{t-1})}{e^h(\mathbf{p}_{t-1}, u^h_{t-1})}$$

is the household's base-period expenditure share on the category.

This is the key hedonic adjustment. A comparison such as

$$\log P_t(Q^h_t) - \log P_{t-1}(Q^h_{t-1})$$

would combine price change with quality change, because the household may choose a different quality level in period t. The cost-of-living object instead holds base-period utility fixed. To a first order, this requires pricing the base-period quality Q^h_{t-1} under the period-t price schedule, $P_t(Q^h_{t-1})$.

A.5.6 General case

We now extend the scalar-quality argument to the general case used in the paper body. There are product categories $j \in \mathcal{J}$. A product in category j is described by a vector of characteristics

$$X \in \mathcal{X}^j \subseteq \mathbb{R}^{K_j}.$$

The period-t hedonic price schedule for category j is

$$P_{jt} : \mathcal{X}^j \to \mathbb{R}_+.$$

Thus, if product i belongs to category $j(i)$ and has characteristics X_{it}, its price is

$$p_{it} = P_{j(i),t}(X_{it}).$$

Let

$$\mathcal{P}_t \equiv \{P_{jt}\}_{j \in \mathcal{J}}$$

denote the collection of hedonic price schedules in period t. This price environment induces the price vector $\mathbf{p}_t$ used in the paper body.

We assume that, in a neighborhood of the relevant bundles, household h's problem can be represented as a choice over a fixed finite set of purchase positions $r \in \mathcal{R}^h$. Each purchase position has a category j_r^h, a quantity $q_r^h \geq 0$, and a characteristic vector $X_r^h \in \mathcal{X}^{j_r^h}$. This notation is only used for the derivation. In the final expression, each base-period purchase position corresponds to a base-period product $i \in \Omega_{t-1}$.

A bundle for household h is

$$b^h \equiv \left\{ q_r^h, X_r^h, j_r^h \right\}_{r \in \mathcal{R}^h}.$$

The period-t cost of this bundle is

$$C_t(b^h) \equiv \sum_{r \in \mathcal{R}^h} q_r^h P_{j_r^h,t}(X_r^h). \tag{A.24}$$

Household h's expenditure function can therefore be written as

$$e^h(\mathbf{p}_t, \bar{u}) = e^h(\mathcal{P}_t, \bar{u}) = \min_{b^h} \left\{ C_t(b^h) : U^h(b^h) \geq \bar{u} \right\}. \tag{A.25}$$

This is the same expenditure function as in the paper body, written in a way that makes explicit that differentiated-product prices are generated by hedonic schedules.

Let

$$b_{t-1}^h = \left\{ q_{r,t-1}^h, X_{r,t-1}^h, j_r^h \right\}_{r \in \mathcal{R}^h}$$

be household h's base-period bundle, and let

$$u_{t-1}^h = U^h(b_{t-1}^h).$$

Under local nonsatiation, the base-period bundle is expenditure-minimizing for the utility level it attains, so

$$e^h(\mathbf{p}_{t-1}, u_{t-1}^h) = C_{t-1}(b_{t-1}^h) = \sum_{r \in \mathcal{R}^h} q_{r,t-1}^h P_{j_r^h,t-1}(X_{r,t-1}^h). \tag{A.26}$$

Let

$$\widetilde{b}_t^h = \left\{ \widetilde{q}_{r,t}^h, \widetilde{X}_{r,t}^h, j_r^h \right\}_{r \in \mathcal{R}^h}$$

denote the period-t Hicksian bundle that attains the base-period utility level u_{t-1}^h. Thus

$$e^h(\mathbf{p}_t, u_{t-1}^h) = C_t(\widetilde{b}_t^h) = \sum_{r \in \mathcal{R}^h} \widetilde{q}_{r,t}^h P_{j_r^h,t}(\widetilde{X}_{r,t}^h). \tag{A.27}$$

Define the changes from the period-t compensated bundle to the base-period bundle as

$$\Delta q_r^h \equiv q_{r,t-1}^h - \widetilde{q}_{r,t}^h, \qquad \Delta X_r^h \equiv X_{r,t-1}^h - \widetilde{X}_{r,t}^h. \tag{A.28}$$

