Decoding Economic Signals

Using Large Language Models to Investigate Sellers' Inflation in Earnings Calls

Markus Lang
University of Heidelberg
markus.lang@mwi.uni-heidelberg.de

ABSTRACT I introduce new supply, demand, cost, and profit margin indices based on U.S. earnings calls in Q1 2023. These indices were constructed with the assistance of GPT 3.5 and will be calculated for all quarters between 2017 Q1 and 2023 Q2 if successfully validated. Already, it is evident that these indices track conditions of corporate pricing in line with a microeconomic perspective on inflation. The ultimate goal of index creation is to devise measures through which coordination among companies can be depicted more directly and promptly than in corporate surveys.

KEYWORDS inflation; prices; profits marginsa; cost; demand; nlp

Introduction

While inflation in the United States is currently decreasing significantly, except for energy prices, questions about its exact trajectory persist. Particularly surprising from the perspective of conventional economic theory is the fact that inflation in the United States increased sharply by mid-2021, even though real GDP had not substantially exceeded its pre-pandemic trend (Mason & Jayadev, 2023).

This unexpected surge not only challenges explanations of inflation that emphasize excessive demand but also underscores a shift in the approach of central bank researchers. It's noteworthy that they have begun to recognize limitations with traditional output gap measures, which they had long relied upon to monitor inflation. In response to these limitations, central bank researchers have increasingly utilized firm data and advanced natural language processing (NLP) techniques to model pricing decisions more accurately.

Examples include the development of demand and supply indices by researchers at the Bank of Canada (Gosselin & Taskin, 2023), supply chain bottleneck sentiment indices by Federal Reserve researchers (Young et al., 2021), and input costs, demand, and final price indices by researchers at the

Reserve Bank of Australia (Windsor & Zang, 2023). These corresponding indices represent an essential step toward a more microeconomically oriented perspective on inflation.

In line with microeconomic theories of inflation, such as those of Isabella Weber and Evan Wasner (2023), central bank researchers attempting to construct indices with firm data increasingly emphasize differences between industries and heterogeneity in pricing decisions. However, what central bank researchers often overlook is the significance of relationships between companies. Unlike Weber and Wasner, who assume shocks in a few strategically important sectors that then propagate through supplier-client relationships, central bank researchers typically focus on economy-wide excess demand. Possible effects of upstream shocks, where companies downstream might see opportunities to secure or increase existing profit margins, are often disregarded.

In this paper, I examine pricing behavior at the micro level using what U.S. companies say about supply, demand, costs, and profit margins in earnings calls. To achieve this, I draw inspiration from similar data sources and NLP methods as the aforementioned central bank projects. My goal is to construct indices using largely similar data and methods to ensure comparability with existing central bank indices. Due to the costs of data collection and processing, I initially limit myself to a single quarter in this paper (2023 Q1). In the long run, the indices developed here for testing purposes are intended to be calculated for all quarters between 2017 and 2023 Q2.

Because of the limitation to a single quarter, it is currently not possible for me to sensibly verify the external validity of the newly calculated measures. To do so, it would be necessary to systematically compare the new index scores at the firm, sector, and economic levels with Bureau of Labor time series and surveys. Instead, in this paper, I focus on testing the internal validity of the newly calculated indices. I do this by examining, similar to central bank researchers, the relationships between the newly constructed supply, demand, cost, and profit margin indices and a newly constructed price sentiment index.

What I find is that positive price sentiment, as expected, appears to be associated with positive sentiment regarding costs and demand. At the same time, I also find that positive price sentiment is associated with intentions expressed in earnings calls to maintain profit margins. Expectations of being able to maintain or even increase profits appear to be unexpectedly important for the prediction and determination of price sentiment. However, central bank researchers have not yet provided indices for profit expectations. This is not particularly surprising, given the difficulties in surveying companies about their profit interests. Therefore, the creation and validation of such indices would be a genuinely new contribution of NLP methods to the study of inflation. This paper represents an initial, cautious step in that direction.

Data

The Earnings Calls Transcripts of U.S. companies are sourced from the provider FMP Cloud. Each transcript comprises both prepared comments at the outset of each call and subsequent Q&A sessions. At the commencement of a typical call, company executives provide an overview of the current business position based on the recently released financial report. Subsequently, company executives respond to questions from analysts and other interested parties. Frequently, they respond more directly than in prepared comments and financial reports, as they must provide on-the-spot answers to analysts' inquiries.

