Theory Meets Textual Analysis: Measuring Firm-Level Labor Cost Pressures and Inflation Pass-Through*

Aakash Kalyani† Serdar Ozkan‡

First Version: July 6, 2025—This Version: December 23, 2025

Abstract

We develop a novel measure of firm-level marginal labor cost and investigate its pass-through to inflation. To construct this measure, we apply textual analysis to earnings calls to identify discussions of labor-related topics such as higher costs, shortages, and hiring. Leveraging the theoretical principle that cost-minimizing firms equate marginal costs across variable inputs, we project changes in firms' intermediate input revenue shares onto the intensity of labor-related discussions to quantify their contributions to marginal labor costs. This approach provides an economically-motivated way to reduce the multidimensional qualitative textual information into a single quantitative measure. An aggregate index from this measure tracks closely with conventional aggregate slack variables and outperforms them in forecasting inflation. The vast majority of variation in labor cost pressures is idiosyncratic rather than systematic, in the sense that it is neither captured by aggregate slack nor by heterogeneous exposure of individual firms to aggregate slack. This granular variation drives the measure's superior predictive power. When aggregated at the industry level, we find a significant but heterogeneous pass-through of marginal labor costs to PPI inflation, with the pass-through highest for the service sector and near-zero for manufacturing. Consistent with the latter fact, firm-level data reveal that investment in automation mitigates the effects of higher labor cost pressures in manufacturing.

Keywords: Wage inflation, automation, textual data, machine learning
JEL codes: E24, J24, J31, J64.

*We thank Mickenzie Bass and Mick Dueholm for excellent research assistance. The views expressed herein are those of the authors and do not reflect the ones of the Federal Reserve Bank of St. Louis, the Federal Reserve System or the Board of Governors.

†St. Louis Fed; aakash.kalyani@outlook.com; www.aakashkalyani.com

‡St. Louis Fed; serdar.ozkan@gmail.com; www.serdarozkan.me

1. Introduction

Labor cost pressures have long been recognized as a fundamental driver of inflation since Phillips (1958). While several aggregate indicators such as unemployment rate, vacancy-to-unemployment ratios, and output gaps are commonly used to gauge marginal labor cost shocks, there is little consensus on the most appropriate measure of slack (e.g., Barnichon and Shapiro (2024)) and no single measure fully captures true marginal cost pressures Galı and Gertler (1999). Moreover, their aggregate nature precludes firm-level analysis.

This paper addresses these limitations by developing a novel measure of labor cost pressures derived from textual analysis of corporate earnings calls combined with financial data from Compustat. We use our measure to examine the pass-through from labor costs to inflation at both the aggregate and industry level, as well as heterogeneous firm-level responses to labor cost pressures. Our measure explains two thirds of post-pandemic inflation surge.

Earnings calls are quarterly conference calls where company executives discuss financial results and business updates with investors and analysts. These calls contain detailed discussions of operational challenges—including hiring difficulties, wage pressures, and workforce management—making them valuable sources of real-time information about firm-level labor market conditions. We analyze 250,167 earnings call transcripts for 6,237 U.S. public companies between 2002 and 2025. About 86% of these earnings calls discuss labor relates issues, with content varying cyclically: tight labor markets feature discussions of hiring, higher costs, and shortages; recessions feature discussions of layoffs and cost reduction.[1]

Our main textual analysis methodology follows a supervised approach. Through manual reading of transcripts, we identify five primary labor-related topics: 1) labor costs, 2) labor shortages, 3) headcount, 4) labor agreements, and 5) labor efficiency. Executives often qualify the discussions of costs, headcount, and efficiency by direction—higher or lower—, leaving us with $J = 8$ topics. To construct dictionaries, we begin with seed keywords directly indicative of each topic such as "labor

[1]Other papers have used earnings calls to infer firms' cost of inputs. For example, Gormsen and Huber (2024) use them to measure firms' perceptions of their cost of capital. In a contemporaneous paper Harford *et al.* (2024) also extract information from earnings calls to develop a measure of labor shortage exposure. Furthermore, Gosselin and Taskin (2024) constructs demand and supply indicators for Canada and United States using earnings call transcripts.

shortage", "labor costs", and "wage inflation" and expand them using embeddings trained on earnings conferences calls following Hassan *et al.* (2024a).To measure firm i's exposure to topic j of labor-related issues in quarter t, Λ_{it}^{j}, we count sentences containing these keywords and normalize it by call length.[2]

While our textual analysis provides rich descriptive evidence of labor market pressures, we need a more structured approach to quantify the economic importance of this qualitative data. We develop this framework through the firm's cost minimization problem that features materials as a variable input and generalized labor costs which capture both direct wage costs and indirect labor expenses, such as job posting, hiring, training, and retention costs. This model recognizes that firms facing labor market pressures incur multiple types of costs as they attempt to expand their effective workforce. We leverage the fact that cost-minimizing firms equate marginal costs across all variable inputs. In our framework, an increase in marginal labor cost—holding revenue per employee and output elasticities constant—should increase the share of spending on materials.

Therefore, we estimate each topic's contribution on marginal labor cost by regressing firms' revenue shares of intermediate inputs on $\Lambda_{it} = [\Lambda_{it}^{1}, \Lambda_{it}^{2}, ..., \Lambda_{it}^{J}]$ while controlling for revenue per employee, and firm and time fixed effects, thereby exploiting the within-firm variation. This approach serves a dual purpose. First, it translates our high-frequency qualitative measure from earnings calls into quantitative estimates of labor cost pressures (ω_{it}). Second, it provides an economically-motivated way to reduce the multidimensional information from earnings calls into a single measure. The regression coefficients effectively determine the optimal weights for combining different labor topics into a unified measure of their impact on firm costs.

We first validate this measure by comparing a sales-weighted average of ω_{it} with traditional aggregate indicators. Our index exhibits strong correlations: 0.76 with labor market tightness, –0.55 with unemployment rate, and 0.63 with the Employment Cost Index (ECI). It also reveals an intriguing nonlinear relationship, sharply increasing when unemployment falls below 5% and labor market tightness exceeds around 1.5. Notably, following a sharp but short-lived collapse during the Great

[2]As a robustness check, we also employ an unsupervised approach to let the data speak more freely. We represent each labor related sentence using pre-trained Sentence-T5 embeddings (Ni *et al.* (2021)) and apply principal component analysis to extract latent topics purely from textual patterns. We then replicate our analysis using these unsupervised topics. Comparing results across both approaches confirms our findings are not driven by particular hand-crafted labels.

Recession, our index remained persistently low throughout the subsequent recovery, consistent with the observed flattening of the Phillips curve Powell (2018). However, it surged dramatically during and after the COVID-19 pandemic, aligning closely with the steepening Phillips curve Domash and Summers (2022). When aggregated at the industry level, our measure maintains a robust correlation with labor market tightness and earnings growth of new hires, further validating its relevance. Despite strong correlations with aggregate and industry-level slack variables, these observables explain very little of the variation in firm-level labor cost pressures, suggesting our measure captures idiosyncratic shocks such as the loss of key talent, localized wage competition, or firm-specific expansion plans.

We next quantify the pass-through from labor cost pressures to inflation using industry-level variation in the Producer Price Index (PPI). We find a strong and statistically significant relationship between increased labor cost pressures (aggregated at the industry level) and subsequent PPI inflation even after controlling for industry and time fixed effects. Specifically, our estimates indicate that a 1.0 percentage point (pp) rise in labor cost pressures corresponds to approximately a 0.4 to 0.6 pp increase in PPI inflation over the following year. Importantly, the strength of this pass-through varies substantially across industries: sectors with higher labor intensity, such as wholesale trade, retail trade, and accommodation and food services, demonstrate stronger pass-through effects, whereas heavily regulated industries, like utilities and healthcare as well as manufacturing exhibit notably lower sensitivity.

Further, we evaluate the predictive power of the aggregate labor cost pressure measure for core Personal Consumption Expenditures (PCE) inflation using a Phillips curve framework, comparing its performance to traditional measures of labor market slack (à la Barnichon and Shapiro (2024)). We aggregate our quarterly industry-level measure using PCE weights and find that our measure outperforms traditional labor market indicators—such as unemployment rates, labor market tightness, the output gap, and the ECI—in explaining inflation dynamics. Our measure can explain 19.5 pp of the 29.3 pp cumulative surge in inflation during the post-pandemic period. Our measure also outperforms a "Raw Text" benchmark that includes the eight topic frequencies directly without the theoretical weights, confirming that the structure imposed by the firm's first-order condition is essential.

When included alongside these traditional measures, our indicator remains statistically significant, effectively subsuming their explanatory power. To investigate

the sources of this predictive power, we decompose our measure into a "systematic" component (co-moving linearly with aggregate slack) and an "idiosyncratic" residual (orthogonal to aggregate slack). We find that the idiosyncratic residual robustly predicts inflation, whereas the systematic component does not. This result highlights the critical role of our measure in capturing firm-level convexities and granular shocks in labor cost pressures. For instance, when the marginal cost curve is non-linear, the dispersion of firm-level constraints generates net inflationary pressure that aggregate slack variables miss.

Finally, we exploit the disaggregated nature of our measure and examine how firms respond to labor cost pressures using firm-level data from Compustat. We analyze changes in several key firm outcomes, including investment rates, R&D spending, and productivity measures, in response to ω_{it}. Our empirical strategy employs a comprehensive set of controls including firm fixed effects, year fixed effects, industry characteristics, and other firm-specific variables to isolate the causal effect of labor cost pressures on firm behavior.

Firms experiencing higher labor cost pressures increase their investment, with this effect being particularly pronounced in industries that heavily employ routine manual tasks. These differential responses translate into productivity outcomes: firms facing labor cost pressures experience faster productivity growth, a one standard deviation increase in labor cost pressures corresponds to a 1.55 percentage point increase in productivity growth. However, this productivity gain is primarily driven by firms in routine-manual task-intensive industries, while firms in non-routine industries do not observe such increases. These findings suggest that labor cost pressures may accelerate automation and technological adoption, particularly in industries where human labor can be more readily substituted with capital. They also explain why the pass-through of marginal labor costs to PPI inflation is highest for service sector and near-zero for manufacturing.

Related Literature. Our paper contributes to two strands of literature using textual data in economics. The first employs textual analysis to quantify previously unmeasurable phenomena such as economic and political uncertainty (Hassan *et al.* (2019, 2024b, 2023) and Baker *et al.* (2016)) and firm attention to macroeconomic variables or financial constraints (Song and Stern (2024) and Buehlmaier and Whited (2018)). Following their methodology, we construct topic-specific dictionaries using

word embeddings trained on earnings calls, iteratively expanding them through manual reading. We extend this method by embedding firm-level textual measures within a theoretical framework to quantify labor cost pressures. This proves essential: raw counts of labor discussions—the prior approach—show no correlation with aggregate slack variables or inflation, whereas our theoretically-grounded measure exhibits strong correlations with both, demonstrating that economic theory substantially improves the predictive power of textual data.

Second, a recent literature uses natural language processing to forecast macroeconomic variables. Most studies time-aggregate their text signal—e.g., averaging daily news-sentiment or collapsing a quarterly central-bank press-conference transcript into a single embedding—so that both left- and right-hand variables share the same frequency (e.g. Ashwin *et al.* (2024); Araujo *et al.* (2025)). With only 140 quarterly or 300 monthly post-1990 observations, this top-down approach faces severe degrees-of-freedom constraints. We instead discipline the text at the firm level, leveraging around 250,000 firm-quarters. Guided by a cost-minimization framework, we convert each firm's labor-related discussions into its marginal labor-cost shock, and only then aggregate using theory-based weights. This micro-to-macro approach preserves the richness of the text while maintaining statistical power.

Our paper also contributes to a large literature using the Phillips curve to forecast inflation. First, while existing work relies almost exclusively on aggregate slack indicators—such as the unemployment rate (Stock and Watson (1999)), the output gap (Galí (2015)), the job-switching rate (Moscarini and Postel-Vinay (2012)), and labor market tightness (Barnichon and Shapiro (2024))—we construct a novel firm-level measure of labor cost pressures from earnings call transcripts. Our measure directly captures executives' concerns about labor market conditions, avoiding assumptions about which aggregate variables reflect these conditions or whether they affect all firms uniformly. When aggregated using PCE weights, our measure outperforms all standard indicators in Phillips curve estimation.

Second, exploiting firm-level granularity, we show that small changes in unemployment or output gaps rarely generate widespread labor cost discussions at the firm level. However, during extremely tight labor markets, these concerns become widespread, translating into higher labor cost pressures and inflation. Furthermore, we show that firm level labor cost shocks are granular and that this granularity should also enter Phillips curve design (e.g., Gabaix (2011)).

This paper also relates to the literature on capital–labor substitution and productivity. Studies such as Acemoglu and Restrepo (2022b) and Graetz and Michaels (2018) show that automation can reduce firms' reliance on labor. We show that labor cost pressures prompt capital-deepening investments—particularly in industries with routine manual work—often spurring productivity gains. Our findings thus reinforce automation as a critical margin through which firms adapt to labor market shocks (e.g., Leduc and Liu (2024)), with implications for aggregate productivity and optimal monetary policy.

The rest of the paper is organized as follows: Section 2 presents the datasets we use in our analysis as well as the details of the empirical methodology. In Section 3, we introduce the simple theoretical framework to fix the ideas. Sections **??** and 5 present estimates of labor cost pressures and its inflation passthrough. In Section 6 we present how firms respond to changes in labor costs. Finally, Section 7 concludes.

2. Earnings Call Data and Empirical Methodology

Our main source of data is earnings conference calls. We use 443,870 transcripts of earnings conference calls held between 2002 and 2025 from S&P Global. For our baseline results, we focus on 248,437 earnings calls by US headquartered firms. We begin by constructing a dictionary of 104 labor-related terms (such as 'personnel', 'wage', 'workforce' etc.) and reading excerpts from these earnings conference calls that discuss these terms. In total, about 88% of earnings conference calls discuss labor-related terms.

Table 1 shows sample excerpts from earnings calls discussing labor issues. These transcripts reveal several important patterns. First, executives discuss not only labor cost increases but also labor shortages and tight labor market conditions. Second, labor cost pressures extend beyond wages to include sign-on bonuses, retention costs, and other hiring expenses—highlighting that executives conceptualize labor costs more broadly than standard wage measures capture. Finally, executives frequently emphasize the unexpected nature of these developments, describing how labor market conditions " hit us faster than we anticipated" or produced "higher than anticipated wage pressure". These excerpts demonstrate that earnings calls provide rich, real-time information about firms' labor cost experiences that may not be fully reflected in traditional aggregate measures.

Figure 1 shows the percentage of earnings calls that contain labor related discussions over time from 2002 to 2024. The time series shows distinct patterns across different economic periods. During the pre financial crisis period 2002 to 2007 around 85 to 90 percent of earnings calls consistently discuss labor issues. After the 2008 to 2009 financial crisis the percentage drops to around 85 percent and remains at this lower level through most of the 2010s. The data show another shift during the Covid pandemic with the percentage of labor related discussions spiking to almost 100 percent in 2021 which reflects the unusual labor market disruptions during this period. After this spike the percentage gradually declines but remains elevated and volatile which suggests ongoing labor market adjustments in the post pandemic period.

