Firm-Level Input Price Changes and Their Effects: A Deep Learning Approach *

Sudheer Chava[†] Wendi Du[‡] Indrajit Mitra[§] Agam Shah[¶]

Linghang Zeng[‖]

Abstract

We develop firm-level measures of input and output price changes based on textual analysis of earnings calls. Our measures establish five facts: (1) The median firm experiences an increase (decrease) in input prices every 7 (30) months. (2) Input price changes are driven by aggregate and firm specific components. Each component contributes equally. (3) Firms pass through input price changes to output prices in the same quarter with a pass through magnitude of 0.55. (4) Our input price change measure predicts future changes in COGS. (5) Firm stock price reaction is negatively related to our input price change measure.

*A version of this paper was previously titled "Measuring Firm-Level Inflation Exposure: A Deep Learning Approach." We thank Viral Acharya, Will Cong, Anthony Diercks, Jan Eeckhout, Norman Guo, Gerard Hoberg, Alex Hsu, Wei Huang, Pete Klenow, Da Li, Jiacui Li, Alejandro Lopez-Lira, Ivan Shaliastovich, Erik Sirri, Wenting Song, Luke Stein, Jerome Taillard, Jincheng Tong, Michael Weber, Michael Weisbach, Nancy Xu, Baozhong Yang, Mao Ye, Tong Yu, Miao Ben Zhang, and participants at AFA, CICF, CFEA, CETAFE, FMA, MFA, SITE, Swedish House of Finance Conference, UF Conference, UT Dallas Conference, Babson, Florida State, Georgia Tech, Texas A&M, Miami for comments. The views expressed here are ours and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

[†]Scheller College of Business, Georgia Institute of Technology, chava@gatech.edu

[‡]Darla Moore School of Business, University of South Carolina, wendi.du@sc.edu

[§]Federal Reserve Bank of Atlanta, indrajit.mitra@atl.frb.org

[¶]College of Computing, Georgia Institute of Technology, ashah482@gatech.edu

[‖]Babson College, lzeng@babson.edu

1 Introduction

The dynamics of input prices is an important source of a firm's fundamental risk. However, very little is known about its empirical properties and its effects on firm policies and firm value. This gap exists in the literature because of the lack of input price data of U.S. firms. In this paper, we perform textual analysis of transcripts of quarterly earnings calls of publicly listed U.S. firms to construct firm level measures of input and output (i.e., product) price changes. We apply our measures to establish properties of input price changes and their effects on two firm policies: price setting decisions and adjustment of total variable input costs. We also analyze the effect of input price changes on a firm's stock price.

Earnings calls offer several advantages for our purpose. First, the universe of firms whose earnings calls we analyze covers a large cross-section of firms. This allows us to analyze heterogeneity in both properties of input price changes across firms and also how firms respond to input price changes. Second, because earnings calls contain information about changes in both input and output prices over time, we are able to estimate dynamic cost pass through policies of individual firms. Third, because we can identify individual firms, we are able to link them to other datasets that allow us to study the response of balance sheet items and stock prices to input price changes.

While earnings calls offer the advantages listed above, their textual analysis is challenging. An important reason is because it is crucial to pay attention to word ordering and context when analyzing the transcript of a call in order to fully extract price change information. This feature of earnings calls rules out using a traditional natural language processing model based on simple rules and/or the frequency with which pre-determined key words are used. Instead, we use a pre-trained deep learning model which is better suited to analyze text with diverse syntax.

The specific deep learning model we use is RoBERTa (Robustly Optimized BERT Pre-training Approach, Liu et al. (2019)). We follow two main steps to generate our firm-level input and output price change measures. First, we use human-generated labels to fine-tune

RoBERTa to identify price change related discussions together with the direction of price change (i.e., increase or decrease) and the type of price change (i.e., input or output). To do so, we manually label all sentences in a sub-sample of earnings calls to indicate price-change information, including direction and type, and then fine-tune the model using these manual labels. Once the fine-tuning is complete, we perform out-of-sample tests to evaluate the model's performance and find RoBERTa to perform well.

In the second step, we use the fine-tuned model to generate labels for our entire sample of earnings calls. We use these labels to construct four numbers for each call. These are the number of sentences in the entire call that mention an input price increase $\#InputUp_{i,t}$, an input price decrease $\#InputDown_{i,t}$, an output price increase $\#OutputUp_{i,t}$, and an output price decrease $\#OutputDown_{i,t}$ for firm i at time t. We use these four numbers to construct our price change measures. Our input price change measure is:

$$InPrChg_{i,t} = \frac{\#InputUp_{i,t} - \#InputDown_{i,t}}{\#Sentences\ in\ Transcript_{i,t}},$$

where $\#Sentences\ in\ Transcript_{i,t}$ is the total number of sentences in that particular earnings call. We define our output price change measure, $OutPrChg_{i,t}$, by replacing $InputUp_{i,t}$ and $InputDown_{i,t}$ with $OutputUp_{i,t}$ and $OutputDown_{i,t}$, respectively.

The measures above literally measure the intensity with which input (output) price change discussions take place in an earnings call. To interpret them as measures that capture changes in input (output) prices, we make two assumptions. First, a firm typically uses various inputs and sells multiple products. Therefore, we assume that when an earnings call discusses price changes, it is referring to a value-weighted average of the growth rates of all their inputs and products. Second, we assume that the intensity of these discussions is proportional to this value-weighted average.

The proportionality assumption above, while admittedly strong, has large potential payoffs. Specifically, it allows us to establish properties of firm-level input price changes and firm's pass through policies that are otherwise challenging to analyze because of lack of publicly available data. We perform several validation checks to verify the validity of the proportionality assumption. For example, we compare suitably aggregated versions of our measure to comparable measures published by the Bureau of Labor Statistics (BLS). We find a high correlation between our measures and the BLS measures.

We use the variables $InPrChg_{i,t}$ and $OutPrChg_{i,t}$ to document five new facts. First, we decompose the variation in $InPrChg_{i,t}$ into aggregate (including an economy-wide and an industry component) and firm-specific components. We find that the aggregate and the firm-specific components contribute roughly equally, with economy-wide inflation contributing only 7% to the variation of $InPrChg_{i,t}$. An implication of this result is that $InPrChg_{i,t}$ contains a non-trivial amount of firm-level price change information that is missed by aggregate price change series published by the BLS.

Second, we use a modified version of $InPrChg_{i,t}$ to estimate the average likelihood of a firm to experience a price change. We find that the median firm experiences an input price increase once every 7 months. Input price decreases are much rarer, occurring once every 30 months for the median firm.

Third, we use our input and output price change measures to estimate firm's dynamic cost pass through policies. We find the magnitude of the contemporaneous pass through to be 0.55, that is, a 10% increase in input cost results in a 5.5% increase in output price that quarter. We document that pass through declines quickly over time, lasting about 2 quarters following an input price change.

Our measures allow us to empirically test theories of cost pass through by firms. To illustrate this, we ask if the pass through magnitude depends on a firm's market power, focusing on industry concentration as a measure of the latter. To carry out our analysis, we rely on the measure of industry concentration constructed by Hoberg and Phillips (2016). We estimate a higher pass through for firms operating in an industry with lower industry concentration. The difference in pass through between firms in the top and bottom terciles of industry concentration is economically large.

Fourth, we find that changes in input prices as captured by $InPrChg_{i,t}$ is positively related to future changes in the firm's cost of goods sold (COGS). Since COGS is a product of the quantity times price of inputs, our result implies that the average firm in our sample has a demand curve for inputs with elasticity less than one.

Fifth, we find a firm's stock price to immediately respond to our input price change measure. The direction of the stock price reaction is intuitive. Quantitatively, a one standard deviation increase (decrease) in $InPrChg_{i,t}$ is followed by a 37 basis point lower (higher) cumulative abnormal stock return (CAR) measured over a three-day window around the earnings call. Our results provide an estimate of the sensitivity of firm value to mentions of changes in input prices in earnings calls.

The primary contribution of our paper is to point out that earnings calls provide a useful source of information about firm-level input and output price changes for a relatively large cross-section of firms and to show how textual analysis helps extract such price change information. Our firm level input and output price change measures contribute to two literatures.

The first is the literature that has emphasized the joint dynamics of input and output prices as an important determinant of a firm's fundamental risk (see e.g., Gorodnichenko and Weber 2016 for evidence). Because of its influence on firm risk, in equilibrium, input and output price dynamics can be expected to impact firm policies and also the cost of capital of firms. D'Acunto et al. (2018), for instance, provides evidence of the effect of price changes on firm policies, focusing on financing decisions. The implications of a mismatch in the cyclicality of input and output prices for the cost of capital of a firm has been extensively studied in the context of firms' labor costs being more rigid relative to its revenues. Danthine and Donaldson (2002) is an early example.

Our use of earnings calls as a data source for input and output price changes offers several advantages over the traditional data sources used in the literature. The latter include the confidential micro data of output prices that underlie the Producer Price Index (PPI) from the Bureau of Labor Statistics, firm's cost of goods sold (COGS) from Compustat, and firm level wage data also from Compustat. While the BLS micro data contains information about output prices only, our measures capture the joint dynamics of input and output price changes. Similarly, while COGS represents a firm's expenditures incurred for production, changes in COGS reflect not only changes in input prices but also changes in the quantity of inputs used by the firm. The latter may be affected by changes in demand for the firm's products rather than changes in input prices. Finally, while Compustat wage data does reflect the price of a production input, by definition, it does not capture changes in other important input costs such as raw materials. Furthermore, while Compustat wage data is available only at an annual frequency, our input price change measure is available at quarterly frequency.

The second strand of literature that our measure contributes to is the one that studies the price setting decisions of firms studied both in macroeconomics and in the industrial organization literatures. The issue of cost pass through is widely studied in the sticky price literature in macroeconomics, specifically, the New Keynesian literature that analyzes the real effects of monetary policy (Woodford, 2011). There is also a recent strand of this literature that studies the implications of the recent rise in market power in shaping industry dynamics and macroeconomic fluctuations (see e.g., Gutierrez and Philippon 2017 and Loecker, Eeckhout, and Unger 2020). This literature models pass through as being shaped by some form of adjustment cost, such as menu costs. Our estimates of firm's dynamic cost pass through policies can be used to discipline calibrations of quantitative models in this literature. Indeed, while much is known about the properties of firm-level output prices for a broad cross-section of firms (see e.g., Nakamura and Steinsson 2008), there is relatively little empirical work documenting how firms pass through input price changes to output prices for a similarly broad cross-section of firms. Existing papers that analyze cost pass through belong to one of three types. The first type studies specific products, such as changes in the price of crude oil and the response of retail gas prices (Borenstein, Cameron, and Gilbert, 1997). The second type analyzes scanner data of consumer products, such as from a single supermarket store (see e.g., Peltzman 2000 and Eichenbaum, Jaimovich, and Rebelo 2011) or a cross-section of retailers (Nakamura, 2008). These studies also do not cover well the broad range of products produced by firms in the U.S. The third type uses the national average prices for individual producer goods that are part of the PPI index together with the make-use tables from the Bureau of Economic Analysis (see e.g., Peltzman 2000). This approach misses heterogeneity in the price paid by different firms for the same input. In contrast to these papers, to the best of our knowledge, our paper is the first study that analyzes input price changes and its consequences in a large cross-section of firms. There is also a substantial industrial organization literature that has studied the role played by market power in determining cost pass through policies of firms. This include analyses of the effect of horizontal competition (see e.g., Berman, Martin, and Mayer 2012, Auer and Schoenle 2016, and Amiti, Itskhoki, and Konings 2019), vertical integration (see e.g., Hastings 2004 and Neiman 2010), and the interaction of horizontal competition and vertical integration (see e.g., Hong and Li 2017) on firms' cost pass through. Our text-based measures of cost pass through opens the door to carry out similar analyses, as long as the focus is on public firms.

In order to extract price change information from earnings calls, we build on recent developments in the field of natural language processing (NLP). Existing papers that also apply NLP methods to capture a particular dimension of a firm's economic environment and analyze the effect of a change in this environment include constructing firm-level measures of sentiment (Hanley and Hoberg, 2010; Loughran and McDonald, 2011; Narasimhan and Wu, 2013; Jiang et al., 2019), financial constraints (Hoberg and Maksimovic, 2015; Buehlmaier and Whited, 2018), product similarity (Hoberg and Phillips, 2016), political risk (Hassan et al., 2019), firm uncertainty (Handley and Li, 2020), corporate culture (Li et al., 2021), risk exposures to systematic risk factors (Lopez-Lira, 2023), reason for firms to change their product prices (Pitschner, 2020), cybersecurity risk (Florackis et al., 2023), climate change exposure (Sauter et al., 2023), emerging technologies (Chava, Du, and Paradkar, 2022), labor shortage(Harford, He, and Qiu, 2024), and managers' anticipated changes in capital expenditures (Jha et al., 2024).

The rest of the paper is organized as follows. We discuss the need for deep learning to classify text in earnings calls in Section 2. We discuss our data and methodology in Section 3 and perform several validation checks of our measures in Section 4. We apply our input and output price change measures to document properties of input price changes and their effects in Section 5 and conclude in Section 6.

2 Why Do We Need Deep Learning Classification?

The key challenge in measuring input and output price changes from earnings conference calls is identifying discussions related to price changes. The difficulty arises because the language patterns for these discussions are complex since they use diverse vocabulary and syntax. As an example, consider a portion of the 2021 Q2 earnings conference call of Sanderson Farms discussing price increases in their business:

> Sanderson Farms operated very well during the second quarter of fiscal 2021 in all areas of our business. Improved poultry markets more than offset **feed grain costs that were significantly higher** compared to last year's record fiscal quarter, resulting in increased operating margins... In addition to improved domestic demand for chicken, export demand also improved during the quarter as a result of **higher crude oil prices**... **Prices paid for corn and soybean meal increased significantly** during the quarter compared to last year... **We have priced all of our soy meal basis** through October **and most of our corn basis** through September.

The bolded text above contains price change related discussions. These sentences highlight the challenges faced by traditional natural language processing (NLP) techniques. First, it is difficult to create a rule to comprehensively identify sentences with such information. For instance, the phrase "Prices paid for corn and soybean meal increased significantly ..." might suggest a rule: if a sentence contains the word "price" or "cost" together with "increase", "decrease", "high", "low", or "change" then label this sentence as being price change related. However, this rule overlooks a sentence such as "We have priced all of our soy meal basis...," which is also price change related. Second, similar to the challenge faced by rule-based approaches, dictionary-based approaches would struggle to capture such nuanced linguistic information because they do not consider the word order information. As a result, an ideal dictionary often only includes words whose mere presence directly indicates price changes, such as "inflation" or "deflation." This limited focus results in missing a big portion of price change content conveyed through syntax and context nuances. If the dictionary is less restrictive in word selection, on the other hand, it increases the occurrence of false positives.

In contrast to rule-based approaches, deep learning approaches do not require explicitly defined rules and are therefore well suited for handling complex text characterized by diverse vocabulary and syntax. Three related reasons contribute to the superior performance of deep learning approaches in this regard.

First, deep learning models are pre-trained on a very large amount of text which makes them powerful in understanding the general meaning of words and sentences in human languages (Jurafsky and Martin, 2014). For example, RoBERTa (which is the model we use) is pre-trained with over 160GB of uncompressed text from BookCorpus and Wikipedia (16GB), CC-News (76GB), OpenWebText (38GB), and Stories (31GB). This enables the model to absorb the general semantic and syntactic knowledge of the English language. When detecting price change information in earnings calls, the pre-training allows the model to extract a word's meaning even if the algorithm has not encountered the vocabulary in the training sample.

Second, deep learning models are further trained on additional *human-labeled* sentences from which the algorithm autonomously learns the language patterns associated with price changes, a step known as fine-tuning. In this step, the model further learns which parts of a sentence are important to focus on to detect price change information. Moreover, the human-labeled training samples used in this fine-tuning step are well suited to train the deep learning model to discern nuanced information, in particular, the *direction* (increase or decrease) and *type* (input or output) of the discussed price changes that we use in our analysis.