Notable limitations in the use of Earnings Calls include the fact that they are typically conducted by large companies, and transcripts of the Earnings Calls are published on a quarterly basis. Furthermore, it is possible that the questions posed by analysts to managers of the respective companies may vary from call to call and from one company to another. These limitations do not impede the creation of the indices in our case but must be taken into account when interpreting the results. In essence, we are always examining precisely the segment of the U.S. economy that regularly conducts Earnings Calls. For this segment, however, we can determine sentiment more promptly and precisely than would be possible with conventional central bank surveys.

My preliminary sample includes a total of 3,191 Earnings Calls Transcripts in Q1 2023 from an equal number of U.S. companies. I split the individual transcripts into a total of 31,427 paragraphs. This is necessary to utilize OpenAI's GPT 3.5 model, as it only allows a context window of 16,000 tokens. The Earnings Calls Transcripts in the sample are, on average, considerably longer than 16,000 tokens, while the vast majority of paragraphs remain under 10,000 tokens. This means that GPT3.5-16K can process each paragraph individually, typically corresponding to an analyst's question or a manager's response.

Index construction

The projects of central bank researchers that employ Large Language Models in the construction of sentiment indices typically draw inspiration from dictionary methods (see Hassan et al. (2019)). The concept behind dictionary-based sentiment analysis involves sliding a window of 10 or 20 words around a focal word, such as "demand" or "prices," and within this window, counting positive and negative words based on a custom or established dictionary. Positive words may include, for example, advantages, boom, and efficiencies. Negative words include, for example, adversities, barriers, and dysfunctions. To construct a sentiment index from this, researchers subtract the sum of negative words from the sum of positive words and divide the result by the number of words in the relevant document, in this case, an Earnings Call.

The primary advantage of this method of index construction lies in its simplicity and transparency. Its chief drawback is that nuances in managerial language are often overlooked. For instance, whether a manager is speaking of demanding conditions or consumer demand cannot typically be distinguished in a straightforward dictionary analysis. This is why it is sensible, like central bank and IMF researchers (Albrizio et al., 2023; Windsor & Zang, 2023), to employ more complex models, such as large language models, in the creation of sentiment indices. The latter promise to handle nuances in managerial language more effectively.

As my use of Large Language Models in this paper is purely exploratory, I refrain from making a comparison between classic dictionary-based sentiment analyses and Large Language Models. My primary aim is to determine whether Large Language Models can create internally valid indices that require a nuanced understanding of language, such as Demand and Profit Margin Indices.

To achieve this, I formulate prompts in an iterative process that instructs GPT 3.5 to assess the extent to which specific topics are discussed in a given earnings call paragraph. The prompt I ultimately use appears as follows:

> Role = system
>
> "Please act as a financial analyst with in-depth knowledge of .x"
>
> Role = user
>
> "Analyze the provided paragraph from .x .y earnings call transcript and assign a probability score between 0 and 1 to indicate the degree to which the paragraph discusses each of the following predefined topics:
>
> 1. Supply Shortages (p =)
> 2. Increasing General Input Costs (p =)
> 3. Decreasing General Input Costs (p =),
> 4. Increasing Final Prices (p =),
> 5. Decreasing Final Prices (p =),
> 6. Increasing Consumer Demand (p =),
> 7. Decreasing Consumer Demand (p =),
> 8. Protection of Profit Margins against Rising Costs (p =).
>
> If the excerpt contains information unrelated to all topics, assign p = 0 to all topics.
>
> Earnings call transcript excerpt for analysis: .z"

During the process of prompt development, I randomly selected 100 earnings call paragraphs from the overall corpus and manually coded them according to the topics in the prompt. Subsequently, I sent this random sample to OpenAI to calculate comparable results using GPT 3.5. I defined a specific earnings call paragraph as successfully assigned to a topic by ChatGPT if the probability calculated by ChatGPT exceeded 0.7. If the topic assignments by ChatGPT matched my manual topic assignments, I retained the words describing a topic in the prompt. If not, I attempted to adjust the topic words within the framework of additional random samples or removed the topics.

Using the adjusted prompt, I ultimately assigned probabilities to all 31,427 paragraphs in the first quarter of 2023 through the OpenAI API. Similar to the researchers at the Reserve Bank of Australia (Windsor & Zang, 2023), I only unambiguously assigned paragraphs to a specific topic if the probability assigned by the model exceeded the 0.7 threshold. For example, if a paragraph discussing the topic "decreasing final prices" had a probability of 0.8, I assigned it a 1; otherwise, I assigned a 0.