Table 1: Sample Excerpts from Earnings Calls

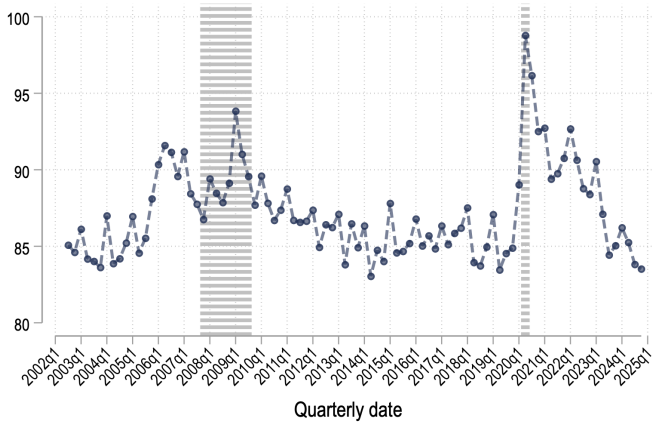
Firm	Excerpt
Basic Energy Services Inc (2011)	We produced a sequential increase in revenue for the quarter but only a modest increases in cash flow and earnings as **labor cost increases** jumped ahead of our ability to move pricing up in some segments.
KBR Inc (2013)	We had done forecast of **labor** availabilities and projects and had anticipated that market getting **tighter** in the first part of 2014. And candidly, it hit us faster than we anticipated.
US Foods Holding Corp (2019)	Third, on the cost side, we made substantial progress on our distribution initiatives. However, the higher than anticipated **wage pressure**, as a result of a very **tight labor** market, did offset some of this progress.
Akumin Inc (2023)	So we have responded by like everybody and paying **more** sign-on **bonuses**, giving more than typical **wage increases**, particularly on the clinical labor side. ... as we discussed, there is a bit of a **headwind** resulting from those **labor costs**.

While these excerpts and their patterns signal the attention executives pay to the labor markets, they do not directly provide a quantitative measure of labor cost pressures. Our next step is to organize these discussions into structured topics and to quantify t each topic's importance for labor cost pressures separately. We employ two complementary approaches: supervised and one unsupervised.

In the supervised approach we use our judgment and manual reading of excerpts to define a small number of interpretable labor topics, such as higher labor costs or labor shortages. We then build dictionaries that map sentences into these topics, following the procedure in Hassan *et al.* (2024a). This supervised strategy produces transparent, and easily interpretable topics with clear links to economic concepts.

In the unsupervised approach we let the data speak more freely. We represent each labor related sentence using pre-trained sentence embeddings and apply principal component analysis to the resulting high dimensional vectors. This yields a set of

Figure 1: Earnings calls with Labor mentions over time, %



Notes: The figure shows the percentage of earnings calls that mention one of labor terms by quarter.

latent labor topics that are estimated purely from textual patterns. We then replicate our main analysis using these unsupervised topics. Comparing results across both approaches confirms that our findings are not driven by particular supervised labels and show that the information we extract is also present in more flexible-data driven representations of labor discussions.

2.1. Supervised Approach

From reading a random sample of transcripts, we identify five core topics: labor costs, labor shortages, headcount, labor efficiency and labor agreements. When discussing labor costs, headcount and efficiency, the executives also qualify the direction of shift—higher or lower—yielding eight topics total.

To label these topics automatically, we develop dictionaries for each topic following Hassan *et al.* (2024a). We begin with a small set of seed keywords that directly relate to each topic. For instance, for the topic of labor shortages, we started with words like labor and shortage. Similarly, for labor costs, we considered terms such as wage and compensation. To expand our initial keyword list, we employ word embeddings trained on the earnings call transcripts. This helped us identify additional words and phrases used in similar contexts. For example, the embedding model suggested terms like personnel shortage and staffing constraints as contextually similar to labor shortage.

In parallel, we manually review excerpts to identify additional frequently used terms missed by the embedding model. To ensure relevance, we extract 10 excerpts containing each candidate keyword from our dataset and manually assess whether they aligned with the intended topic. We retain a keyword in the dictionary for a topic if at least 70% of the sampled excerpts are correctly classified under that topic (i.e., true positives). If a keyword is found to be frequently ambiguous or misclassified, it is either discarded or reassigned to a different topic where it is more relevant. We repeat this process iteratively, adding to the keyword lists with each iteration until no additional keywords emerge. Table A.1 shows the top 20 keyword combinations for each topic from a total of 13,566 keyword combinations across all topics.

Our classification model operates at the sentence level. A sentence is assigned a topic if it contains all the constituent words from one of that topic's keyword combinations. For instance, a sentence is categorized under 'labor shortage' if it includes both the keywords 'labor' and 'shortage', regardless of the order or distance between them. We then aggregate across the earnings call transcript, defining firm i's exposure to topic j of labor-related issues in quarter t Λ_{it}^{j}, as the percentage of sentence in an earnings call transcript which mention a topic j.

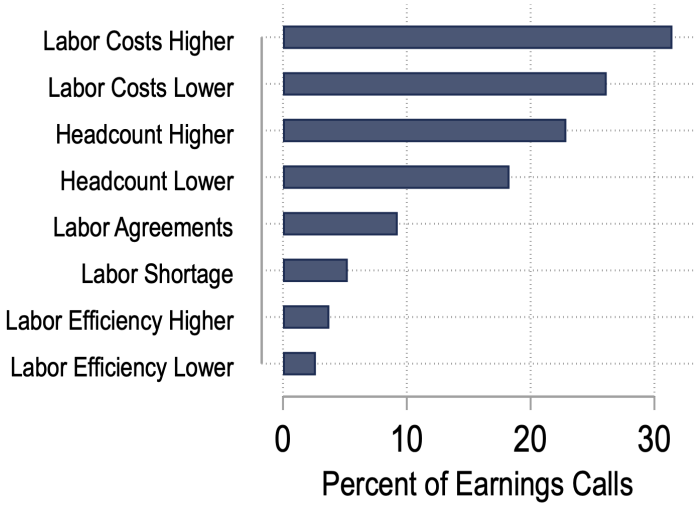
Figure 2 shows the incidence of these topics in earnings conference calls. These topics have significant overlaps as executives discuss multiple topics in a call. At least one of these topics is mentioned in 62% of earnings conference calls, accounting for 70.4% of overall executive discussions of labor-related terms. The four most prominent topics are higher and lower labor costs, and higher and lower head counts (hiring and firing). These four topics themselves are mentioned in 60% of earnings calls.

Validation of Labor Topic Discussions

Figure A.1 plots the evolution of each labor discussion topic from 2002 to 2025, revealing distinct patterns across economic cycles. While discussions of higher labor costs were elevated before the Great Recession, mentions of lower labor costs and headcount reductions peaked during the downturn. A similar pattern emerged during the Covid-19 pandemic recession. The post-pandemic period, in contrast, is defined by an unprecedented surge in labor shortage discussions in 2021-2022, with mentions reaching levels nearly four times higher than any previous peak, highlighting the unique severity of labor market tightness and validating that our textual measures capture critical, time-varying firm concerns.

Figure A.2 confirms these topics capture meaningful economic signals by corre-

Figure 2: Percentage of Earnings Calls, by Topic



Notes: The figures show percentage of earnings calls by labor topic between 2002 and 2025.

lating them with key macro variables. As expected, discussions of labor shortages, higher costs, and increased headcount are positively correlated with labor market tightness, vacancy to unemployment (V/U) ratio, and wage growth captured by the ECI, and negatively with the unemployment rate. The opposite holds for discussions of lower costs and headcount reductions.

Finally, the bottom-right panel shows correlations with the output gap. Labor cost increases and efficiency decreases show the strongest positive correlations, suggesting particular sensitivity to the overall state of the economy. Notably, labor shortage discussions show relatively modest correlation with the output gap compared to their strong correlations with direct labor market measures, indicating they capture labor market-specific rather than general economic conditions. These correlation patterns validate our textual analysis approach by demonstrating that discussions of labor issues in earnings calls systematically align with traditional macroeconomic measures of labor market conditions and overall economic activity. The consistent and intuitive patterns suggest that earnings calls provide reliable signals about labor market conditions.

Table 2: Discussion of labor topics and wage growth of new hires, at industry level

	$\Delta log(Earnings_{n,t})$		
	(1)	(2)	(3)
Labor Costs Higher (std.)$_{j,t}$	1.158***	1.572***	0.620***
	(0.249)	(0.281)	(0.208)
Labor Costs Lower (std.)$_{j,t}$	–1.388***	–1.194***	–0.469
	(0.333)	(0.293)	(0.281)
Headcount Higher (std.)$_{j,t}$	0.593***	–0.251	–0.121
	(0.180)	(0.326)	(0.329)
Headcount Lower (std.)$_{j,t}$	–1.096***	–1.352***	–0.948***
	(0.290)	(0.293)	(0.300)
Labor Shortage (std.)$_{j,t}$	0.677***	0.730***	0.206*
	(0.174)	(0.174)	(0.107)
Labor Efficiency Higher (std.)$_{j,t}$	0.078	0.242	0.029
	(0.155)	(0.183)	(0.159)
Labor Efficiency Lower (std.)$_{j,t}$	–0.215	–0.249	–0.155
	(0.268)	(0.236)	(0.177)
Labor Agreement (std.)$_{j,t}$	0.122	–0.009	0.129
	(0.198)	(0.251)	(0.201)
R^2	0.061	0.089	0.439
N	4,298	4,298	4,298
Time FE	N	N	Y
Industry FE	N	Y	Y

Notes: The table shows regression of changes in on discussion of labor topics by industry n in quarter t. Earnings denotes earnings for new hires observed in the quarterly workforce indicates aggregated over industry n at time t. Industry is at the NAICS 3-digit level. Each observation denotes a industry n and year t. To construct labor topic observations at the industry x year level we take averages across all quarters in a year. We exclude firms in financial and administrative industries (NAICS 52, 53 and 56). Topic variables are standardized. Regression is weighted by number of earnings calls in the industry. Standard errors are clustered by industry.

We also investigate how different labor-related discussions in earnings calls relate to wage growth for new hires at the industry level with time and industry fixed effects (Table 2). Our results show economically significant relationships between labor-related discussions and wage growth. Discussions of higher labor costs are positively associated with wage growth, with coefficients ranging from 0.62 to 1.58 across specifications, all statistically significant. Conversely, mentions of lower labor costs are negatively associated with wage growth, with coefficients between –0.47 and –

1.39. Similarly, discussions about decreasing headcount show negative associations (coefficients –0.95 to –1.35). These results are robust across specifications and remain significant even after controlling for both time and industry fixed effects. The R-squared increases substantially from 0.06 to 0.44 when including both time and industry fixed effects, indicating that industry-specific and time-varying factors explain a significant portion of the variation in wage growth.

Discussions of labor shortages and efficiency show smaller coefficients, but their signs align with economic theory: labor shortages and higher efficiency discussions are positively associated with wage growth, while lower efficiency discussions show a negative association. Labor agreements is not significantly correlated with wage growth. Recall that firms also mention these topics less, suggesting these discussions may capture aspects of labor market conditions that are not directly related to wage growth.

2.2. Unsupervised labor topics from sentence embeddings

The supervised topics are transparent and easy to interpret but they rely on human labels and a predefined set of categories. As a complementary approach, we also construct unsupervised labor topics using sentence embeddings.

We employ pre-trained sentence embeddings to convert each labor-related sentence into a high-dimensional vector that captures its meaning. These vectors can be viewed as coordinates in a space where semantically similar sentences lie close to each other. The pre-trained model used in our analysis is Sentence T5 by Google Research (Ni *et al.*, 2021). This model maps each sentence into a 768 dimensional vector and has been shown to perform well on semantic textual similarity tasks and to outperform Sentence BERT Reimers and Gurevych (2019).

Let x_{itl} denote the 768-dimensional embedding of sentence l in firm i at time t. To reduce number of dimensions, we apply principal component analysis to a stack all such vectors for all labor related sentences across firms and time and to the resulting matrix. For a given number K we interpret the first K principal components as K orthogonal latent labor topics. The loading of sentence l on component k measures how strongly that sentence expresses latent topic k.

We consider a range of values for K equal to 2, 4, 6, 8, 10, 20, 30, 40, and 50. For each choice we compute call-level topic shares by averaging the component scores for all sentences in a call and we standardize these shares. This gives a set of unsupervised

topic variables that play the same role as the supervised topics.

We then replicate our full analysis using these unsupervised topics in place of the supervised ones: constructing firm level and aggregate measures of labor cost pressures, estimating the same forecasting regressions, and analyzing the inflation passthrough. The comparison of results across the two approaches confirms our main findings are not driven by the choice of supervised labels and that the information content of labor discussions is also present in purely data driven topic structures derived from sentence embeddings. In the main text, we focus primarily on the supervised approach, highlighting differences with the unsupervised approach where relevant. See Appendix for full set of results from the unsupervised approach.

3. Theoretical Framework

While our textual analysis of earnings calls provides rich descriptive evidence of labor market pressures, we need a more structured approach to quantify their economic importance on a firm's marginal costs. In this section we develop a framework through the firm's profit maximization problem with generalized labor costs, which capture both direct wage costs and indirect labor expenses, such as job posting, hiring, training, and retention costs. This comprehensive view recognizes that firms facing labor market pressures incur multiple types of costs as they attempt to expand their effective workforce as we have seen in the discussions of labor issues in the previous section.

We consider a firm's static profit maximization problem with general labor costs capturing both direct wage costs and indirect labor expenses, such as job posting, hiring, training, and retention costs. Firm i in quarter t simultaneously chooses between its quantity of materials, $M_{i,t}$, and its target headcount, $\bar{L}_{i,t}$. The key friction is that effective labor input, $L_{i,t} = \mu_{i,t}\bar{L}_{i,t}$, where $0 \leq \mu_{i,t} \leq 1$ is the idiosyncratic 'job filling' shock:

$$\max_{M_{i,t},\bar{L}_{i,t}} P(Y_{it})F_t(M_{it}, \mu_{i,t}\bar{L}_{it}, \Omega_{it}) - x_t^M M_{it} - w(\bar{L}_{it})\mu_{i,t}\bar{L}_{it} - C(\bar{L}_{it}) \tag{1}$$

Output (Y_{it}) is produced using two inputs, intermediate materials (M_{it}) and effective labor ($L_{it} = \mu_{i,t}\bar{L}_{it}$), according to the common production function (F_t). Ω_{it} denotes

idiosyncratic productivity shocks. The firm faces a demand curve that determines the price (P_{it}) as a function of its output level. [3] On the cost side, intermediate inputs are the flexible input (ła Gandhi *et al.* (2020)) and represent materials, supplies, and other variable inputs excluding labor and are purchased at price, x_t^M. However, the labor decision is more nuanced than in standard models. Firms must distinguish between two labor-related quantities: the number of positions they create or workers they aim to hire ($\bar{L}_{it}$) and the effective labor input that actually contributes to production ($L_{i,t} = \mu_{i,t}\bar{L}_{it}$). For labor costs, the wage function $w(\bar{L}_{it})$ determines the base wage costs and may reflect the firm specific labor supply curve and it's influence on wages, particularly in markets where it has significant hiring power. The function $C(\bar{L}_{it})$ captures additional labor-related expenses such as those for recruitment, vacancy posting, training, signing bonuses, etc.

Unlike the typical firm problem with only direct wage costs of labor, the marginal cost of labor in this framework is given by:

$$MCL_{it} = w(\bar{L}_{it}) + \underbrace{\frac{w_{\bar{L}}\left(\bar{L}_{it}\right) L_{i,t}}{\mu_{i,t}}}_{\text{Wage pressure}} + \underbrace{\frac{C_{\bar{L}}\left(\bar{L}_{it}\right)}{\mu_{i,t}}}_{\text{Hiring costs}} = w(\bar{L}_{it}) + \lambda_{i,t}\left(\bar{L}_{it}, \mu_{i,t}\right) \tag{2}$$

where $\lambda_{i,t}$ is the additional marginal cost of hiring workers over and above the wage costs. Changes in MCL_{it} is a key object for many firm decisions. For example, from the first order condition (FOC) with respect to labor input L_{it}, one can show that under flexible prices the pass-through rate of changes in marginal cost of labor to prices is one:

$$\log\left(P_{it}\right) = \underbrace{\log\left(\frac{\epsilon_{it}}{\epsilon_{it}-1}\right)}_{\text{Markup }(\eta_{it})} + \underbrace{\log\left(w\left(\bar{L}_{it}\right) + \lambda_{it}^{L}\right)}_{\text{Marginal cost of labor}} - \underbrace{\log\left(Y_{it}/L_{it}\right)}_{\text{Labor productivity}} - \underbrace{\log\left(\alpha_{it}^{L}\right)}_{\text{Output elasticity}},$$

$$\Delta\log\left(P_{it}\right) = \Delta\log\left(w\left(\bar{L}_{it}\right) + \lambda_{it}^{L}\right) - \Delta\log\left(Y_{it}/L_{it}\right) - \Delta\log\left(\alpha_{it}^{L}\right) - \Delta\log\left(\eta_{it}\right) \tag{3}$$

where ϵ_{it} is the price elasticity of demand and $\alpha_{it}^{L} = \frac{\partial Y_{it}}{\partial L_{it}}\frac{L_{it}}{Y_{it}}$ is the output elasticity of labor input.