Third, deep learning models can understand multiple connotations of the *same* phrase depending on how the phrase is used. This contrasts with models such as Word2Vec (Mikolov et al., 2013) and Glove (Pennington, Socher, and Manning, 2014), which assume a unique meaning for each phrase.

We compare performances of a dictionary approach, two rule-based approaches, and the deep-learning approach employed in our paper in Section IA.3 of the Internet Appendix. We find the deep-learning approach outperforms both the rule-based and dictionary approaches (see further details in that section and also Section 3.2.2 for further discussion). Next, we describe how we construct our input and output price change measures.

3 Data and Methodology

Earnings conference calls are known to provide critical corporate information to the market (see Sauter et al. (2023) for a recent example). In this section, we describe how we use conference calls to extract firm-level measures of changes in the price of inputs used by firms in production (henceforth input price changes) as well as changes in prices charged by firms to their customers (henceforth output price changes).

3.1 Data sources

We collect data on 178,547 earnings conference call transcripts from January 2007 to July 2021 from SeekingAlpha. We then merge the transcripts with CRSP and Compustat based on the identification information of the stock ticker, the company name, the title of the event, and the earnings conference call date. We remove financial and utilities firms and also firms with missing SIC code. We further restrict our sample to ordinary common shares (share codes 10 or 11) and stocks trading on the NYSE, AMEX, and NASDAQ exchanges (exchange code of 1, 2, or 3). We obtain financial variables from Compustat, I/B/E/S, and CRSP. In our baseline regression, there are 81,473 earnings call observations after dropping those with missing financial variables. We describe how we match earnings calls to Compustat and CRSP in Section IA.1 of the Internet Appendix. We also describe the various filters we add to our sample before analysis in Table IA.1 of the Internet Appendix.

3.2 Constructing measures of input and output price changes

We construct our text-based input and output price change measures in three steps. We first construct the training sample. Next, we train four deep learning models on the training sample to generate four price change related labels. Finally, we use the trained models to process all the earnings call transcripts and generate firm-level measures of input and output price changes. We explain each of the steps below.

3.2.1 Constructing the training sample

The first step in building our deep learning model is to construct a high-quality training sample. Because we had a limited pool of human resources available for transcript reading, we aimed at choosing a training sample that contained transcripts with a very high number of price change-related sentences per transcript. To this end, we adopted the following two-step approach.

First, we limited ourselves to a high-inflation period subsample: January 1, 2021 —June 30, 2021. We based this choice on the premise of a relatively strong signal in this period, that is, companies are more likely to discuss price fluctuations in this high inflationary period than in other periods.

Second, from among all the transcripts within this period, we selected a subset of 50 transcripts with a relatively high likelihood of containing price change information. To detect this subset of transcripts, we constructed a list of target words related to price change information, such as "inflation," "deflation," "price," "cost," "margin," "labor," "wage," "expense," and "payment." The full set of target words are included in Internet Appendix Table IA.2. In choosing these target words, we erred on the side of casting a wide net, knowing that some sentences may generate false positives (i.e., contain target words but are actually unrelated to price changes). We then ranked all earnings call transcripts in our sub-sample period (January 1, 2021 – June 30, 2021) according to the overall frequency of target words. To achieve uniformity across industries, we picked the top five transcripts (ranked by the frequency of occurrence of target words) from each of the 12 Fama-French industries, excluding finance and utilities. This gave us our training sample of 50 earnings call transcripts.

We then manually labeled every sentence in each of these 50 transcripts, 28,932 sentences in total, with four labels. The first label flags if there is mention of a price change (input, output, or both)

$$L_1 : \text{price change related or not}\,. \tag{1}$$

For instance, the first sentence of the Sanderson Farms transcript in Section 2, "Sanderson Farms operated...business" contains no information about price changes. We therefore set L_1 to zero for this sentence. The next sentence in the above call "Improved poultry markets...margins" clearly discusses price changes. We therefore set L_1 to 1 for this sentence. For those sentences that are labeled as being price change related, we use three additional labels:

$$L_2 : \text{price increase or not}\,, \quad L_3 : \text{input price or not}\,, \quad L_4 : \text{output price or not}\,. \tag{2}$$

Note that we use separate labels L_3 and L_4 for input and output prices, respectively, to allow for the possibility that the same sentence may contain mention of both a change in input prices as well as a change in output prices. Such sentences do occur and can be seen, for example, in the Sanderson Farms call in the sentence "Improved poultry markets...margins". L_2, L_3, and L_4 are all set to 1 in this case. As discussed in Section 3.2.2 below, these human labels are crucial both in training the deep learning models we use and in performing out-of-sample tests to assess the performance of these models. The Appendix Section A.1 provides a detailed description of our labeling procedures.

Before proceeding further, it is reasonable to ask if target words alone can be used to construct measures of firm-level changes in input and output prices from earnings calls, completely bypassing deep learning models. The answer turns out to be negative. The results in panel A of Table 1 help us see this. From this table, we see that $3,430/4,710 = 73\%$ of sentences containing target words do not actually convey price-change-related information.[1] That is, target words cast too wide a net in picking out price-change related sentences.

3.2.2 Training models

Next, we used the training sample to train three related deep learning models and then chose a winner based on their performance in accurately detecting price change related sentences. We define accuracy in equation (3) below. The three candidate deep learning models we analyzed were: BERT (Devlin et al., 2018), RoBERTa (Liu et al., 2019), and FinBERT (Araci, 2019). We picked these three models because of their ability to effectively capture contextual information (Niu, Zhong, and Yu, 2021). This capability translates into improved performance across text processing tasks like machine translation, sentiment analysis, and question answering.

All three candidate models mentioned above already came pre-trained on vast text datasets giving them a solid understanding of human language. This saved us the effort

[1] We find a similar result in a random sample that we constructed as a robustness check on the accuracy of our deep learning model. The results are reported in panel B of Table 1, where we find that $671/750 = 89.5\%$ of sentences containing target words do not actually convey price-change-related information.

of building models from scratch. To tailor these models for our specific objective of identifying sentences related to price changes, we fine-tuned them using our labeled dataset. During this fine-tuning process, the model adapts its focus to relevant input parts to better identify whether a sentence contains price change information. For details about the fine-tuning steps, see Internet Appendix Section IA.2.

During the fine-tuning process, we adopt a computational efficiency enhancing step: we drop all sentences without any target words. This improves computational efficiency significantly as the majority of sentences in any transcript do not contain target words. In dropping sentences without target words, however, we need to make sure that we are not losing too many sentences that actually contain price-change related information. The results in Table 1 assures us that this is unlikely to be the case. Panel A shows that, in the training sample, only $55/24,222 = 0.23\%$ of sentences without a target word are human-labeled as price-change related. Panel B, which are results from a random sample that we constructed as a robustness check on the accuracy of our deep learning model, show similar results: only about $2/750 = 0.27\%$ of sentences without a target word are human-labeled as price-change related.

After the fine-tuning was completed, we assessed the accuracy of each of the three models (BERT, RoBERTa, and FinBERT) in identifying language related to price changes using an out-of-sample test using a portion of the training sample that was not used for fine-tuning. We define accuracy as follows:

$$Accuracy = \frac{TruePositives + TrueNegatives}{TruePositives + FalsePositives + TrueNegatives + FalseNegatives}, \quad (3)$$

where true positives and true negatives are instances where the model's label for a sentence as being price change related or not coincides with the sentence's human label, a false positive is an instance where the model labels a sentence as being price change related but the human label does not, and a false negative is an instance where the model labels a sentence as not being price change related but its human label indicates it as being price change related. The results are reported in the Internet Appendix Table IA.3. From this table, we see that all models perform well, achieving around 90% accuracy.[2] Notably, RoBERTa outperforms others with a test accuracy of 90.44%.

The accuracy test described above was conducted on our training sample which covered earnings calls in the high inflationary sub-period of our full sample. As a robustness check, we repeated the out-of-sample accuracy test described above on a random sample constructed as follows. We first divided all sentences in earnings calls in our full sample into two groups: with/without target words. Then, for each year, we randomly choose 50 sentences that contained a target word and 50 sentences that did not contain a target word, giving us a sample of 1500 sentences. We ran our accuracy test on this random sample. The results, shown in Table IA.3, show even greater accuracy of RoBERTa in this random sample compared to our baseline training sample (95.47% versus 90.44%). We therefore chose RoBERTa as our model for training and generating measurements.

We also compared the performance of RoBERTa with a dictionary-based model, two rule-based models, a bidirectional long short-term memory (Bi-LTSM), a supervised machine learning model that has been used for text classification (Support Vector Machine), and sentence-RoBERTa. The results for all the models are listed in the Internet Appendix Table IA.3. Our RoBERTa model outperforms all these models.

The RoBERTa model discussed above was fine-tuned to identify price-change related information, that is, it used the label L_1 in (1). We will henceforth refer to this RoBERTa model as Model 1. In order to produce the remaining three labels L_2 through L_4, we fine-tune three additional RoBERTa models that generate these labels. We will henceforth refer to these three models as Model 2 through Model 4, respectively. We report the accuracy of all four RoBERTa models in the Internet Appendix in Table IA.4; they range from 90.44% to 96.09%.

[2]To understand how RoBERTa works, we employ the visualization technique introduced by Alammar (2021) on our trained model in Internet Appendix Section IA.4.

3.2.3 Generating the measure of input and output price changes

We use Models 1 through 4 to extract measures of input and output price changes on all earnings calls in our full sample. Similar to our fine-tuning process, for each earnings call transcript, we keep all sentences with target words.

Next, we generate the four price-change labels L_1, L_2, L_3, and L_4 for these sentences. To do so, we first use Model 1, which generates label L_1, and then feed only those sentences that Model 1 labels as being price change related (i.e., those sentences with $L_1 = 1$) into Models 2, 3, and 4 in sequence. These three models produce the remaining three labels L_2, L_3, and L_4. We use these labels to classify price-change related sentences as being: (i) "InputUp" – these are sentences with $L_2 = 1$ and $L_3 = 1$, (ii) "InputDown" – these are sentences with $L_2 = 0$ and $L_3 = 1$, (iii) "OutputUp" – these are sentences with $L_2 = 1$ and $L_4 = 1$, and (iv) "OutputDown" – these are sentences with $L_2 = 0$ and $L_4 = 1$. Finally, we aggregate the sentence classifications to the transcript-level by counting the number of such sentences in each transcript to generate: $\#InputUp_{i,t}$, $\#InputDown_{i,t}$, $\#OutputUp_{i,t}$, and $\#OutputDown_{i,t}$ where i is an index that labels the firm and t denotes the fiscal quarter referred to by the earnings call.[3]

We use these four numbers to construct our measures of input and output price changes. Our baseline measure of input price change is:

$$InPrChg_{i,t} = \frac{\#InputUp_{i,t} - \#InputDown_{i,t}}{\#Sentences\ in\ Transcript_{i,t}}, \tag{4}$$

where $\#Sentences\ in\ Transcript_{i,t}$ is the total number of sentences in firm i's transcript at time t. We subtract $\#InputDown_{i,t}$ from $\#InputUp_{i,t}$ to compute the net input price change. If a firm does not discuss input price changes in a quarter, we set $InPrChg_{i,t} = 0$ in that quarter.

[3]For instance, Apple held its earnings call on April 28, 2021, to report financial performance for the fiscal quarter ending in March 2021. We attribute the text-based measure from this earnings call to Q1 2021, rather than Q2 2021.

Let us relate our price change measure $InPrChg_{i,t}$ to actual changes in input prices at the firm level. First, we define a firm level Laspeyres index $\mathcal{P}_{i,t}^{Input}$ for the basket of the firm's input goods between periods $t-1$ and t:

$$\mathcal{P}_{i,t}^{Input} = \frac{\sum_{g=1}^{M} P_{g,t} Q_{i,g,t-1}}{\sum_{g=1}^{M} P_{g,t-1} Q_{i,g,t-1}}, \tag{5}$$

where $P_{g,t}$ is the price of good g at time t, $Q_{i,g,t}$ is the quantity of good g at time t used by firm i, and M is the total number of goods in the economy. Two comments are in order. First, note that $\mathcal{P}_{i,t}^{Input}$ measures the gross growth rate in total input costs between periods $t-1$ and t, holding *fixed* the quantity of inputs at their $t-1$ values.[4] Second, we note that goods not used in production by firm i have the corresponding quantities $Q_{i,g,t-1}$ set to zero. Next, let us define

$$\pi_{i,t}^{Input} \equiv \mathcal{P}_{i,t}^{Input} - 1 \tag{6}$$

It is easy to show that $\pi_{i,t}^{Input}$ is a weighted sum of the net growth rates of prices of individual goods:

$$\pi_{i,t}^{Input} = \sum_{g=1}^{M} w_{i,g,t}^{Input} \pi_{g,t}, \tag{7}$$

where $\pi_{g,t} = P_{g,t}/P_{g,t-1} - 1$ is the net growth rate of the price of good g between times $t-1$ and t and the weight $w_{i,g,t}^{Input} = P_{g,t-1}Q_{i,g,t-1}/\left(\sum_{g=1}^{M} P_{g,t-1}Q_{i,g,t-1}\right)$ is good g's share of total cost at time $t-1$. Note that $\sum_{g=1}^{M} w_{i,g,t}^{Input} = 1$ for all i and t.

Next, we assume that our input price index is proportional to $\pi_{i,t}^{Input}$:

$$InPrChg_{i,t} = a^{Input} \pi_{it}^{Input}, \tag{8}$$

where a^{Input} is a time-invariant constant of proportionality that is the same for all firms. We will refer to assumption (8) as the "proportionality assumption" below.

[4] $\mathcal{P}_{i,t}^{Input}$ is different from the gross growth rate of COGS of firm i between times $t-1$ and t because $\mathcal{P}_{i,t}^{Input}$ assumes input quantities to be held fixed between $t-1$ and t while COGS does not.

We proceed exactly in the same way for output prices. Our baseline output price measure is:

$$OutPrChg_{i,t} = \frac{\#OutputUp_{i,t} - \#OutputDown_{i,t}}{\#Sentences\ in\ Transcript_{i,t}}. \tag{9}$$

We set $OutPrChg_{i,t} = 0$ if a firm does not discuss output price changes in a quarter. Similar to the assumption (8), we assume that:

$$OutPrChg_{i,t} = a^{Output} \pi_{it}^{Output}, \tag{10}$$

where a^{Output} is a time-invariant constant of proportionality that is the same for all firms and

$$\pi_{i,t}^{Output} = \sum_{g=1}^{M} w_{i,g,t}^{Output} \pi_{g,t}, \tag{11}$$

where $w_{i,g,t}^{Output}$ is the weight of good g in the basket of output goods produced by firm i at time $t-1$.

Assumptions (8) and (10) are based on the premise that a firm experiencing a change in input (output) prices in a particular quarter will discuss input (output) price changes with an intensity that is proportional to the size of the change in input (output) price indices. We will henceforth refer to both assumptions (8) and (10) as the "proportionality assumption". We test the validity of the proportionality assumption in Section 4. In addition, some of our results in Section 5.2, and Section 5.5 provide further validation of this assumption.