To create sentiment indices from these assignments, I summed the "increase" assignments per earnings call and subtracted the sum of "decrease" assignments. I then divided the result by the number of paragraphs in the respective earnings call. An "increase" assignment in this context might be something like "increasing consumer demand," while a "decrease" assignment could be "decreasing consumer demand." In the case of topics like "protection of margins against rising costs," which cannot be clearly assigned to an increase or decrease, I simply summed the paragraphs with probabilities over 0.7 at the earnings call level and divided by the number of paragraphs.

An initial look at the indices

A preliminary examination of the index scores for individual companies suggests that ChatGPT 3.5 does, indeed, appear capable of capturing the extent of topic-specific discussions. Out of all 3,191 U.S. companies in the dataset, Apple receives the second-highest overall supply shortage score. If one were to research Apple's supply chain issues in Q1 2023 in business publications, you would indeed find indications of significant problems in the delivery of the iPhone 14 Pro/Pro Max, attributed to supply shortages from Apple's supplier Foxconn due to Covid restrictions (see Welch, 2023). In the earnings call itself, Apple CEO Tim Cook stated:

> The Pro has been a — the 14 Pro and the 14 Pro Max have done extremely well up until the point where we had a supply shortage and couldn't provide them — couldn't provide the total of the demand.

Due to the paragraph in which Cook makes this statement, GPT 3.5 assigns not only a high probability for supply shortages but also for demand increases to the paragraph. It seems, therefore, that GPT 3.5 can successfully distinguish between discussions of various nuanced topics within the same paragraph.

In addition to the results at the company level, the results at the sector level also appear to reflect real discussion intensity. When calculating the index average across all NAICS Subsectors (3-digit), one finds the second-most intensive discussion of supply shortages in the subsector *Motor Vehicle and Parts Dealer* (see Figure 1).

This aligns with reports in the business press that supply shortages related to semiconductors will persist for the subsector until early 2023 (see Gitlin, 2022). Autonation's Manager Mike Manley described supply shortages in Q1 2023 as follows:

> Yes. I mean unlike the previous recessions where all markets where being served with plenty of supply, and it was pure demand. The fleet

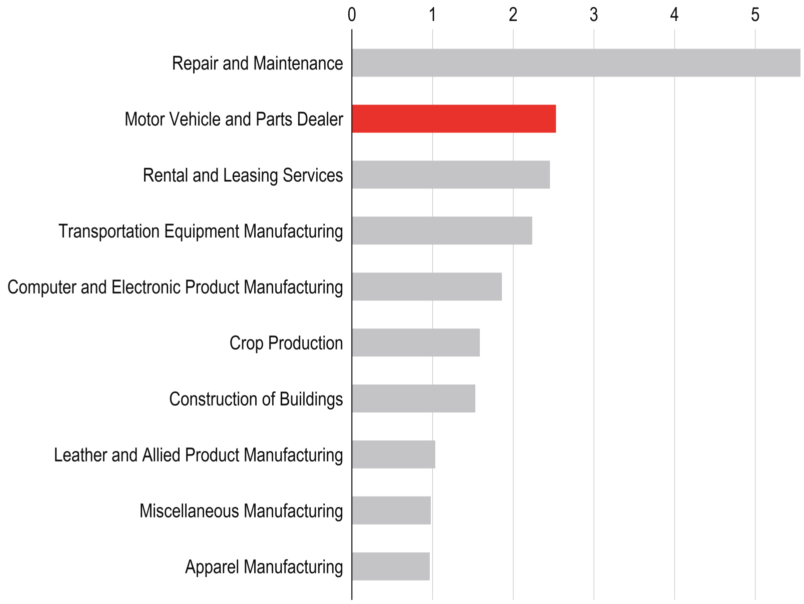

Figure 1: NAICS subsectors with the top 10 supply shortage index scores. Standardized and mean-centered index scores, Q1 2023.

> market obviously has been starved of supply during this period as the OEMs have prioritized their retail channels. So I think as OEMS are beginning to return, they are obviously looking at their margin (…).

With regards to this paragraph, GPT 3.5 is capable of discerning that Autonation is only currently grappling with supply shortages, and the mention of demand relates to past recessions. Additionally, the discussion about the margins of original equipment manufacturers (OEMs) is not associated with Autonation. Consequently, the paragraph is only coded as a supply shortage paragraph.