Because both b_{t-1}^h and $\widetilde{b}_t^h$ deliver utility u_{t-1}^h, a first-order Taylor expansion of U^h around $\widetilde{b}_t^h$ gives

$$\begin{aligned} 0 &= U^h(b_{t-1}^h) - U^h(\widetilde{b}_t^h) \\ &= \sum_{r\in\mathcal{R}^h} U_{q_r}^h(\widetilde{b}_t^h)\Delta q_r^h + \sum_{r\in\mathcal{R}^h} \nabla_{X_r} U^h(\widetilde{b}_t^h)'\Delta X_r^h + O\left(\|\Delta b_t^h\|^2\right), \end{aligned} \tag{A.29}$$

where

$$\Delta b_t^h \equiv \left\{\Delta q_r^h, \Delta X_r^h\right\}_{r\in\mathcal{R}^h}.$$

Here $\nabla_{X_r} U^h(\widetilde{b}_t^h)$ is the $K_{j_r^h}$-dimensional gradient of utility with respect to the characteristics of purchase position r:

$$\nabla_{X_r} U^h(\widetilde{b}_t^h) = \left(\frac{\partial U^h(\widetilde{b}_t^h)}{\partial X_{r1}}, \ldots, \frac{\partial U^h(\widetilde{b}_t^h)}{\partial X_{rK_{j_r^h}}}\right)'.$$

The period-t compensated problem is

$$\min_{b^h}\left\{\sum_{r\in\mathcal{R}^h} q_r^h P_{j_r^h,t}(X_r^h) : U^h(b^h) \geq u_{t-1}^h\right\}.$$

Assume the relevant choices are interior and that U^h and P_{jt} are twice continuously differentiable in a neighborhood of $\widetilde{b}_t^h$. Let μ_t^h be the multiplier on the utility constraint. The first-order conditions at $\widetilde{b}_t^h$ are, for each purchase position r,

$$P_{j_r^h,t}(\widetilde{X}_{r,t}^h) = \mu_t^h U_{q_r}^h(\widetilde{b}_t^h), \tag{A.30}$$

and

$$\widetilde{q}_{r,t}^h \nabla_X P_{j_r^h,t}(\widetilde{X}_{r,t}^h) = \mu_t^h \nabla_{X_r} U^h(\widetilde{b}_t^h). \tag{A.31}$$

Component by component, (A.31) says that for each characteristic dimension $k = 1, \ldots, K_{j_r^h}$,

$$\widetilde{q}_{r,t}^h \frac{\partial P_{j_r^h,t}(\widetilde{X}_{r,t}^h)}{\partial X_{rk}} = \mu_t^h \frac{\partial U^h(\widetilde{b}_t^h)}{\partial X_{rk}}. \tag{A.32}$$

Thus the gradient of the hedonic price schedule equals the household's local marginal willingness to pay for each quality dimension, scaled by the quantity of the good.

Multiplying (A.29) by μ_t^h and using (A.30) and (A.31) yields

$$0 = \sum_{r\in\mathcal{R}^h} P_{j_r^h,t}(\widetilde{X}_{r,t}^h)\Delta q_r^h + \sum_{r\in\mathcal{R}^h} \widetilde{q}_{r,t}^h \nabla_X P_{j_r^h,t}(\widetilde{X}_{r,t}^h)'\Delta X_r^h + O\left(\|\Delta b_t^h\|^2\right). \tag{A.33}$$

This is the multidimensional analogue of the cancellation condition in the single-quality case. The first-order change in utility from moving from the period-t compensated bundle to the base-period bundle is zero. Therefore, after using the household's first-order conditions, the corresponding first-order change in cost is also zero.

Next consider the period-t cost of the base-period bundle:

$$C_t(b_{t-1}^h) = \sum_{r\in\mathcal{R}^h} q_{r,t-1}^h P_{j_r^h,t}(X_{r,t-1}^h). \tag{A.34}$$

Using

$$q^h_{r,t-1} = \widetilde{q}^h_{r,t} + \Delta q^h_r, \qquad X^h_{r,t-1} = \widetilde{X}^h_{r,t} + \Delta X^h_r,$$

we can write

$$C_t(b^h_{t-1}) = \sum_{r\in\mathcal{R}^h} \left(\widetilde{q}^h_{r,t} + \Delta q^h_r\right) P_{j^h_r,t}\left(\widetilde{X}^h_{r,t} + \Delta X^h_r\right). \tag{A.35}$$

