Russell Tran

Professor Chris Makler

CS 191W

17 March 2022

Information asymmetry in online soap ingredient markets and opportunities for improved efficiency in ecommerce

Executive summary

Five artisan soap-making small business owners from the United States were interviewed for their perspectives on soap ingredient sourcing challenges to identify general trends in the small business chemical supply chains. An analysis of price variation in ingredient markets was then used to validate their claims. A custom set of web scraping and browser automation tools were prototyped both to facilitate the collection of unstructured data for the analysis and to collect user feedback from demonstrations. The results show greater price competition among suppliers when their prices are listed in a central platform (Amazon) as opposed to when they are listed on independent web storefronts on the open Internet. All of this has implications for opportunities to improve market efficiency in ecommerce where otherwise overlooked.

Background

The Law of One Price in economics suggests that the prices of an identical commodity from different locations should converge in a frictionless market where there is perfect information (Potters). The state of the Internet and general Internet literacy today tend to imply that ecommerce has successfully achieved this Platonic ideal. For instance, Google Search liberates tremendous amounts of information that buyers and sellers would otherwise not have

access to, and at virtually zero cost. Moreover, an industry of "dropshippers" exists which conducts arbitrage across the Internet by reselling products offered by websites that are more obscure to the general public. And yet price disparities on the Internet are still surprisingly common. The project described in this report effectively demonstrates that Google Search does not fully liberate price information and that dropshippers have not achieved full saturation.

Motivation

A friend who makes artisan soaps complained to the author that it was difficult to 1) find soap making supplies on the Internet, and 2) do price comparisons among identical ingredients from different sources. Because he made his soap in small, ad hoc batches for his product demonstrations, the per-unit price listings for ingredients in small quantities always incorporated high markups. The variances in these markups were sufficient to justify doing tedious price comparisons in spreadsheets. Moreover, because of pandemic-induced supply chain disruptions, the online suppliers tended to impose 2-3 week shipping delays, raise prices abruptly, or mark regular products out of stock.

These complaints matched the ones prevalent on soapmaking Facebook groups that number in excess of 30,000 members. A key archetype of the "soapmaking struggle" was having 20 browser tabs open to compare prices and availability among providers for a single identical product. The Facebook groups also dedicated significant discussion to recommending to their peers suppliers with a reliable reputation and warning them of suppliers who are untrustworthy. These complaints imply that friction still exists on the Internet.

Lastly, inflation in the economy has encouraged ecommerce stores to implement automated price changes to their products that operate in real-time in response to changing

market conditions (Wakabayashi). This represents a threat to buyers who are then caught in information asymmetry because they had previously mapped the landscape of seller prices by slow, manual comparison. Therefore they would be unable to keep up and make price comparisons in a timely manner. It suggests that commensurate automation technology should exist on behalf of buyers to level the asymmetry.

Stakeholder research

Five soapmaking small business owners were interviewed to confirm the hypothesis of friction in the marketplace. The interview highlights are listed below.

Natalie J. of Alabama:

- There is sufficient price fluctuation as of January 2022 that as a home business, Natalie has to make a new spreadsheet from scratch each month to do price comparisons. Price comparisons take 10 hours for a total ingredient purchase of $1,500-$2,500.
- Sourcing is punctuated in synchronization with soap batch-making timelines.
- Buying at the bulk size of 55 gallon drums is sometimes justified for a home business to save on costs.

Benjamin M. of Illinois:

- Soapmaker businesses have different supplier needs at three different stages of growth: hobbyist (sell to friends), home business (sell to events), and enterprise scale (sell to retailers).

- The most significant growing pain that occurs between each stage is the need to form a new supply chain from scratch that can accommodate greater volumes and offer bulk discounts commensurate with the new size of the business.
- At the enterprise scale, supplier price information remains entirely off the Internet.
- Distributors who are abroad as opposed to local (e.g., those on Alibaba) are disfavored even though they may offer better prices because it makes a particular supply line more vulnerable to shocks.
- Even though he already has established his supply chain, Benjamin would love a hypothetical price aggregating software to "check the rearview mirror" and ensure his suppliers are pricing him fairly against the market.

Frencesca M. of California:

- Sources at least once per month, and takes two full business days to complete it each time.
- Determines her batch ingredient quantities using an automated software tool called SoapMaker, but nonetheless relies on manual searches and word of mouth to find suppliers.
- Regards the sourcing process with "dread" and "stress," even after having completed it, because she is uncertain whether her purchases were actually optimal.
- In response to a demonstration of the web-scraping prototype, suggested that approximate shipping time be incorporated into the data, and to standardize quantity units, since suppliers differed between gallons, liters, pounds, kilograms, etc.