Regarding other indices, one can observe intuitively plausible results as well. For instance, as Figure 2 show, companies in the NAICS subsector *Food Manufacturing* engage in relatively intensive discussions about profit margins. This aligns with reports in the business press that resilience in the Food Manufacturing sector increased in Q1 2023 (see Yu (2023)). An example of this can be found in Kellogg's Earnings Call:

> So we feel very good about our topline growth momentum and outlook. We also feel good about restoring our profit margins. We said that this would be a year in which we stabilize and even improve our margins after being pressured the last couple of years by soaring input cost inflation, and inefficiencies in costs related to bottlenecks and shortages.

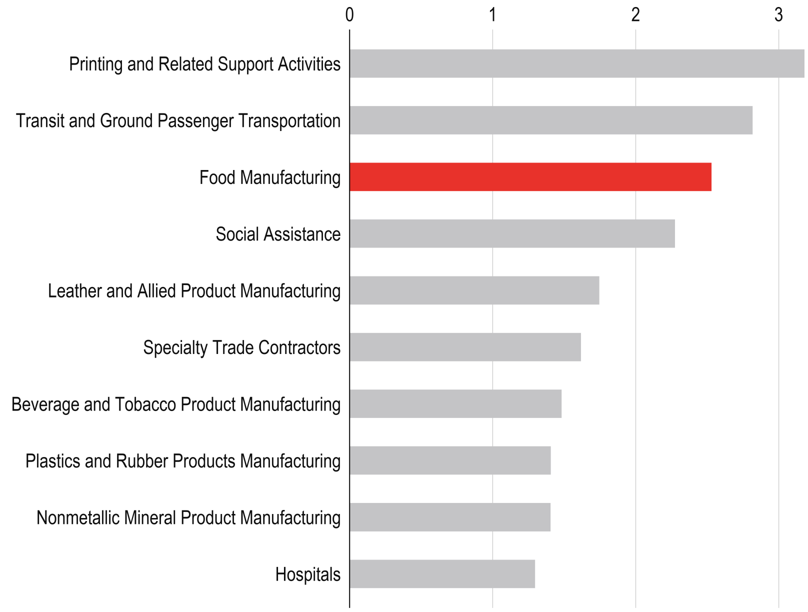

Figure 2: NAICS subsectors with the top 10 profit margins index scores. Standardized and mean-centered index scores, Q1 2023.

Here, too, GPT 3.5 is capable of distinguishing between mentions of profits in the current quarter and increases in input costs in past quarters. There appear to be some indications that simple topic intensity indices, such as the supply shortage index and the profit margins index, adequately represent earnings call discussions. Does this also hold true for the sentiment indices?

Following the consumer demand index in Figure 3, discussions about demand increases in Q1 2023 are primarily observed in the *Educational Services* subsector. This seems plausible in light of the rapid increase in enrollments in the United States, according to a May 2023 report from the Institute of Education Services (Albrizio et al., 2023), and the new potential for digitalization in "Edtech". Discussions about demand increases with regard to building, accommodation, and air transportation also seem reasonable.

Similar considerations apply to sector-level scores in the general input cost index (see Figure 4). Among the subsectors discussing cost increases are those in the manufacturing sectors that are facing cost increases due to the Russia-Ukraine war. These include *Plastics and Rubber Products Manufacturing* and *Nonmetallic Mineral Product Manufacturing*, among others. Conversely, GPT 3.5 appears capable of accurately identifying earnings calls with discussions about cost decreases. Both the *Apparel Manufacturing* and the *Leather and Allied Product Manufacturing* sectors saw discussions about cost reductions and experienced actual cost reductions due to declining cotton, polyester, and leather prices.