Despite its central role in firm's decisions, measuring changes in marginal costs of labor presents a challenge since financial data does not provide such detailed break-

[3] Capital is pre-determined in the previous period and for parsimony we only focus on firm's static decision.

downs. To this end, we exploit the standard optimality conditions with respect to intermediate input and labor inputs:

$$\frac{\partial Y_{i,t}/\partial L_{i,t}}{\partial Y_{i,t}/\partial M_{i,t}} = \frac{MCL_{i,t}}{x_t^M}$$

After re-arranging:

$$\begin{aligned}
\frac{x_t^M}{P_{i,t}\frac{\partial Y_{i,t}}{\partial M}\cdot\mu_{it}} &= \frac{w\left(\bar{L}_{it}\right)+\lambda_{it}^L}{P_{i,t}\frac{\partial Y_{i,t}}{\partial L}\cdot\mu_{it}} \\
\frac{x_t^M M_{it}}{P_{i,t}Y_{it}} &= \frac{\left(w\left(\bar{L}_{it}\right)+\lambda_{it}^L\right)L_{it}}{P_{i,t}Y_{it}}\cdot\frac{\alpha_{it}^M}{\alpha_{it}^L}
\end{aligned}$$

$$\underbrace{\Delta\log s_{it}^M}_{M\text{-share}} = \underbrace{\Delta\log\left(w\left(\bar{L}_{it}\right)+\lambda_{it}^L\right)}_{\text{Marginal cost of labor}} - \underbrace{\Delta\log\frac{P_{i,t}Y_{it}}{L_{it}}}_{\text{Revenue per worker}} + \underbrace{\Delta\log\frac{\alpha_{it}^M}{\alpha_{it}^L}}_{\text{Output elasticities}}, \tag{4}$$

where $\alpha_{it}^M = \frac{\partial Y_{it}}{\partial M_{it}}\frac{M_{it}}{Y_{it}}$ is the output elasticity of intermediate input and s_{it}^M is the revenue share of intermediate input cost.

Our theoretical framework provides a clear, testable prediction: when a firm faces rising marginal labor costs (our unobservable w_{it}), it will substitute away from labor and towards other variable inputs, namely materials. This substitution will, all else equal, increase the share of revenue spent on materials (S_{it}^M). This relationship is the key to estimating our measure. While Compustat does not report marginal labor costs, our textual analysis of earnings calls provides a rich, high-frequency proxy for the pressures driving them (Λ_{it}).

We can therefore quantify the economic significance of these discussions by directly estimating their impact on the materials share. This leads to our core empirical specification, which rearranges the first-order condition from Equation (4):

$$\Delta\log\left(w\left(\bar{L}_{it}\right)+\lambda_{it}^L\right) = f(\Lambda_{it})+\varsigma_{it}, \tag{5}$$

where $\Lambda_{it} = [\Lambda_{it}^1, \Lambda_{it}^2, ..., \Lambda_{it}^8]'$ is an array of firm i's exposure to topic k of labor-related issues in quarter t—Λ_{it}^k—, which is measured as the count of instances of keywords normalized by the length of the call (see Section 2). Therefore, in our

empirical specification we project changes in firms' intermediate input revenue shares onto their exposure to labor topics and estimate $\omega_{it} = f(\Lambda_{it})$ in the below equation:

$$\underbrace{\Delta \log s_{it}^{M}}_{M\text{-share}} = \underbrace{f(\Lambda_{it})}_{\omega_{it}} - \underbrace{\Delta \log \frac{P_{i,t} Y_{it}}{L_{it}}}_{\text{Revenue per worker}} + \epsilon_{it}, \tag{6}$$

where $\epsilon_{i,t}$ denotes the error term which absorb changes in output elasticities and measurement error as well as possible variation in utilization of inputs.

This regression approach converts our textual data from earnings calls into quantifiable estimates of labor cost pressures. Second, it provides an economically grounded method for condensing the complex, multidimensional information in earnings calls into a single meaningful measure. By using the regression coefficients as weights, we create a composite measure that captures how different labor-related discussions contribute to overall firm costs. This approach yields a comprehensive labor cost pressure measure that weights each topic's mention according to its estimated impact on firms' cost structures, combining the rich qualitative information from earnings calls into a single, economically interpretable metric.

One might consider an alternative 'residual approach', where labor cost pressures are defined as the unobserved component from a regression of materials' revenue share on revenue per worker. However, such an approach suffers from a critical drawback. While our method uses external, independently measured topic signals on the right hand side and only requires that this signal is not systematically correlated with other unobservables after including fixed effects. The residual approach requires a much stronger and less plausible assumption. Specifically, it assumes that any variation in materials' revenue share not explained by revenue per worker is the labor cost pressures shock, thereby mechanically conflating the signal with all sources of noise and misspecification.

4. Estimating labor cost pressures, ω_{it}

4.1. Data construction

Our estimation of each topic's contribution to the marginal cost of labor in equation 6 relies on financial data from Compustat. However, Compustat lacks direct measures of intermediate inputs, so we employ two alternative measurement approaches following

the previous literature.

Our primary approach, following Demirer (2022), creates a broad measure of variable inputs. We calculate this as the sum of cost of goods sold (COGS) and selling, general, and administrative expenses (SGA), from which we subtract depreciation (DP) and estimated wage expenditures. While this approach is designed to capture a wide array of non-labor variable costs, the SGA component may contain some of the very indirect labor costs (e.g., recruitment, training) that our textual analysis is designed to identify. To ensure our results are not mechanically driven by this overlap, we construct a key variant of this measure that excludes SGA entirely. As we show, our findings are robust to this exclusion.

Our second approach, following Keller and Yeaple (2009), employs a narrower measure: year-end raw materials inventory. The advantage of this approach is its clearly captures material expenditures. However, it represents a stock rather than a flow and excludes other important variable inputs like energy or purchased services. Reassuringly, our core estimates of labor cost pressures remain remarkably consistent across these two distinct measurement strategies.

For implementation, we note two additional details. First, since firm-level wage bills are not directly reported, we approximate them by multiplying a firm's total employees by the average industry earnings from the Quarterly Census of Employment and Wages (QCEW). Second, because our input measures are constructed from annual financial statements, our baseline estimation is performed at an annual frequency. Finally, our sample excludes firms in financial and administrative industries (NAICS 52, 53, and 56) and uses firm-level data only through 2019 for the baseline estimation to avoid the unique confounding effects of the COVID-19 pandemic on cost structures.

4.2. Labor cost pressures (ω_{it}) estimates

We regress firms' revenue shares of variable intermediate inputs on labor issue topics while controlling for firm and time fixed effects. We also choose to flexibly control for employment growth and sales growth to account for measurement error in reporting employment. And, to account for flexible output elasticities—which have been shown to be functions of inputs and to vary over time (Hubmer *et al.* (2024))—we control for industry-year fixed effects.[4] In particular, we estimate equation 6 at the annual

[4]This specification choice is consistent with the marginal costs as specified in Gagliardone *et al.* (2023) who specify firm level marginal costs for firm i at time t as $MC_{i,t} = C_{i,t}A_{i,t}Y_{i,t}^{v}$where v

level using the following regression specification.

$$\Delta \log s_{i,t}^{M} = \underbrace{\sum_{k} \beta_{k}^{topic} \Lambda_{i,t}^{k}}_{\omega_{it}} + \beta_1 \Delta log(Emp_{i,t}) + \beta_2 \Delta log(Sales_{i,t}) + \delta_{jt} + \varepsilon_{it}. \tag{7}$$

where $s_{i,t}^{M}$ is the share of intermediate input and $Emp_{i,t}$ and $Sales_{i,t}$ are employee count and sales reported by firm i in year t. Each topic score for topic k by firm i in year t is constructed by averaging over the topic scores calculated quarterly for each earnings conference call. $\delta_{j,t}$ denotes various controls such as firm-level risk and sentiment scores, and industry, firm, and time fixed effects.

In this section we show the regression results when we use Demirer (2022)'s material measure (i.e., the sum of COGS and SGA minus DP and wage expenditures). The regression results for other measures of materials — COGS minus imputed wage expenditures or COGS and material inventory additions — are remarkably similar (Table A.3). Table 3 shows the regression coefficients for different labor-related discussions in earnings calls relate to firms' cost structures. We present five different specifications with progressively more stringent controls, ranging from a basic specification to one that includes firm fixed effects, time fixed effects, and additional controls for risk and sentiment[5]. Using these coefficients, we calculate the estimated change in marginal costs of labor using:

$$\omega_{it} = \sum_{k} \hat{\beta}_{k}^{topic} \Lambda_{k,i,t} \tag{8}$$

The results reveal overall robust and intuitive relationships between labor cost discussions and firms' cost structures. Discussions around higher labor costs, higher headcount, labor shortages and lower labor efficiency are associated with higher cost pressures. Whereas discussions around lower labor costs, lower headcount, and higher labor efficiency are associated with lower cost pressures.

Magnitudes and statistical significance of these coefficients exhibit interesting variations. Specifically, discussions of higher labor costs are consistently strongly associ-

is a scaling parameter and $A_{i,t}$is a firm level productivity shifter. This type of functional form assumption allows for more general production functions and nests Cobb-Douglas and CES cases.

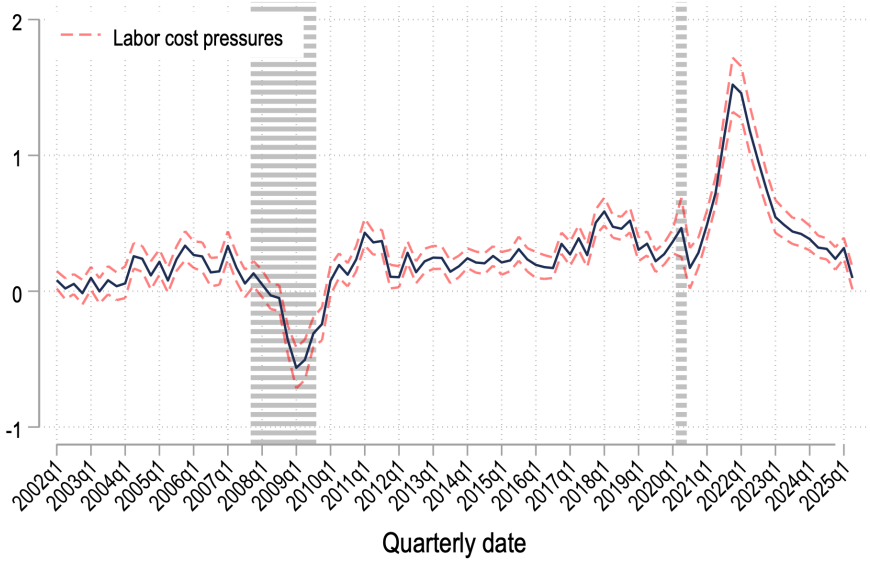
[5]We find nearly identical coefficients when including firm fixed effects along with industry-year fixed effects.

Table 3: Estimation of labor cost pressures: Labor topics and variable input cost share, at firm x year level

	$\Delta log(Materials/Sales)_{i,t}$				
	(1)	(2)	(3)	(4)	(5)
Labor Costs Higher$_{i,t}$	4.270***	3.981***	3.427***	3.580***	3.854***
	(0.613)	(0.604)	(0.675)	(0.686)	(0.731)
Labor Costs Lower$_{i,t}$	–6.075***	–6.437***	–6.739***	–6.685***	–6.497***
	(0.973)	(0.971)	(1.008)	(1.044)	(1.014)
Headcount Higher$_{i,t}$	–0.788	–0.529	0.263	0.392	0.882
	(0.724)	(0.727)	(0.733)	(0.730)	(0.930)
Headcount Lower$_{i,t}$	–0.981	–0.365	–0.188	0.444	–0.764
	(1.032)	(1.029)	(1.038)	(1.075)	(1.162)
Labor Shortage$_{i,t}$	2.042	2.446	1.234	1.231	2.019
	(1.808)	(1.890)	(2.118)	(2.154)	(2.486)
Labor Efficiency Higher$_{i,t}$	–1.061	–1.838	–2.899	–2.883	–3.636
	(2.821)	(2.774)	(2.928)	(3.011)	(3.146)
Labor Efficiency Lower$_{i,t}$	6.885**	7.447**	7.494**	8.318**	1.373
	(3.291)	(3.287)	(3.464)	(3.608)	(3.398)
Labor Agreement$_{i,t}$	1.013	0.813	–0.425	–0.696	–0.441
	(1.102)	(1.082)	(1.112)	(1.167)	(1.247)
Residual category$_{i,t}$	0.374	0.286	0.520*	0.556*	1.056**
	(0.282)	(0.285)	(0.312)	(0.333)	(0.421)
R^2	0.050	0.060	0.099	0.099	0.223
N	23,790	23,790	23,714	21,500	21,240
Baseline Controls	Y	Y	Y	Y	Y
Time FE	N	Y	Y	Y	Y
Industry-Year FE	N	N	Y	Y	N
Sentiment and Risk	N	N	N	Y	Y
Firm FE	N	N	N	N	Y

Notes: The table shows regression of changes in on labor topics. Each observation denotes a firm i and year t. To construct labor topic observations at the firm x year level we take averages across all quarters in a year. All specifications include controls for yearly changes in sales and employment. Columns (4) and (5) include controls for risk and sentiment. We exclude firms in financial and administrative industries (NAICS 52, 53 and 56). We only include firm level data till 2019 to exclude Covid-19 pandemic period. Standard errors are clustered by firm.

Figure 3: Labor Cost Pressure Index ($\bar{\omega}_t$)



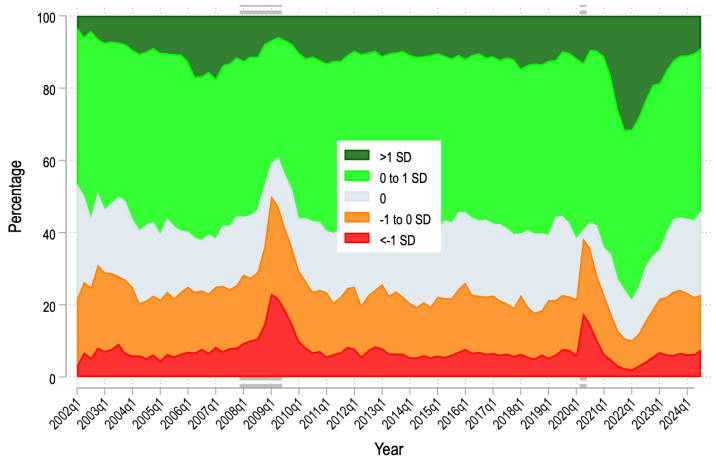
Notes: The figure shows the average change in estimated marginal cost of labor across firms by quarter along with bootstrapped confidence intervals. The estimation uses coefficients from Table 3 and topic specific scores by firm and by quarter. Bootstrapped confidence intervals are calculated by randomly leaving out 10% of the sample one at a time for the estimation procedure for 200 samples.

ated with increases in the materials-to-sales ratio. The estimated coefficients range from 4.270 (s.e. = 0.613) in the baseline model without any controls to 3.854 (s.e.= 0.731) in our most restrictive specification that includes firm and time fixed effects. In our most restrictive specification, the coefficient remains highly statistically significant at the 1% level. Similarly, mentions of lower labor costs are linked to significant negative coefficients, ranging from -6.075 (s.e.=0.973) to -6.497 (s.e.=1.014).