Kari C. of Montana:

- Lack of factories or major distributors in Montana results in her sourcing almost exclusively from the East and West Coasts.
- Describes the sourcing process as “awful”; as of January 2022 spends time sourcing daily.
- Due to pandemic economics, had to sink more of her working capital into supplies because many suppliers became only willing to ship by the pallet instead of in smaller denominations.
- Sourcing is relevant both for R&D and batch making, especially because home businesses such as hers constantly launch new product offerings.
- The greatest frustrations with suppliers arise from their turnaround, shipping, price, and oil quality.

Laura H. of Texas:

- Sourcing problems have become so frustrating that she is going to drop her business altogether if she cannot figure out a way to make sourcing more streamlined.
- Does not use spreadsheets for price comparisons, but rather scratch paper notes.

Information from the interviewees is also aggregated in the following graph (Figure 1):

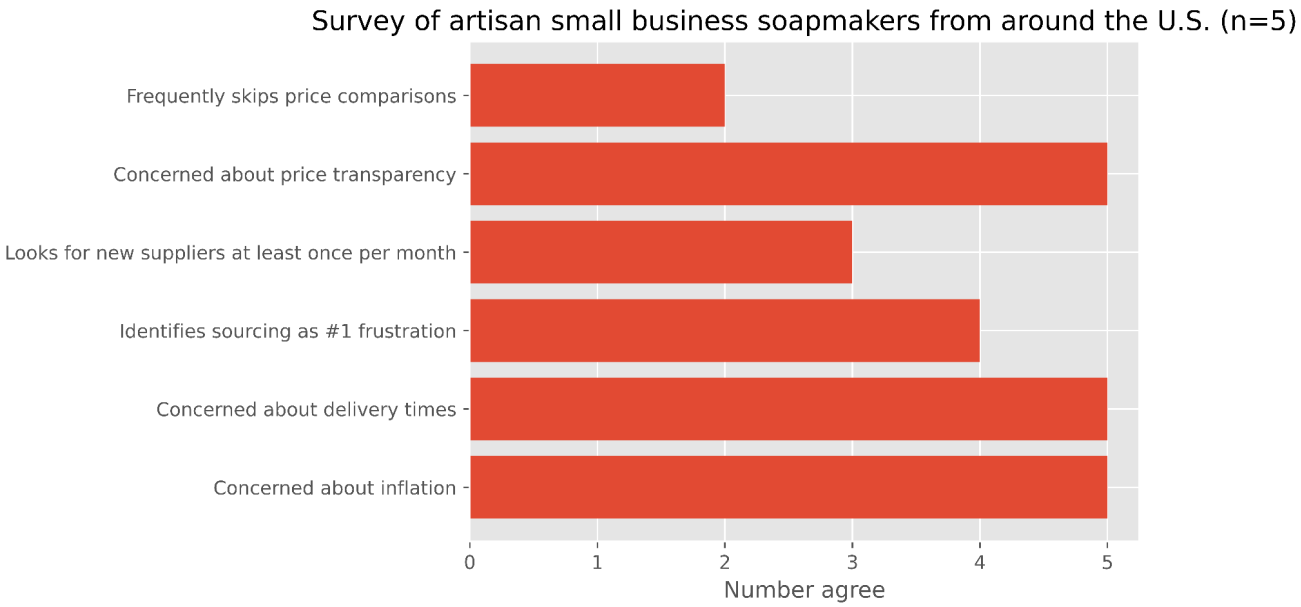

Figure 1

Methodology

A novel web-scraping pattern was written in Python and devised on top of the browser automation framework Selenium. A "Bot" and "Menu" class interface was defined such that for any product URL which belonged to a particular predefined web domain and a target quantity, the Bot would 1) begin with a fresh browser instance free of cookies; 2) interpret the menu of available quantity denominations; 3) determine the denomination closest to the target quantity (by shortest absolute distance, converting units if necessary); 4) add the target quantity to the "shopping cart" and "proceed to checkout" as if a human; 5) input a shipping address such that the ecommerce store generates tax and shipping rates in addition to the subtotal; 6) scrape the shipping rate and total.