Our price change measures $InPrChg_{i,t}$ and $OutPrChg_{i,t}$ capture changes in input and output prices that are driven by aggregate inflation π_t as well as relative price changes. To see this, note that π_{gt} can be decomposed as:

$$\pi_{g,t} = \pi_t + r_{g,t}, \tag{12}$$

where π_t reflects the growth rate of input prices that correspond to changes in the numeraire while $r_{g,t}$ is the relative price change of good g (Reis and Watson, 2007, 2010). Because

$r_{g,t}$ are relative price changes, they add to zero when aggregated over all M goods in the economy

$$\sum_{i=1}^{M} r_{g,t} = 0 \tag{13}$$

for all t.[5] Equations (7), (8), and (12) imply that

$$InPrChg_{i,t} = a^{Input}\pi_t + a^{Input}\sum_{g=1}^{M} w_{i,g,t}^{Input} r_{g,t}, \tag{14}$$

where we have used the fact that the weights $\sum_{g=1}^{M} w_{i,g,t}^{Input} = 1$ for all i and t. The first term on the right-hand side of (14) explicitly shows that $InPrChg_{i,t}$ is affected by aggregate inflation π_t, while the second term on the right-hand side of (14) captures the effect of relative price changes. A similar logic applies to our output price change measure to give:

$$OutPrChg_{i,t} = a^{Output}\pi_t + a^{Output}\sum_{g=1}^{M} w_{i,g,t}^{Output} r_{g,t}. \tag{15}$$

While (4) and (9) are our baseline price change measures, we also use variants of these measures to estimate the likelihood of price changes. Specifically, we construct price change indicator variables which flag a mention of a price change in an earnings call. There are four such indicators:

$$\chi_{it}^{InputUp} = \mathbb{1}_{\#InputUp_{i,t}>0}, \tag{16}$$

$$\chi_{it}^{InputDown} = \mathbb{1}_{\#InputDown_{i,t}>0}, \tag{17}$$

$$\chi_{it}^{OutputUp} = \mathbb{1}_{\#OutputUp_{i,t}>0}, \tag{18}$$

$$\chi_{it}^{OutputDown} = \mathbb{1}_{\#OutputDown_{i,t}>0}. \tag{19}$$

where (16), (17), (18), and (19) are flags for input price increase, input price decrease, output price increase, and output price decrease, respectively, at firm i at time t.

[5] See e.g., the comment in footnote 1 of Reis and Watson (2010) and also Reis and Watson (2007).

While these indicator variables literally flag a mention/or not of an input (output) price change in an earnings call, our proportionality assumptions (8) and (10) imply that a firm mentions input or output price changes in an earnings call only when the firm actually experiences a change in input or output prices that quarter. This, in turn, implies that price change indicators can be used to compute the likelihood of price changes. For instance, the time-series mean of $\chi_{it}^{InputDown}$ estimates the likelihood that firm i experiences an increase in input prices each quarter.

Table 2 presents the descriptive statistics for the 81,473 earnings conference calls in our sample, covering the period from 2007 to 2021. The sample mean of $InPrChg_{i,t}$ is 0.526%, while the average vlaue of $OutPrChg_{i,t}$ is 0.546%.

4 Validation: Comparison with Standard Measures

The ability of our text-based measures of firm-level input and output price changes to reflect actual price changes rely on the proportionality assumption (see equations (8) and (10)). In this section, we test the validity of this assumption by comparing the time-series dynamics of various cross-sectional aggregates of output price changes $OutPrChg_{i,t}$ with existing, commonly used price change measures compiled by the Bureau of Labor Statistics (BLS). We are unable to test the validity of our input price change measure $InPrChg_{i,t}$ along the above lines because the BLS index of input price costs for producers is a much shorter time-series starting in December 2018. Instead, in Section 5.5, we perform a high-frequency analysis of firm-level responses of stock prices to input price changes and show the response to be consistent with interpreting $InPrChg_{i,t}$ as actual input price changes experienced by the firm.

4.1 Aggregate level

We first show that under assumption (10), the cross-sectional mean of $OutPrChg_{i,t}$ is expected to be approximately proportional to aggregate inflation. To see this, let us first define the cross-sectional mean of our output price measure $\mu_t^{OutPrChg} \equiv \frac{1}{N}\sum_{i=1}^{N} OutPrChg_{i,t}$. Using equations (10), (11), (12), and the fact $\sum_{g=1}^{M} w_{i,g,t}^{Input} = 1$, we have:

$$\mu_t^{OutPrChg} = a^{Output}\pi_t + a^{Output}\frac{1}{N}\sum_{i=1}^{N}\sum_{g=1}^{M} w_{i,g,t}^{Output} r_{g,t}\,, \tag{20}$$

where N is the number of firms in our sample. If the weights $w_{i,g,t}^{Output}$ are uncorrelated to the relative price changes $r_{g,t}$, then the last term on the right-hand side of (20) is zero, and the cross-sectional mean of output price change measure $\mu_t^{OutPrChg}$ would be perfectly correlated with aggregate inflation π_t. Whether the sum in (20) is close to zero or not is an empirical question. We proceed with our validation exercise by testing the joint hypotheses—the proportionality assumption (10) is valid and the sum on the right-hand side of equation (20) is approximately zero.

We use the growth rate of the PPI as a proxy for inflation π_t in order to test (20) that is implied by our proportionality assumption. Although the CPI would also work, we use the PPI because the latter is a wholesale price index and therefore is closer in spirit to our firm-level output price measure.

Figure 1 shows the result. The solid line shows $\mu_t^{OutPrChg}$ and the dot-dash line shows the year-over-year growth rate of PPI. The correlation between the two series is 0.8. Moreover, $\mu_t^{OutPrChg}$ successfully captures episodes of large increases/decreases in PPI growth rates. For instance, $\mu_t^{OutPrChg}$ increases significantly in 2008, successfully capturing the 2008 inflationary period driven by large increases in gas prices. Similarly, $\mu_t^{OutPrChg}$ increases in 2011 when PPI increased due to surging food and energy prices. Finally, $\mu_t^{OutPrChg}$ successfully captures the oil price plunge driven decline in PPI in 2014–2015. The high correlation between $\mu_t^{OutPrChg}$ and the growth rate in PPI supports the proportionality assumption (10).

While our text-based measure constructed using earnings calls performs well in capturing aggregate price changes, a similar measure using 10-K filings does not. To show this, we constructed a similar aggregate measure by applying our trained model to the 10-K filings' Management Discussion and Analysis (MD&A) sections of our sample firms. The aggregate trend is shown in Figure IA.4. The correlation between the aggregate 10-K-based output price measure and PPI growth rate is only 0.30, suggesting earnings conference calls are better than 10-K filings in capturing price changes.

4.2 Industry level

We repeat our analysis from Section 4.1 above, but now using disaggregate industry level data. We obtain the industry-level producer price indices from the BLS. We use their industry classification scheme with nine industries that closely resemble the two-digit NAICS codes. The nine industries are: mining, utilities, manufacturing, wholesale trade, retail trade, transportation and warehousing, information, health care, and other selected services less trade, transportation, and warehousing.[6] From this list, we remove two industries: "utilities" since this industry is excluded in our earnings call sample and "other selected services less trade, transportation, and warehousing" since this includes financial companies which are excluded in our earnings call sample. We are thus left with 7 industries. We obtain our text-based measure of output price changes for each of these seven industries by first mapping each industry to a two-digit NAICS code and then computing the cross-sectional mean of $OutPrChg_{i,t}$ for all firms i belonging to each of the seven NAICs industries.

Figure 2 shows the results. Panel A of this figure reports the standard deviation of the year-over-year growth rate at quarterly frequency for each of the seven industries' PPI. Panels B through H of this figure compares our industry-level output price measure (solid line) with the industry-level PPI year-over-year growth rate (dot-dash line). We have sorted the figures in panels B through H according to the volatility of the year-over-year growth

[6] BLS does not provide data for two-digit NAICS industries.

rate of PPI, that is, by the strength of the price change signal (see panel A of Figure 2). From these figures, we see that our text-based measure and industry-level PPI growth rates are highly correlated in industries with a relatively high volatility of its PPI growth rate, that is, in those industries where the price change signal is strong. For instance, Mining has the strongest price change signal and we find the correlation between our text-based measure and PPI growth rate to be 0.74. Similarly, Manufacturing, which has the largest share of firms (41%) in the Compustat universe excluding financial and utility sectors, has the second highest volatility of PPI growth rate. The correlation between our text-based measure and PPI growth rate for Manufacturing is even higher at 0.83.

5 Applications of the Text-Based Measure

In this section, we apply our text-based measure of firm-level price changes derived from earnings calls to first establish properties of input price changes experienced by individual firms. We then establish properties of output price changes. Next, we analyze how firms respond to input price changes. We focus on two firm policies: how they pass through cost changes to output prices and how they adjust the amount they spend on inputs (i.e., cost of goods sold). We conclude by analyzing the immediate reaction of a firm's stock price to input price changes.

Properties of input price changes, firm policies on pass through of input price changes, and the stock price response to input price changes, to the best of our knowledge, are new to the literature. We expect these moments to discipline quantitative analyses of heterogeneous firm models exploring the implications of market power or sticky prices. Properties of output price changes have been studied in the past. The existing literature contains estimates of key moments of output price changes. In these cases, we report the existing estimates along with our own. We find our estimates to be in line with existing estimates in the literature, even though we use a different data source and methodology.

5.1 Properties of input price changes

We use our text-based measure to establish the following properties of input price changes faced by firms: (1) the median firm experiences an increase in input prices once every 7 months. Input price decreases are rarer than increases and occur once every 30 months for the median firm. (2) There is significant cross-sectional heterogeneity both in the likelihood and in the size of input price changes $InPrChg_{i,t}$. These differences can be attributed to an economy-wide common component, an industry component, and a firm-specific component with the firm-specific component explaining about half of the variation in $InPrChg_{i,t}$. (3) The probability of the average firm experiencing an input price increase (decrease) in two successive quarters is 0.53 (0.45).

5.1.1 Likelihood of input price changes

We estimate the likelihood with which firm i experiences an increase or decrease in input prices per quarter to be $\chi_i^{InputUp} = \frac{1}{T_i}\sum_{t=1}^{T_i} \chi_{it}^{InputUp}$ and $\chi_i^{InputDown} = \frac{1}{T_i}\sum_{t=1}^{T_i} \chi_{it}^{InputDown}$, respectively, where T_i is the total number of earnings calls for firm i in our sample and the indicator variables $\chi_{it}^{InputUp}$ and $\chi_{it}^{InputDown}$ are defined in (16) and (17), respectively.

Panel A of Figure 3 shows the distribution of $\chi_i^{InputUp}$. This figure shows that $\chi_i^{InputUp} = 0.44$ for the median firm, which implies that it experiences an input price increase once every $1/0.44 = 2.3$ quarters, that is, once every 7 months. Panel B of Figure 3 shows the distribution of $\chi_i^{InputDown}$. Comparing with panel A, we see that input price decreases are much rarer than increases. For the median firm, $\chi_i^{InputDown} = 0.1$, which translates into a decrease in input prices once every 10 quarters, that is, once every 30 months.

Panel C of Figure 3 is a bin-scatter plot of $\chi_i^{InputUp}$ versus $\chi_i^{InputDown}$. This figure shows that firms which experience more frequent input price increases also experience more frequent input price decreases. In order to determine the relative frequency of input price decreases versus increases, we compute the ratio $\chi_i^{InputDown}/\chi_i^{InputUp}$ for each firm in our sample. We find the median value of this ratio to be 0.33, that is, input price decreases are about a third as frequent as input price increases.[7]

5.1.2 Heterogeneity in frequency and size of input price changes

In this section we show that the frequency and size of input price increases and decreases varies substantially across firms. This can be seen from panel A of Figure 3. The standard deviation of $\chi_i^{InputUp}$ is 0.34. Similar to input price increases, there is also significant heterogeneity across firms. This can be seen from panel B of Figure 3. The standard deviation of $\chi_i^{InputDown}$ is 0.26.

Next, we investigate whether input price changes arise due to aggregate shocks or whether they are mostly firm-specific in nature. Because input price increases are much more frequent than input price decreases, we focus our analysis on input price increases. As a first step, we construct the likelihood of an increase in input prices per quarter for industry J, by computing $\chi_J^{InputUp} = \frac{\sum_{i \in J} \chi_{i,t}^{InputUp}}{\sum_{i \in J} T_i}$ where the sum is over all firms i in industry J. Table 3 shows results for the ten Fama-French industries (i.e., the twelve FF industries excluding finance and utilities). Columns (1) and (2) of this table show substantial heterogeneity across industries in the frequency of both input price increases and decreases. That is, part of the variation in $\chi_i^{InputUp}$ across firms are attributable to across-industry differences. For instance, the firms in the Chemical industry experience input price increases quite frequently, with a likelihood of 0.89 per quarter. In contrast, firms in Business Equipment industry have a likelihood of an input price increase of 0.32 per quarter. We obtain a similar result when we analyze the continuous variable $InPrChg_{i,t}$ from equation (4). For the latter, we compute the industry level measure by computing the equal weighted average of $InPrChg_{i,t}$ (see equation (4)) for all firms i that belong to the industry being analyzed. The results reported in column (3) of Table 3 show substantial heterogeneity in the size of input price changes.

[7] In computing the median, we only considered values of the ratio $\chi_i^{InputDown}/\chi_i^{InputUp}$ that were defined. That is, we dropped firms (about 12% of all firms in our sample) for which $\chi_i^{InputUp} = 0$, that is, firms which never discussed input price increases in our sample. Also, note that the median of the ratio $\chi_i^{InputDown}/\chi_i^{InputUp}$, 0.33, is different from the ratio of the medians of $\chi_i^{InputDown}$ and $\chi_i^{InputUp}$, $7/30 = 0.23$. This is because the latter ratios may correspond to different firms.

Decomposing $InPrChg_{i,t}$ into aggregate, industry, and firm-specific. Next, we decompose $InPrChg_{i,t}$ into aggregate and firm-specific components. To do so, we first assume that a change in input prices experienced by a firm is the sum of the following three components: (1) an economy-wide aggregate component, (2) an industry component including a fixed and also a time-varying part, and (3) a firm-specific component, also including a fixed and a time-varying part. Our decomposition is similar to the one in Reis and Watson (2010) who decompose the change in prices of consumer goods into their aggregate and idiosyncratic components. To quantify the contribution of each component, we follow Hassan et al. (2019) and regress $InPrChg_{i,t}$ on each component in a staggered fashion that we outline below and report the increase in R^2 from adding each component.

We first regress $InPrChg_{i,t}$ on YearQtr fixed effects to determine the contribution of economy-wide inflation. We define the latter as being a common, equiproportional change in all prices, that is, what Reis and Watson (2010) call "pure inflation". The first row in Table 4 shows that this factor explains about 7% of firm-level input price changes.

Next, we estimate the contribution of the industry component. To do so, we must choose an industry classification scheme. The second row of Table 4 shows the contribution of the industry factor for three classification schemes: column (1) uses the Fama-French 12-industry classification scheme, while columns (2) and (3) use the 2-digit and the 3-digit SIC codes, respectively. For expositional ease, we discuss results using the 3-digit SIC code here; the complete results including the other two classication schemes are in Table 4. We see that the industry fixed effect explains 26% of the variance in input price changes; including industry $\times$ YearQtr fixed effects explain an additional 22% of the variance.

The results so far show that the aggregate components (inflation and industry) explain 55% of the variation in $InPrChg_{i,t}$. Of the remaining 45% that is idiosyncratic, 14% is due to a firm fixed effect while the time-varying, firm-specific component explains the remaining 31% of the variation.

To summarize, we find that about half of firm-level variation in input prices is firm-specific containing price change information that is not captured by traditional aggregate price change measures at the economy-wide level or the industry level (e.g., PPI measures). Put differently, our firm-level input price change measure $InPrChg_{i,t}$ contains firm-level input price information missed by aggregate price change series published by the BLS.

5.1.3 Persistence of input price changes

A common approach in reporting the persistence of a random variable is to model it as an AR(1) process and report the estimated AR(1) coefficient. We do not adopt this approach here. Instead, we report a measure of persistence that is model-free because it does not assume a particular stochastic process for the random variable, such as an AR(1) process. To construct this measure, we estimate the probability that a price increase shock lasts for exactly 1, 2, 3, 4, and 5 or more quarters.

To obtain these estimates, we restrict our attention to earnings calls that satisfy all of the following 3 conditions. We keep an earnings call if it: (1) mentions an input price increase at least once, (2) the previous quarter's earnings call does not mention an input price increase, and (3) our sample (which is a subset of all earnings calls that actually occurred) contains earnings calls by this firm for at least five consecutive quarters including this call. The second condition ensures that we start counting the length of consecutive price increases from a call which mentions an input price increase for the first time in at least the past two quarters. The third condition ensures that we are able to estimate the probability that a price increase shock lasts for exactly 1 through 5 or more quarters.