Appendix table A.4 shows regressions with individual topics. The table shows that individually discussion of higher and lower labor costs, and decreasing headcounts are significantly associated with changes in materials-to-sales ratio. However, these associations are sufficiently summarized in the discussion of higher and lower labor cost pressures and therefore are not statistically significant in our baseline table.

The stability of key coefficients across specifications for labor costs discussions, suggests these relationships are robust and not driven by omitted variables at the firm, industry, or time level.

Figure 4: Distribution of Labor Cost Pressures across firms (ω_{it})



Notes: The figure shows the distribution of change in estimated marginal cost of labor across firms by quarter. The estimation uses coefficients from Table 3 and topic specific scores by firm and by quarter. Different colors show the distributions of labor cost pressures separated by 1 standard deviation.

4.3. Time series variation of labor cost pressure (ω_{it})

We first construct an aggregate measure of labor cost pressures ($\bar{\omega}_t$) by creating a sales-weighted average of our firm-level measure (ω_{it}) to validate it against traditional macroeconomic indicators. Figure 3 shows our aggregate index. Following a sharp but short-lived collapse during the Great Recession, the index remained persistently low throughout the subsequent recovery. Consequently, it aligns more closely with the observed flattening of the Phillips curve (Powell (2018)). Furthermore, the index surged dramatically during and after the COVID-19 pandemic, surpassing its pre-recession peaks, a movement that corresponds to the steepening of the Phillips curve observed during this time (Domash and Summers (2022)). Figure A.5 shows that we find very similar aggregate time series pattern when we follow the same methodology with 2, 4, 8 or 50 *unsupervised* topics. Series that use number of *unsupervised* topics (more blue) exhibit a higher correlation with our baseline.

Table 4 presents correlations between our measure and four key macroeconomic indicators: unemployment rate, output gap, and the Employment Cost Index (ECI) (along with previously discussed labor market tightness). The results show strong and statistically significant correlations across all measures, with particularly robust

relationships with labor market tightness (0.76) and the ECI (0.63). The positive correlation with different slack variables confirm that our measure effectively captures labor market conditions in a way that aligns with traditional indicators. However, the imperfect correlation suggests our measure captures critical information that conventional indicators might overlook.

Table 4: Correlation with labor cost pressure index, $\bar{\Lambda}_t$

Tightness	Unemp. Rrate	Output gap	ECI
0.76***	–0.55***	0.29***	0.63***
(0.06)	(0.09)	(0.10)	(0.07)

4.4. Distribution of labor cost pressure $\omega_{i,t}$

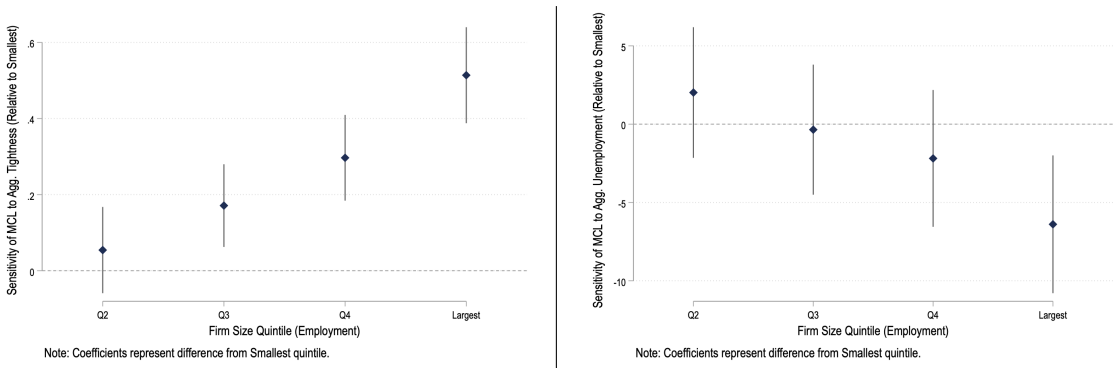
Do all firms experience similar changes in labor cost pressures, or are these pressures concentrated among a small subset of firms? Figure 4 shows the distribution of labor cost pressures across firms over time. For each quarter, labor cost pressures at the firm level are split into five categories based on standard deviations. Notably, for more than 75% of of firms, labor cost pressures remain negligible between -1 to 1 standard deviations, indicating that substantial labor-related cost concerns are concentrated within a relatively small subset of firms at any given point.

Firm's labor cost pressure ω_{it}. How do firms' labor cost pressures change over the business cycles? During periods of tight labor markets—particularly evident around 2005–2006 and again in 2021–2022—the number of firms experiencing meaningful labor cost pressures (more than one standard deviation to either side) increases more than two fold. Similarly, the slack in labor market after the Great recession between 2011 and 2019 was characterized by more firms not mentioning labor cost pressures. These patterns suggest that labor market tightness not only intensifies labor cost pressures among affected firms but also broadens their reach, causing more firms to encounter meaningful labor-related cost challenges.[6]

Are some firms more sensitive to aggregate labor market conditions than others?

[6]These findings echo those in Hassan *et al.* (2019), who use earnings conference calls to show that a large portion of the variation in political risk is at the firm quarter level rather than aggregate level. Similarly, Hassan *et al.* (2024b) show that the effects of Brexit varied largely by firm with large variation in winners, losers and unaffected firms.

Figure 5: Heterogeneous Sensitivity to Labor Market Slack



Notes: The figure plots the interaction coefficients of firm size quintiles (based on employment) with aggregate labor market tightness (Panel A) and the unemployment rate (Panel B). The smallest quintile is the omitted base category. Confidence intervals represent the 95% level.

To investigate this question consider the following equation:

$$\omega_{it} = \alpha_i + \beta_i s_t + u_{it}$$

where s_t is an aggregate slack measure. In a representative firm model we would have $\beta_i = \beta$ for all firms. Our estimates reject this restriction. We find that β_i varies systematically with firm size. In particular, when unemployment falls the increase in labor cost pressures is stronger for large firms than for small firms. In other words the slope of the relation between ω_{it} and slack is steeper for large firms.

We make this pattern transparent by estimating

$$\omega_{it} = \alpha_i + \varpi_t + \sum_{q=2}^{5} \beta_q \left(\mathbb{I}_{i \in q} \times Slack_t \right) + \varrho_{it} \tag{9}$$

where $Slack_t$ is either labor market tightness V_t/U_t or the unemployment rate and size quintiles based on employment are indexed by q with the smallest quintile omitted. Figure 5 shows the interaction coefficients. We find a monotonic pattern. Large firms are significantly more sensitive to aggregate labor market conditions than small firms. When tightness rises or unemployment falls large firms report much larger increases in labor cost pressures than small firms.

These results suggest that the effect of slack on marginal costs is both state de-

pendent and firm specific. One way to formalize this idea is

$$\omega_{it} = \alpha_i + \beta_i(s_t)s_t + u_{it}$$

where the slope $\beta_i(s_t)$ can depend on the level of slack and on firm characteristics such as size. Our evidence is consistent with models with a job ladder and frictional labor markets in which high productivity large employers adjust employment and wages more strongly over the business cycle than small firms (Moscarini and Postel-Vinay, 2012; Bilal *et al.*, 2022; Berger *et al.*, 2022, 2024). In those models tight labor markets lead large firms to expand and to bid up wages through poaching which raises their marginal labor costs more than for small firms.

Granularity of firm level labor cost shocks

After showing that marginal costs are both state-dependent and firm-specific, we now investigate the extent to which aggregate variables miss idiosyncratic cost pressures such as the loss of key talent, localized wage competition, or firm-specific expansion plans. We quantify the information content of these shocks using a variance decomposition.

Table 5 reports the R^2 from regressions of ω_{it} on progressively granular fixed effects and controls. Column (1) shows that aggregate time fixed effects explain only 3.0% of the variation in ω_{it}, meaning 97% of the variation is cross sectional. Labor cost pressures are not a uniform wave hitting all firms simultaneously, but rather a highly dispersed phenomenon.

In Column (2) we add industry-by-time fixed effects. If labor cost pressures were primarily driven by sectoral bargaining agreements or industry-specific supply chains, the R^2 would increase strongly. It reaches 10.6%, indicating that almost 90% of the variation occurs within industries.

Adding firm fixed effects in Column (3) increases the R^2 by about 11 percentage points to 0.215. This increase demonstrates that some persistent differences exist across firms—factors such as permanent workforce composition or bargaining power matter. However, these time-invariant characteristics explain only a small part of the total variation.

Columns (4) and (5) include observable time varying firm shocks. We control for uncertainty risk sentiment and growth shocks using stock returns. These variables increase the R^2 by only about 2 percentage points. Our measure is therefore

Table 5: Variance Decomposition of Labor Cost Pressures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Labor Cost Pressures (ω_{it})						
$\text{Risk}_{i,t}$				–0.006***	–0.006***	–0.007***	–0.007***
				(0.001)	(0.001)	(0.001)	(0.001)
$\text{Net Sentiment}_{i,t}$				0.001***	0.001***	0.001***	0.001***
				(0.000)	(0.000)	(0.000)	(0.000)
$\text{Stock Return}_{i,t-1}$					0.010***	0.010***	0.006***
					(0.001)	(0.001)	(0.002)
$\text{Stock Return}_{i,t}$					–0.001	0.000	0.001
					(0.001)	(0.001)	(0.001)
R^2	0.030	0.106	0.215	0.233	0.234	0.311	0.370
N	227,783	227,376	227,376	184,673	135,877	129,305	123,813
Time	Y	N	N	N	N	N	N
Industry × Time	N	Y	Y	Y	Y	–	–
Firm	N	N	Y	Y	Y	Y	Y
Ind × Size × Time	N	N	N	N	N	Y	Y
Ind × Growth × Time	N	N	N	N	N	N	Y

Notes: The table reports results from regressing firm-level labor cost pressures (ω_{it}) on various fixed effects and firm-specific controls. Columns (1)–(5) include time, industry × time, and firm fixed effects. Columns (6) and (7) include granular fixed effects, where *Size* and *Growth* quintiles are constructed based on the pooled distribution of firm employment and lagged stock returns, respectively. *Risk* and *Net Sentiment* are text-based measures constructed following Hassan *et al.* (2024a). Standard errors are clustered at the firm level.

not simply a proxy for general firm performance, management sentiment, or growth opportunities.

Columns (6) and (7) impose the strictest tests. We include granular fixed effects: Industry × Size Quintile × Time and Industry × Growth Quintile × Time. These specifications absorb all shocks common to groups such as large retailers in 2022 or high-growth software firms in 2023. Even with these saturated controls, the R^2 reaches only 0.370. More than 60% of the variation in ω_{it} remains unexplained, reflecting firm specific shocks that are orthogonal to observable characteristics.

5. Labor Cost Pressures and Inflation Pass-through

5.1. Estimating pass-through to PPI inflation

We begin by quantifying the pass-through of labor cost pressures (ω_{it}) to Producer Price Index (PPI) inflation. To do so, we specify an estimation equation by combining

the marginal cost calculated at the firm level, and then taking a sales weighted average at the industry level. From Equation 5 with a log-linearized version of the pricing equation 3:

$$\hat{P}_{i,t} = \beta_p \omega_{it} - \widehat{(Y_{it}/L_{it})} - \hat{\alpha}^L_{it} - \hat{\eta}_{it},$$

where β_p represents the pass-through elasticity of labor cost pressures to prices. To leverage the industry-level variation in PPI data, we aggregate this equation under the assumption of Cobb-Douglas preferences:

$$\hat{P}_{n,t} = \beta_p \sum_{i \in n} \theta_{i,t} \omega_{it} - \widehat{(Y_{nt}/L_{nt})} - \hat{\alpha}^L_{nt} - \hat{\eta}_{nt},$$

where $log(P_{n,t}) = \sum_{i,t} \theta_{i,t} log(P_{i,t})$ is the aggregate price index for industry n, and $\theta_{i,t}$ denotes the sales share of firm i in industry n at time t.

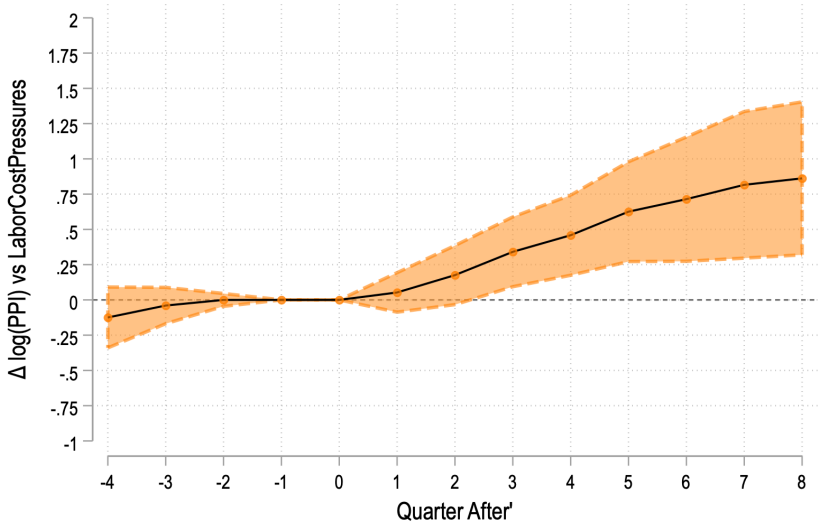
To estimate β_p and understand the dynamic response of prices to these pressures, we employ the following Jordà local projection specification:

$$log(PPI_{n,t+h}) - log(PPI_{n,t}) = \alpha_h + \beta_h \bar{\omega}_{nt} + \sum_{k=1}^{3} \gamma_k \Delta log(PPI_{n,t-k}) + \delta_{n,h} + \delta_{t,h} + \epsilon_{n,t+h}$$

where $\bar{\omega}_{n,t}$ is the sales-weighted average of ω_{it} across firms in industry n. The dependent variable denotes cumulative PPI inflation over horizon h. We include industry (δ_n) and time (δ_t) fixed effects. The time fixed effects are crucial for identification: as noted by Fitzgerald *et al.* (2014), Drechsel *et al.* (2019), and Hazell *et al.* (2022), these control for aggregate shocks (such as changes in monetary policy or aggregate demand) that might simultaneously affect labor costs and inflation. We exclude heavily import-dependent industries (apparel and computer-electronics) following Borusyak and Jaravel (2021) to isolate domestic cost pressures.

In our analysis we use a panel of 2,832 industry-quarter observations (Table A.5). Figure 6 shows that these labor cost pressures take about 6-8 quarters to pass through 86 percent of the increase in labor cost pressures to PPI inflation. Table A.5 shows a consistently positive and significant relationship between industry-level labor cost pressures and subsequent PPI inflation within the next four quarters across various specifications with and without fixed effects. The coefficient on the labor cost pressure measure ranges from 0.387 to 0.493, indicating that a one percentage point increase

Figure 6: Dynamic response of industry inflation to labor cost pressures

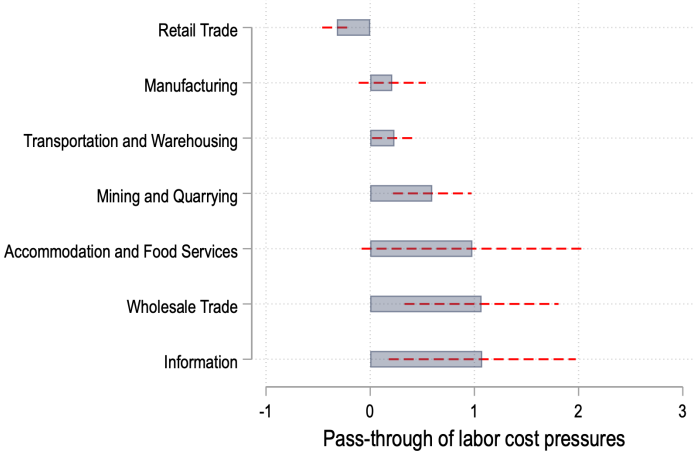


Notes: The figure plots estimated pass-through of labor cost pressures to PPI inflation (β_h) for horizons $h = 0$ to 8. Shaded areas represent 95% confidence intervals. Standard errors are clustered by industry. The regression is weighted by the number of firms observed within the industry. We exclude financial and administrative industries (NAICS 52, 53 and 56). Inflation is winsorized at the 2nd and 98th percentiles.

in labor cost pressures is associated with roughly 0.493 percentage points higher PPI inflation four quarters later. Table A.6 shows that these patterns remain largely unchanged with other measures of labor cost pressures calculated using alternative proxies for materials or calculated using *unsupervised approach*. Figure A.4 shows that the pass-through estimates are quite stable for any chose number of principal components greater than 6.