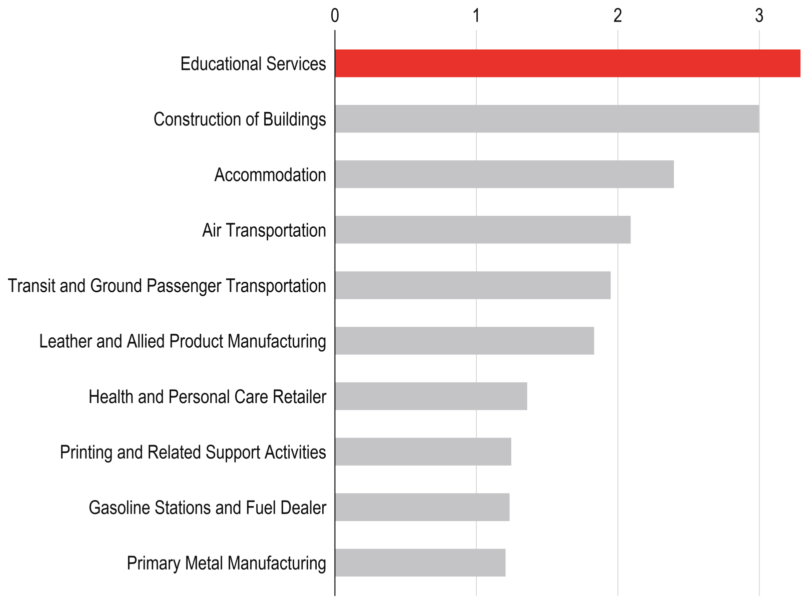

Figure 3: NAICS subsectors with the top 10 consumer demand index scores. Standardized and mean-centered index scores, Q1 2023.

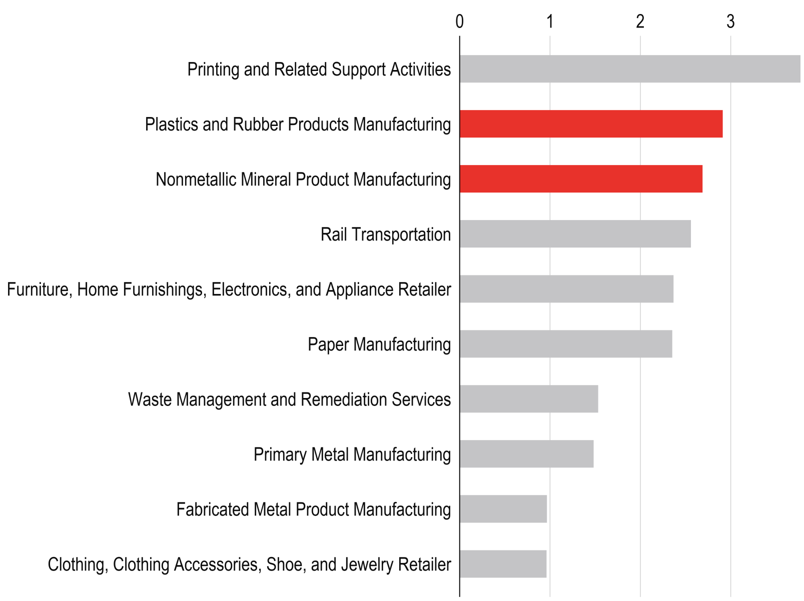

Figure 4: NAICS subsectors with the top 10 input cost index scores. Standardized and mean-centered index scores, Q1 2023.

All in all, the indices appear to be highly valuable in identifying and deciphering trends at the company and sector levels through the scope and direction of topic-specific discussions.

Relationships Between the Indices

However, a question that remains is whether the relationships between demand and price sentiment or cost and price sentiment, identified by central bank researchers using their indices, can also be identified using the indices developed here. To explore this, I estimate simple cross-sectional regressions using the company-level index scores.

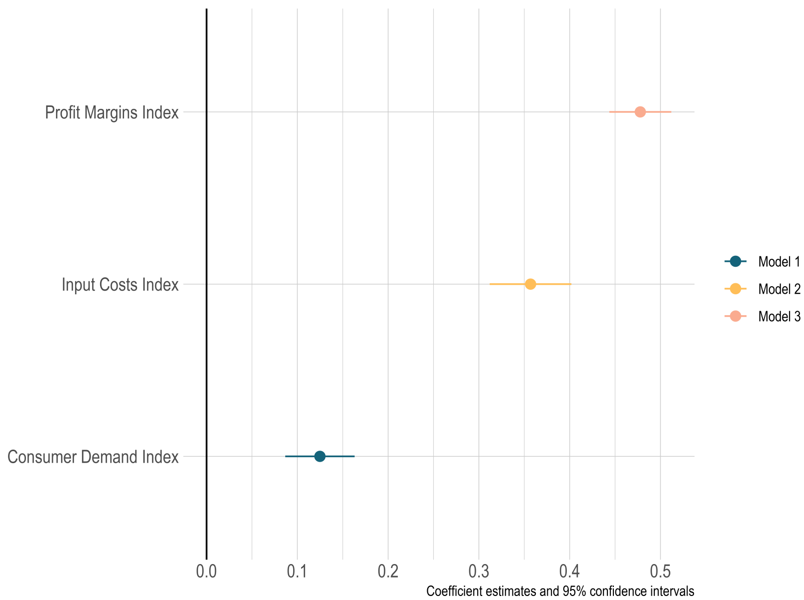

Figure 5: Linear regression models of final prices index.

