Online Appendix

Measuring Economic Sentiment from Open-Ended Survey Comments Using Large Language Models

Pascal Seiler
ETH Zurich

September 2025

Abstract

This article develops a novel economic sentiment indicator (LLM-ESI) by applying large language models to open-ended responses from Swiss business tendency surveys. Using a BERT-based transformer model, it extracts firm-level sentiment from free-text survey comments and aggregates it into a high-frequency indicator of macroeconomic conditions. The LLM-ESI closely tracks the business cycle and performs on par with, or better than, traditional benchmarks in nowcasting GDP. These results highlight the potential of large language models and open-ended survey responses to deliver timely and nuanced signals for real-time economic analysis.

JEL classification: C55, C53, E32, E37, E66

Keywords: Economic Sentiment, Large Language Model, Business Tendency Surveys, Survey Comments, Textual Analysis, Forecasting

A. Methodology

A.1 Questionnaire of the monthly KOF business tendency survey in the manufacturing sector

KOF Business tendency survey
Industry

KOF Swiss Economic Institute
ETH Zürich, LEE F 101, 8092 Zürich
http://www.kof.ethz.ch
Tel: 044 632 43 26
ind@kof.ethz.ch

12034

Survey INU
Company-ID
Contact-ID
Sector-ID

Sector name: _____
clasification: _____

Please note

- Your responses should refer only to the branch named above
- The questions refer to the activities of domestic branches
- Do not use a red pencil
- Tick the appropriate box ☒ ☒
- The notes are on the back of the sheet
- Your responses are treated strictly confidential.

Review and Assessment of the Current Situation

1. Incoming orders

a) In the past month compared to the previous* they have
○ increased ○ remained the same ○ declined

b) Compared to the same past month one year ago they were
○ higher ○ the same ○ lower

2. Order backlog ☐ no order backlog

a) In the past month compared to the previous* month orders have
○ increased ○ remained the same ○ declined

b) How would you assess the present order backlog* overall? As
○ large ○ normal ○ too low

c) How would you assess the present order backlog* for exports? As ☐ no exports
○ large ○ normal ○ too low

3. Production

a) In the past month compared to the previous* it has
○ increased ○ not changed ○ decreased

b) Compared to the same past month one year ago it was
○ higher ○ the same ○ lower

4. Intermediate products inventory ☐ no inventory

a) In the past month compared to the previous* it has been
○ higher ○ the same ○ lower

b) How would you assess the intermediate product inventory* ? As
○ too high ○ normal ○ too low

5. Finished products inventory ☐ no inventory

a) In the past month compared to the previous* it has
○ increased ○ the same ○ dropped

b) How would you assess the finished product inventory* ? As
○ too high ○ normal ○ too low

6. Employment levels

We would assess the current number of employees* as
○ too large ○ normal ○ too small

7. Business situation

a) How would you assess your current overall business situation*?
○ good ○ satisfactory ○ poor

b) In **the next 6 months*** our business situation will
○ improve ○ remain the same ○ get worse

c) To predict the **future development** of our business situation is currently
○ easy ○ rather easy ○ rather difficult ○ difficult

d) The **uncertainty** about the **future development** of our business situation is currently
○ higher than usual ○ normal/as usual ○ lower than usual

Expectations

8. It is likely that in the next 3 months

a) incoming orders will*
○ increase ○ remain the same ○ decrease

b) export orders will* ☐ no export
○ increase ○ remain the same ○ decrease

c) production will*
○ increase ○ remain the same ○ decrease

d) the purchase of intermediate products* will
○ increase ○ remain the same ○ decrease

e) the number of employees (FTEs) will*
○ increase ○ remain the same ○ decrease

f) our selling prices will*
○ increase ○ remain the same ○ decrease

g) our purchase prices will*
○ increase ○ remain the same ○ decrease

* Excluding seasonal fluctuations

Continue on the back page

Comments

A.2 Firm-level Comments in Business Tendency Surveys

Table A.1. Distribution of comments by firm characteristics

	Share of firms (%)
Language	
German	77.1
English	0.1
French	17.2
Italian	5.5
Sector group	
Manufacturing	18.8
Construction	11.8
Retail trade	20.4
Wholesale trade	4.1
Hotel and catering	14.7
Financial and insurance activities	3.9
Project engineering	11.0
Other service activities	15.3
Firm size	
L	5.9
M	17.7
S	76.4
Region	
Central Switzerland	10.4
Eastern Switzerland	14.7
Espace Mittelland	17.5
Northwestern Switzerland	11.9
Region Lemanique	14.2
Ticino	6.4
Zurich	25.0
Gender of respondent	
Female	18.0
Male	82.0

Notes: This table shows the distribution of comments by firm characteristics: language, sector group, firm size, region, and gender of respondents. Sector group "other service activities" includes all services excluding retail and wholesale trade, hotel and catering, financial and insurance activities and project engineering. Size classes differentiate between large (employing more than 250 employees, "L"), medium-sized (employing more than 50 employees but less than or equal to 250 employees, "M"), and small firms (employing fewer than 50 employees but more than 1 employee, "S").