Panel A of Table 5 shows the results. We see that $4,595$ calls satisfy all three conditions above. From these calls, we estimate that the probability that an input price increase will last exactly 1, 2, 3, or 4 quarters is 46.66%, 18.78%, 9.60%, and 5.79%, respectively, while there is a 19.17% probability that the input price increase will persist for 5 or more quarters. The table implies that the probability of a firm experiencing an input price increase at time $t+1$ given that it experienced an input price increase at t is $(863+441+266+881)/4595 = 0.53$.

To estimate the persistence of input price decreases, we mimic our approach for input price increases, except that in selecting the sub-sample of earnings calls, we use the same three conditions listed above, but with “input price increase” replaced by “input price decrease” in the first two conditions. Panel B of Table 5 shows the results for input price decreases. We estimate the probability that an input price decrease will last exactly 1, 2, 3, or 4 quarters to be 55.10%, 17.41%, 8.27%, and 4.44%, respectively, while there is a 14.78% probability that the input price decrease will persist for 5 or more quarters. The table implies that the probability of a firm experiencing an input price decrease at time $t+1$ given that it experienced an input price decrease at t is $(808+384+206+686)/4,641 = 0.45$. Comparing these numbers with our estimates for input price increases, we see that the persistence of input price increases and decreases are similar to each other.

5.2 Properties of output price changes

In this section, we report key summary statistics of the frequency of output price changes. Our headline results in this section are: (1) the median frequency of output price increase at the firm-level is once every 6 months. (2) Output price decreases are much rarer, occurring once every 21 months. (3) There is substantial cross-sectional heterogeneity in the frequency with which firms change their output prices. This heterogeneity persists when firms are aggregated to the industry level. The above properties have been estimated in prior literature. We find out estimates to be close to existing estimates even though we use a different data source and methodology. Our results in this section, while not new, serve as important validations of our text-based measure of output price changes that are at a more disaggregated level than those discussed in Section 4.

Likelihood of output price changes. We estimate the likelihood with which firm i experiences an increase or decrease in output prices per quarter to be $\chi_i^{OutputUp} = \frac{1}{T_i}\sum_{t=1}^{T_i}\chi_{it}^{OutputUp}$ and $\chi_i^{OutputDown} = \frac{1}{T_i}\sum_{t=1}^{T_i}\chi_{it}^{OutputDown}$, respectively, where T_i is the total number of earnings calls for firm i in our sample and the indicator variables $\chi_{it}^{OutputUp}$ and $\chi_{it}^{OutputDown}$ are defined in equations (18) and (19), respectively.

Panels A and B of Figure 4 show the distributions of $\chi_{i}^{OutputUp}$ and $\chi_{i}^{OutputDown}$, respectively. The likelihood of output price increase for the median firm is 0.5, which implies an increase in output price once every 2 quarters or 6 months. There are estimates of price stickiness in the existing literature. Those were calculated using confidential microdata that the Bureau of Labor Statistics (BLS) uses in constructing the producer price index (PPI). Those estimates of price stickiness range from a change once every 4.3 months (Bils and Klenow, 2004) to once every $8-11$ months (Nakamura and Steinsson, 2008). Our estimates are within this range and close to the estimate in Gorodnichenko and Weber (2016) who find a mean duration of price change once every 6.54 months.

Comparing panels A and B of Figure 4, we see that output price increases are more common than output price decreases. The median likelihood of output price decrease is 0.143. This implies a frequency of output price decrease once every 7 quarters or 21 months.

Panel C of this figure is a bin-scatter plot that compares the frequency of output price increases and decreases at the same firm. Similar to input prices, firms which experience more frequent output price increases also experience more frequent output price decreases. In order to determine the relative frequency of output price decreases versus increases, we compute the ratio $\chi_{i}^{OutputDown}/\chi_{i}^{OutputUp}$ for each firm in our sample. We find the median value of this ratio to be 0.39.[8] Our result is similar to Nakamura and Steinsson (2008) who use product level data and also find that a third of all price changes are price decreases.

Panels A and B also highlight considerable heterogeneity in $\chi_{i}^{OutputUp}$ and $\chi_{i}^{OutputDown}$ across firms. The standard deviations of $\chi_{i}^{OutputUp}$ and $\chi_{i}^{OutputDown}$ are 0.344 and 0.269, respectively. Our result on heterogeneity in the frequency of output price changes is not new to the literature and has been previously documented by D'Acunto et al. (2018). Using the confidential BLS microdata, they find that the standard deviation of price changes is 0.14

[8] Similar to our computation for input prices, in computing the median, we only considered values of the ratio $\chi_{i}^{OutputDown}/\chi_{i}^{OutputUp}$ that were defined.

per month, or 0.42 per quarter, which is comparable to our estimate.

5.3 Properties of cost pass through

Our text-based measures of input and output price changes allow us to analyze individual firms' dynamic pricing policy in response to changes in input costs. Estimation and characterization of firms' cost pass through policies is the subject of substantial work in macroeconomics and industrial organization, especially given the recent rise in product market competition (see e.g., Gutierrez and Philippon 2017 and Loecker, Eeckhout, and Unger 2020). There is also an emerging literature in finance that establishes the link between a firm's price setting policy and the firm's cost of capital (see, e.g., Dou, Ji, and Wu 2021). Our contribution to this literature is to provide estimates of pass through using a relatively large cross-section of firms.

We begin by estimating properties of the dynamic cost pass through policy of the average firm in our sample. We then illustrate how our text-based price change measures can be used to better understand the relation between a firm's competitive environment and its cost pass through policy.

5.3.1 Properties of cost pass through in the data

For pass through estimation, we restrict our analysis to firms whose average frequency of input price changes is less than or equal to once a year. We use this filter to make sure that our estimates are not confounded by input price changes experienced by a firm in the recent past. That is, the observed response of $OutPrChg_{i,t}$ of firm i at time t is, in general, the response to an input price change at time t and also input price changes experienced at $t-1$, $t-2$, and so on. By focusing on firms which, on average, experience less than one input price change in a year, we reduce the confounding effects of past input price changes. A drawback of our filter, however, is that our pass through estimates do not cover firms which receive frequent input price changes, say every quarter.

We estimate the magnitude of firm's cost pass through by analyzing the behavior of $OutPrChg_{i,t}$ in the contemporaneous quarter and subsequent quarters following an input price change. To estimate the magnitude of the pass through, we first assume that a^{Input} of Equation (8) and a^{Output} of Equation (10) are equal to each other, and then use the following regression specification:

$$OutPrChg_{i,t+h} = \theta_i + \beta_h InPrChg_{i,t} + Controls_{i,t} + \epsilon_{i,t}, \tag{21}$$

for $h = 0, 1, 2, 3$ (i.e., we run separate regressions for each horizon h). We include firm fixed effects θ_i to control for potential systematic differences in firms' communication strategies. Additionally, we also control for the following firm characteristics: size, market-to-book ratio, earnings surprise, pre-event return, uncertainty, overall sentiment, leverage, cost of goods sold, and return on assets (see definitions in Table A.1). We do not include time (i.e. year-quarter) fixed effects in our baseline regression in order to interpret β_h as the sensitivity of a firm's output price to a change in its input price, where the input price change includes aggregate inflation.[9] We do, however, report results of all of our main regressions using both firm and year-quarter fixed effects in the Internet Appendix Table IA.5 and find our results to remain unchanged from our baseline results.

Table 6 reports the results. The coefficient of interest β_h estimates pass through as an elasticity since $OutPrChg$ and $InPrChg$ both capture the fractional change in output and input prices, respectively. Column (1) shows that the contemporaneous pass through is 0.55, that is, a 10% change in input prices in a quarter shows up as a 5.5% change in the output price in that quarter. This result is statistically significant at the 1% level. Our estimate is similar to that obtained by Amiti, Itskhoki, and Konings (2019) who use Belgian firm data and estimate a cost pass through of 0.6. Our estimate is also not far from the estimates in Peltzman (2000), who uses PPI indices for several categories of output goods, together with

[9] Adding a time-fixed effect would remove the common component of input price change, that is, aggregate inflation.

the PPI indices corresponding to the input goods used to produce those output goods, to find a pass through of about 0.4.[10] Finally, columns (2)–(4) of Table 6 shows that the pass through declines over time lasting about two quarters following an input price shock.

5.3.2 A firm's competitive environment and its pass through

In this section, we analyze how cost pass through varies across firms operating in different competitive environments. We begin by describing a simple, static toy model that helps us understand how a firm's market power shapes its pass through policy. We then report our empirical results.

Model. We use the framework of Weyl and Fabinger (2013) (henceforth "WF") because it covers a wide class of models commonly used in the industrial organization literature including Cournot, Bertrand, and monopolistic competition (e.g., Atkeson and Burstein 2008).

Each firm in the economy uses a common input good for production and sells a homogeneous output good. Firms face symmetric demand and cost functions. Let c be the constant marginal cost of production and q_i be the quantity of good sold by firm i. We focus on the symmetric equilibrium in which all firms sell the same quantity $q_1 = \cdots = q_n = q$ at the common price $p(q)$.

An individual firm's profit maximization condition can be written in the form:

$$p(q) - c = \theta\mu(p)\,, \tag{22}$$

where the left-hand side of equation (22) is the absolute markup charged by the firm. The function $\mu(p) \equiv p/\epsilon_D$ is the ratio of the price to the elasticity of the market demand function

[10]More precisely, Table 4 of Peltzman (2000) reports pass throughs of 0.371 and 0.430 within 2 and 4 months, respectively, following an input price change. Since we estimate the contemporaneous pass through as occuring in the same quarter, we report the estimate in Peltzman (2000) as the average of 0.371 and 0.430.

$\epsilon_D = -\frac{p}{qp'}$ where the derivative $p' \equiv dp(q)/dq$. In equation (22), θ reflects the extent of a firm's market power—varying from 0 for perfect competition to 1 for a monopolist, and taking intermediate values such as $1/n$ for Cournot competition. WF show that θ does not vary with market conditions in a wide class of competitive environments. We will therefore assume θ to be a constant parameter.

Pass through. Let us define the *absolute* cost pass through $\rho \equiv \frac{dP}{dc}$. Note that the definition of ρ is different from the one we used in our empirical estimates where we estimated pass through defined in *relative* terms $\frac{\Delta p/p}{\Delta c/c} = \frac{d \log p}{d \log c}$. We prefer using absolute pass through in the model section because the expression for absolute pass through is algebraically simpler than that for relative pass through (see equation (23)). Additionally, since p and c are both positive, the measures of relative and absolute pass through are either both increasing or both decreasing in θ. That is, we obtain the same qualitative comparative static result using either absolute or relative cost pass through.

Implicit differentiation of equation (22) with respect to c gives an expression for the absolute cost pass through:

$$\rho = \frac{1}{1 - \theta\mu'(p)}\,. \tag{23}$$

Equation (23) shows that the dependence of pass through on market power is, in general, ambiguous and depends on the sign of $\mu'(p)$, or equivalently, on the sign of the curvature of the log demand function $\log q(p)$.[11] For instance, if the log demand function is concave, then $\mu'(p) < 0$, and equation (23) implies that firms with more market power (i.e., higher values of θ) will be associated with a smaller pass through ρ.

Empirical results. We analyze the dependence of pass through on market power by focusing on industry concentration as a measure of market power. To determine the relation

[11] $\mu'(p)$ is related to $(\log q(p))''$ by the relation $\frac{\mu'(p)}{\mu^2(p)} = (\log q)''$. The latter relation follows from first writing $\mu(p) = p/\epsilon_D = -\frac{q}{dq/dp}$ which implies that $-\frac{1}{\mu(p)} = (\log q)'$, and then differentiating with respect to p to obtain $\frac{\mu'(p)}{\mu^2(p)} = (\log q)''$.

between pass through and industry concentration, we assign firms into one of three groups (high, medium, and low) based on terciles of industry concentration within each quarter. The measure of industry concentration we use is based on the 10-K text-based network (TNIC) industry concentration from Hoberg and Phillips (2016). For each group, we separately estimate cost pass through by estimating:

$$OutPrChg_{i,t} = \theta_i + \beta InPrChg_{i,t} + Controls_{i,t} + \epsilon_{i,t}. \tag{24}$$

We focus on the contemporaneous pass through in this analysis because our results in Table 6 show that the pass through occurs mostly in the same quarter as the input price change.

Table 7 shows the results. We see that pass through is negatively related to industry concentration, that is, pass through is lower for firms with greater market power. If we make the common assumption that the high concentration group has a higher value of θ, then, our empirical results imply that $\mu'(p) < 0$ (see equation (23)), or equivalently, that the log demand function is concave. Quantitatively, the coefficient β is 20% higher for the high concentration group compared to the low concentration group (0.622 versus 0.501). The large difference in cost pass through between firms in low and high concentration industries suggests that it might be important to account for this heterogeneity in quantitative analyses in which cost pass through by firms plays an important role in determining aggregate outcomes. The inverse relation between pass through and market power that we obtain is in line with existing studies in the industrial organization literature which has found a similar inverse relation in the context of exchange-rate pass through (see e.g., Auer and Schoenle 2016 and Berman, Martin, and Mayer 2012) and for Belgian firms (Amiti, Itskhoki, and Konings, 2019). To the best of our knowledge, our result for a relatively large cross-section of U.S. firms is new to the literature.

5.4 Input price changes and cost of goods sold

In this section, we analyze how firms respond to input price changes by changing the amount spent on inputs as measured by their reported value of cost of goods sold (COGS). In particular, we find that input price increases (decreases) experienced by a firm are followed, on average, by an increase (decrease) in the firm's COGS and this increase persists for five quarters. Our result implies that the average firm in our sample has an inelastic demand curve for inputs. To the best of our knowledge, this result is new.

We analyze the response of firm-level COGS to input price changes using the following regression specification:

$$Y_{i,t+h} = \theta_i + \beta InPrChg_{i,t} + Controls_{i,t} + \epsilon_{i,t} \tag{25}$$

for $h = 0, 1, 2, 3, 4$ quarters, where $Y_{i,t+h} = \left(COGS_{i,t+h} - COGS_{i,t-1}\right)/AT_{i,t-1}$. We add a firm-fixed effect and use the same controls as in (21), namely, size, market-to-book ratio, earnings surprise, pre-event return, uncertainty, overall sentiment, leverage, cost of goods sold, and return on assets. Similar to Section 5.3.2, this analysis focuses on firms that experience input price changes no more than once a year.

Table 8 shows the results. We see that changes in input price are positively related to future changes in COGS. This, in turn, implies that the elasticity of the average firm's demand for its basket of inputs in response to changes in the price of inputs is less than one. To see this, first note that, COGS is a product of the quantity of input bought times the input price. Our finding of a positive relation between changes in input price and COGS implies that a 1% increase in input price index $P_{i,t}^{Input}$, for example, results in a decline in the quantity of input bought that is less than 1% (so that the product, which is COGS, increases). In short, the average firm's demand for inputs has an elasticity less than one.

5.5 Stock price reaction

In this section, we analyze the immediate reaction of a firm's stock price following input price changes. The hypothesis we want to test is the following: a greater intensity of discussion of input price increases by a firm (as captured by a high realization of $InPrChg_{i,t}$) is expected to negatively affect the firm's stock price because an increase in input costs reduces future firm profits and dividends. For this analysis, we do not restrict our analysis to firms which experience less than one input price change per year on average. This is because the confounding effect of input price changes in past quarters are not expected to affect a firm's stock price in the current quarter.