(1) started a new instance of Firefox using Selenium. (2) required manual Selenium scripting of a particular website's html quantity menu archetype for its products, in combination

with the Python NLP library Quantulum3 to be able to extract the quantity from unstructured text (frivolous words and typos were often common). (3) used the Python quantity library Pint to ingest the Quantulum3 output and compute absolute distances. (4) used manual scripting. (5) used manual scripting in conjunction with “user shipping profile” information from the Bot abstraction. (6) used manual scripting.

The reason for this pattern was to capture the “true price” of products, because ecommerce stores tend to impose additional fees or mark up shipping fees to improve their margins, and this information is only available to the user once they complete the checkout process. This represents a “dark pattern” in user interface design whereby the user is expected to jump through hoops in order to get the true cost of their purchase. This sunk cost, which genuinely costs the user time and attention, is what adds significant friction to the current process of price comparison by our interviewees. That is, ecommerce storefronts which claim to offer “free shipping” are in their view offering a perk, but this perk is merely the willingness to provide their buyers transparent price information by incorporating the shipping cost into the easily-readable list price.

The Bot class was instantiated (hardcoded) for 10 supplier websites which were deemed to be the top 10 most popular soap ingredient suppliers by our interviewees: Brambleberry, Bulk Apothecary, Jedwards, Chemistry Store, Essential Depot, Mountain Rose Herbs, Nurture Soap, Rustic Escentuals, and Scent Sational Supply. 63 unit tests were written to test whether the hardcoded Bots were viable over time, or whether they would rot as supplier websites changed their (html) designs.

In order to obtain price information on a large scale, and to demonstrate a prototype to the interviewees, the Bot class was also connected to Google Sheets via simple scripting using the

Python gspread Google Sheets API wrapper. A list of any arbitrary product URL, provided they belonged to one of the 10 websites, could be provided to the first column of the spreadsheet, and the prototype would take this column and populate the adjacent columns with information such as chosen quantity, price per unit, tax, shipping, total, etc.

Another embodiment of the prototype involved crawling the 10 websites using Elasticsearch (pooling the contents into a miniature search engine) and integrating the Elasticsearch API with the Google Sheets scripting, but this was not used in the demonstrations because the vanilla Elasticsearch did not return sufficiently relevant results for the 10 sites. Further tuning of Elasticsearch would have been required, or a custom backend system to uprank and downrank results based on manual intervention.

For the price analysis, 17 different soapmaking ingredients with identical specs were chosen among the 10 supplier websites and run through the integrated Google Sheets system. The mean and standard deviation of these ingredients' per-kilogram price, selected by product denominations as close to 7 lbs as possible, were calculated. These data were also compared against the mean and standard deviation of the commodities as listed on Amazon; these data were collected manually. Generally speaking, a given soapmaking commodity of identical specs was listed by about 5 of the 10 suppliers, and so all mean and standard deviations were held to a sample size of n=5.