Figure A.1. Word clouds of firm comments

Notes: Word clouds visualizing the hundred most frequently used terms in firms' comments to the KOF Business Tendency Surveys across the four languages German, French, Italian, and English. The size of each word reflects its relative frequency in the sample.

A.3 Construction of the LLM-Based Economic Sentiment Indicator

Table A.2. Examples of comments with the highest and lowest sentiment scores

Comment	Sentiment Score
March was excellent, with a clear upturn in activity in the building trades. The 1st quarter was also very good.	0.938
Another good month! Competitive prices and the desire to buy once again contributed to a positive month.	0.936
December is expected to be the best month of the year. The joy of consumption will show with the Christmas business.	0.936
We're delighted to report another increase. Always good surprises at the end of the month!	0.935
June in particular was a very pleasing month for us!	0.933
⋮	⋮
Big drop in catering. However, it is difficult for us to close the restaurant.	-0.954
Unfair competition. Undeclared work.	-0.954
Legal instability is unbearable.	-0.955
This war is a human and economic catastrophe. 40% of our customers will go bankrupt. We are very worried.	-0.956
COVID-related absences manageable. Low/negative interest rates still a problem.	-0.960

Notes: This table shows the five comments with the highest and lowest sentiment scores, respectively. The comments have been translated into English.

B. Empirical Analysis of the LLM-ESI

Table B.1. Unconditional time series moments of the LLM-ESI

	Sentiment
Mean	-0.24
Median	-0.24
Variance	0.02
Volatility	-0.57
Skewness	-0.10
Kurtosis	3.68
AR(1)	0.25
Half-life	0.50

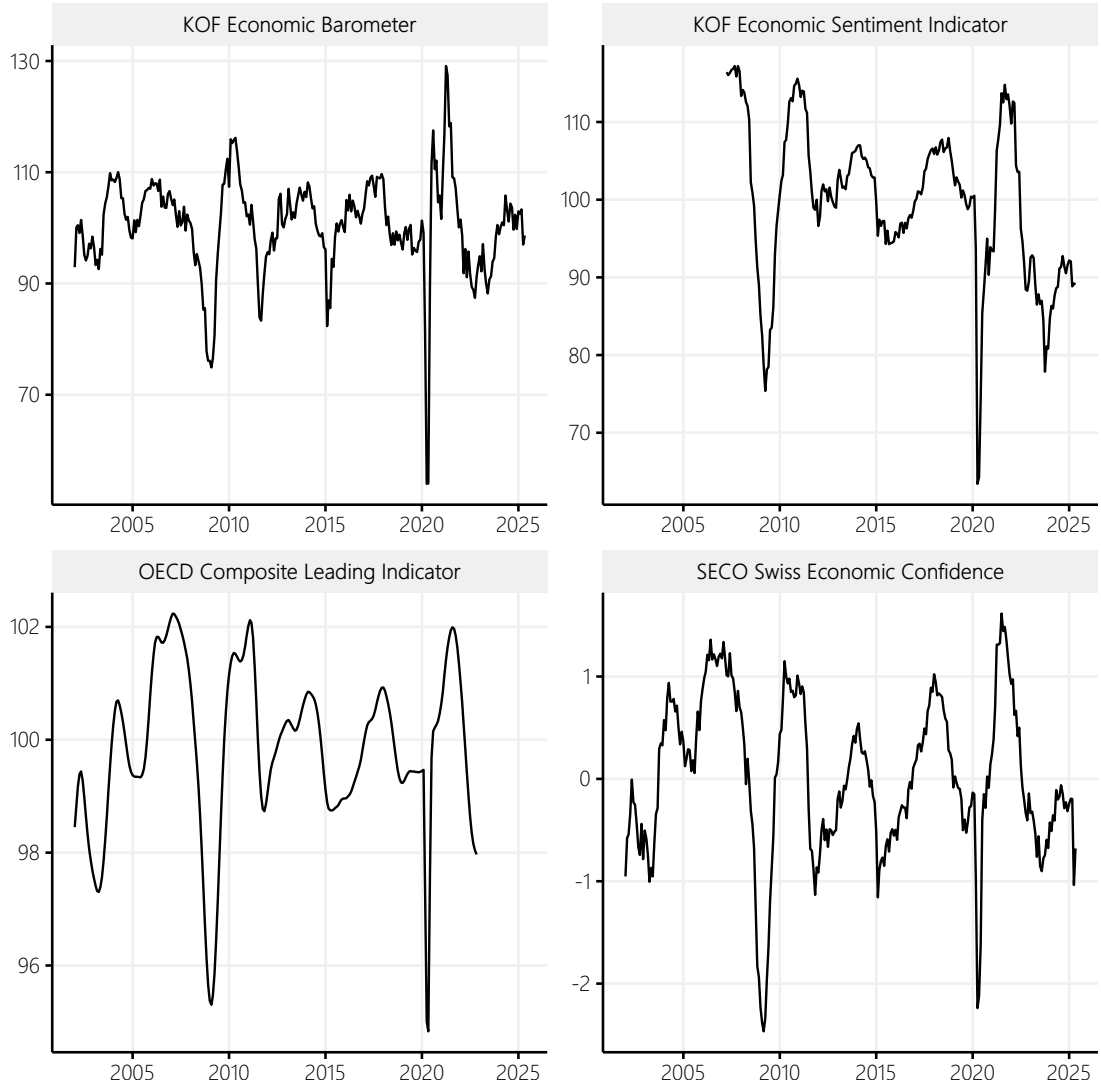
Notes: This table shows unconditional time series moments of the LLM-ESI. The sample periods is 2002:01–2025:05. Volatility is the coefficient of variation. AR(1) is the first-order autocorrelation coefficient. Half-life estimates the half-life of an aggregate innovation from an univariate autoregression as $(\ln(0.5) / \ln(|\text{AR}(1)|))$.