We estimate the immediate reaction of the stock price of firm i to an input price change mentioned in the time t earnings call using the empirical specification:

$$Y_{i,t} = \theta_i + \beta InPrChg_{i,t} + Controls_{i,t} + \epsilon_{i,t}. \tag{26}$$

The dependant variable $Y_{i,t}$ is the firm's cumulative abnormal return (CAR) measured over a three-day window starting one day before the call and ending one day after the call. The CAR is measured relative to the Capital Asset Pricing Model (CAPM). Specifically, the abnormal return for firm i on day t is calculated as $AR_{i,t} = R_{i,t} - (\alpha_i + \beta_i R_{M,t})$, where $R_{i,t}$ is the stock return for firm i on day t and $R_{M,t}$ is the return of value-weighted market index on day t. Then the cumulative abnormal return over the three day window is calculated as $CAR_{i,t}[-1,+1] = \sum_{t=-1}^{+1} AR_{i,t}$. We estimate α_i and β_i in the CAPM model from a 255 trading-day estimation period ending 91 trading days before the earnings call. We remove an observation from the sample if the stock has fewer than 15 days of return data in the estimation period. We add a firm-fixed effect and use the same controls as in (21). We report robust t-statistics that are double clustered at the firm and year-quarter levels.

Table 9 presents the findings. In Column (1), we include all the control variables without adding any fixed effects. We find that the coefficient estimate on $InPrChg_{i,t}$ is negative and significant at the 1% level. In column (2), we add firm fixed effects to account for the possibility that certain firms may, on average, discuss price change information more/less than other firms. We find that the coefficient estimate remains negative and significant at the 1% level. The effect is economically significant—a one standard deviation increase (decrease) in input price is associated with a $36.3 \times 1.026 = 37$ basis points negative CAR.

Our result of an immediate stock price reaction following input price change discussions in earnings calls underscores the fact that our input price change measure $InPrChg_{i,t}$ contains a part that is a surprise to investors. Indeed, if $InPrChg_{i,t}$ was entirely anticipated by investors, we would not expect to see a firm's stock price react to $InPrChg_{i,t}$. Additionally, as a robustness check, we control for the predictable component of $InPrChg_{i,t}$ by including lagged input price change as a control in (26). The results are shown in column (3) of Table 9. We see that the coefficient estimate hardly changes from the baseline specification (see columns (2) and (3) of Table 9). Specifically, while our baseline specification estimates the cumulative abnormal return to a one standard deviation change in $InPrChg_{i,t}$ to be 37 bps, the estimate controlling for lagged input price change is negative $32.7 \times 1.026 = 33.6$ bps.

We perform two robustness checks on our result relating the immediate stock price reaction to $InPrChg_{i,t}$. First, we make sure that the abnormal return around an earnings call are not driven by announcements in the call that indicated a change in firm risk. To do so, we control for the measure of firm-level risk by Hassan et al. (2019). Their measure is a text-based measure of overall risk and is constructed by counting the number of mentions of synonyms for risk or uncertainty in earnings conference calls divided by the length of the transcript (see equation (2) of that paper). Column (4) of Table 9 shows results when we control for this measure of firm risk. We see that the slope coefficient does not change much and the result remains statistically significant at the 1% level.

Our second robustness check addresses the concern that our result is driven by periods of high price volatility. We do so by including only the earnings calls reported during years of low realized price volatility. We define realized price volatility for a given year as the standard deviation of month-over-month CPI within that year. Years that fall into the bottom half of the price volatility distribution are classified as low price volatility periods.[12] Column (5) of Table 9 shows that the economic magnitude of the slope coefficient is actually larger than the full sample and remains statistically significant at the 1% level. Our results are therefore not driven by periods of high price volatility.[13]

Our results on the immediate stock price reaction following mentions of input price changes serve two purposes. First, they serve as an important validation exercise for our input price measures. This is useful for our input price measure because unlike our output price measure, we were unable to compare our input price measure with those from the BLS because of the lack of a sufficiently long time series for input price indices. Second, our results provide an estimate of the sensitivity of firm value to announcements of changes in input prices. To the best of our knowledge, these results are new.

Our final test is more granular. Here, we analyze the immediate stock price response following an input price increase and an input price decrease separately. Specifically, we run the regression specification (26), but with the key independent variables being $Frac_{i,t}^{InputUp} = \frac{\#InputUp_{i,t}}{\#Sentences\ in\ Transcript_{i,t}}$ and $Frac_{i,t}^{InputDown} = \frac{\#InputDown_{i,t}}{\#Sentences\ in\ Transcript_{i,t}}$. Column (6) of Table 9 shows our results. We see that the stock price declines following an increase in discussions of input price increase in earnings calls. The result is statistically significant at the 1% level and also economically large. A one standard deviation increase in $Frac_{i,t}^{InputUp}$ is accompanied by a $-40.5 \times 1.105 = -45$ bps CAR. The point estimate for $Frac^{InputDown}$ is positive, suggesting that an increase in discussions of input price decreases are accompanied by a positive CAR. However, this result is not statistically significant, likely because there are about three times less observations for input price decreases relative to input price increases (see Section 5.1.1).

[12] In our sample from 2007-2021, the years 2007, 2010, 2011, 2014, 2016, 2017, 2018, and 2019 exhibit low price volatility.

[13] In Table IA.6 of the Internet Appendix, we further investigate long horizon versions of our baseline results in Table 9. We find that the stock price response does not persist for long horizons.

6 Conclusion

We propose firm level measures of input and output price changes constructed by textual analysis of earnings calls. Our measures cover a broad cross-section of publicly listed U.S. firms and we establish several properties of input price changes and their effects. We analyze the effect of input price changes on two firm policies: their price setting decisions in response to input price changes and how they adjust their total variable input costs (i.e., cost of goods sold). We also analyze the effect of input price changes on a firm's stock price.

Our specific input and output price change measures have several potential applications. For instance, they can be used to analyze the effect of input prices on real (e.g., hiring) and financial decisions made by firms. Moreover, because our measures track input and output price changes for a large cross-section of firms over varying macroeconomic conditions (e.g, varying rates of aggregate inflation) they can be used to discipline, or potentially rule out, models of sticky prices along the lines of Alvarez, Lippi, and Oskolkov (2020).

References

Alammar, J. 2021. Ecco: An open source library for the explainability of transformer language models. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: System Demonstrations*. Association for Computational Linguistics.

Alvarez, F., F. Lippi, and A. Oskolkov. 2020. The macroeconomics of sticky prices with generalized hazard functions. *Quarterly Journal of Economics* 137:989–1038.

Amiti, M., O. Itskhoki, and J. Konings. 2019. International shocks, variable markups, and domestic prices. *The Review of Economic Studies* 86:2356–402.

Araci, D. 2019. Finbert: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063* .

Atkeson, A., and A. Burstein. 2008. Pricing-to-market, trade costs, and international relative prices. *American Economic Review* 98:1998—-2031.

Auer, R. A., and R. S. Schoenle. 2016. Market structure and exchange-rate pass-through. *Journal of International Economics* 98:60—-77.

Berman, N., P. Martin, and T. Mayer. 2012. How do different exporters react to exchange rate changes? *Quarterly Journal of Economics* 127:437—-492.

Bils, M., and P. J. Klenow. 2004. Some evidence on the importance of sticky prices. *The Journal of Political Economy* 112:947–85.

Borenstein, S., A. C. Cameron, and R. Gilbert. 1997. Do gasoline prices respond asymmetrically to crude oil price changes? *Quarterly Journal of Economics* 112:305–39.

Buehlmaier, M. M., and T. M. Whited. 2018. Are financial constraints priced? evidence from textual analysis. *The Review of Financial Studies* 31:2693–728.

Chava, S., W. Du, and N. Paradkar. 2022. More than buzzwords? firms' discussions of emerging technologies in earnings conference calls. *Working Paper* .

D'Acunto, F., R. Liu, C. Pflueger, and M. Weber. 2018. Flexible prices and leverage. *Journal of Financial Economics* 129:46–68.

Danthine, J.-P., and J. B. Donaldson. 2002. Labour relations and asset returns. *Review of Economic Studies* 69:41–64.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* .

Dou, W. W., Y. Ji, and W. Wu. 2021. Competition, profitability, and discount rates. *Journal of Financial Economics* 140:582–620.

Eichenbaum, M., N. Jaimovich, and S. Rebelo. 2011. Reference prices, costs, and nominal rigidities. *American Economic Review* 101:234–62.

Florackis, C., C. Louca, R. Michaely, and M. Weber. 2023. Cybersecurity risk. *The Review of Financial Studies* 36:351–407.

Gorodnichenko, Y., and M. Weber. 2016. Are sticky prices costly? evidence from the stock market. *American Economic Review* 106:165–99.

Gutierrez, G., and T. Philippon. 2017. Declining competition and investment in the u.s. *NBER Working Paper* .

Handley, K., and J. F. Li. 2020. Measuring the effects of firm uncertainty on economic activity: New evidence from one million documents. *NBER Working Paper* .

Hanley, K. W., and G. Hoberg. 2010. The information content of ipo prospectuses. *Review of Financial Studies* 23:2821–64.

Harford, J., Q. He, and B. Qiu. 2024. Firm-level labor-shortage exposure. *Working Paper* .

Hassan, T. A., S. Hollander, L. Van Lent, and A. Tahoun. 2019. Firm-level political risk: Measurement and effects. *The Quarterly Journal of Economics* 134:2135–202.

Hastings, J. S. 2004. Vertical relationships and competition in retail gasoline markets: Empirical evidence from contract changes in southern california. *American Economic Review* 94:317—-328.

Hoberg, G., and V. Maksimovic. 2015. Redefining financial constraints: A text-based analysis. *The Review of Financial Studies* 28:1312–52.

Hoberg, G., and G. Phillips. 2016. Text-based network industries and endogenous product differentiation. *Journal of Political Economy* 124:1423–65.

Hong, G. H., and N. Li. 2017. Market structure and cost passthrough in retail. *The Review of Economics and Statistics* 99:151—-166.

Jha, M., J. Qian, M. Weber, and B. Yang. 2024. Chatgpt and corporate policies. *NBER Working Paper* .

Jiang, F., J. Lee, X. Martin, and G. Zhou. 2019. Manager sentiment and stock returns. *Journal of Financial Economics* 132:126–49.

Jurafsky, D., and J. H. Martin. 2014. Speech and language processing. vol. 3.

Li, K., F. Mai, R. Shen, and X. Yan. 2021. Measuring corporate culture using machine learning. *Review of Financial Studies* 34:1–35.

Liu, Y., M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692* .

Loecker, J. D., J. Eeckhout, and G. Unger. 2020. The rise of market power and the macroeconomic implications. *The Quarterly Journal of Economics* 135:561—-644.

Lopez-Lira, A. 2023. Risk factors that matter: Textual analysis of risk disclosures for the cross-section of returns. *Working Paper* .

Loughran, T., and B. McDonald. 2011. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *Journal of Finance* 66:35–65.

Mikolov, T., I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, 3111–9.

Nakamura, E. 2008. Pass-through in retail and wholesale. *American Economic Review* 98:430—-437.

Nakamura, E., and J. Steinsson. 2008. Five facts about prices: A reevaluation of menu cost models. *The Quarterly Journal of Economics* 123:1415–64.

Narasimhan, J., and D. Wu. 2013. Word power: A new approach for content analysis. *Journal of Financial Economics* 110:712–29.

Neiman, B. 2010. Stickiness, synchronization, and passthrough in intrafirm trade prices. *Journal of Monetary Economics* 57:295—-308.

Niu, Z., G. Zhong, and H. Yu. 2021. A review on the attention mechanism of deep learning. *Neurocomputing* 452:48–62.

Peltzman, S. 2000. Prices rise faster than they fall. *Journal of Political Economy* 108:466–502.

Pennington, J., R. Socher, and C. D. Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 1532–43.

Pitschner, S. 2020. How do firms set prices? narrative evidence from corporate filings. *European Economic Review* 124:103406–.

Reis, R., and M. W. Watson. 2007. Measuring changes in the value of the numeraire. *Working Paper* .

———. 2010. Relative goods' prices, pure inflation, and the phillips correlation. *American Economic Journal: Macroeconomics* 2:128–57.

Sauter, Z., L. van Lent, G. Vilkov, and R. Zhang. 2023. Firm-level climate change exposure. *The Journal of Finance* 78:1449–98.

Weyl, E. G., and M. Fabinger. 2013. Pass-through as an economic tool: Principles of incidence under imperfect competition. *Journal of Political Economy* 121:528–83.

Woodford, M. 2011. *Interest and prices: Foundations of a theory of monetary policy*. Princeton University Press.

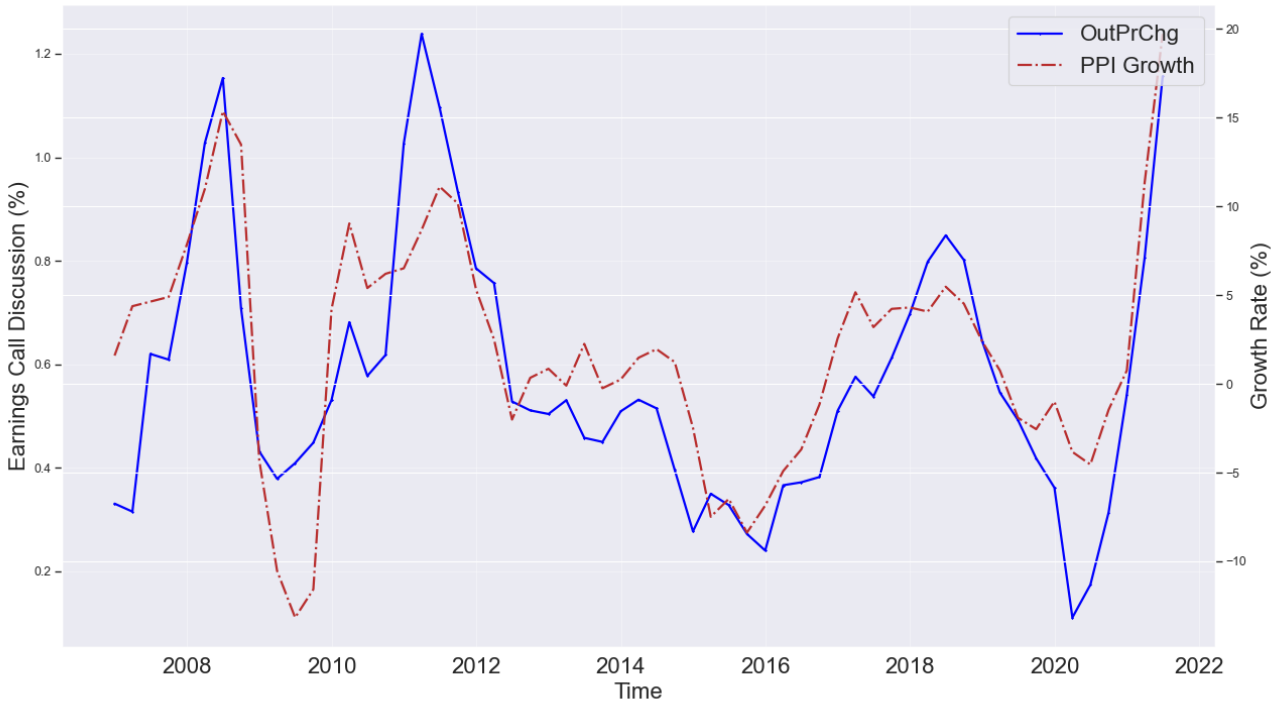

Figure 1: Aggregate PPI and the aggregate text-based output price change measure. This figure shows the year-over-year growth rate of the aggregate Producer Price Index (PPI) by commodity (all commodities) and the cross-sectional mean of the text-based aggregate output price change measure. Both lines are at quarterly frequency.