Results

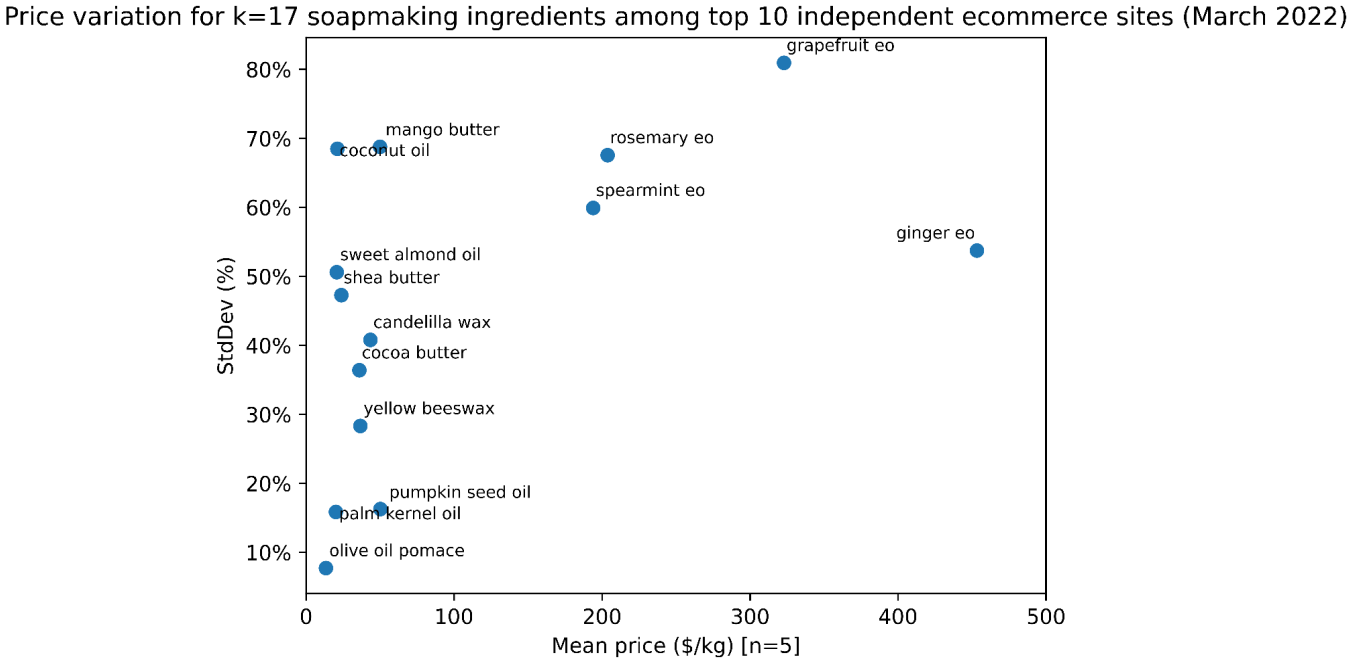

Figure 2

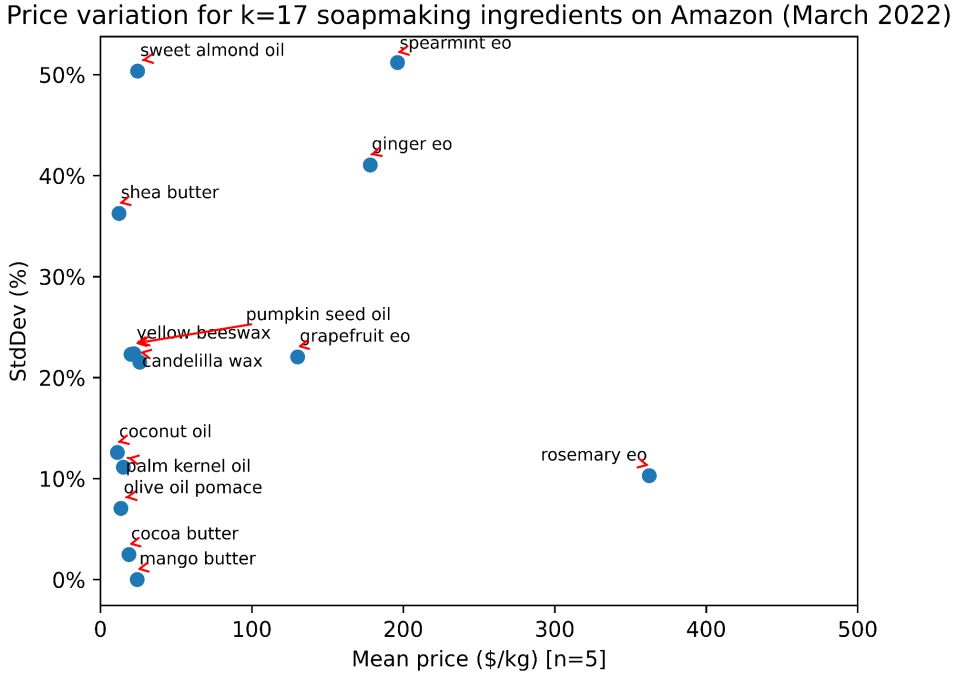

Figure 3

For the 10 independent ecommerce sites, more than half of the product offerings had a per-kilogram cost that exceeded a standard deviation of 40% (Figure 2). On the other hand, for Amazon significantly more than half the product offerings had a per-kilogram cost less than a standard deviation of 40% (Figure 3).