However, this average effect masks significant heterogeneity. Figure 7 plots the interaction of the labor cost pressure coefficient with industry dummies. Industries with high labor intensity—such as Information, Wholesale Trade, and Accommodation and Food Services—exhibit significantly higher pass-through. Conversely, the pass-through for the manufacturing sector is near zero. These results align with Heise *et al.* (2022), who find significant wage-price pass-through in services but not in manufacturing. While they attribute the manufacturing disconnect to import competition and market concentration, our findings in Section 6 suggest that investment in automation provides an alternative explanation: manufacturing firms mitigate cost pressures through capital-labor substitution rather than price increases.

Figure 7: Pass-through of labor cost pressures by Industry



Notes: The figure plots estimated pass-through of labor cost pressures by 2-digit NAICS industry. The plot shows estimates and 95% confidence intervals. Standard errors are clustered by industry.

5.2. Aggregate Labor Cost Pressures and PCE Inflation

Having established the link between our measure and producer prices, we now examine its implications for aggregate consumer inflation. We construct an aggregate Labor Cost Pressure Index ($\bar{\omega}_t$) by weighting our industry-level measures using Personal Consumption Expenditures (PCE) industry weights.

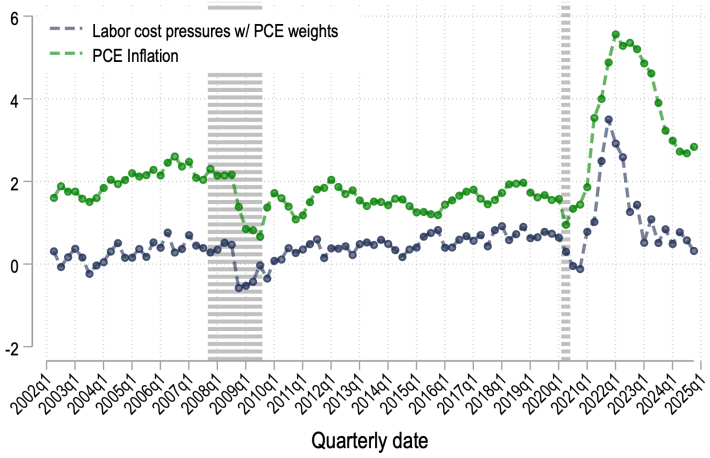
Figure 8 plots this aggregate index alongside core PCE inflation. The correlation is striking. Our index captures the short-lived disinflation during the Great Recession and the rapid surge during the post-pandemic recovery. Quantitatively, our measure can explain 19.5 percentage points of the 29.3 percentage point cumulative surge in inflation during the post-pandemic period.

Phillips Curve Estimation

We formally test the explanatory power of our measure within a standard New Keynesian Phillips Curve framework. Following Barnichon and Shapiro (2024), we regress changes in core PCE inflation on a forcing variable ($\hat{x}_t$), controlling for long-run inflation expectations and past inflation:

$$\pi_{t,t+4} = \alpha + \beta_x \hat{x}_t + \beta_\pi E_t \pi_\infty + \beta_1 \pi_{t-4} + v_t,$$

Figure 8: Labor Cost Pressure Index ($\bar{\omega}_t$) and PCE Inflation



Notes: The figure shows our labor cost pressure index (weighted with PCE weights) and the PCE inflation. To aggregate with PCE weights, we use industry level aggregate at the NAICS 3 digit level and then aggregate up to quarterly level using PCE weights for each industry.

where $\pi_{t,t+4}$ is the year-on-year percentage change in core PCE inflation, $E_t\pi_\infty$ denotes long-run inflation expectations (from the Livingston survey), and π_{t-4} represents lagged inflation. Following Atkeson *et al.* (2001) and Stock and Watson (2008), we include lagged inflation to control for the persistence of price setting. We standardize all forcing variables ($\hat{x}_t$) to facilitate comparison.

We compare our Labor Cost Pressure Index ($\bar{\omega}_t$) against standard slack measures: the unemployment rate, labor market tightness (V/U), the CBO output gap, and the Employment Cost Index (ECI).

Table 6 presents the results. In single-variable regressions (Panel A), our measure exhibits the strongest relationship with future inflation ($\beta = 0.312, p < 0.01$) and the highest R^2 (0.424). Traditional measures such as the output gap and ECI are statistically insignificant. In the "horse race" specifications (Panel B), our measure remains robust and statistically significant (coefficients ranging from 0.289 to 0.382) even when controlling for other slack variables. Notably, the traditional measures remain insignificant when included alongside our index, suggesting that our measure effectively subsumes the inflationary signal contained in standard aggregate indicators. Table A.7 shows that these patterns remain largely unchanged with other measures of labor cost pressures calculated using alternative proxies for materials or

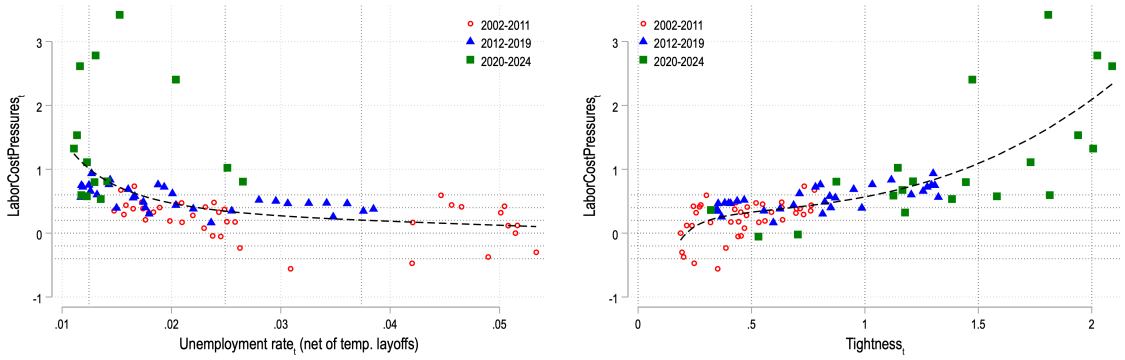
Table 6: Phillips Curve Estimation

Panel A: Push variables and Inflation						
	PCE Inflation$_{t,t+4}$					
	(1)	(2)	(3)	(4)	(5)	(6)
Labor Cost Pressures$_t$	0.343***					
	(0.119)					
Tightness$_t$ (std.)		0.078				
		(0.114)				
Unemployment Rate$_t$ (std.)			0.056			
			(0.080)			
Output Gap$_t$ (std.)				–0.157		
				(0.116)		
ECI$_t$ (std.)					0.287	
					(0.196)	
E-E-Flows$_t$ (std.)						–0.266
						(0.219)
R^2	0.430	0.385	0.385	0.398	0.401	0.397
N	88	88	88	88	88	88

Panel B: Comparison with Labor Cost Pressures						
	PCE Inflation$_{t,t+4}$					
	(1)	(2)	(3)	(4)	(5)	(6)
Labor Cost Pressures$_t$	0.343***	0.427***	0.382***	0.367***	0.320***	0.338***
	(0.119)	(0.145)	(0.120)	(0.118)	(0.121)	(0.120)
Tightness$_t$ (std.)		–0.188				
		(0.145)				
Unemployment Rate$_t$ (std.)			0.124			
			(0.081)			
Output Gap$_t$ (std.)				–0.189*		
				(0.102)		
ECI$_t$ (std.)					0.227	
					(0.191)	
E-E-Flows$_t$ (std.)						–0.254
						(0.210)
R^2	0.430	0.437	0.440	0.450	0.440	0.442
N	88	88	88	88	88	88

Notes: The table reports OLS estimates of the Phillips Curve specification. The dependent variable is PCE inflation over the next four quarters. The labor cost pressure measure is the PCE-weighted average of estimated changes in marginal labor costs. Standard errors are robust.

Figure 9: Non-linear relationship between labor cost pressures and slack



Notes: This figure plots the unemployment rate (net of temporary layoffs) and labor market tightness against our labor cost pressure measure at the quarterly level.

calculated using *unsupervised approach*[7].

Figure 9 helps explain this outperformance. It plots the relationship between our measure and traditional slack variables. The relationship is non-linear: labor cost pressures inferred from executive discussions remain muted over a large range of unemployment and tightness but spike sharply when unemployment falls below 4% and tightness exceeds 1.5. This convexity captures the non-linear nature of supply constraints that linear aggregate variables miss.

5.3. Out-of-Sample Forecasting

To test the predictive power of our measure, we conduct a recursive pseudo-out-of-sample forecasting exercise. We employ a direct forecasting method with a horizon of $h = 4$ quarters.

For every quarter t in our evaluation period (2015Q1–2025Q1), we estimate the forecasting model using a rolling window of the most recent $W = 40$ quarters (10 years) of data available at time t. Specifically, the information set available at time t includes labor cost pressures and controls up to time t, but realized 4-quarter-ahead inflation ($\pi_{s,s+4}$) is only observed up to $s = t - 4$. Therefore, the estimation sample for a forecast originating at time t consists of observations $\tau \in [t - W - 4, t - 4]$.

We estimate the following predictive regression for our proposed measure:

$$\pi_{\tau,\tau+4} = \alpha_t + \beta_{\omega,t}\bar{\omega}_\tau + \beta_{\pi,t}E_\tau\pi_\infty + \beta_{lag,t}\pi_{\tau-4,\tau} + v_\tau, \tag{10}$$

[7]Table A.7 uses 25 principal components.

where the subscript t on the parameters ($\beta_{.,t}$) indicates that they are re-estimated for each rolling window. We then generate the out-of-sample forecast for time $t+4$ using the current realized values:

$$\hat{\pi}_{t+4|t} = \hat{\alpha}_t + \hat{\beta}_{\omega,t}\bar{\omega}_t + \hat{\beta}_{\pi,t}E_t\pi_\infty + \hat{\beta}_{lag,t}\pi_{t-4,t}.$$

Theory vs. Raw Text. A key contribution of our paper is utilizing the firm's cost minimization problem to aggregate the eight labor-related topics into a single sufficient statistic, ω_{it}. To validate this theoretical restriction, we compare our model against a "Raw Text" benchmark that includes the eight topic frequencies directly in the forecasting equation without the theoretical weights:

$$\pi_{\tau,\tau+4} = \alpha_t + \sum_{k=1}^{8} \gamma_{k,t}\Lambda_\tau^k + \beta_{\pi,t}E_\tau\pi_\infty + \beta_{lag,t}\pi_{\tau-4,\tau} + \eta_\tau, \tag{11}$$

where Λ^k represents the aggregate frequency of topic k (e.g., "Labor Shortage", "Wage Inflation"). While Equation 11 utilizes the same textual information, it requires estimating eight separate coefficients ($\gamma_{k,t}$), potentially leading to overfitting. In contrast, our measure $\bar{\omega}_t$ projects these topics onto material shares first, effectively "pre-weighting" them based on their economic impact on marginal costs.

We evaluate performance using the Root Mean Squared Error (RMSE) over the evaluation sample T_{eval}:

$$RMSE = \sqrt{\frac{1}{T_{eval}} \sum_{t \in T_{eval}} \left(\pi_{t,t+4} - \hat{\pi}_{t+4|t}\right)^2}$$

Table 7 reports the results. Our structural measure achieves an RMSE of 0.646, significantly outperforming the "Raw Text" model, which has an RMSE of 2.176. This massive divergence confirms that the structure imposed by the firm's first-order condition is essential; without it, the high-dimensional textual data leads to severe overfitting and poor out-of-sample performance.

Furthermore, our measure consistently outperforms traditional slack indicators. It reduces forecast error by 25% relative to the Unemployment Rate (0.861) and by roughly 30% relative to Labor Market Tightness (0.923) and the Output Gap (0.885). Finally, our measure also outperforms the "Theory Residual" approach

Table 7: Forecasting Performance: Root Mean Squared Error (2015–2025)

Model	RMSE	Relative to Unemp.
Labor Cost Pressures	**0.645**	**0.75**
Raw Text (8 Topics)	1.975	2.29
Theory Residual	0.807	0.94
Tightness (V/U)	0.923	1.07
Unemployment Rate	0.861	1.00
Output Gap	0.885	1.03
Employment Cost Index	1.203	1.40
Labor Cost Pressures with 25 PCA topics	0.865	1.00
Labor Cost Pressures with 50 PCA topics	0.751	0.87

Notes: The table reports the Root Mean Squared Error (RMSE) for core PCE inflation forecasts four quarters ahead. The evaluation period is 2015–2025. Models are estimated using a rolling 10-year window. The "Relative to Unemp." column reports the ratio of the model's RMSE to the RMSE of the Unemployment Rate model.

(RMSE 0.807)—which defines labor cost pressures purely as the residual of the materials share equation—confirming that the textual projection successfully isolates the specific labor cost signal from other unobserved production shocks.

5.4. Decomposing Inflation: Systematic vs. Idiosyncratic Pressures

Our variance decomposition (Table 5) showed that the majority of variation in labor cost pressures is firm-specific. A key question is whether this idiosyncratic variation aggregates to affect inflation, or if only the systematic component driven by the business cycle matters. Empirically, we isolate these components by regressing ω_{it} on aggregate slack and estimating firm level betas and error terms using $\omega_{it} = \alpha_i + \beta_i Slack_t + u_{it}$, as we did in Section 4.4. The fitted value $(\hat{\beta}_i \times Slack_t)$ captures the Systematic Component. The residual (ϵ_{it}) absorbs both the idiosyncratic term, which can capture the non-linear response of marginal costs to labor constraints or nonlinear inflation dynamics in menu cost economies (e.g., Blanco *et al.* (2024)).

Table 8 presents the results of forecasting PCE inflation using these components. The results yield three specific insights into the nature of inflationary pressures.

Table 8: Inflation Decomposition: Systematic vs. Idiosyncratic Pressures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Total	Unemployment Decomposition			Tightness Decomposition		
Total Labor Cost Pressures	0.343***						
	(0.119)						
Systematic ($\beta \times U_t$)		–0.069		0.056			
		(0.221)		(0.224)			
Idiosyncratic Residual			0.397***	0.402***			
			(0.124)	(0.125)			
Systematic ($\beta \times V/U_t$)					0.154		0.211
					(0.167)		(0.161)
Idiosyncratic Residual						0.396**	0.412***
						(0.155)	(0.152)
R^2	0.430	0.384	0.440	0.440	0.387	0.427	0.434
N	88	88	88	88	88	88	88
Controls							
Exp & Lag Inf	Y	Y	Y	Y	Y	Y	Y

Notes: The table reports results from regressing PCE inflation on the systematic and idiosyncratic components of the Labor Cost Pressure index. The Systematic Component is constructed by aggregating firm-level predicted values based on aggregate slack (Unemployment or Tightness). The Idiosyncratic Component is constructed by aggregating firm-level residuals. All independent variables are standardized. Standard errors are robust.

First, the **idiosyncratic residual** is a robust and economically large driver of inflation. In Column (3), the coefficient on the residual is 0.397 and statistically significant at the 1% level. Second, the **systematic component** provides no incremental predictive power beyond standard controls. In Column (2), the coefficient on the systematic component (constructed via the unemployment rate) is -0.069 and statistically indistinguishable from zero. Similarly, in Column (5), the systematic component constructed via labor market tightness is insignificant (0.154). Table A.8 shows that these systematic components are statistically significant when we commit lagged inflation controls. The disappearance of significance suggests that the information contained in the standard linear business cycle is already fully captured by the persistence of inflation itself (π_{t-4}).