Table B.2. Macroeconomic data and leading indicator

	Frequency	Source	Comment
GDP	Quarterly	SECO	
GDP vintages	Quarterly	SECO	From Indergand and Leist (2014) and retrieved from ALFRED (CPMNACSAB1GQCH)
Recession dummy	Monthly	OECD	Retrieved from FRED (CHERECD)
KOF Economic Barometer	Monthly	KOF	Leading composite indicator (300+ economic time series)
KOF Economic Sentiment Indicator	Monthly	KOF	Composite index combining results from the KOF Business Tendency Surveys and the SECO Consumer Confidence Survey, following the method employed by the EU Commission to calculate the European ESI.
Composite Leading Indicator (CLI)	Monthly	OECD	Leading indicator, aggregating various forward-looking economic variables.
Swiss Economic Confidence	Monthly	SECO	Composite indicator of 30 domestic survey indicators

Notes: The table provides details of the macroeconomic data and leading indicators used in the analysis.

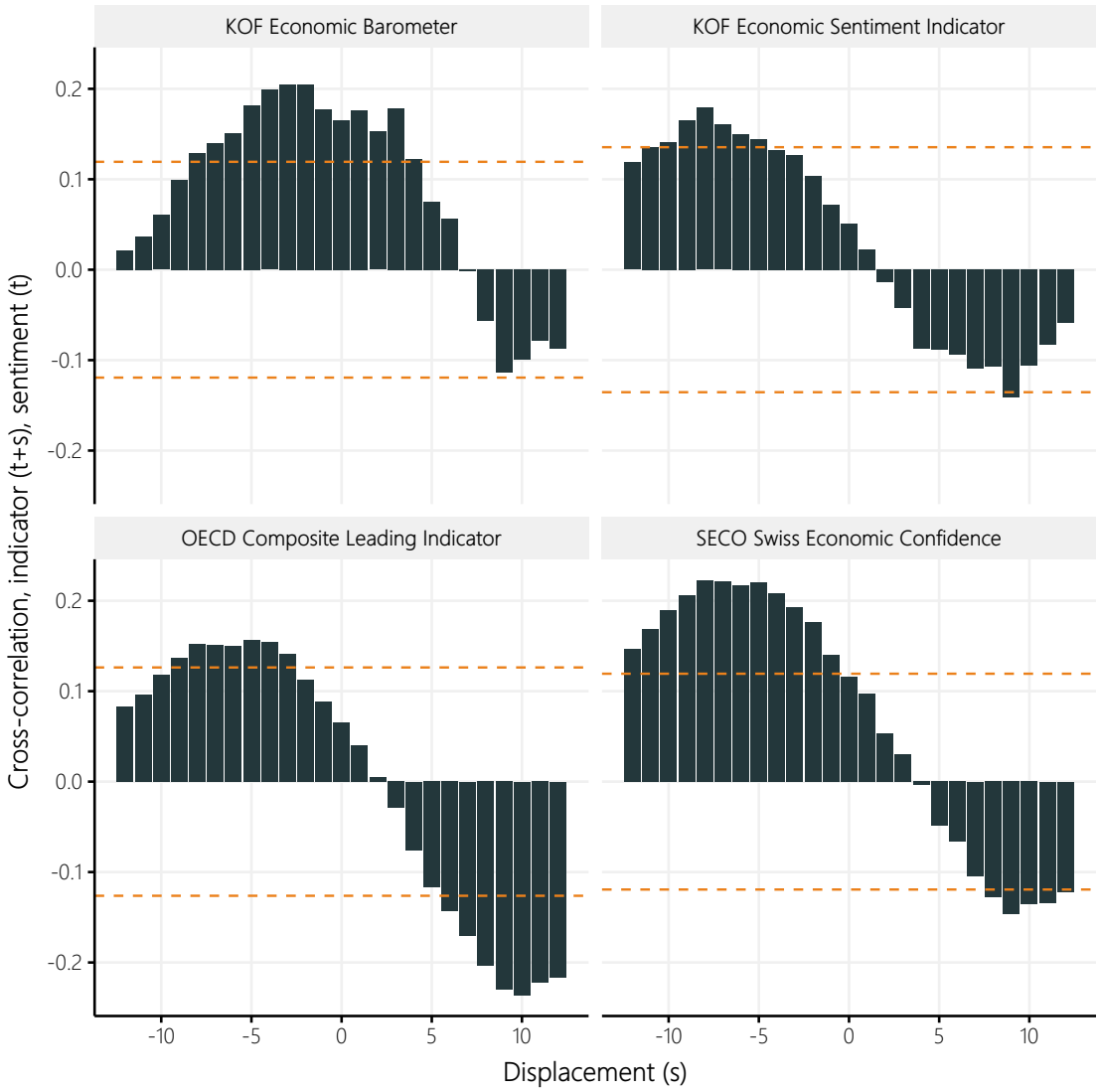
Figure B.1. Prominent leading and economic sentiment indicators



Notes: The figure shows plots for the four leading indicators used in the pseudo out-of-sample analysis. The indicators are described in Table B.2.

Figure B.2 presents cross-correlations of the LLM-ESI with four leading indicators: the KOF Economic Barometer, the KOF Economic Sentiment Indicator, the OECD Composite Leading Indicator, and the SECO Swiss Economic Confidence. It shows that the LLM-ESI exhibits both significant coincident and leading relationships. Most notably, it leads the KOF Economic Sentiment Indicator, the OECD CLI and the Swiss Economic Confidence Indicator by more than six months, on average. With the KOF Economic Barometer, it shows leading, coincident, and lagging correlations.

Figure B.2. Cross-correlation with other prominent leading and economic sentiment indicators