A first-order Taylor expansion of each hedonic schedule gives

$$P_{j^h_r,t}\left(\widetilde{X}^h_{r,t} + \Delta X^h_r\right) = P_{j^h_r,t}(\widetilde{X}^h_{r,t}) + \nabla_X P_{j^h_r,t}(\widetilde{X}^h_{r,t})'\Delta X^h_r + O\left(\|\Delta X^h_r\|^2\right). \tag{A.36}$$

Substituting (A.36) into (A.35) gives

$$\begin{aligned} C_t(b^h_{t-1}) &= \sum_{r\in\mathcal{R}^h} \left(\widetilde{q}^h_{r,t} + \Delta q^h_r\right)\left[P_{j^h_r,t}(\widetilde{X}^h_{r,t}) + \nabla_X P_{j^h_r,t}(\widetilde{X}^h_{r,t})'\Delta X^h_r\right] + O\left(\|\Delta b^h_t\|^2\right) \\ &= \sum_{r\in\mathcal{R}^h} \widetilde{q}^h_{r,t} P_{j^h_r,t}(\widetilde{X}^h_{r,t}) \\ &\quad + \sum_{r\in\mathcal{R}^h} P_{j^h_r,t}(\widetilde{X}^h_{r,t})\Delta q^h_r \\ &\quad + \sum_{r\in\mathcal{R}^h} \widetilde{q}^h_{r,t}\nabla_X P_{j^h_r,t}(\widetilde{X}^h_{r,t})'\Delta X^h_r \\ &\quad + O\left(\|\Delta b^h_t\|^2\right). \end{aligned} \tag{A.37}$$

The omitted terms include products such as

$$\Delta q^h_r \nabla_X P_{j^h_r,t}(\widetilde{X}^h_{r,t})'\Delta X^h_r,$$

as well as Hessian terms from the Taylor expansion of the hedonic schedule; all are second order.

The first term in (A.37) is the exact compensated expenditure in period t:

$$\sum_{r\in\mathcal{R}^h} \widetilde{q}^h_{r,t} P_{j^h_r,t}(\widetilde{X}^h_{r,t}) = e^h(\mathbf{p}_t, u^h_{t-1}).$$

The next two terms are zero up to second-order error by (A.33). Therefore,

$$C_t(b^h_{t-1}) = e^h(\mathbf{p}_t, u^h_{t-1}) + O\left(\|\Delta b^h_t\|^2\right). \tag{A.38}$$

Equivalently, by Equation (A.34),

$$e^h(\mathbf{p}_t, u^h_{t-1}) = \sum_{r\in\mathcal{R}^h} q^h_{r,t-1} P_{j^h_r,t}(X^h_{r,t-1}) + O\left(\|\Delta b^h_t\|^2\right). \tag{A.39}$$

Equation (A.39) is the key result. It says that the period-t expenditure required to attain base-period utility is, to a first order, the period-t cost of the base-period bundle, where each base-period product is repriced using the period-t hedonic schedule evaluated at its base-period characteristics.

Recall that the object of interest is,

$$COLI_t \equiv \sum_h \lambda_h \log \frac{e^h(\mathbf{p}_t, u^h_{t-1})}{e^h(\mathbf{p}_{t-1}, u^h_{t-1})} \tag{A.40}$$

The denominator of the household-level COLI is exact. Since b^h_{t-1} is the bundle chosen by household h in period $t-1$,

$$e^h(\mathbf{p}_{t-1}, u^h_{t-1}) = \sum_{r\in\mathcal{R}^h} q^h_{r,t-1} P_{j^h_r,t-1}(X^h_{r,t-1}). \tag{A.41}$$

Therefore,

$$\log \frac{e^h(\mathbf{p}_t, u^h_{t-1})}{e^h(\mathbf{p}_{t-1}, u^h_{t-1})} = \log \frac{\sum_{r\in\mathcal{R}^h} q^h_{r,t-1} P_{j^h_r,t}(X^h_{r,t-1})}{\sum_{r\in\mathcal{R}^h} q^h_{r,t-1} P_{j^h_r,t-1}(X^h_{r,t-1})} + O\Big(\|\Delta b^h_t\|^2\Big). \tag{A.42}$$

The second-order remainder remains second order after taking logs, provided household expenditure is bounded away from zero.