Third, the joint specifications (Columns 4 and 7) confirm that the granular signal is orthogonal to the business cycle. When both components are included simultaneously, the idiosyncratic coefficient remains stable and large (≈ 0.40), while the systematic component remains insignificant. For instance, in Column (7), the idiosyncratic residual (0.412^{***}) clearly dominates the tightness-based systematic component (0.211).

This dominance implies that the unique value-added of our text-based measure stems entirely from its ability to capture the convex costs of granular shocks—information that is, by definition, absent from aggregate slack variables.

6. How do firms respond to labor cost pressures?

Capital–labor substitution in production is central for many questions in economics, such as factor income shares (e.g., Karabarbounis and Neiman (2014)) or earnings distribution (e.g., Krusell *et al.* (2000)). Most models of labor demand imply that when firms face higher labor costs they adopt automation technologies (e.g., Acemoglu and Autor (2011); Acemoglu and Restrepo (2022a); Leduc and Liu (2024)).

We now exploit the disaggregated nature of our measure and examine how firms respond to labor cost pressures using a large panel of firm-quarter observations from Compustat data. We analyze changes in several key firm outcomes, including investment rates, R&D spending, and productivity measures, in response to ω_{it}. Our empirical strategy employs a comprehensive set of controls including firm fixed effects, year fixed effects, industry characteristics, and other firm-specific variables to isolate the causal effect of labor cost pressures on firm behavior. The results reveal intriguing patterns in how firms adapt to labor cost pressures, with notable differences across industry types. We use the following specification:

$$log(CapEx_{i,t,t+4}) = \alpha + \beta\omega_{i,t} + \gamma log(assets_{i,t-1}) + \chi_{i,t} + \delta_i + \delta_t + v_t,$$

where ω_{it} denotes labor cost pressures at the firm level observed in quarter t. $log(CapEx_{i,t,t+4})$ denotes log capital expenditures between quarter t and t+4. $\delta_i$and $\delta_t$are firm and time fixed effects. We control for risk and sentiment in expressed in earnings conference calls following Hassan *et al.* (2024a). We focus on heterogeneity across industries with different levels of routine manual task intensity (Table 9 Panel A). The baseline specification (column 1) indicates that a percentage point increase in labor cost pressures is associated with a 2.58% increase in capital expenditure. The inclusion of various fixed effects and controls helps isolate the effect of labor cost pressures from other factors that might influence investment decisions, such as industry trends, macroeconomic conditions, and firm-specific characteristics.

Columns 2-3 reveal interesting heterogeneity across industry types based on their routine manual task intensity. High routine manual (HRM) industries show the

Table 9: Labor Cost Pressures and Investment, Firm x time Level

Panel A: Capital Investment				
	log(Capital Exp.$_{i,t:t+4}$) * 100			
	All	LRM	MRM	HRM
	(1)	(2)	(3)	(4)
Labor Cost Pressures$_{i,t}$	2.581***	2.238***	1.890***	4.203***
	(0.193)	(0.363)	(0.252)	(0.351)
R^2	0.920	0.889	0.934	0.914
N	148,510	45,360	61,459	41,532
Panel B: R&D Investment				
	log(R&D Exp.$_{i,t:t+4}$) * 100			
	All	LRM	MRM	HRM
	(1)	(2)	(3)	(4)
Labor Cost Pressures$_{i,t}$	0.556***	0.157	0.617**	0.967***
	(0.175)	(0.318)	(0.251)	(0.362)
R^2	0.959	0.960	0.960	0.957
N	53,582	15,624	22,968	14,966
Controls (Risk and Sentiment)	Y	Y	Y	Y
Time FE	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y

Notes: The table shows regression of log(Capital expenditures) reported by firm i between quarters t+1 and t+4 on labor topic scores of earnings conference calls held by firm i in quarter t. . Each observation denotes a firm i and year t. All columns include controls for risk, sentiment and log assets. We exclude firms in financial and administrative industries (NAICS 52, 53 and 56). Standard errors are clustered by firm.

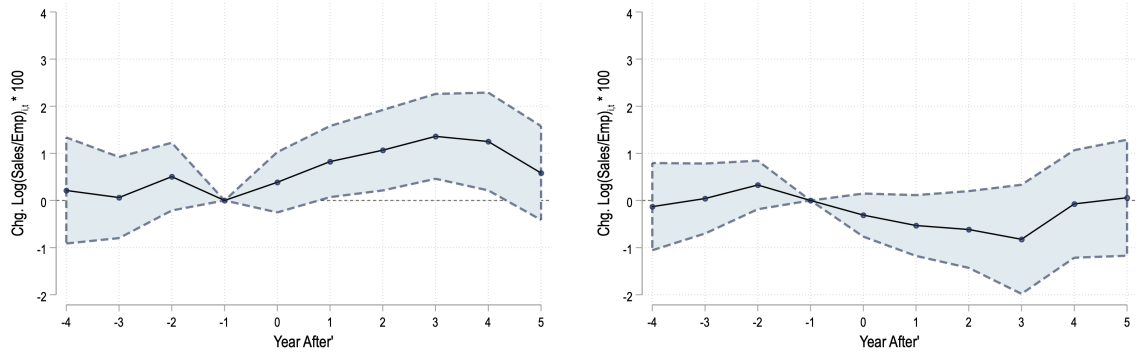
strongest investment response to labor cost pressures (4.20%), followed by low routine manual (LRM) industries (2.24%), while medium routine manual (MRM) industries show the smallest response (1.89%). This pattern suggests that firms in industries with high routine manual task content are more likely to respond to labor cost pressures by increasing capital investment, consistent with greater opportunities for automation and capital-labor substitution in these industries.

We next examine how firms adjust their R&D expenditures in response to labor

Figure 10: Productivity response to labor cost pressures, firm x time level

(a) High Routine Manual

(b) Low Routine Manual



Notes: This figure plots the response of labor productivity growth measured as change in log of sales per employee. The sample includes an yearly panel of firms in Compustat. We control for firm and year fixed effects, and earnings call risk and net sentiment. Standard errors are clustered by firm.

cost pressures, using a similar framework as the capital expenditure analysis (Table 9 Panel B). For the full sample of firms (columns 1-3), labor cost pressures show a positive and significant relationship with R&D spending. The effect is substantial, with a one percentage point increase in labor cost pressures associated with a 0.56% increase in R&D expenditure in the baseline specification. We find similar results to those observed for capital expenditures: firms in high routine manual industries invest in R&D at 7 times the rate compared to low routine manual industries. This finding, combined with the heterogeneous responses across industry types, indicates that firms' technological capabilities and industry characteristics significantly influence their choice between capital investment and R&D as responses to labor cost pressures.

How does an increase in capital investment and automation in response to labor cost shock affect labor productivity? Figure 9 presents an event study analyses showing how labor productivity (measured as revenue per employee) responds to labor cost pressure shocks from 4 years before to 5 years after the shock, with confidence intervals shown by the dotted lines. Panel A, focusing on HRM industries, reveals a clear pattern of productivity gains following labor cost pressure shocks. While there is no significant pre-trend before the shock (years -4 to 0), productivity begins to increase notably around year 1 and continues to rise, reaching a peak of nearly 1%

higher productivity by year 3. The effect remains positive and statistically significant through year 5, suggesting persistent productivity improvements in these industries following labor cost pressures.

In contrast, Panel B shows that low routine manual industries exhibit no significant productivity response to labor cost pressure shocks. The point estimates fluctuate around zero throughout the post-shock period, and the confidence intervals consistently include zero. This stark difference between HRM and LRM industries aligns with the earlier findings on investment and automation responses, suggesting that firms in HRM industries are more successful at converting their technological responses to labor cost pressures into actual productivity gains, likely through successful automation and capital-labor substitution.

These productivity patterns align closely with and help explain our earlier findings about firms' differential responses to labor cost pressures across industry types. HRM firms successfully automate and achieve productivity gains. This heterogeneity in responses and outcomes highlights the important role of industry characteristics in determining the effectiveness of different strategies for dealing with labor cost pressures.

7. Conclusion

This paper develops and validates a novel approach to measuring labor cost pressures by combining textual analysis of earnings calls with Compustat data. Our measure offers several advantages over traditional indicators. First, it provides granular, firm-level data that captures both direct wage costs and indirect labor expenses. Second, it shows strong predictive power for inflation in both aggregate and industry-level analysis. Third, its disaggregated nature allows us to study how firms respond to labor cost pressures, revealing important heterogeneity across industries.

Our findings make three main contributions. First, we show that earnings calls contain valuable real-time information about labor market conditions and labor cost pressures. Second, we demonstrate that our measure outperforms traditional slack variables in Phillips curve estimations, suggesting it better captures inflationary pressures. Third, we document how labor cost pressures drive technological adoption differently across industries. Firms in routine manual industries respond through automation and achieve productivity gains.

These results have important implications for understanding both inflation dynamics and technological change. Our measure could help policymakers better anticipate inflationary pressures, while our firm-level findings suggest that labor cost pressures may accelerate automation and technological adoption, particularly in industries where human labor can be more readily substituted with capital. This highlights how labor market conditions can shape the pace and direction of technological change across different sectors of the economy.

References

Acemoglu, D. and Autor, D. H. (2011). Skills, tasks and technologies: Implications for employment and earnings. *Handbook of Labor Economics*, **4**, 1043–1171.

— and Restrepo, P. (2022a). Demographics and automation. *The Review of Economic Studies*, **89** (1), 1–44.

— and — (2022b). Tasks, automation, and the rise in us wage inequality. *Econometrica*, **90** (5), 1973–2016.

Araujo, D., Bokan, N., Comazzi, F. and Lenza, M. (2025). Word2prices: embedding central bank communications for inflation prediction.

Ashwin, J., Kalamara, E. and Saiz, L. (2024). Nowcasting euro area gdp with news sentiment: a tale of two crises. *Journal of Applied Econometrics*, **39** (5), 887–905.

Atkeson, A., Ohanian, L. E. *et al.* (2001). Are phillips curves useful for forecasting inflation? *Federal Reserve bank of Minneapolis quarterly review*, **25** (1), 2–11.

Baker, S. R., Bloom, N. and Davis, S. J. (2016). Measuring economic policy uncertainty. *The quarterly journal of economics*, **131** (4), 1593–1636.

Barnichon, R. and Shapiro, A. H. (2024). Phillips meets beveridge. *Journal of Monetary Economics*, **148**, 103660.

Berger, D., Herkenhoff, K., Kostøl, A. R. and Mongey, S. (2024). An anatomy of monopsony: Search frictions, amenities, and bargaining in concentrated markets. *NBER Macroeconomics Annual*, **38** (1), 1–47.

—, — and Mongey, S. (2022). Labor market power. *American Economic Review*, **112** (4), 1147–1193.

Bilal, A., Engbom, N., Mongey, S. and Violante, G. L. (2022). Firm and worker dynamics in a frictional labor market. *Econometrica*, **90** (4), 1425–1462.

Blanco, A., Boar, C., Jones, C. and Midrigan, V. (2024). *Non-linear inflation dynamics in menu cost economies*. w32094.

Borusyak, K. and Jaravel, X. (2021). *The distributional effects of trade: Theory and evidence from the united states*. Tech. rep., National Bureau of Economic Research.

Buehlmaier, M. M. and Whited, T. M. (2018). Are financial constraints priced? evidence from textual analysis. *The Review of Financial Studies*, **31** (7), 2693–2728.

Demirer, M. (2022). Production function estimation with factor-augmenting technology: An application to markups.

Domash, A. and Summers, L. H. (2022). *How tight are US labor markets?* Tech. rep., National Bureau of Economic Research.

Drechsel, T., McLeay, M. and Tenreyro, S. (2019). Monetary policy for commodity booms and busts.

Fitzgerald, T. J., Nicolini, J. P. *et al.* (2014). *Is there a stable relationship between unemployment and future inflation?: Evidence from US cities*. Citeseer.

Gabaix, X. (2011). The granular origins of aggregate fluctuations. *Econometrica*, **79** (3), 733–772.

Gagliardone, L., Gertler, M., Lenzu, S. and Tielens, J. (2023). *Anatomy of the Phillips curve: micro evidence and macro implications*. Tech. rep., National Bureau of Economic Research.

Galí, J. (2015). *Monetary policy, inflation, and the business cycle: an introduction to the new Keynesian framework and its applications*. Princeton University Press.

Gali, J. and Gertler, M. (1999). Inflation dynamics: A structural econometric analysis. *Journal of monetary Economics*, **44** (2), 195–222.

Gandhi, A., Navarro, S. and Rivers, D. A. (2020). On the identification of gross output production functions. *Journal of Political Economy*, **128** (8), 2973–3016.

Gormsen, N. J. and Huber, K. (2024). *Firms' Perceived Cost of Capital*. Tech. rep., National Bureau of Economic Research.

Gosselin, M.-A. and Taskin, T. (2024). *Textual Economic Indicators for Canada and the United States: Insights From Earnings Calls*. Tech. rep., Bank of Canada Staff Discussion Paper.

Graetz, G. and Michaels, G. (2018). Robots at work. *Review of economics and statistics*, **100** (5), 753–768.

Harford, J., He, Q. and Qiu, B. (2024). Firm-level labor-shortage exposure. *Available at SSRN 4410126*.

Hassan, T. A., Hollander, S., Kalyani, A., van Lent, L., Schwedeler, M. and Tahoun, A. (2024a). *Economic Surveillance using Corporate Text*. Tech. rep., National Bureau of Economic Research.

—, —, Van Lent, L. and Tahoun, A. (2019). Firm-level political risk: Measurement and effects. *The Quarterly Journal of Economics*, **134** (4), 2135–2202.

—, —, van Lent, L. and Tahoun, A. (2023). Firm-level exposure to epidemic diseases: Covid-19, SARS, and H1N1. *Review of Financial Studies*, **34** (6), 2409–2449.

—, —, — and — (2024b). The global impact of Brexit uncertainty. *Journal of Finance*, **79**, 413–458.

Hazell, J., Herreno, J., Nakamura, E. and Steinsson, J. (2022). The slope of the phillips curve: evidence from us states. *The Quarterly Journal of Economics*, **137** (3), 1299–1344.

Heise, S., Karahan, F. and Şahin, A. (2022). The missing inflation puzzle: The role of the wage-price pass-through. *Journal of Money, Credit and Banking*, **54** (S1), 7–51.

Hubmer, J., Chan, M., Ozkan, S., Salgado, S. and Hong, G. (2024). *Scalable versus Productive Technologies*. Tech. rep., Penn Institute for Economic Research, Department of Economics, University of Pennsylvania.

Karabarbounis, L. and Neiman, B. (2014). The global decline of the labor share. *The Quarterly journal of economics*, **129** (1), 61–103.

Krusell, P., Ohanian, L. E., Ríos-Rull, J.-V. and Violante, G. L. (2000). Capital-skill complementarity and inequality: A macroeconomic analysis. *Econometrica*, **68** (5), 1029–1053.

Leduc, S. and Liu, Z. (2024). Automation, bargaining power, and labor market fluctuations. *American Economic Journal: Macroeconomics*, **16** (4), 311–349.

Moscarini, G. and Postel-Vinay, F. (2012). The contribution of large and small employers to job creation in times of high and low unemployment. *American Economic Review*, **102** (6), 2509–2539.

Ni, J., Abrego, G. H., Constant, N., Ma, J., Hall, K. B., Cer, D. and Yang, Y. (2021). Sentence-t5: Scalable sentence encoders from pre-trained text-to-text models. *arXiv preprint arXiv:2108.08877*.

Phillips, A. W. (1958). The relation between unemployment and the rate of change of money wage rates in the united kingdom, 1861-1957. *economica*, **25** (100), 283–299.