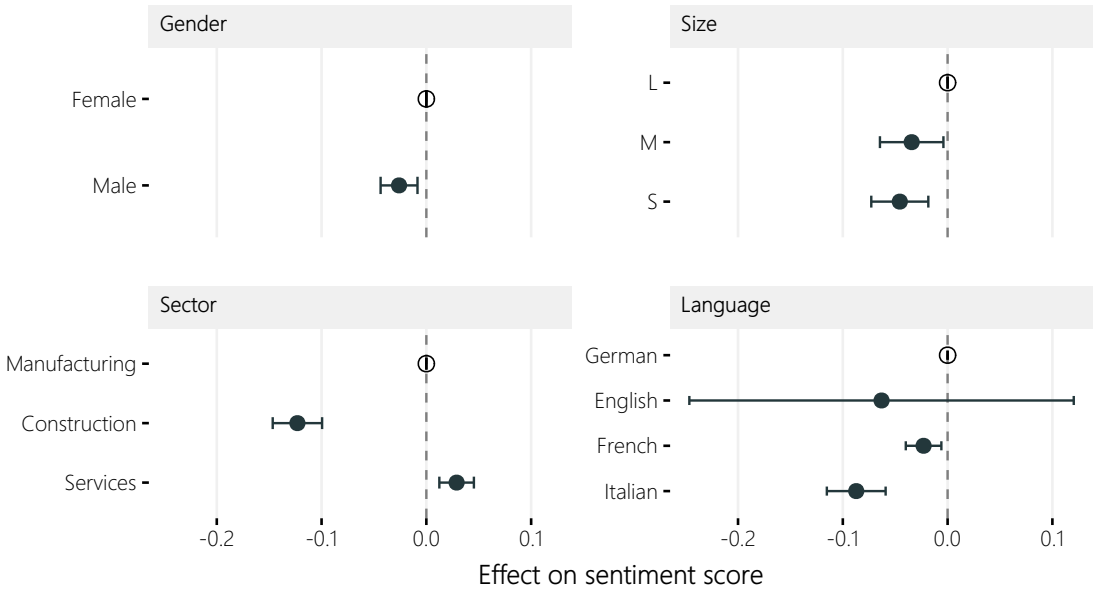
Notes: Cross-correlation between the LLM-ESI and other prominent leading and economic sentiment indicators. All data are at monthly frequency. The dashed lines give 95% confidence intervals. A bar outside of the interval suggests a statistically significant correlation between the indicator at a lead/lag of s. Before computing the cross-correlation, the series have been pre-wightened with an AR(p) model (Neusser, 2016). The lag order has been determined using the Bayesian information criterion. The sample period is 2002:01–2025:05.

B.1 Heterogeneity in firm-level sentiment expression

To complement the aggregate time series analysis of sentiment, I explore heterogeneity in sentiment scores at the level of individual survey comments. Specifically, I examine whether and how sentiment varies with firm and respondent characteristics as well as with the timing of the response. This analysis is motivated by the possibility that certain structural or behavioral factors–such as firm size, sectoral environment, respondent demographics, or timing of response–might influence how economic conditions are perceived and articulated in textual comments.

To assess heterogeneity along firm and participant dimensions, I regress the sentiment score of each comment on a set of immutable categorical variables capturing gender, firm size, sector, and questionnaire language. The results are presented in Figure B.3. Male respondents