Define household h's base-period expenditure share on product r as

$$s^h_{r,t-1} \equiv \frac{q^h_{r,t-1} P_{j^h_r,t-1}(X^h_{r,t-1})}{\sum_{\ell\in\mathcal{R}^h} q^h_{\ell,t-1} P_{j^h_\ell,t-1}(X^h_{\ell,t-1})}. \tag{A.43}$$

Also define the hedonic log price relative for base-period product r as

$$\Delta^H \log p_{rt} \equiv \log P_{j^h_r,t}(X^h_{r,t-1}) - \log P_{j^h_r,t-1}(X^h_{r,t-1}). \tag{A.44}$$

Using (A.43), the first term on the right-hand side of (A.42) can be written as

$$\begin{aligned}\log \frac{\sum_{r\in\mathcal{R}^h} q^h_{r,t-1} P_{j^h_r,t}(X^h_{r,t-1})}{\sum_{r\in\mathcal{R}^h} q^h_{r,t-1} P_{j^h_r,t-1}(X^h_{r,t-1})} &= \log\left[\sum_{r\in\mathcal{R}^h} s^h_{r,t-1} \frac{P_{j^h_r,t}(X^h_{r,t-1})}{P_{j^h_r,t-1}(X^h_{r,t-1})}\right] \\ &= \log\left[\sum_{r\in\mathcal{R}^h} s^h_{r,t-1} \exp\Big(\Delta^H \log p_{rt}\Big)\right]. \end{aligned} \tag{A.45}$$

A first-order expansion of the log-sum expression around $\Delta^H \log p_{rt} = 0$ for all r gives

$$\log\left[\sum_{r\in\mathcal{R}^h} s^h_{r,t-1} \exp\Big(\Delta^H \log p_{rt}\Big)\right] = \sum_{r\in\mathcal{R}^h} s^h_{r,t-1} \Delta^H \log p_{rt} + O\Big(\|\Delta^H \log \mathbf{p}^h_t\|^2\Big). \tag{A.46}$$

Combining (A.42) and (A.46), the household-level log COLI is therefore

$$\log \frac{e^h(\mathbf{p}_t, u^h_{t-1})}{e^h(\mathbf{p}_{t-1}, u^h_{t-1})} = \sum_{r\in\mathcal{R}^h} s^h_{r,t-1} \Delta^H \log p_{rt} + R^h_t, \tag{A.47}$$

where

$$R^h_t = O\Big(\|\Delta b^h_t\|^2\Big) + O\Big(\|\Delta^H \log \mathbf{p}^h_t\|^2\Big). \tag{A.48}$$

The first term in R^h_t is the second-order error from holding fixed the base-period bundle rather than allowing the household to reoptimize. The second term is the second-order error from replacing the exact log Laspeyres expression with an expenditure-weighted average of log price relatives.

We now aggregate across households. The aggregate log COLI is

$$\begin{aligned} COLI_t &= \sum_h \lambda_h \log \frac{e^h(\mathbf{p}_t, u^h_{t-1})}{e^h(\mathbf{p}_{t-1}, u^h_{t-1})} \\ &= \sum_h \lambda_h \sum_{r\in\mathcal{R}^h} s^h_{r,t-1} \Delta^H \log p_{rt} + \sum_h \lambda_h R^h_t. \end{aligned} \tag{A.49}$$

Let $\mathcal{R}_{t-1}$ denote the set of all products purchased in the base period, and set $q^h_{r,t-1} = 0$ for households that do not purchase product r. Then the aggregate base-period expenditure share on product r is

$$s_{r,t-1} \equiv \sum_h \lambda_h s^h_{r,t-1}. \tag{A.50}$$

Using this definition, (A.49) becomes

$$COLI_t = \sum_{r \in \mathcal{R}_{t-1}} s_{r,t-1} \Delta^H \log p_{rt} + R_t, \tag{A.51}$$

where

$$R_t \equiv \sum_h \lambda_h R_t^h.$$

Substituting the definition of the hedonic log price relative gives

$$COLI_t = \sum_{r \in \mathcal{R}_{t-1}} s_{r,t-1} \log \frac{P_{j_r,t}(X_{r,t-1})}{P_{j_r,t-1}(X_{r,t-1})} + R_t. \tag{A.52}$$

This is the population hedonic Laspeyres approximation. It says that, up to second-order terms, the aggregate log COLI equals a base-period expenditure-share weighted average of the log change in the hedonic price of each base-period product's characteristics.