In the first model in Figure 5, the dependent variable is the price sentiment index, and I regress price sentiment on demand sentiment (see also Table 1). The model predicts that for companies with a one-standard-deviation more positive consumer demand, there is a 0.12 increase in positive price sentiment. This is consistent with the positive and significant coefficient found by the Reserve Bank of Australia researchers for their demand index (Windsor & Zang, 2023).

Also in line with the model of the Reserve Bank of Australia researchers, the second model in Figure 5 predicts that companies with a one-standard-deviation higher cost index have over 0.36 more positive price sentiment than companies with a lower cost index. This aligns with survey findings that suggest a substantial pricing strategy for companies involves setting prices as a markup over costs (see Windsor & Zang, 2023).

Moving on to Model 3 in Figure 5, which focuses on the profit margins index, the model predicts that for companies with one standard deviation more discussion on profit margins, there is 0.48 significantly more positive price sentiment. However, it's important to note that this prediction cannot be causally interpreted, as discussions on profit margins should not be independent of the cost and demand situations.

To credibly determine the causal effect of discussing profit margins, it will be necessary to create and use time-series data. Nonetheless, by controlling the profit margin index with the demand and cost indices, I can already test whether there is still an effect of profit margins on price sentiment in a cross-sectional regression. Model 4 in Figure 6 indicates that this is indeed the case. Companies discussing profit margin retention with one standard deviation more have 0.42 higher price sentiment than companies with the same demand and cost sentiment. The lower profit margin index score than in the model before suggests that demand and cost sentiment not only directly influence price sentiment but also indirectly through companies' profit considerations.

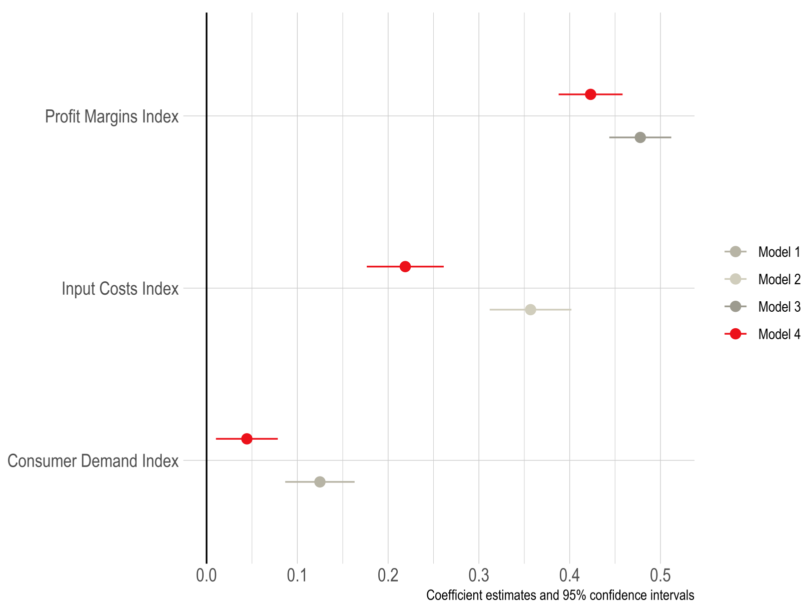

Figure 6: Linear regression models of final prices index.

In general, these preliminary tests of the internal validity of the new indices suggest that it is worthwhile to expand index construction and testing over time. All previous test results are consistent with similar projects by central bank researchers and provide clear indications that efforts to secure profit margins may be more important than apparent in the index development projects of central banks.

Potential Contributions

The central contribution of the new indices could be in making the relationships between demand, costs, profit strategies, and pricing more analyzable than before. One significant issue with surveys is that researchers often do not or cannot ask questions about the maintenance or increase of profit margins, either due to reluctance or for the successful completion of the surveys (see Blinder et al., 1998).