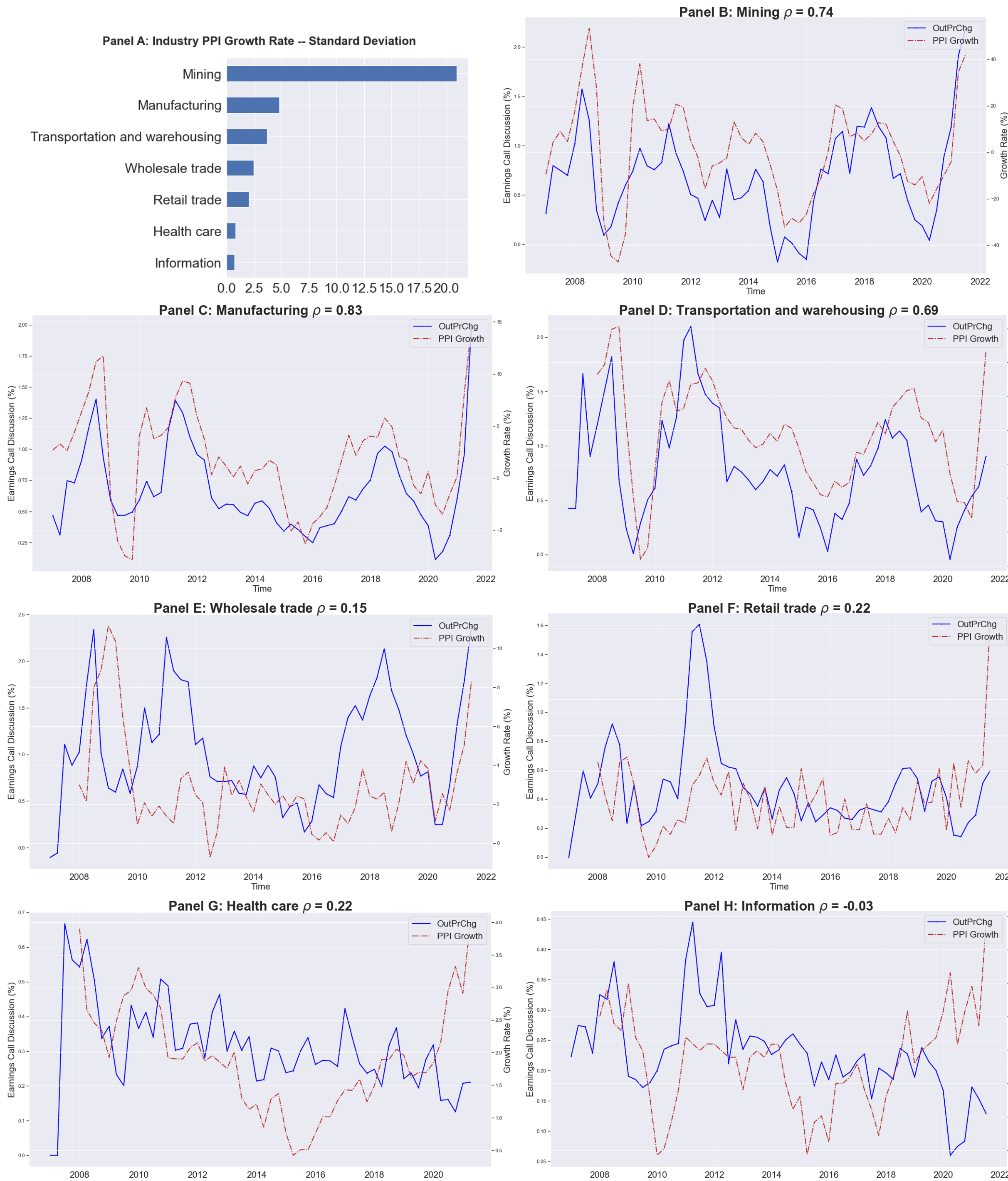

Figure 2: Industry-level PPI and text-based output price changes. Panel A shows the standard deviation of industry-level year-over-year PPI growth rate measured at quarterly frequency. Panels B through H compare the year-over-year growth rate of the BLS reported PPI for each industry and the cross-sectional mean of the text-based output price changes across earnings call transcripts for that industry in each quarter. Panels B through H are sorted in decreasing order of the standard deviation of industry-level PPI growth rate. The correlation between the two time series are reported in the title of each figure.

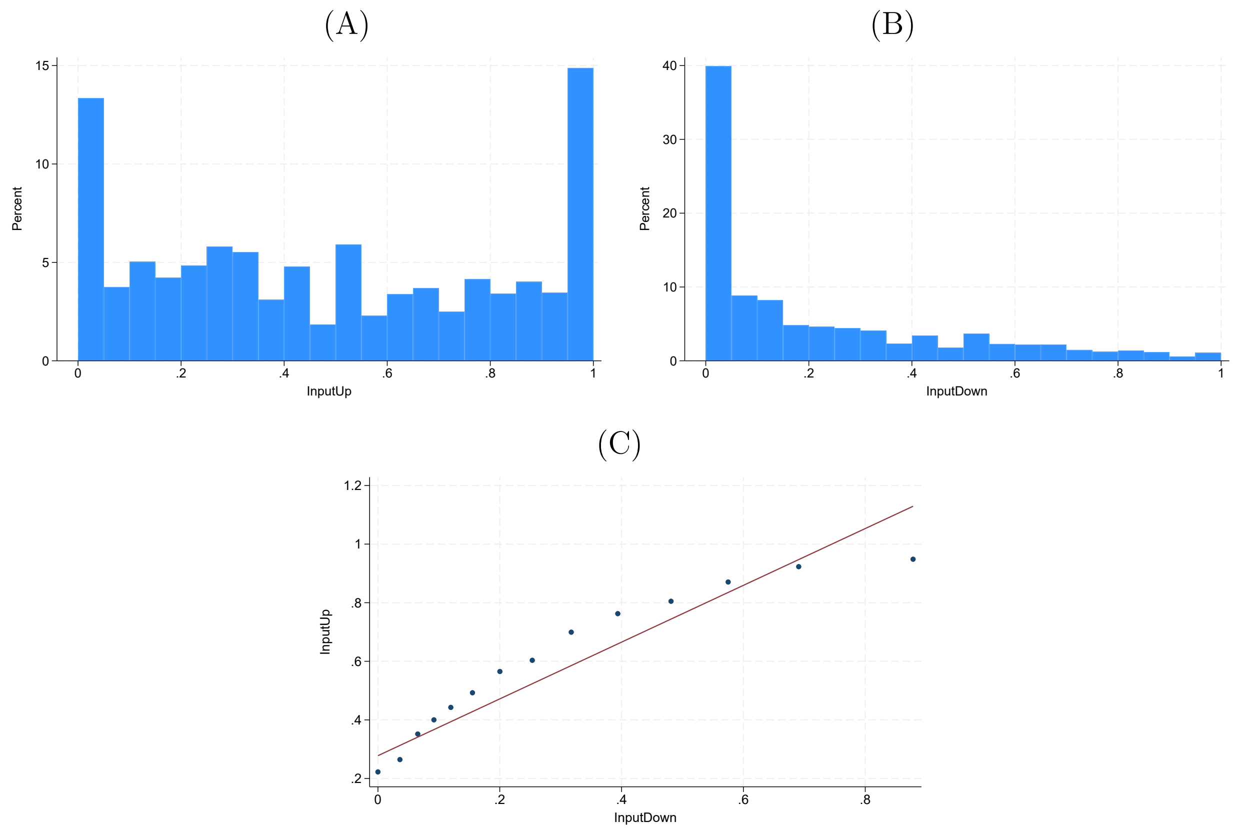

Figure 3: Distribution of likelihood input price increases, decreases, and their relation. Panels A and B plot the firm-level distribution of the likelihood of input price increases and decreases per quarter, respectively. Panel C is a bin-scatter plot which shows the relation between the likelihood of input price increases (y-axis) and input price decreases (x-axis) at the individual firm level.

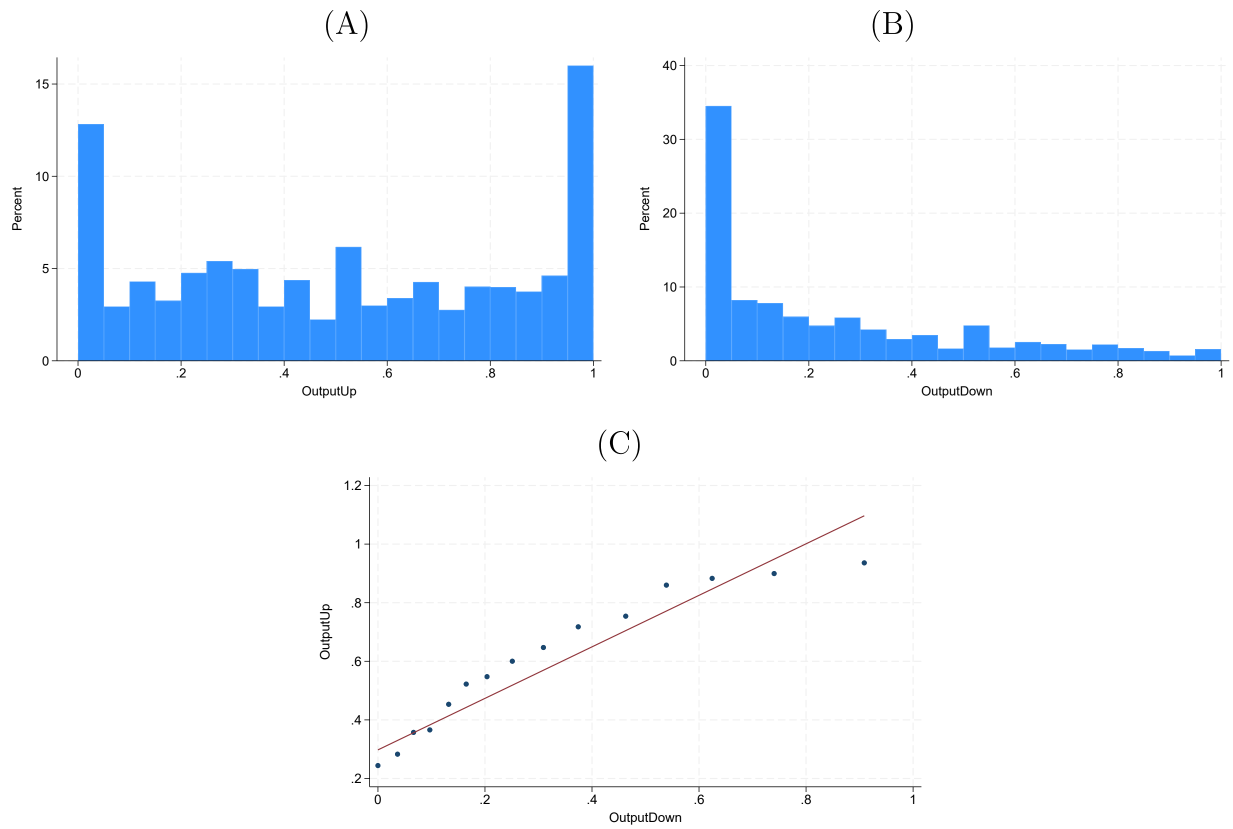

Figure 4: Distribution of likelihood output price increases, decreases, and their relation. Panels A and B in the top row plots the firm-level distribution of the likelihood of output price increases and decreases, respectively. Panel C in the top row is a bin-scatter plot which shows the relation between the likelihood of output price increases (y-axis) and the output price decreases (x-axis) at the individual firm level.

Table 1: Properties of training and random sample

Panel A panel reports the number of sentences under each category in the labeled training sample, which consists of 50 earnings call transcripts. Panel B reports the number of sentences under each category in the randomly sampled data. In the randomly sampled data, we sample 100 sentences from each year – 50 with a target word and 50 with no target words.

Panel A: Main Training Data			
	Target Words	**No Target Words**	Sum
Price Change	1,280	55	1,335
No Price Change	3,430	24,167	27,597
Total	4,710	24,222	28,932
Panel B: Randomly Sampled Data			
	Target Words	**No Target Words**	Sum
Price Change	79	2	81
No Price Change	671	748	1,419
Total	750	750	1,500

Table 2: Descriptive statistics

This table presents the descriptive statistics of the characteristics of earnings conference calls and the characteristics of the firms participating in these earnings calls. All continuous variables are winsorized at the 1% and 99% levels.

	Mean	Median	Std. Dev.
Input Price-Change Discussion in Earnings Calls			
InPrChg (%)	0.526	0.000	1.026
$Frac^{InputUp}$ (%)	0.674	0.224	1.105
$Frac^{InputDown}$ (%)	0.145	0.000	0.334
Outcome variables			
OutPrChg (%)	0.546	0.074	1.078
CAR[−1,+1] (%)	0.052	0.092	9.370
Control variables			
Earnings surprise (%)	0.043	0.065	1.391
PreEvent Return	0.001	0.001	0.004
Uncertainty (%)	0.993	0.973	0.237
SentimentOverall (%)	0.702	0.700	0.585
Size	7.479	7.419	1.811
MTB	2.291	1.709	1.713
Leverage	0.241	0.211	0.216
COGS	0.157	0.115	0.148
ROA	-0.003	0.009	0.052

Table 3: Industry heterogeneity in input price changes

This table shows the industry averages of the likelihood of input price increases, decreases, and the baseline input price change measure $\chi^{InputUp}$, $\chi^{InputDown}$, and $InPrChg$, respectively. These averages for each industry J are calculated by taking the mean of each measure across all earnings calls for firms within industry J. The industries listed are 10 Fama-French industries (i.e., the 12 FF industries excluding finance and utilities).

Industry	$\chi_J^{InputUp}$ (1)	$\chi_J^{InputDown}$ (2)	$InPrChg_J$ (%) (3)
Chems	0.89	0.62	1.31
NoDur	0.78	0.37	1.10
Manuf	0.77	0.45	0.92
Shops	0.74	0.34	0.90
Durbl	0.76	0.36	0.88
Enrgy	0.78	0.51	0.49
Other	0.48	0.19	0.39
Hlth	0.38	0.12	0.20
Telcm	0.45	0.11	0.20
BusEq	0.32	0.12	0.15

Table 4: Decomposition of input price changes into aggregate, industry, and firm-specific components

This table provides a decomposition of the firm-level measure of input price changes. The entries in the table report the incremental R^2 from adding a specific fixed effect.

	FF12	2-digit SIC	3-digit SIC
Incremental R^2	(1)	(2)	(3)
YearQtr FE	7.15%	7.15%	7.15%
Industry FE	11.27%	17.38%	25.66%
Industry × YearQtr FE	5.53%	10.58%	21.72%
Firm-level	76.05%	64.89%	45.47%
Permanent differences across firms within industries (Firm FE)	29.71%	22.89%	14.25%
Idiosyncratic input price changes (residual)	46.34%	42.00%	31.22%

Table 5: Persistence of input price increases and decreases: Event studies

This able reports the persistence of input price increases (decreases) using event studies. The total number of events for input price increases (decreases) is 4,595 (4,641) where these events are defined to be the sub-set of all conference calls that satisfy the following three criteria: 1) the call mentions input price increases (decreases); 2) the previous call of the same firm did not mention input price increases (decreases); 3) our sample contains calls by this firm for at least five consecutive quarters, including this call. This table reports the number of the events where there are 1, 2, 3, 4, or more than 4 consecutive quarters with input price increases (decreases) since the initial input price increases (decreases).