In the empirical implementation, the period-t counterfactual price $P_{j_r,t}(X_{r,t-1})$ may not be observed because product r may no longer be sold in period t. We therefore replace it with the fitted value from the estimated period-t hedonic schedule, $\widehat{P}_{j_r,t}(X_{r,t-1})$. Since $P_{j_r,t-1}(X_{r,t-1}) = p_{r,t-1}$ is observed for all base-period products, the implemented index is

$$\widehat{COLI}_t = \sum_{r \in \mathcal{R}_{t-1}} s_{r,t-1} \log \frac{\widehat{P}_{j_r,t}(X_{r,t-1})}{P_{j_r,t-1}(X_{r,t-1})}. \tag{A.53}$$

Thus, apart from estimation error in the hedonic schedules, the hedonic Laspeyres index differs from the true log COLI only by second-order terms: those due to household reoptimization and those due to the log-linear Laspeyres approximation.

A.6 Extracting Structured Features from the Data

Our analysis in Section 4.1.1 requires extracting detailed structured characteristics about products from free text data. First, we must identify important characteristics about the products of interest. We select 28 technology-intensive goods to extract attributes for, which we list in Table A.2.

For each of the 28 product categories, we query an LLM[21] to identify features that are important determinants of the price and quality of that product. We also provide the LLM a sample of 5 randomly selected products for each year of the data to guide its attribute selection. Our exact prompt is included below,

```
### SYSTEM ROLE
You are an expert Data Archivist and Product Analyst specializing in historical
{category_name} from the period {years}. Your goal is to define a structured schema
based on unstructured text to help collectors and consumers make purchasing
decisions.

### TASK
Analyze the provided list of `product_descriptions`. Generate a comprehensive list of
distinct attributes (schema) that can be used to describe these items.

### DEFINITIONS
```

[21] We use Gemini 3 Pro

```
* **Binary Attributes:** True/False values (e.g., `is_plastic`,
`has_original_packaging`).
* **Numeric Attributes:** Measurable values (e.g., `weight_kg`, `lens_diameter_mm`).

### GUIDELINES
1.  Ignore any products that do not strictly belong to the category: {category_name}.
2.  All attributes must be either binary or numeric. Categorical attributes should be
split into several binary attributes for the main categories (e.g. instead of having
a `material` attribute use `is_plastic`, `is_wood`, etc. )
3.  Attributes must be objectively verifiable (fact-based). Ignore subjective
descriptors like 'beautiful' or 'rare.' Ignore brand names.
4.  Avoid synonyms. If `width_mm` exists, do not add `width_cm`.
5.  Include both major characteristics and niche features, even if they only apply to
a subset of products (e.g., obsolete features).
6.  All attribute names must be in `snake_case`.

### OUTPUT FORMATTING
* **Format:** Raw CSV string.
* **Quoting:** Wrap cells containing commas in double quotes ("").
* **Strictness:** Do NOT include markdown formatting (like ```csv), code blocks, or
conversational text. Return *only* the text.

### CSV COLUMNS
1.  `attribute_name`: (snake_case)
2.  `attribute_description`: (Brief explanation of what this measures/indicates)
3.  `is_binary`: (True for binary, False for numeric)

### INPUT DATA
`product_descriptions`:
{product_descriptions}
```

This methodology yields approximately 50 attributes for each good. We then manually review each attribute list to ensure that there are no mistakes and the attributes are comprehensive and not redundant.