express slightly more negative sentiment compared to female respondents. With respect to firm size, sentiment is more negative in small and medium-sized firms than in large firms. Sentiment is significantly more negative in the construction sector relative to the manufacturing sector, while comments from the service sector tend to be more positive than those from manufacturing. Language-related differences are also pronounced: sentiment is more negative in French-language responses than in German, and more negative still in Italian compared to French, while English-language responses show no statistically significant difference. These findings point to meaningful variation in sentiment expression that aligns with both structural firm characteristics and the linguistic context in which responses are formulated.

Figure B.3. Heterogeneity in sentiment scores by firm and participant characteristics



Notes: Variation in sentiment scores by firm and participant characteristics: gender (in the top left panel), firm size (top right), sector (bottom left), and language of the questionnaire (bottom right). Dots with vertical lines indicate point estimates with cluster-robust 95% confidence intervals from OLS regressions. The hollow dots on the zero line denote the reference category. Table B.3 contains the numerical estimates.

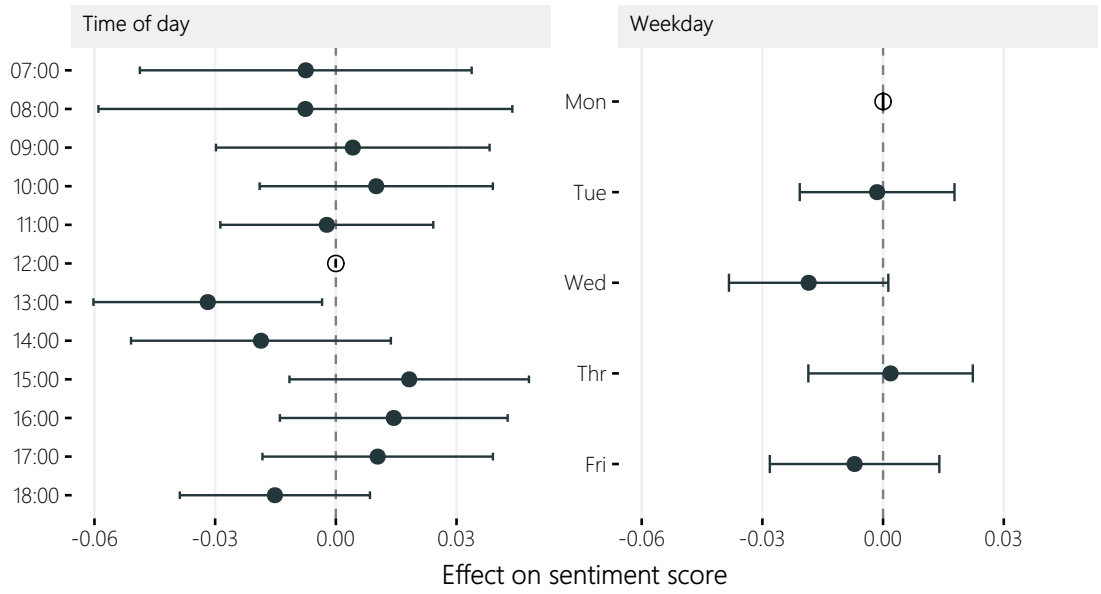
Table B.3. Heterogeneity in sentiment scores by firm and participant characteristics