In earnings calls, on the other hand, managers frequently speak more openly about how they unilaterally or in accordance with competitors raise prices when the general perception allows it due to rising costs. Large language models make it possible for the first time to identify and systematically compare relatively nuanced expressions in thousands of earnings calls transcripts. Restricting earnings calls to similar or identical tasks as surveys could largely leave the potential of large language models in economics untapped.

It is currently not entirely clear what specific profitable uses large language models can have in economics (see Ash & Hansen, 2023), and it is likely that the literature tends to fit the new models into old theoretical and methodological frameworks. However, for a microeconomic perspective on inflation, it is important that methods that appear to yield useful results at the company and sector levels are not immediately re-aggregated at the level of the entire economy. For these reasons, the internal and external validation of demand or profit indices at the company, sector, and economic levels could be a significant contribution of this project to a microeconomic perspective on inflation.

Appendix: Regression tables

Table 1: Linear regression models of final prices index (without combined model).

	Model 1	Model 2	Model 3
Consumer Demand Index	0.125***		
	(0.019)		
Input Costs Index		0.357***	
		(0.023)	
Profit Margins Index			0.478***
			(0.017)
Num.Obs.	2824	2824	2824
R2	0.014	0.079	0.211

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2: Linear regression models of final prices index (with combined model).

	Model 1	Model 2	Model 3	Model 4
Consumer Demand Index	0.125***			0.044*
	(0.019)			(0.017)
Input Costs Index		0.357***		0.219***
		(0.023)		(0.022)
Profit Margins Index			0.478***	0.423***
			(0.017)	(0.018)
Num.Obs.	2824	2824	2824	2824
R2	0.014	0.079	0.211	0.240

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

References

Albrizio, S., Dizioli, A. G., & Simon, P. V. (2023). Mining the Gap: Extracting Firms' Inflation Expectations From Earnings Calls. https://www.imf.org/en/Publications/WP/Issues/2023/09/28/Mining-the-Gap-Extracting-Firms-Inflation-Expectations-From-Earnings-Calls-539617

Ash, E., & Hansen, S. (2023). Text Algorithms in Economics. *Annual Review of Economics*, *15*, 659–688.

Blinder, A. S., Canetti, E. R. D., Lebow, D. E., & Rudd, J. B. (1998). *Asking About Prices: A New Approach to Understanding Price Stickiness*. Russell Sage Foundation.

Gitlin, J. M. (2022). Automative chip shortages to continue throughout 2023, industry says. *arsTechnica*. https://arstechnica.com/cars/2022/12/automotive-chip-shortages-to-continue-throughout-2023-industry-says/

Gosselin, M.-A., & Taskin, T. (2023). What Can Earnings Calls Tell Us About the Output Gap and Inflation in Canada? *Bank of Canada Staff Discussion Paper*. https://www.bankofcanada.ca/wp-content/uploads/2023/06/sdp2023-13.pdf

Hassan, T. A., Hollander, S., van Lent, L., & Tahoun, A. (2019). Firm-Level Political Risk: Measurement and Effects. *The Quarterly Journal of Economics*, *134*(4), 2135–2202.

Mason, J., & Jayadev, A. (2023). Rethinking Supply Constraints. *Review of Keynesian Economics*, 232–251.

Weber, I. M., & Wasner, E. (2023). Sellers' Inflation, Profits and Conflict: Why can Large Firms Hike Prices in an Emergency? https://scholarworks.umass.edu/cgi/viewcontent.cgi?article=1348&

Welch, C. (2023). Apple's iPhone 14 Pro supply problems sank its holiday revenues. https://www.theverge.com/2023/2/2/23583244/apple-iphone-14-q1-2023-earnings-pro-shortage-mac

Windsor, C., & Zang, M. (2023). Firm' Price-setting Behaviour: Insights from Earnings Calls. https://www.rba.gov.au/publications/rdp/2023/pdf/rdp2023-06.pdf

Young, H. L., Monken, A., Haberkorn, F., & van Leemput, E. (2021). Effects of Supply Chain Bottlenecks on Prices using Textual Analysis. *FED Notes*. https://www.federalreserve.gov/econres/notes/feds-notes/effects-of-supply-chain-bottlenecks-on-prices-using-textual-analysis-20211203.html

Yu, D. (2023). The Food and Beverage Industry Showing Signs of Resilience Despite Declining U.S. VC Deal Volume. *Forbes*. https://www.forbes.com/sites/douglasyu/2023/04/13/the-food-and-beverage-industry-showing-signs-of-resilience-despite-declining-us-vc-deal-volume/