Once we compile attribute lists for each product category, we query the LLM once more to extract values for each attribute, for each product in the category. The exact prompt we use is provided below,

```
### Role & Context
You are an expert Data Archivist and Product Analyst specializing in historical
{category_name} catalog data from the year {year}. Your specific expertise lies in
structuring unstructured product descriptions into machine-readable datasets.

### Task
You will be provided with a list of products. Each product entry contains:
1.  Unique ID (Format: "X_Y").
2.  Product Name & Version Name (identifying a specific SKU).
3.  Product Description (Text that may cover multiple versions of the product
simultaneously).
```

```
You will also be provided with Target Attributes (A specific list of attributes you
must look for). These are separated into Numeric Attributes and Binary Attributes.
Each attribute is in snake case and is accompanied by a short description.


Your goal is to analyze the text and extract the precise value for every requested
attribute for each specific version.

### Extraction Rules

1. If a product does not belong to the category '{category_name}', you MUST return an
empty JSON object for that product {{}}.

2. Version Disambiguation (Crucial):
   - The description often describes a base product and then lists variations (e.g.,
   "The Standard model features X, while the Deluxe model adds Y").
   - If a feature is described generally (e.g., "All models made of oak"), apply it
   to all versions.
   - If a feature is specific (e.g., "No. 2 is 12 inches wide"), apply it only to
   that specific version.

3. Data Typing:
   - Numeric Attributes: Extract the number only as a float or integer. Remove units
   (e.g., if text says "10 inches", extract `10`).
   - Binary Attributes: Use `true` or `false`.
     - True: The text explicitly states or implies the feature exists.
     - False: The text explicitly states or implies the feature does not exist.
   - Null Values: If the text does not mention the attribute, and it cannot be
   logically inferred from the historical context of {year}, return `null` (do not
   use string "nan"). Do not guess.
If the product matches '{category_name}', the output JSON must include every key
listed in product_attributes (use null if the value is missing). If the product does
not match, return {{}}.

3. Historical Context:
   - Use your knowledge of {year} to interpret vague terms (e.g., in a 1900 catalog,
   "standard metal" implies steel or iron, not aluminum, unless specified).

### Input Data
[Insert your input data here]

### Output Format
Return a single JSON object. The keys must be the Unique IDs ("X_Y"). The values must
be dictionaries of the attributes.
Example Output:
{{
  "1_1": {{
    "length_ft": 4,
    "material_oak": true,
    "has_carved_legs": true
```

```
  }},
  "1_2": {{
    "length_ft": 6,
    "material_oak": true,
    "has_carved_legs": false
  }}
}}

`product_descriptions`
{product_descriptions}

`product_attributes`
{product_attributes}
```

Table A.2: Technology-Intensive Goods

Product Name	HS Codes
Air Conditioners	841510, 841581, 841582, 841583
Refrigerators And Ice Boxes	841810, 841821, 841822, 841829
Freezers	841830, 841840
Stoves	851660, 732111, 732112, 732113
Heaters	732181, 732182, 732183, 732211 732219, 732290, 841911, 841919 851610, 851621, 851629
Water Purifyers	842121
Dishwashers	842211
Scales	842310, 842330, 842381, 842382
Lawn Mowers	843311, 843319
Wahing Machines	845011, 845012, 845019, 845020 845090
Drying Machines	845121, 845129
Ironing Machines	845130, 851640
Sewing Machines	845210
Typewriters	846910, 846921, 846929, 846931 846939
Calculators	847010, 847021, 847029, 847030
Batteries	850611, 850612, 850613, 850619 850620
Vacuum Cleaners	850910
Garbage Disposals	850930
Microwaves	851650
Coffee Makers	851671
Toasters	851672
Loudspeakers	851821, 851822, 851829
Headphones	851830
Sound Reproduction Devices	851921, 851929, 851931, 851939 851991, 851999
Cameras	852530, 900610, 900620, 900630 900640, 900651, 900652, 900653 900659, 900711, 900719
Radios	852711, 852719, 852721, 852729 852731, 852732, 852739
Tvs	852810, 852820
Projectors	900721, 900729