	Dependent variable:			
	Sentiment Score $S_{i,t}$			
	(1)	(2)	(3)	(4)
Gender				
Male	-0.026^{***}			
	(0.009)			
Size				
M		-0.034^{**}		
		(0.015)		
S		-0.046^{***}		
		(0.014)		
Sector				
Construction			-0.123^{***}	
			(0.012)	
Services			0.029^{***}	
			(0.008)	
Language				
English				-0.063
				(0.094)
French				-0.023^{***}
				(0.009)
Italian				-0.087^{***}
				(0.014)
Constant	-0.242^{***}	-0.218^{***}	-0.264^{***}	-0.250^{***}
	(0.008)	(0.013)	(0.007)	(0.004)
Observations	17,642	19,839	19,862	19,862
R^2	0.0005	0.001	0.011	0.002
Adjusted R^2	0.0004	0.0005	0.011	0.002
Residual Std. Error	0.458	0.459	0.456	0.458
F Statistic	8.425^{***}	5.842^{***}	111.464^{***}	13.800^{***}

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Variation in sentiment scores by firm and participant characteristics relative to the respective reference category: gender (reference category: female), firm size (ref. cat.: L), sector (ref. cat.: manufacturing), and language of the questionnaire (ref. cat.: German). OLS regressions over the sample period 2002:01–2025:05.

In a second step, I investigate heterogeneity related to the timing of response. Using the same micro-level regression framework, I examine whether sentiment systematically varies across hours of the day or days of the week. The corresponding estimates are shown in Figure B.4. While sentiment does not differ systematically across most weekdays, it tends to dip in the middle of the week, with Wednesday showing slightly more negative sentiment compared to Monday. There is also some indication that sentiment expressed in early afternoon responses–especially around 1 p.m.–is slightly more negative than sentiment expressed at other times. However, the magnitude of these temporal effects remains modest. Overall, these results suggest that while aggregate sentiment is a robust indicator, individual sentiment expression does exhibit heterogeneity along observable dimensions, which may be important for applications focused on subgroups or high-frequency dynamics.

Figure B.4. Heterogeneity in sentiment scores by response time



Notes: Variation in sentiment scores during the workday (in the left panel) and the week (in the right panel). Dots with vertical lines indicate point estimates with cluster-robust 95% confidence intervals from OLS regressions. The hollow dots on the zero line denote the reference category. Table B.4 contains the numerical estimates.

Table B.4. Heterogeneity in sentiment scores by response time

	Dependent variable:	
	Sentiment Score $S_{i,t}$	
	(1)	(2)
Time of day		
07:00	−0.007	
	(0.021)	
08:00	−0.008	
	(0.026)	
09:00	0.004	
	(0.017)	
10:00	0.010	
	(0.015)	
11:00	−0.002	
	(0.014)	
13:00	−0.032**	
	(0.014)	
14:00	−0.019	
	(0.016)	
15:00	0.018	
	(0.015)	
16:00	0.014	
	(0.014)	
17:00	0.010	
	(0.015)	
18:00	−0.015	
	(0.012)	
Weekday		
Tue		−0.001
		(0.010)
Wed		−0.018*
		(0.010)
Thr		0.002
		(0.010)
Fri		−0.007
		(0.011)
Constant	−0.256***	−0.256***
	(0.009)	(0.007)
Observations	19,856	18,721
R^2	0.001	0.0003
Adjusted R^2	0.0004	0.00004
Residual Std. Error	0.458	0.458
F Statistic	1.809**	1.198

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Variation in sentiment scores during the workday and the week relative to the respective reference categories: 12:00 (for time of day) and Monday (for weekday). OLS regressions over the sample period 2002:01–2025:05.