Panel A: Input price increases

#Consecutive shocks	1	2	3	4	≥ 5
#Events	2,144	863	441	266	881
Percentage	46.66%	18.78%	9.60%	5.79%	19.17%

Panel B: Input price decreases

#Consecutive shocks	1	2	3	4	≥ 5
#Events	2,557	808	384	206	686
Percentage	55.10%	17.41%	8.27%	4.44%	14.78%

Table 6: Pass through of input price changes

This table estimates the response of output price changes to input price changes using equation (21). Robust t-statistics, double clustered at the firm and year-quarter levels, are presented in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

OutPrChg (%)	t (1)	$t+1$ (2)	$t+2$ (3)	$t+3$ (4)
InPrChg (%)	0.550***	0.074***	0.033	0.001
	(16.28)	(2.92)	(1.15)	(0.06)
Observations	17,090	12,749	11,883	11,004
Adjusted R-squared	0.258	0.186	0.195	0.177
Controls	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓

Table 7: Industry concentration and pass through

This table presents results for the heterogeneous response of output price changes to input price changes. The dependent variables below, are the $OutPrChg$ at time t. We split the sample into three groups based on industry concentration. Robust t-statistics, double clustered at the firm and year-quarter levels, are presented in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Low (1)	Medium (2)	High (3)
InPrChg (%)	0.622*** (10.44)	0.564*** (9.31)	0.501*** (9.74)
Observations	5,305	5,269	5,281
Adjusted R-squared	0.315	0.265	0.263
Controls	✓	✓	✓
Firm FE	✓	✓	✓

Table 8: Cost of goods sold and input price changes

This table shows the relation between the input price change measure $InPrChg$ and changes in cost of goods sold over five quarters estimated using equation (25). Robust t-statistics, double clustered at the firm and year-quarter levels, are presented in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	$\Delta COGS(t)$	$\Delta COGS(t+1)$	$\Delta COGS(t+2)$	$\Delta COGS(t+3)$	$\Delta COGS(t+4)$
	(1)	(2)	(3)	(4)	(5)
InPrChg (%)	0.004***	0.006***	0.007***	0.006*	0.004
	(2.66)	(3.76)	(3.26)	(1.96)	(1.13)
Observations	17,002	16,851	16,586	16,280	15,974
Adjusted R-squared	0.104	0.071	0.125	0.186	0.206
Controls	✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓	✓

Table 9: Immediate stock price response to input price changes

This table presents results for the cumulative abnormal return (CAR) of a firm's stock around announcement of an input price change in an earnings call using equation (26). The CAR is measured over a three-day window, starting a day before the call and ending a day after the call. We use the CAPM to model the expected stock return. Robust t-statistics, double clustered at the firm and year-quarter levels, are presented in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	CAR[−1,+1] (%)					
	(1)	(2)	(3)	(4)	(5)	(6)
InPrChg (%)	-0.263***	-0.363***	-0.327***	-0.366***	-0.377***	
	(-4.35)	(-5.38)	(-4.84)	(-5.37)	(-6.20)	
$Frac^{InputUp}$ (%)						-0.405***
						(-5.70)
$Frac^{InputDown}$ (%)						0.063
						(0.38)
Observations	81,473	81,473	66,946	76,976	43,195	81,473
Adjusted R-squared	0.065	0.103	0.108	0.103	0.112	0.103
Controls	✓	✓	✓	✓	✓	✓
Firm FE		✓	✓	✓	✓	✓
$InPrChg_{i,t-1}$			✓			
Risk				✓		
Low price volatility periods					✓	

Appendix

Table A.1: Definitions of control variables

This table describes the construction of control variables used in all the regression specifications in the paper. All continuous variables are winsorized at the 1% and 99% levels.

Earnings Surprise (%)	Actual earnings per share (EPS) from IBES minus the consensus (median) of EPS forecasts issued or reviewed in 90 days before the earnings announcement date. The difference is scaled by the stock price at the end of the quarter and multiplied by 100.
PreEvent Return	Average stock return in window [−71,−11] in terms of trading days relative to the earnings conference call date.
Size	Natural logarithm of the market cap at the end of the quarter.
MTB	Market cap plus book value of liabilities scaled by total assets at the end of the quarter.
Leverage	Long-term debt divided by total assets at the end of the quarter.
COGS	Cost of goods sold divided by total assets at the end of the quarter. The cost of goods sold is treated as missing when it is non-positive.
ROA	Income before extraordinary items divided by total assets at the end of the quarter.
Uncertainty (%)	Percentage of uncertain words in the earnings call transcript based on Loughran and McDonald (2011) dictionary and the code from Loughran and McDonald's website.
SentimentOverall (%)	Percentage of positive words minus the percentage of negative words in the earnings call transcript based on Loughran and McDonald (2011) dictionary and the code from Loughran and McDonald McDonald's website.

A.1 Training Sample Labeling

We have the following rules when labeling the sentences of the earnings conference calls in the training sample:

1. Sentences have to be self-contained. No contextual information is required for the related labels.
2. If one sentence contains the information of both input and output or the entire market, we consider this sentence as both input-related and output-related.
3. If we are not sure about how to label a sentence, we skip that sentence.
4. Sentences which discuss demand and supply but do not mention price changes are not labeled as price-change-related.
5. If a sentence is about price increases of competitors' products, we label it as output-related price-change information, as it is often observed that managers use such discussions to justify their decision to raise output prices.
6. Sentences about general costs are not treated as price-change related. For example, the general cost decreases for one firm could be due to improved efficiency, instead of a decline of input costs.
7. Business-strategy sentences are not considered as price-change related.
8. "Price action" or "pricing action" is viewed as information of product price increase.

Internet Appendix for
"Firm-Level Input Price Changes and Their Effects:
A Deep Learning Approach"

Sudheer Chava, Wendi Du, Indrajit Mitra, Agam Shah, Linghang Zeng

Appendix IA.1 Matching Earnings Conference Calls to GVKEY

We downloaded 200,587 transcripts in HTML format from SeekingAlpha from Jan. 2007 to Jul. 2021. Each transcript contains identification information of title, stock ticker, event date, and the date when the transcript is posted on the website. We identify the 178,547 earnings conference call transcripts based on their title containing "earning", fiscal quarter information, but without "webcast".

We notice that the stock ticker from SeekingAlpha suffers from the "backfill" problem as discussed by Li, Mai, Shen, and Yan (2021) for the earnings call transcripts from Thomson Reuters' StreetEvents (SE) database. That is when one company changes its ticker, for example due to name change or being acquired, the SeekingAlpha backfills with the new company's stock ticker or the ticker of the acquirer's ticker. Fortunately, the SeekingAlpha earnings call transcripts store the historical stock tickers, in addition to the historical company names, in the title and the first sentence of each transcript. Thus, our matching process starts from matching with the historical tickers, and then we do company name matching for the remaining transcripts.

We use python code to extract the historical ticker in the title, in the first sentence, and the potentially backfilled ticker of the transcript. To make sure the historical tickers are accurate, we get the final historical ticker by setting up three rules:

1. If the historical tickers in the title and the first sentence are non-missing and the same;
2. Else, if the historical ticker in the title and the potentially backfilled ticker are non-missing and the same;
3. Else, if the historical ticker in the first sentence and the potentially backfilled ticker are non-missing and the same.

The basic idea is we have three tickers, we pick the one on which at least two of the three agree on it. This step helps us to tackle some coding errors from SeekingAlpha website. The latter two cases imply that for those transcripts, there is no backfilling problem.

For the 178,547 earnings conference call transcripts, we get the accurate historical ticker from case 1 for 118,460 transcripts (66.35%), from case 2 for 314 transcripts (0.18%), from Case 3 for 51,896 transcripts (29.07%). There are 7,877 transcripts (4.41%) satisfying none of the cases.

IA.1.1 Matching to CRSP PERMNO

We download CRSP *dsenames* data, which stores the link between the historical ticker and PERMNO of a stock.[14] By using the historical ticker and the event date of the earnings call, we match each transcript with the corresponding PERMNO. If multiple PERMNOs satisfy the requirement (around 0.5% of transcripts), which is often the case that one company has multiple shares traded in the market, we sort the records by share classes and starting date of the record, and select the top one record.

Using CRSP *dsenames* data, we have 152,656 transcripts matched to PERMNO, which is 85.5% of earnings calls and 89.4% of the transcripts with accurate historical tickers. There are 18,014 transcripts with accurate historical tickers, but not matched to PERMNO.

For the remaining 25,891 transcripts without the matched PERMNO, including those with and without the accurate historical tickers, we continue with the name matching method. We extract the historical company name from each transcript's title, and standardize the historical company names in earnings call transcripts and the CRSP *desnames* data. For each standardized company name of the earnings call transcript, we find the closest matched CRSP company name by using Python package of *fuzzywuzzy*. Then, for the matched names selected by *fuzzywuzzy*, we further request the first 25 characters (without space) of the two names should be the same. With the matched company name and the event date of the earnings call, we get another 6,891 transcripts matched with PERMNO after manual checking.

In total, we get 159,547 transcripts matched with PERMNO. We drop 599 duplicated transcripts with the same PERMNO and event date, caused by multiple versions of the same earnings call transcript. Overall, we get 158,948 transcripts matched with CRSP PERMNO.

IA.1.2 Matching to Compustat GVKEY

By using CRSP-Compustat link table, we get 157,751 transcripts (99.2%) matched with GVKEY. Then, for each earnings call transcript, we find the closest prior earnings announcement date (*rdq*) from Compustat Quarterly data since 2006. There are 157,705 transcripts after removing the ones with missing *rdq*. Based on Bochkay, Hales, and Chava (2020), the earnings call date is within one week after the earnings announcement date. Thus, we keep 154,570 (98.01%) transcripts which satisfy this requirement. Then, we drop 107 transcripts duplicated at GVKEY-earnings announcement date (rdq) level[15], and get 154,463 earnings call transcripts.

Appendix IA.2 Fine-tuning and Performance of ML Models

We evaluate three models (BERT, RoBERTa, and FinBERT) as potential candidates for our task. By using pre-trained models provided by their respective research teams, we fine-tune

[14]The *dsenames* data we downloaded with the *nameendt* max at 2020-12-31. We assume the *nameendt* will extend to the Jul. 26, 2021, the date after the latest data collection date Jul. 19, 2021.

[15]For each GVKEY-rdq pair, we keep the transcripts with the earliest event date and highest share class.

them with our human-labeled training data. During this process, the deep neural networks adjust their weights to accurately predict our human labels. To optimize performance, we search for the optimal hyperparameter configuration for each model. This involves experimenting with various combinations of batch sizes and learning rates. Specifically, we run each model with three different seeds (5768, 78516, and 944601), across three batch sizes (2, 4, and 8), and with three different learning rates (1e-5, 1e-6, and 1e-7). This approach helps to identify the most effective setup for our deep learning models.

In the fine-tuning step, we use *Transformers* library available on huggingface. We run our experiments on NVIDIA V100 GPU. The annotated dataset is split into three parts of 70-10-20 for training-validation-testing. We use AdamW optimizer in our training. We train our model for maximum of 100 epochs. To avoid overfitting, at each epoch of training we calculate accuracy on cross-validation set. If cross-validation accuracy doesn't improve by more than 10^{-2} for 7 consecutive epochs, training will be stopped early to avoid overfitting of the model.

To select model, we measure performance of model based on test accuracy and F-1 score on a task to identify whether sentences have price-change-related information or not. The performance for all three models is listed in the Table IA.3. We also list the best hyperparameters found for the model. Based on the results, we select RoBERTa as our model for all supervised training and prediction.

Our methodology involves a sequential training process. Initially, the model determines if there's a price change or not. For sentences indicating such changes, it then identifies the direction and type of the change. This approach ensures non-overlapping classification labels. Performance metrics across all four tasks can be found in Table IA.4.

Appendix IA.3 Additional Models

In order to establish the value addition of using the transformer-based model in our study, we consider a dictionary-based model, two rule-based models, SVM with TF-IDF and Bi-LSTM as additional methods. The results for all the models are listed in Table IA.3. RoBERTa model outperforms other models.

Dictionary-based: For the dictionary-based approach, we create the word list by excluding words associated with "margin," "labor," and "payment" from the target word list presented in Table IA.2. A sentence is classified as price-change-related if it contains any word from this edited list.

Rule-based (Method 1): For the rule-based approach, we take a simple approach where a sentence is classified as price-change-related if it contains any target word listed from Table IA.2 and at least one word from the following list: ["anchor", "cut", "subdue", "decline", "decrease", "reduce", "low", "drop", "fall", "fell", "decelarate", "slow", "pause", "pausing", "stable", "non-accelerating", "downward", "tighten", "ease", "easing", "rise", "rising", "increase", "expand", "improve", "strong", "upward", "raise", "high", "rapid"].

Rule-based (Method 2): As an alternative rule-based approach, we set a simpler rule that a sentence is classified as price-change-related if it contains at least one word from the list ["price", "cost"] and at least one word from the following list: ["increase", "decrease", "high", "low", "change"].

Bi-LSTM: Long Short-Term Memory (LSTM) is a recurrent neural network architecture that was popular for text analysis. Bidirectional LSTM (Bi-LSTM) is an enhanced approach to LSTM that processes input in both forward and backward directions. In this study, we employ a two-layer Bi-LSTM model to assess the efficacy of RNNs. Instead of training it from scratch, we initialize the embedding layer of the Bi-LSTM using 300-dimensional GloVe embeddings trained using Common Crawl. Similar to BERT-based models, we run the model for three different seeds (5768, 78516, and 944601) with three different batch sizes (2, 4, and 8) and three different learning rates (1e-5, 1e-6, and 1e-7).

SVM with TF-IDF: Support Vector Machine (SVM) is a supervised machine learning model that has been extensively used for text classification tasks. Coupled with Term Frequency-Inverse Document Frequency (TF-IDF), SVM can efficiently classify textual data by transforming it into a high-dimensional feature space where sentences are represented by the weight of each term. In our study, we harness the power of SVM along with TF-IDF feature extraction. The TF-IDF technique is employed to convert the sentences into weight-based features, emphasizing the importance of specific terms in relation to the entire corpus. To fine-tune our SVM model, we perform a grid search over hyperparameters including regularization strengths ('C': [0.1, 1, 10, 100, 1000]), kernel gamma values ('gamma': [1, 0.1, 0.01, 0.001, 0.0001]), and kernel types ('kernel': ['rbf', 'linear']). This systematic hyperparameter optimization ensures the most suitable configuration is chosen for our classification task, reinforcing the reliability of our findings.

Sentence-BERT: Sentence-BERT model is a modification of the pretrained BERT network to generate semantically meaningful sentence embeddings. In this test, we use the Sentence-BERT model available at HuggingFace.[16] We report the model performance in Tale IA.3 with the best learning rate of $1e^{-}5$ and best batch size of 2.

Appendix IA.4 Deep Learning Model Details and Visualization

In this section, we detail the BERT-based deep learning models used in this paper: BERT (Devlin et al., 2018), RoBERTa (Liu et al., 2019), and FinBERT (Araci, 2019). These models utilize attention mechanisms that allow neural networks to focus on important input elements, thereby enhancing performance on various natural language processing (NLP) tasks (Niu, Zhong, and Yu, 2021). By weighing the importance of different words, attention mechanisms enable individual words in a sentence to interact and collectively produce a dense vector representation, effectively capturing contextual information.

[16] `https://huggingface.co/sentence-transformers/stsb-bert-base`

We use pre-trained models to leverage the vector representations generated from their extensive training on large text corpora, avoiding the need to build models from scratch. To adapt these models for classifying price-change related sentences, we fine-tune them using our human-labeled data. Specifically, the vector representations from the pre-training step are passed through feed-forward neural network layers to predict price-change labels. During this fine-tuning process, the network adjusts the weights of the top layer to best predict our human labels.

To understand how our trained RoBERTa model works, we employ the visualization technique introduced by Alammar (2021) to pinpoint neuron groups detecting price-change discussions. As an example, consider the Sanderson Farms sample discussions (Section 2). We break down neuron activations into eight factors and color-coding words when specific neuron groups activate.[17]

Figure IA.1 shows the excitement levels of factor 7, directly relating to price change patterns, with darker colors indicating more intense neuron firing. As shown in the figure, this group of neurons fires intensely during sequences like "price paid ... increased significantly". This observation persists when we conduct factor analysis with seven or nine factors, shown in Figures IA.2 and IA.3.

[17]Although our analysis is at sentence-level, for this illustrative example, we use the sample paragraph from Sanderson Farms because it contains diverse types of price-change discussions. This serves as an effective example to visualize the functioning of our deep learning model.