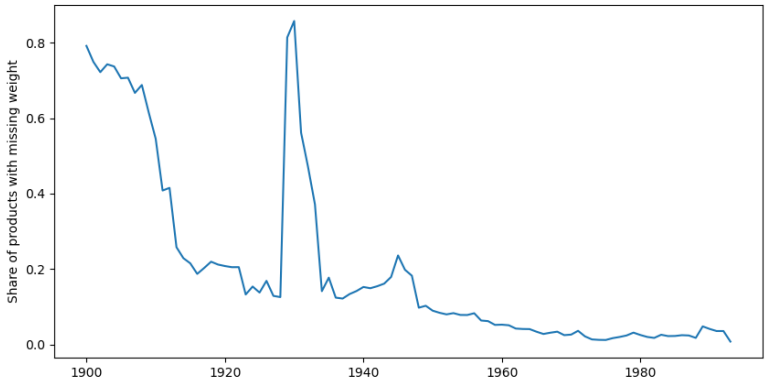
Table A.3: Air Conditioner Attributes

Attribute Name	Attribute Description	Is Binary
cooling_capacity_btu	The cooling power of the unit measured in British Thermal Units per hour	False
heating_capacity_btu	The heating power of the unit measured in British Thermal Units per hour (for heat pumps or units with heaters)	False
cooling_capacity_tons	The cooling power of the unit measured in tons (common for central systems)	False
airflow_cfm	The volume of air the unit moves, measured in Cubic Feet Per Minute	False
coverage_area_sq_ft	The recommended room size the unit can effectively cool, in square feet	False
dehumidification_pints_daily	The amount of moisture removed from the air in 24 hours	False
fan_speeds_count	The number of distinct speed settings available for the fan	False
energy_efficiency_ratio_eer	The ratio of cooling capacity (BTU) to power input (Watts)	False
seasonal_energy_efficiency_ratio_seer	The total cooling output during a typical cooling-season divided by the total electric energy input	False
voltage_v	The electrical voltage required to operate the unit	False
amperage_amps	The electrical current drawn by the unit during operation	False
wattage_watts	The electrical power consumed by the unit	False
horsepower_hp	The power output of the unit's motor	False
weight_lbs	The shipping or unit weight in pounds	False
height_inches	The vertical dimension of the unit in inches	False
width_inches	The horizontal dimension of the unit in inches	False
depth_inches	The measurement from front to back of the unit in inches	False
min_window_width_inches	The minimum window width required for installation	False
max_window_width_inches	The maximum window width accommodated by the installation kit	False
cord_length_feet	The length of the electrical power cord	False
compressor_warranty_years	The duration of the warranty specifically covering the compressor	False
is_window_unit	Indicates if the unit is designed to be mounted in a window	True
is_central_system	Indicates if the unit is part of a whole-home central cooling system	True
is_portable	Indicates if the unit is designed to be moved easily between rooms (e.g., has handles, lightweight)	True
is_automotive_unit	Indicates if the unit is designed for installation in a vehicle	True
is_evaporative_cooler	Indicates if the unit uses water evaporation for cooling (swamp cooler) rather than a refrigerant compressor	True
is_heat_pump	Indicates if the unit can reverse its cycle to provide heating as well as cooling	True
is_through_the_wall	Indicates if the unit is designed to be installed through a wall sleeve	True
is_casement_window_unit	Indicates if the unit is specifically designed for tall, narrow sliding or casement windows	True
is_gas_powered	Indicates if the unit (or its heating component) runs on natural gas	True
has_thermostat	Indicates if the unit has a built-in thermostat for temperature control	True
has_automatic_humidity_control	Indicates if the unit has a humidistat to control humidity levels independent of cooling	True
has_humidifier	Indicates if the unit has a feature to add moisture to the air	True
has_electronic_air_cleaner	Indicates if the unit includes an electronic filtration system	True
has_fresh_air_intake	Indicates if the unit can draw fresh air from outside	True
has_exhaust_vent	Indicates if the unit can pump stale room air outside	True
has_slide_out_chassis	Indicates if the internal components can slide out of the cabinet for easier installation or service	True
has_power_saver_switch	Indicates if the unit has a feature to cycle the fan off when the compressor turns off	True
has_automatic_timer	Indicates if the unit has a timer to schedule on/off operations	True
has_filter_monitor	Indicates if the unit has a light or gauge to signal when the filter needs cleaning	True
has_adjustable_louvers	Indicates if the unit has vents that can be adjusted to direct airflow	True
has_oscillating_louvers	Indicates if the unit has powered louvers that sweep air back and forth automatically	True
has_woodgrain_finish	Indicates if the exterior features a simulated wood grain appearance	True
has_concealed_controls	Indicates if the control panel is hidden behind a door or panel	True
is_steel_cabinet	Indicates if the outer housing is made of steel	True
is_plastic_cabinet	Indicates if the outer housing is made of plastic	True
is_flush_mount	Indicates if the unit is designed to sit flush with the interior wall/curtain line	True

A.7 Interpolating Weight

A large majority of the product listings in our data report shipping weight, since this was used to calculate shipping costs for goods. However, shipping weight data is particularly sparse in the first five years of the data as well as in 1929-1930. The share of products missing shipping weight is shown in Figure A.16. The missing shipping weight creates problems in estimating Equation (3.11), where we want shipping weight to be un-penalized, since it has strong predictive power for the years when it is available.

Figure A.16: Share of Data Without Shipping Weight, Over Time



The figure shows the share of data in the Sears catalog each year that do not report shipping weight. The share is high for the first ten years of our sample but then falls to about 20% for 1910-1930. At the onset of the Great Depression, the share surges again since shipping costs were canceled during these years, presumably to stimulate demand. In the later half of the 1930s, shipping cost was again reported for most items and continued to be reported for most items for the rest of the century.