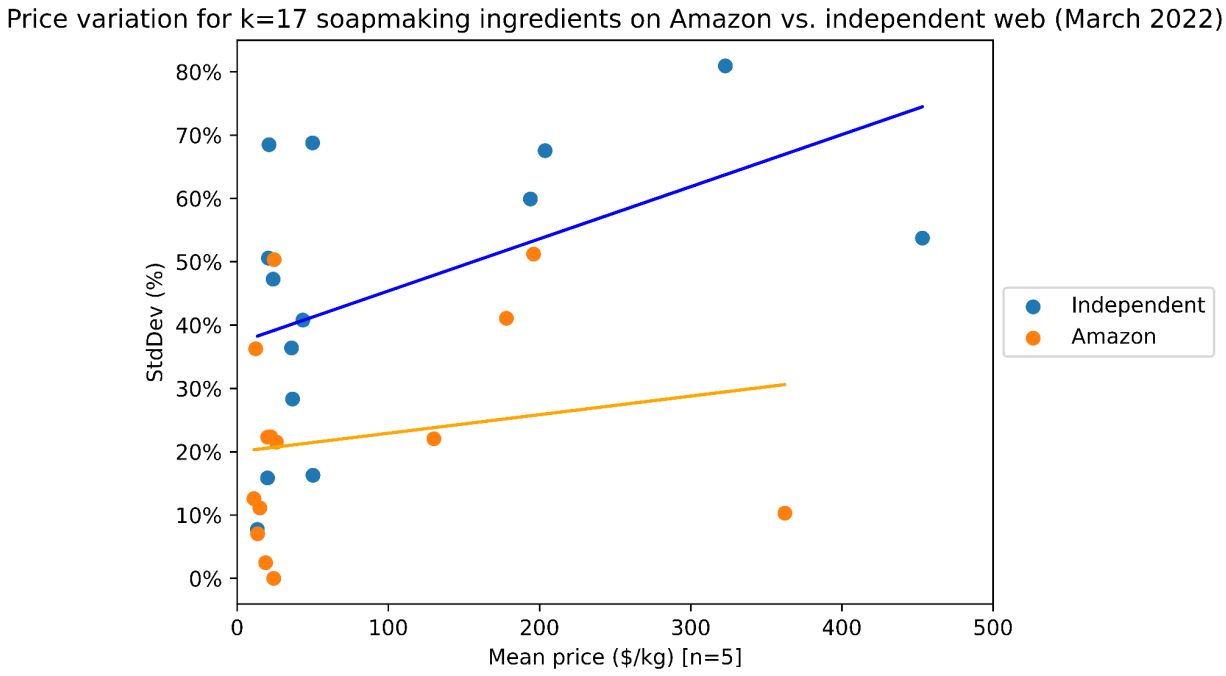

Figure 4

When the results from the independent ecommerce sites (Figure 2) and Amazon's listings (Figure 3) are superimposed, it is more than apparent that price variation is higher on the independent web (Figure 4).

Discussion

The results of the price analysis suggest at least two things: 1) The Internet today is not sufficiently frictionless to truly embody the Law of One Price; 2) the Law of One Price is generally supported by the results.

It is clear that some combination of information-overwhelm, results missed by Google, and dark patterns by contemporary ecommerce sites prevent buyers from seeking the most optimal supplier and keep sellers complacent from having to compete on price. It may be the case that through branding and the unique layout of each supplier's website that they feel "defended" from price competition such that they are not compelled to converge on prices.

At the same time, it is clear that systems of aggregation whereby competitors are held directly next to each other provides some downward pressure on prices, since Amazon had lower standard deviations. At the same time, Amazon has its own sophisticated algorithms for dynamic pricing (and to achieve price discrimination), which may conflate the cause of the results, since the suppliers themselves may not be the agents who are bringing their prices to parity on Amazon. Nonetheless, the result is the same, which is that a centralized platform drives down the price.

An interesting point of discussion is that all 5 interviewees reported that they only use Amazon as a last resort in sourcing, which contradicts the results shown here that demonstrate that Amazon generally has better per unit prices. Some of the reasons they gave were that Amazon does not offer true bulk sizes on the order of drums and pallets, or that in their experience Amazon was most expensive. Moreover, interviewees placed additional premium on brand value (goodwill), subjective quality of the oil, customer service, shipping reliability, and

payment terms. In any of these dimensions Amazon is likely to be inferior to a supplier which caters specifically to soapmaking businesses.