Powell, J. H. (2018). Monetary policy and risk management at a time of low inflation and low unemployment. *Business Economics*, **53**, 173–183.

Reimers, N. and Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.

Song, W. and Stern, S. (2024). Firm inattention and the efficacy of monetary policy: A text-based approach. *Review of Economic Studies*, p. rdae102.

Stock, J. H. and Watson, M. W. (1999). Forecasting inflation. *Journal of monetary economics*, **44** (2), 293–335.

— and — (2008). Phillips curve inflation forecasts.

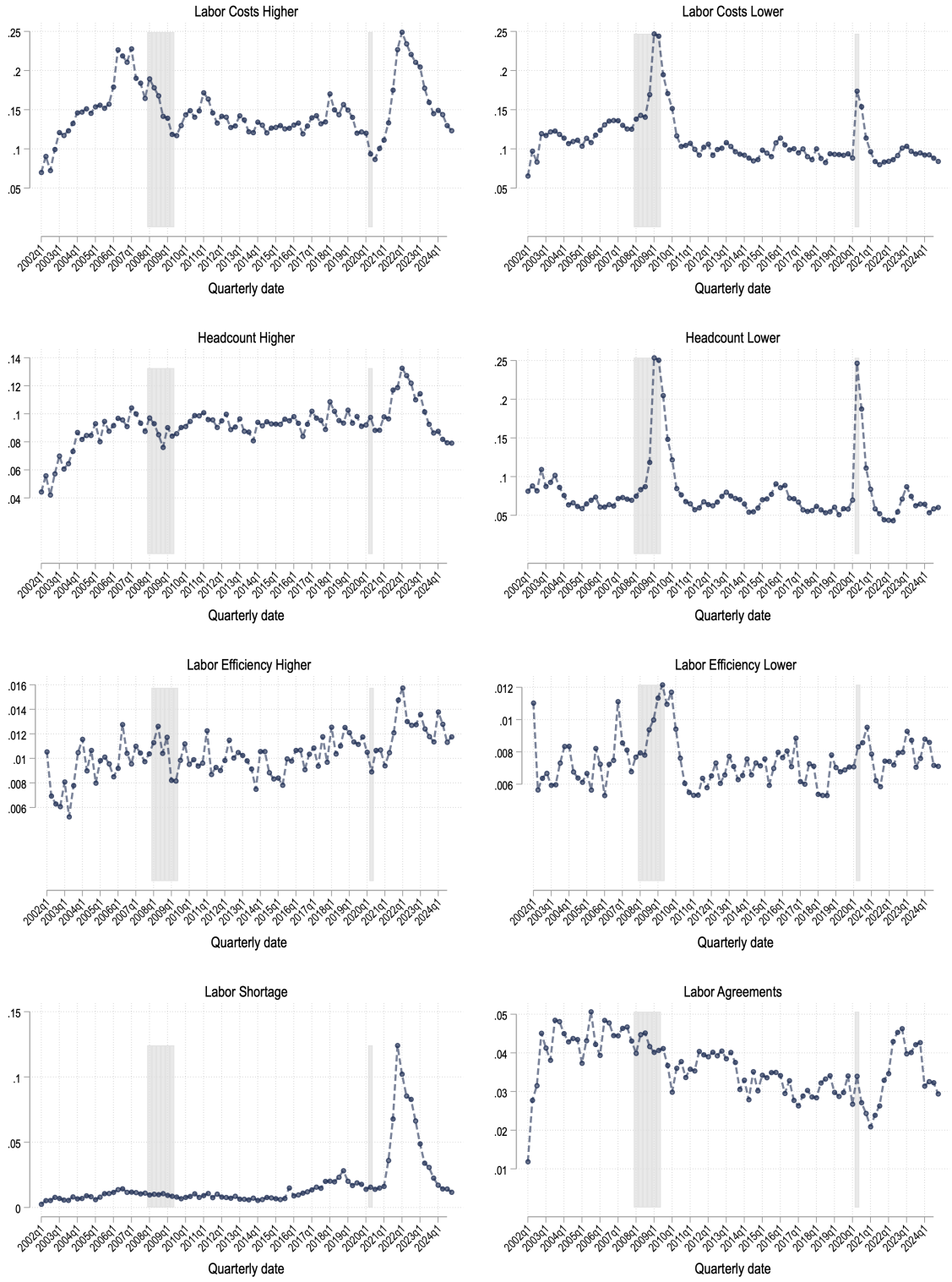
Appendix

Table A.1: Top 20 keywords by topic

Topic	Keywords
headcount_higher	training; hiring; hired; hire; staffing; recruiting; hires; recruitment; recruit; headcount increase; headcount growth; headcount increased; employees growth; employees increase; talent growth; headcount higher; employees increased; employees higher; headcount increases; workforce growth
headcount_lower	headcount reduction; headcount reductions; headcount reduced; headcount down; headcount lower; layoffs; workforce reduction; employees down; employees reduction; headcount reduce; employees less; employees lower; furlough; workforce reductions; employees reduced; headcount reducing; furloughs; headcount decrease; employees reduce; workforce reduced
labor_agreements	labor contract; labor contracts; labor agreement; employees contract; contractor contract; employees agreement; labor agreements; compensation agreement; labor union; labor negotiations; compensation contract; staff contract; contractors contract; employees contracts; contractors contracts; employees union; labor unions; employee contract; compensation contracts; personnel contract
labor_costs_higher	compensation increase; compensation higher; compensation increased; wage inflation; wage increases; compensation increases; labor costs higher; wage increase; salary increase; salary increases; salaries increase; wage higher; labor cost higher; labor costs increased; labor cost inflation; personnel costs increase; personnel expenses increase; labor costs increase; labor cost increase; wages increase
labor_costs_lower	compensation lower; compensation decrease; compensation reduction; compensation down; compensation decreased; compensation reduced; labor costs lower; compensation decline; labor cost lower; compensation reductions; bonus lower; compensation savings; compensation improvement; headcount cost reduction; wage lower; personnel costs lower; salaries lower; salary lower; headcount cost reductions; salary reductions
labor_efficiency_higher	labor productivity higher; labor productivity increase; labor productivity increased; labor efficiencies higher; headcount productivity growth; labor efficiency increase; labor efficiency higher; labor productivity increases; labor efficiencies increase; labor efficiencies increased; labor productivity growth; labor efficiency increased; employee productivity increase; headcount productivity increase; employee productivity increased; employee productivity higher; labor efficiencies growth; labor productivity increasing; employees productivity increase; labor efficiencies increases
labor_efficiency_lower	labor productivity lower; labor efficiencies lower; labor efficiency lower; labor inefficiencies higher; labor productivity reduce; labor productivity reduction; headcount efficiency reduction; labor efficiency reduce; labor productivity reduced; headcount efficiencies reductions; labor productivity down; headcount efficiencies reduction; labor efficiencies reduction; headcount efficiency reductions; headcount productivity reduction; labor productivity reductions; labor productivity reducing; labor efficiency reducing; labor efficiencies reductions; labor efficiency reduction
labor_shortage	labor shortages; labor tight; labor shortage; labor constraints; staffing shortages; labor lack; labor tightness; labor tightening; labor tighter; staff shortages; labor constraint; labor scarcity; staff shortage; hiring tight; workers shortage; employees tight; staffing constraints; headcount tight; labor bottlenecks; staffing shortage

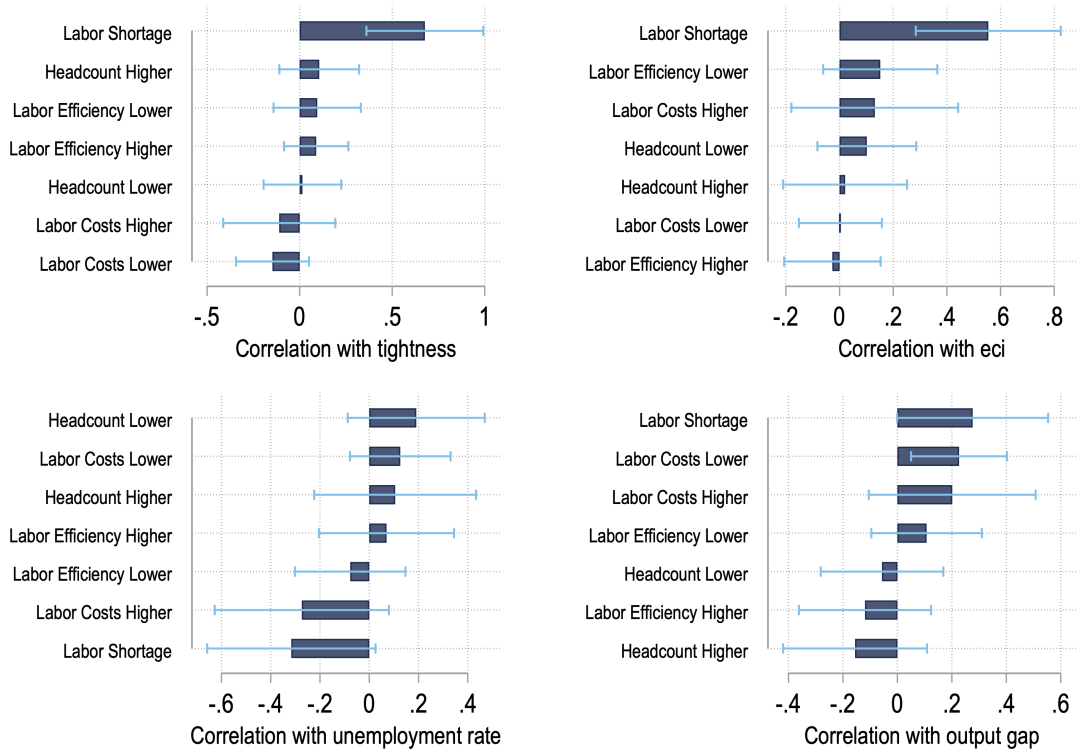
Notes: This table shows top 20 keywords used for each of our labor topics.

Figure A.1: Discussion of Labor Topics over Time



Notes: The figures show the average percent share of sentences of a labor topic mentioned in earnings calls by quarter.

Figure A.2: Correlation of Labor Topic Discussions with Aggregate Variables



Notes: The table shows results of regressions of tightness (calculated as the ratio of postings to vacancies), unemployment rate, employment cost index, and output gap on labor topics. Observations are quarterly. To construct labor topic observations at the quarter level we take sales weighted averages across firms who hold earnings calls in a quarter. We exclude firms in financial and administrative industries (NAICS 52, 53 and 56). Standard errors are robust.

Table A.2: Summary Stats

	Mean	SD	p1	p50	p99	N
Panel A: Firm x Quarter Level						
labor costs higher$_{i,t}$	0.17	0.34	0.00	0.00	1.52	248,437
labor shortage$_{i,t}$	0.02	0.12	0.00	0.00	0.53	248,437
headcount higher$_{i,t}$	0.10	0.26	0.00	0.00	1.19	248,437
labor efficiency higher$_{i,t}$	0.01	0.07	0.00	0.00	0.34	248,437
labor costs lower$_{i,t}$	0.13	0.28	0.00	0.00	1.23	248,437
headcount lower$_{i,t}$	0.08	0.24	0.00	0.00	1.11	248,437
labor efficiency lower$_{i,t}$	0.01	0.06	0.00	0.00	0.29	248,437
labor agreements$_{i,t}$	0.03	0.16	0.00	0.00	0.68	248,437
labor cost pressures$_{i,t}$	0.33	1.95	-5.81	0.23	6.07	248,437

Notes: This table shows summary statistics - mean, standard deviation, 1st percentile, median, 99th percentile and number of observations - for the firm x quarter level (in panel A), industry x quarter level (in panel B) and quarter level (in panel C) data used in empirical results.

Table A.3: Estimation of labor cost pressures: Labor topics and variable input cost share calculated using material inventories, at firm x year level

	$\Delta log(MaterialInventories/Sales)_{i,t}$		
	baseline using cogs + sga	using cogs	using cogs + inventories
	(1)	(2)	(3)
Labor Costs Higher$_{i,t}$	3.854***	2.155***	2.708***
	(0.731)	(0.743)	(0.777)
Labor Costs Lower$_{i,t}$	–6.497***	–3.471***	–3.043***
	(1.014)	(1.026)	(1.149)
Headcount Higher$_{i,t}$	0.882	–0.584	–0.025
	(0.930)	(0.978)	(0.984)
Headcount Lower$_{i,t}$	–0.764	0.367	–1.931*
	(1.162)	(1.210)	(1.121)
Labor Shortage$_{i,t}$	2.019	2.290	2.625
	(2.486)	(2.409)	(2.026)
Labor Efficiency Higher$_{i,t}$	–3.636	–1.542	1.437
	(3.146)	(3.247)	(2.913)
Labor Efficiency Lower$_{i,t}$	1.373	1.582	8.506**
	(3.398)	(4.066)	(3.563)
Labor Agreement$_{i,t}$	–0.441	–0.725	0.118
	(1.247)	(1.029)	(1.081)
R^2	0.223	0.204	0.273
N	21,240	21,224	17,091
Time FE	Y	Y	Y
Controls (Risk and Sentiment)	Y	Y	Y
Firm FE	Y	Y	Y

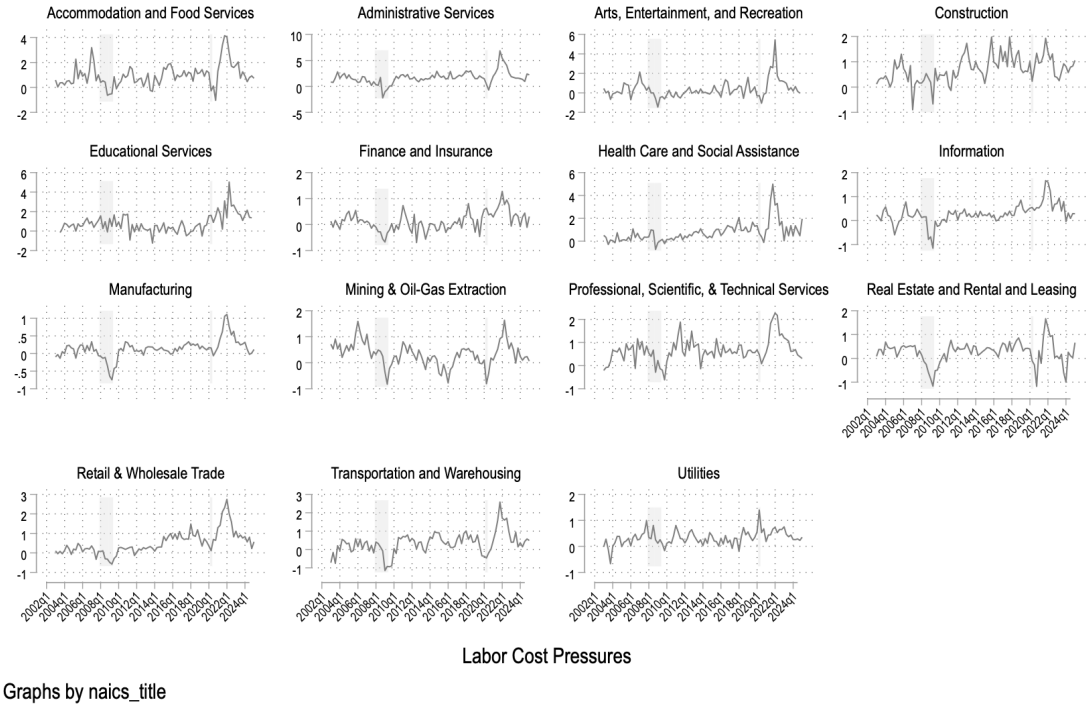
Notes: The table shows regression of changes in $log(MaterialInventories/Sales)_{i,t}$ on labor topics. Each observation denotes a firm i and year t. To construct labor topic observations at the firm x year level we take averages across all quarters in a year. All specifications include controls for changes in $log(Sales)_{i,t}$ and changes in $log(Emp)_{i,t}$. Columns (4) and (5) include controls for risk and sentiment. We exclude firms in financial and administrative industries (NAICS 52, 53 and 56). We only include firm level data till 2019 to exclude Covid-19 pandemic period. Standard errors are clustered by firm.