Therefore, we interpolate weight when it is missing, using a flexible ridge specification. In particular, we estimate a regression for each 6-digit HS code, j, and $\tau \in [1900, 1990]$,

$$\log W_{ijt} = \lambda_j^{(\tau)} + \phi_{ijt}'\gamma_j^{(\tau)} + \psi_j^{(\tau)} t + \lambda_j^{(\tau)} t^2, \quad t \in [\tau - 5, \tau + 5] \tag{A.54}$$

where ϕ_{ijt} is the matrix of high-dimensional text embeddings, which is described in detail in Section 3. We also include a quadratic in time t in case there are systematic patterns in how the weight of a good within a 6-digit HS code, j, changes over time. The regressions are estimated using 11-year overlapping windows, $[\tau - 5, \tau + 5]$. As in Section 3, we estimate Equation (A.54) using a Ridge regression, since the text embeddings are large and highly correlated. We then obtain a fitted estimate of weight for each observation, i at time t,

$$\hat{W}_{ijt}^{(t)} = \hat{\lambda}_j^{(t)} + \phi_{ijt}'\hat{\gamma}_j^{(t)} + \hat{\psi}_j^{(t)} t + \hat{\lambda}_j^{(t)} t^2$$

When weight is unobserved, we interpolate it using $\hat{W}_{ijt}$. The interpolated weight measure is available for 99.2% of the observations in our sample, which means that we can now estimate Equation (3.11) using the vast majority of the catalog data.

A.8 Ridge Regression

We are interested in estimating a regression of the form,

$$\log P_{ijt} \;=\; \alpha_{jt} + \boldsymbol{\phi}_{ijt}'\beta_{jt} + Z_{ijt}'\beta_{jt}^Z + \varepsilon_{ijt}, \tag{A.55}$$

where α_{jt} is a constant, $\beta_{jt} \in \mathbb{R}^d$ are the embedding coefficients, and β_{jt}^Z are the coefficients on covariates Z. ϕ_{ijt}' are high-dimensional text embeddings, as we described in Section 3. We implement a ridge

estimator, which solves

$$(\widehat{\alpha}_{jt}(\lambda), \widehat{\beta}_{jt}(\lambda), \widehat{\beta}^{Z}_{jt}(\lambda)) \in \arg\min_{\alpha,\beta,\beta^Z} \left\{ \frac{1}{N} \sum_i \left(\log P_{ijt} - \alpha - \boldsymbol{\phi}'_{ijt}\beta - Z'_{ijt}\beta^Z \right)^2 \; + \; \lambda \|\beta\|_2^2 \right\}, \tag{A.56}$$

where we impose the penalty only on β, while α and β^Z are left unpenalized.

Computationally, we implement (A.56) via a Frisch–Waugh–Lovell (FWL) decomposition. Let D denote the matrix concatenating a constant and the matrix Z and define the residual-maker matrix

$$M_D \;=\; I - D(D'D)^{-1}D'.$$

We partial out the matrix D from both the outcome and the embeddings by forming

$$\widetilde{y} \equiv M_D \log P, \qquad \widetilde{\Phi} \equiv M_D \, \Phi,$$

where Φ is the $N \times d$ matrix stacking $\boldsymbol{\phi}'_{it}$. Intuitively, $(\widetilde{y}, \widetilde{\Phi})$ retain only de-meaned variation controlling for covariates Z, so any subsequent regression uses identifying variation orthogonal to D.

Given these residualized variables, we estimate the penalized coefficients by Ridge on the within-year components:

$$\widehat{\beta}_j(\lambda) \in \arg\min_{\beta} \left\{ \frac{1}{N} \|\widetilde{y} - \widetilde{\Phi}\beta\|_2^2 + \lambda \|\beta\|_2^2 \right\}. \tag{A.57}$$

We choose λ by K-fold cross-validation, minimizing out-of-sample mean squared error for predicting $\log P_{it}$, and we standardize embedding coordinates within each estimation window so that the penalty treats dimensions comparably.

Finally, conditional on $\widehat{\beta}(\lambda)$, we recover the *unpenalized* coefficients by OLS:

$$(\widehat{\alpha}_j(\lambda), \widehat{\beta}^Z_j(\lambda)) = \arg\min_{\alpha,\beta^Z} \frac{1}{N} \sum_i \left(\log P_{ijt} - \boldsymbol{\phi}'_{it}\widehat{\beta}(\lambda) - Z'_{ijt}\beta^Z - \alpha \right)^2 \tag{A.58}$$

Equations (A.57)–(A.58) produce the same fitted values as directly solving (A.56).