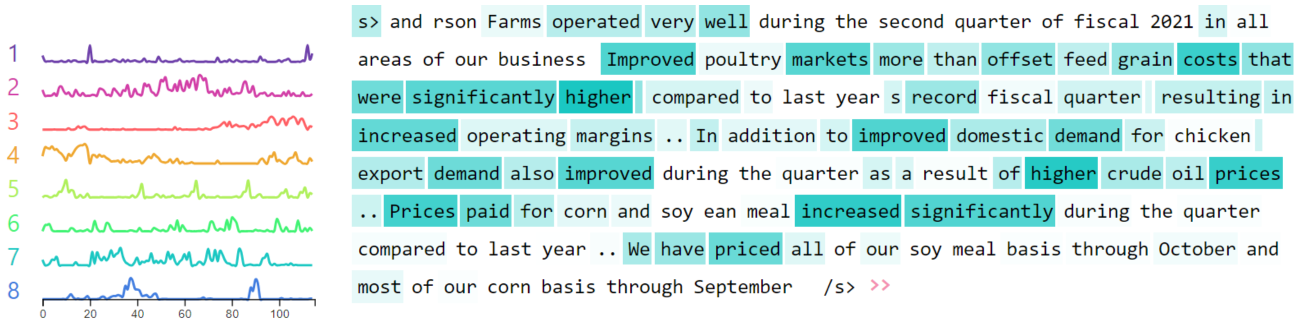

Figure IA.1: Factor analysis of our trained deep-learning model This figure displays the neuron activation factor analysis performed on our trained baseline model, RoBERTa, with the example text shown in Section 2 of our main draft. We utilize the Ecco explainability technique (Alammar, 2021) to gain insights into how our RoBERTa model works in processing discussions related to price changes in earnings calls. By decomposing the activations of neurons into eight factors, we observe that the factor 7 corresponds to the linguistic pattern associated with price changes. This finding is persistent when choosing seven or nine activation factors (see Figure IA.2 and IA.3). Darker colors indicate more intense firing within the neuron group of the factor. The sparklines on the left shows each factor's excitement level throughout the entire sequence.

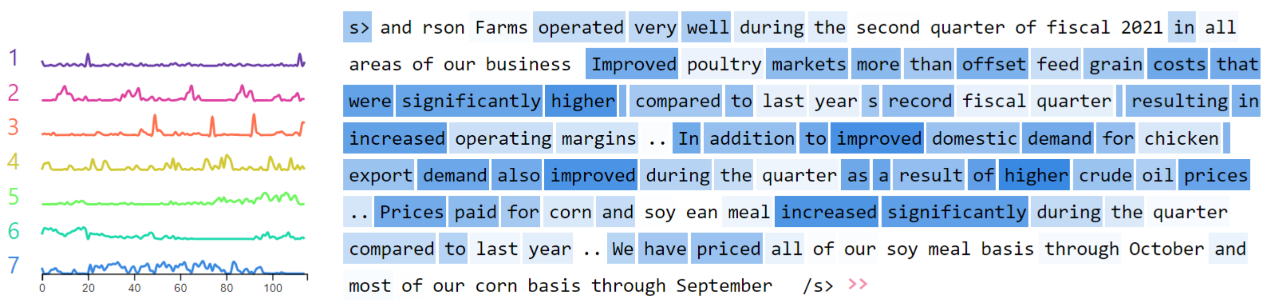

Figure IA.2: Factor analysis of our trained deep-learning model: seven factors This figure displays the neuron activation factor analysis performed on our trained baseline model, RoBERTa, with the example text shown in Section 2 of our main draft. We utilize the Ecco explainability technique (Alammar, 2021) gain insights into how our RoBERTa model works in processing discussions related to price changes in earnings calls. By decomposing the activations of neurons into seven factors, we observe that the factor 7 corresponds to the linguistic pattern associated with price changes. Darker colors indicate more intense firing within the neuron group of the factor. The sparklines on the left shows each factor's excitement level throughout the entire sequence.

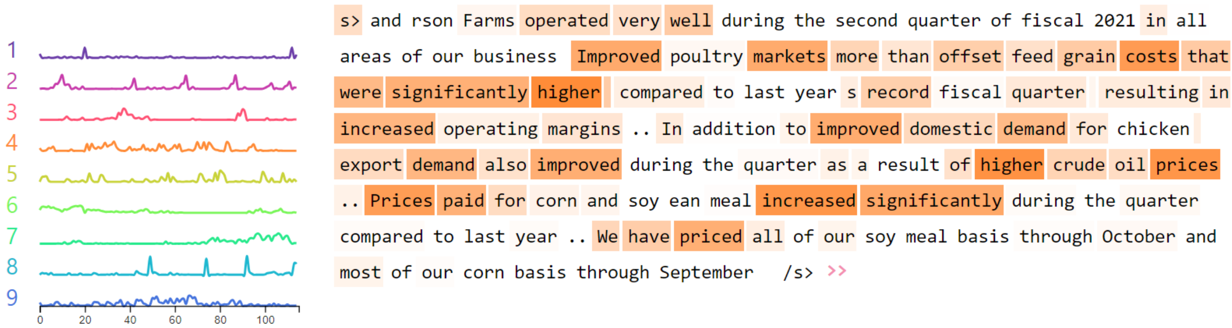

Figure IA.3: Factor analysis of our trained deep-learning model: nine factors This figure displays the neuron activation factor analysis performed on our trained baseline model, RoBERTa, with the example text shown in Section 2 of our main draft. We utilize the Ecco explainability technique (Alammar, 2021) gain insights into how our RoBERTa model works in processing discussions related to price changes in earnings calls. By decomposing the activations of neurons into nine factors, we observe that the factor 4 corresponds to the linguistic pattern associated with price changes. Darker colors indicate more intense firing within the neuron group of the factor. The sparklines on the left shows each factor's excitement level throughout the entire sequence.

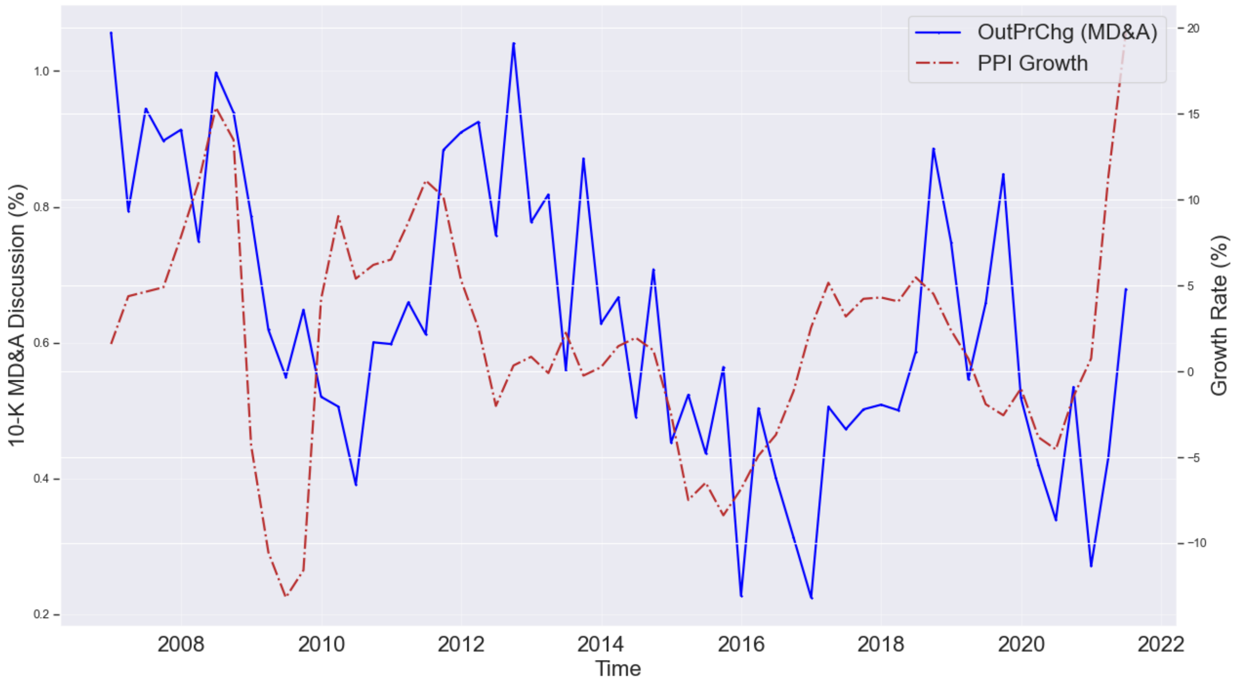

Figure IA.4: Trend of inflation index and text-based aggregate price changes This figure shows the trend of official PPI growth rate and the text-based aggregate output price changes generated from SEC 10-K filings. Specifically, our deep learning models processed the Management Discussion and Analysis (MD&A) section of the 10-K filings reported by our sample firms. The solid blue line represents the text-based aggregated output price change measure, which is the average of *OutPrChg* across all 10-K filings filed in each quarter. The dashed red line represents the quarterly (end of period) measure of the percent change from year ago for the Producer Price Index (PPI) by commodity: all commodities.

Table IA.1: Earnings conference call sample creation

This table reports the impact of various data filters and matching on earnings conference call transcripts. *rdq* represents the firm's earnings announcement date from Compustat.

Steps	**Sample Size**	**#Removed**
Transcripts from SeekingAlpha	200,587	
Earnings call transcripts	178,547	
Transcripts matched with PERMNO	158,948	
Including		
Match with PERMNO with historical ticker	152,656	
Match with PERMNO with historical company name	6,891	
Drop duplicates at PERMNO-date level		599
Transcripts matched with GVKEY	154,463	
Processing		
Match with GVKEY with link table	157,751	
Keep transcripts with matched *rdq*	157,705	46
Keep transcripts within one week after *rdq*	154,570	3,135
Drop duplicates at GVKEY-*rdq* level		107
Non-missing SIC code	154,295	168
Share code of 10 or 11	120,052	34,243
Exchange code of 1, 2, or 3	119,978	74
Drop Financial and Utilities Industries	102,112	17,866
Non-missing financial variables	81,473	20,639

Table IA.2: Target word list

This table provides the detailed words we include in the target word list we used for training sample selection.

Topic	**Target Words**
Inflation	inflation, inflationary, inflate, inflable, inflated, inflates, inflating, inflator, inflators
Deflation	deflation, deflationary, deflate, deflable, deflated, deflates, deflating, deflator, deflators
Price	price, priced, pricing, prices, pricey, pricy
Cost	cost, costs, costing, costed, costly
Margin	margin, margins, margining, margined
Labor	labor, labors, laboring, labored, laborer, laborers, labour, labours, labourer, labourers, laboured, labouring
Wage	wage, wages, waging, waged
Expense	expense, expenses, expensing, expensed, expensive, expensively, expensiveness, expendable, expenditure, expenditures, expend, expends, expending, expended
Payment	pay, pays, paid, paying, payment, payments, payable, payables, payload, payloads, paycheck, paychecks

Table IA.3: Accuracy analysis of three candidate models and additional methods

This table shows the model performance on detecting the price-change information and the best set of hyperparameters for each model. All the values are averaged over three different seeds. For all the deep learning models (i.e., BERT, FinBERT, RoBERTa), we do not perform any pre-training, we only fine-tune them on the labeled data. The additional methods in comparison are defined in Internet Appendix Section IA.3. Following the baseline procedure, we evaluate the model's performance specifically on sentences containing target words. For the definition of accuracy, see equation (3). The weighted F-1 score is defined as: Weighted F-1 Score = (weight for positive class * F-1 Score for positive class + weight for negative class * F-1 Score for negative class) / (weight for positive class + weight for negative class). Results in panel A correspond to the baseline training sample while results in panel B correspond to a random sample (see Panel B of Table 1 for details).

Model	Learning Rate	Batch Size	Test Accuracy	Test Weighted F-1 Score
Panel A: Transformer Encoder Models				
Dictionary-based	NA	NA	56.28%	0.5677
Rule-based (Method 1)	NA	NA	59.60%	0.6176
Rule-based (Method 2)	NA	NA	74.52%	0.7314
SVM with TF-IDF	NA	NA	85.10%	0.8478
Bi-LSTM	1e-5	4	84.36%	0.8445
Sentence-BERT	1e-5	2	88.36%	0.8830
BERT-base	1e-5	8	89.60%	0.8963
FinBERT-base	1e-6	4	89.81%	0.8995
RoBERTa-base	1e-5	8	**90.44%**	**0.9055**
Panel B: Random Test Sample				
RoBERTa-base	1e-5	8	95.47%	0.9265

Table IA.4: Test accuracy for all four classification tasks

This table shows the RoBERTa model's performance on the 4 classification tasks related to price change. The number of observations of each task is also included in the table.

Model Task	**Dataset Size**			**Test Accuracy**
	Train	**Valid**	**Test**	
M_1: price change related or not	3,297	471	942	90.44%
M_2: price increase or not	896	128	256	96.09%
M_3: input price or not	896	127	255	92.94%
M_4: output price or not	896	127	255	95.69%

Table IA.5: Robustness checks: Firm and year-quarter fixed effects

Panels A, B, C, and D of this table presents regression results on the robustness checks of Table 6, 7, 8, and 9, respectively. This analysis includes firm and year-quarter fixed effects. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Pass through of input price changes

OutPrChg	t (1)	$t+1$ (2)	$t+2$ (3)	$t+3$ (4)
InPrChg (%)	0.549*** (16.02)	0.070*** (2.71)	0.031 (1.10)	0.001 (0.03)
Observations	17,090	12,749	11,883	11,004
Adjusted R-squared	0.258	0.187	0.196	0.179
Controls	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓
Year-Quarter FE	✓	✓	✓	✓

Panel B: Industry concentration and pass through

	Low (1)	Medium (2)	High (3)
InPrChg (%)	0.626*** (10.36)	0.561*** (9.49)	0.493*** (9.77)
Observations	5,305	5,269	5,281
Adjusted R-squared	0.316	0.266	0.264
Controls	✓	✓	✓
Firm FE	✓	✓	✓
Year-Quarter FE	✓	✓	✓

Panel C: Cost of goods sold and input price changes

	$\Delta COGS(t)$ (1)	$\Delta COGS(t+1)$ (2)	$\Delta COGS(t+2)$ (3)	$\Delta COGS(t+3)$ (4)	$\Delta COGS(t+4)$ (5)
InPrChg (%)	0.004** (2.62)	0.006*** (3.62)	0.007*** (3.21)	0.006** (2.26)	0.005 (1.57)
Observations	17,002	16,851	16,586	16,280	15,974
Adjusted R-squared	0.121	0.081	0.136	0.193	0.217
Controls	✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓	✓
Year-Quarter FE	✓	✓	✓	✓	✓

Panel D: Immediate stock price response to input price changes

	CAR[−1,+1] (%)					
	(1)	(2)	(3)	(4)	(5)	(6)
InPrChg (%)	-0.263***	-0.339***	-0.298***	-0.337***	-0.382***	
	(-4.35)	(-5.70)	(-4.50)	(-5.42)	(-6.45)	
$Frac^{InputUp}$ (%)						-0.384***
						(-6.66)
$Frac^{InputDown}$ (%)						-0.002
						(-0.01)
Observations	81,473	81,473	66,946	76,976	43,195	81,473
Adjusted R-squared	0.065	0.108	0.114	0.109	0.115	0.108
Controls	✓	✓	✓	✓	✓	✓
Firm FE		✓	✓	✓	✓	✓
Year-Quarter FE		✓	✓	✓	✓	✓
$InPrChg_{i,t-1}$			✓			
Risk				✓		
Low price volatility periods					✓	

Table IA.6: Long-run abnormal stock price response

This table documents the long-run abnormal stock price response to input price changes measured from earnings conference calls. The dependent variables in Columns (1)–(3) report results for progressively longer horizons after the earnings calls. The key independent variable is *InPrChg*, which is computed as the difference between the total number of input-price-up sentences and the total number of input-price-down sentences in an earnings conference call scaled by the total number of sentences in the earnings call. All control variables are described in the Appendix. Robust T-statistics, double clustered at the firm and year-quarter levels, are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels.

	CAR[+2,+15] (%)	CAR[+2,+30] (%)	CAR[+2,+45] (%)
	(1)	(2)	(3)
InPrChg (%)	-0.002	-0.161	-0.237
	(-0.01)	(-1.04)	(-1.32)
Observations	81,472	81,472	81,472
Adjusted R-squared	0.041	0.063	0.077
Controls	✓	✓	✓
Firm FE	✓	✓	✓