Table A.4: Firm level regressions, topics entered individually

	Change in materials share proxy$_{i,t}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Higher labor costs	1.953***							
	(0.620)							
Lower labor costs		-5.117***						
		(0.863)						
Higher headcount			1.266					
			(0.890)					
Lower headcount				-2.942***				
				(1.084)				
Labor shortages					3.033			
					(2.462)			
Higher efficiency						-3.818		
						(2.858)		
Lower efficiency							-4.765	
							(2.982)	
Labor agreements								-1.084
								(1.286)
Labor exposure	0.799**	1.370***	0.923**	1.160***	0.945**	1.051***	1.027**	1.056***
	(0.407)	(0.410)	(0.409)	(0.411)	(0.406)	(0.403)	(0.404)	(0.408)
R^2	0.220	0.221	0.220	0.220	0.220	0.220	0.220	0.220
N	21240	21240	21240	21240	21240	21240	21240	21240

Notes: Each column reports a separate regression of the dependent variable on a single labor topic, controlling for labor exposure, sentiment, risk, and firm growth controls. All specifications include firm and year fixed effects, and standard errors are clustered by firm. The sample is restricted to years up to 2019.

Figure A.3: Labor Cost Pressure Index ($\bar{\omega}_t$) across Industries



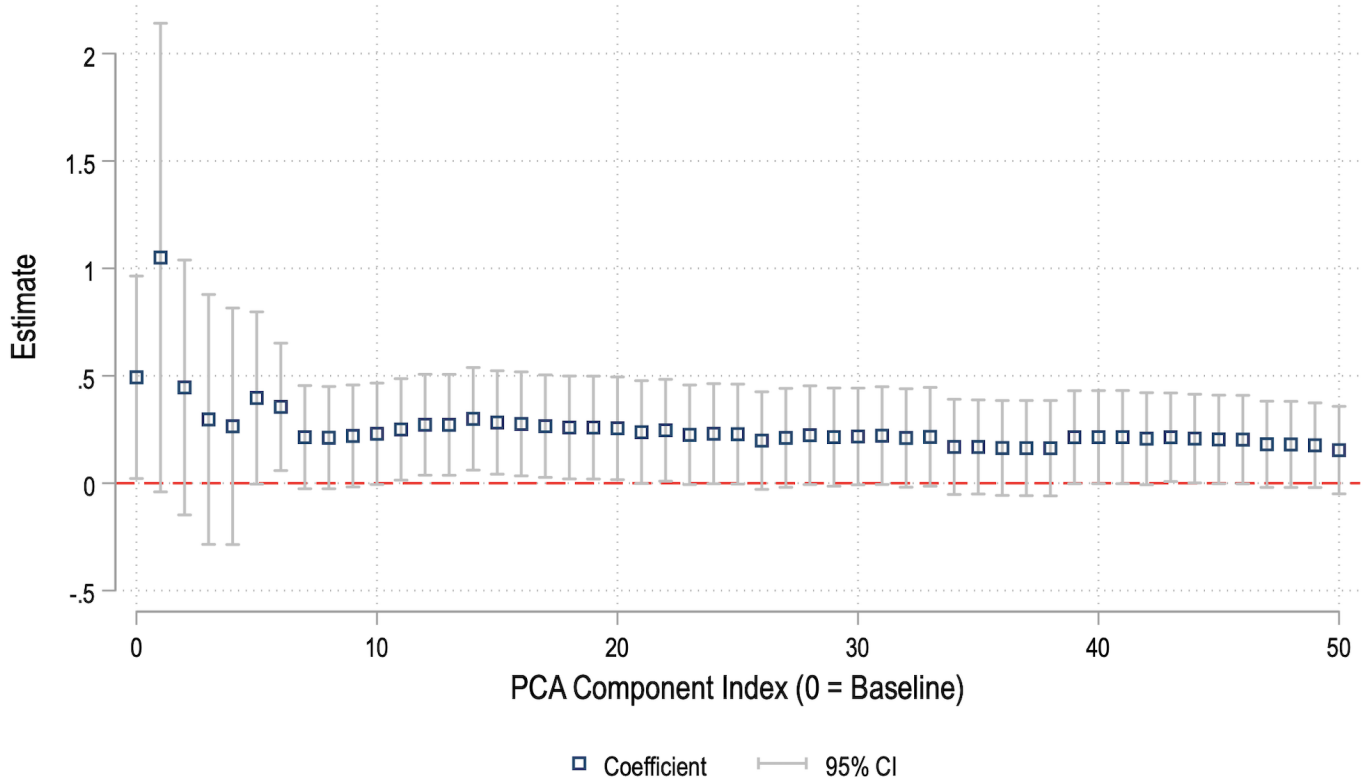
Notes: The figures show the average change in estimated marginal cost of labor across firms by industry and by quarter. The estimation uses coefficients from Table 3 and topic specific scores by firm and by quarter. We then aggregate these by taking a sales weighted average across firms by industry and quarter.

Table A.5: Labor Cost Pressures and Inflation, Industry x time Level

	PPI Inflation$_{n,(t-4,t)}$		
	(1)	(2)	(3)
Labor Cost Pressures$_{n,t}$	0.387*	0.642***	0.493**
	(0.227)	(0.234)	(0.241)
R^2	0.001	0.338	0.362
N	2,832	2,832	2,832
Industry FE	N	N	Y
Time FE	N	Y	Y

Notes: The table shows regression of PPI inflation observed in industry n between quarters and on estimated changes in MCL for industry in quarter . Industry is one of NAICS 3-digit industries. Estimated changes in MCL for industry in quarter is calculated by taking a sales weighted average of estimated changes in MCL across all firms in Compustat in the industry. We exclude financial and administrative industries (NAICS 52, 53 and 56). Regression is weighted by number of firms observed within the industry. Standard errors are clustered by industry. Inflation is winsorized at 2nd and 98th percentile.

Figure A.4: Robustness: Phillips Curve Estimation using Unsupervised Methodology (PCA)



Notes: This figure plots the coefficient estimates and 95% confidence intervals from a series of regressions. The dependent variable is PPI inflation observed in industry n between quarters t and $t+4$. The x-axis represents different specifications: Index 0 is the baseline Labor Cost Pressure measure ($MCL_{n,t}$), while indices 1 through 50 represent specifications using the first 50 Principal Components derived from the underlying text data. The underlying MCL is calculated as a PCE-weighted average of estimated changes in labor cost pressures across US-headquartered firms in Compustat that hold earnings conference calls in quarter t. We exclude financial and administrative industries (NAICS 52, 53, and 56). All regressions include industry (NAICS3) and time (Date) fixed effects. Standard errors are robust.

Table A.6: Industry level Phillips Curve robustness to alternative labor cost pressure measures

	Industry PPI inflation$_{n,t,t+4}$			
	(1)	(2)	(3)	(4)
Labor cost pressures$_{n,t}$ (baseline COGS + SGA)	0.493**			
	(0.241)			
Labor cost pressures$_{n,t}$ (COGS + material inventory)		0.863**		
		(0.337)		
Labor cost pressures$_{n,t}$ (COGS)			0.946**	
			(0.448)	
Labor cost pressures$_{n,t}$ (unsupervised PCA)				0.355**
				(0.151)
R^2	0.362	0.363	0.362	0.364
N	2832	2832	2832	2832

Notes The table reports industry level regressions of four quarter ahead year on year PPI inflation on alternative measures of labor cost pressures, using NAICS three digit industries as the cross section. All specifications include industry and quarter fixed effects and are weighted by mean number of firms holding earnings conference calls. Standard errors are robust.

Table A.7: Robustness of Phillips Curve to alternative labor cost pressure measures

	PCE Inflation$_{t,t+4}$			
	(1)	(2)	(3)	(4)
Labor cost pressures$_t$ (baseline COGS + SGA)	0.343***			
	(0.119)			
Labor cost pressures$_t$ (COGS + inventory)		0.241		
		(0.183)		
Labor cost pressures$_t$ (COGS)			0.471***	
			(0.106)	
Labor cost pressures$_t$ (unsupervised PCA)				0.696***
				(0.148)
R^2	0.430	0.402	0.484	0.503
N	88	88	88	88

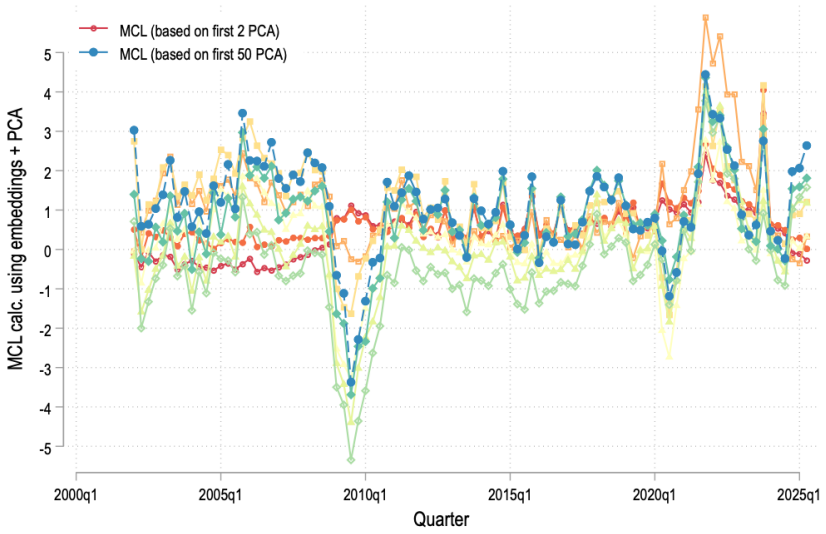
Notes: The table reports OLS estimates of the Phillips Curve specification using alternative measures of labor cost pressures. The dependent variable is PCE inflation over the next four quarters. Column (1) uses the baseline SGA based measure, column (2) the inventory based measure, column (3) the alternative materials flow measure, and column (4) the unsupervised PCA based measure. All regressions control for lagged Michigan inflation expectations. Standard errors are robust.

Table A.8: Inflation Decomposition: Systematic vs. Idiosyncratic Pressures, without lagged inflation control

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Total	Unemployment Decomposition			Tightness Decomposition		
Total Labor Cost Pressures	0.533***						
	(0.118)						
Systematic ($\beta \times U_t$)		0.562***		0.564***			
		(0.135)		(0.133)			
Idiosyncratic Residual			0.519***	0.520***			
			(0.159)	(0.148)			
Systematic ($\beta \times V/U_t$)					0.557***		0.637***
					(0.140)		(0.134)
Idiosyncratic Residual						0.211	0.384**
						(0.209)	(0.156)
R^2	0.372	0.270	0.324	0.372	0.344	0.236	0.385
N	88	88	88	88	88	88	88

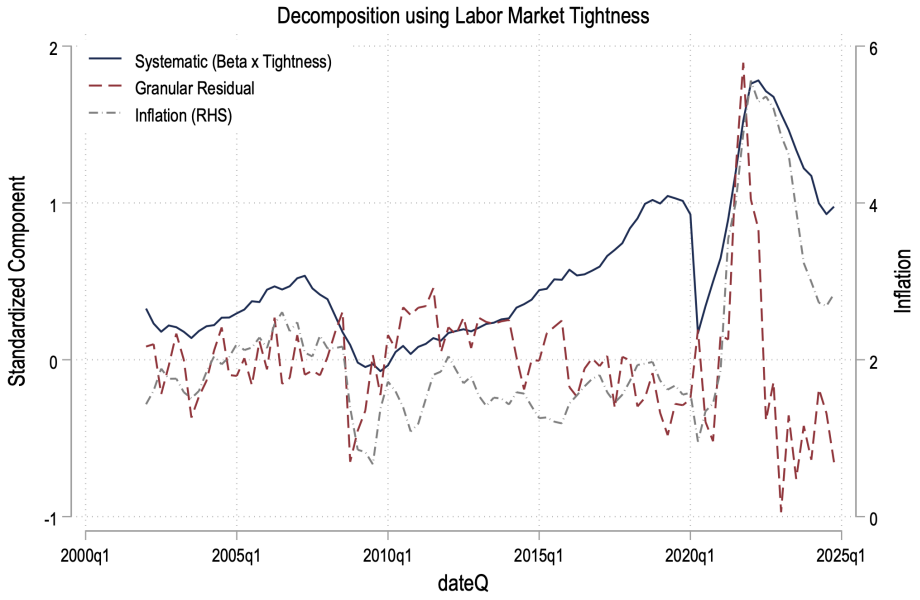
Notes: The table reports results from regressing PCE inflation on the systematic and idiosyncratic components of the Labor Cost Pressure index. The Systematic Component is constructed by aggregating firm-level predicted values based on aggregate slack (Unemployment or Tightness). The Idiosyncratic Component is constructed by aggregating firm-level residuals. All independent variables are standardized. Standard errors are robust.

Figure A.5: Marginal cost of labor estimates calculated using unsupervised methodology

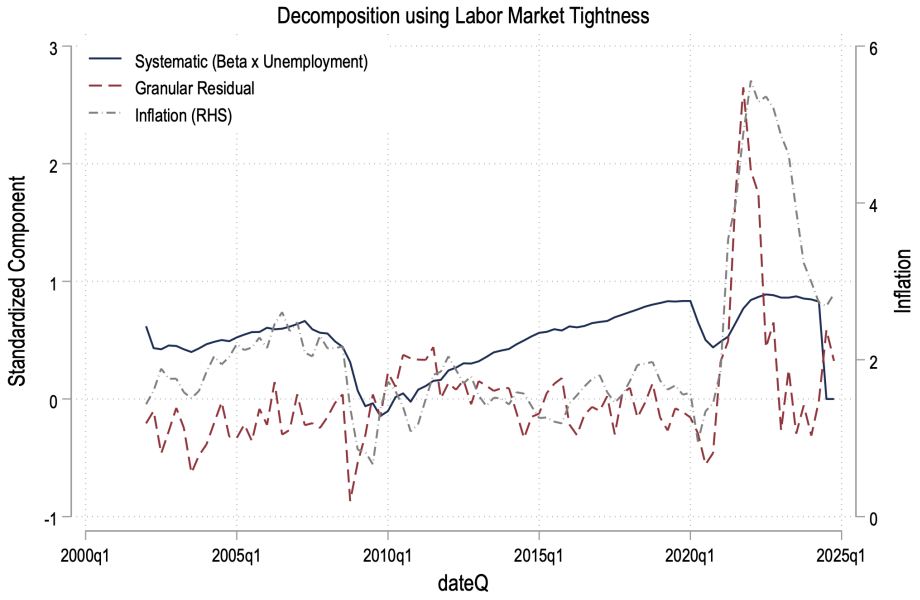


Notes: The figure plots the aggregate labor cost pressure index constructed using the unsupervised sentence embedding approach. For each choice of the number of latent topics $K \in \{2, 4, 6, 8, 10, 20, 30, 40, 50\}$ we apply principal component analysis to Sentence T5 embeddings of labor related sentences, treat the first K components as topics, and then replicate the baseline mapping from topic variation to changes in marginal labor costs to obtain an aggregate series. Each line corresponds to one value of K. Lines based on fewer topics are shown in warmer colors, while lines based on more topics are shown in cooler colors.

Figure A.6: Decomposition of Labor Cost Pressures: Systematic vs. Idiosyncratic



(a) Decomposition using Labor Market Tightness



(b) Decomposition using Unemployment Rate

Notes: The figure decomposes the aggregate Labor Cost Pressure index into a Systematic component (predicted by firm-level sensitivity to aggregate slack) and a Granular Residual (the unexplained firm-level variation). Panel (a) defines the systematic component using Labor Market Tightness (V/U), while Panel (b) uses the Unemployment Rate. Both components are standardized.