Also, evident in this project was that suppliers attempt to differentiate their offerings away from being considered commodities. Suppliers differentiated their products in the following ways (which at times made it difficult to compare products on identical specs):

- Changing the name from the industry standard, or assigning a brand to the commodity
- Organic vs non-organic
- Premium vs standard
- Color (pink vs white, yellow vs white)
- Shape (pastilles, wafers, granules)
- Refined vs unrefined/raw
- Country of origin (Spain)
- Plant variety or strain (Roman, Moroccan)
- Planet component source (seed, nut, fruit, leaf)
- Sustainable vs unsustainable (RSPO)
- Fair trade vs not fair trade
- Melting point (76F vs 92F)
- Fragrance (essential vs nonessential)
- Scent manipulation (deodorized)
- Concentration
- Additives (% alcohol or alcohol-free)

- Extraction method (pressed vs distillate; twice distilled; nonvirgin vs virgin vs extra virgin; filtered, twice filtered)

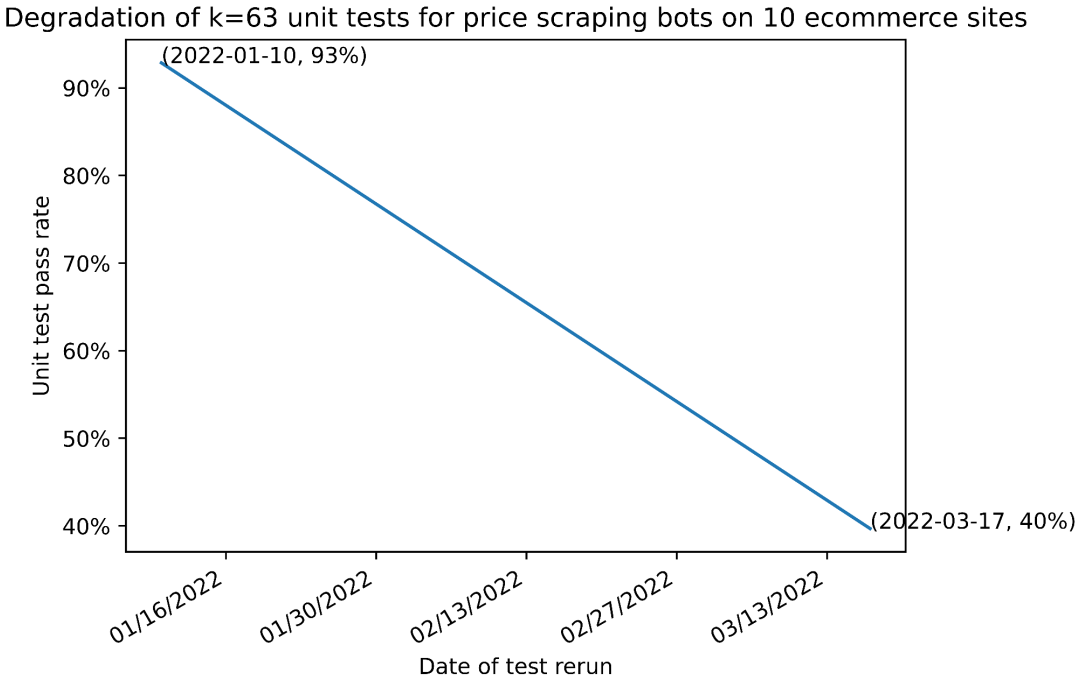

Figure 5

Lastly, this project demonstrates the potential for there to exist a new software service or search engine which can aggregate price information in more sophisticated manners as attempted here (incorporating shipping costs, coupons, delay times, etc.). The Bot class implemented here was particularly time consuming (approximately 40 hours for the 10 sites, including unit testing and refactoring when the scripts rotted), and a more generalizable technology is requisite in order to cover the entire ecommerce ecosystem. Moreover, because suppliers choose to update their website designs, add popups and banners, and employ anti-bot technologies, the reliability of the

hardcoded scripts quickly declines over time. Scripts that were rerun after 2 months began to fail (Figure 5).

In the future, it would make sense for a full web app to access price listings via a standard data protocol or API, or to devise a machine learning model which can interact with a browser and can generalize the process of checking out a product to obtain its true cost. The former may be difficult due to perverse incentives on the seller side as highlighted earlier, but the latter is certainly feasible.

Closing

Thank you to Professor Chris Makler, Professor John Ousterhout, Jill O'Nan of the Technical Communication Program, and Aaron Feigelman for your input and support.

References

Potters, Charles. "Law of One Price Definition". Investopedia. https://www.investopedia.com/terms/l/law-one-price.asp.

Wakabayashi, D. (2022, February 26). *Does anyone know what paper towels should cost?* The New York Times. Retrieved March 17, 2022, from https://www.nytimes.com/2022/02/26/technology/amazon-price-swings-shopping.html