

---

# STANDARDIZATION

UNDERSTANDING THE RELATIONSHIP BETWEEN RESPONSE STANDARDIZATION AND SENTIMENT IN CUSTOMER-COMPANY CONVERSATIONS ON TWITTER

BELLINI ALESSANDRO FRANCESCO

CAPUTI GIULIO

CIRÒ GIACOMO

DORSI TANCREDI

MOROSINI ALESSANDRO

VACCA FRANCESCO

# INTRODUCTION

## RESEARCH OVERVIEW



Our goal is to help managers to make informed **decisions** on how the firm should behave on **social networks**, especially when it comes to interactions with customers. The ultimate goal of our managers is to take decisions that are able to increase sales, brand image and overall firm success.



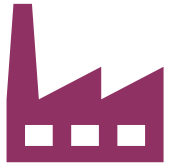
The main independent variable is the degree of **Standardization** in the firm's responses to customer's tweets. We measure its impact on firm's success through the user's **Feedback Sentiment**, measured from the last tweet he posted in the conversation, which is our dependant variable.



Our research question can be formulated as follows:  
«How does the degree of **standardization** in the firm's **responses** to customers' tweets **impacts** customers' **sentiment** once the conversation is concluded?»

# INTRODUCTION

## STANDARDIZATION



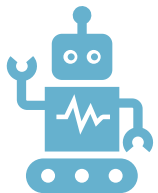
**Business Process Standardization (BPS)** refers to creating company-wide standard procedures which have been proved to improve operational performance, facilitate communication, reduce errors and associated costs.



BPS is also applied to **customer interactions**, leading to faster responses, lower costs, broader availability, and many other **advantages**.



Can also be **detrimental**, as answering a complex complaint with a vague and pre-determined response can lead to a dissatisfied customer



In the era of AI, it is increasingly more common to interact with **conversational agents**, and many researchers transpose the question of standardization to how chatbot personalities should be shaped to increase user satisfaction

# INTRODUCTION

## FEEDBACK SENTIMENT



It serves as a direct measure of **customer satisfaction**, a key factor influencing brand loyalty and the likelihood of repeat business.



The sentiment expressed in customer feedback reflects not only their satisfaction but also their perception of the firm's **brand image**.



Social networks are platforms for direct interaction between customers and the firm. Positive sentiment in feedback suggests effective **engagement** and communication.



Positive sentiment in customer feedback is likely to be associated with higher customer **retention rates** and long-term relationships

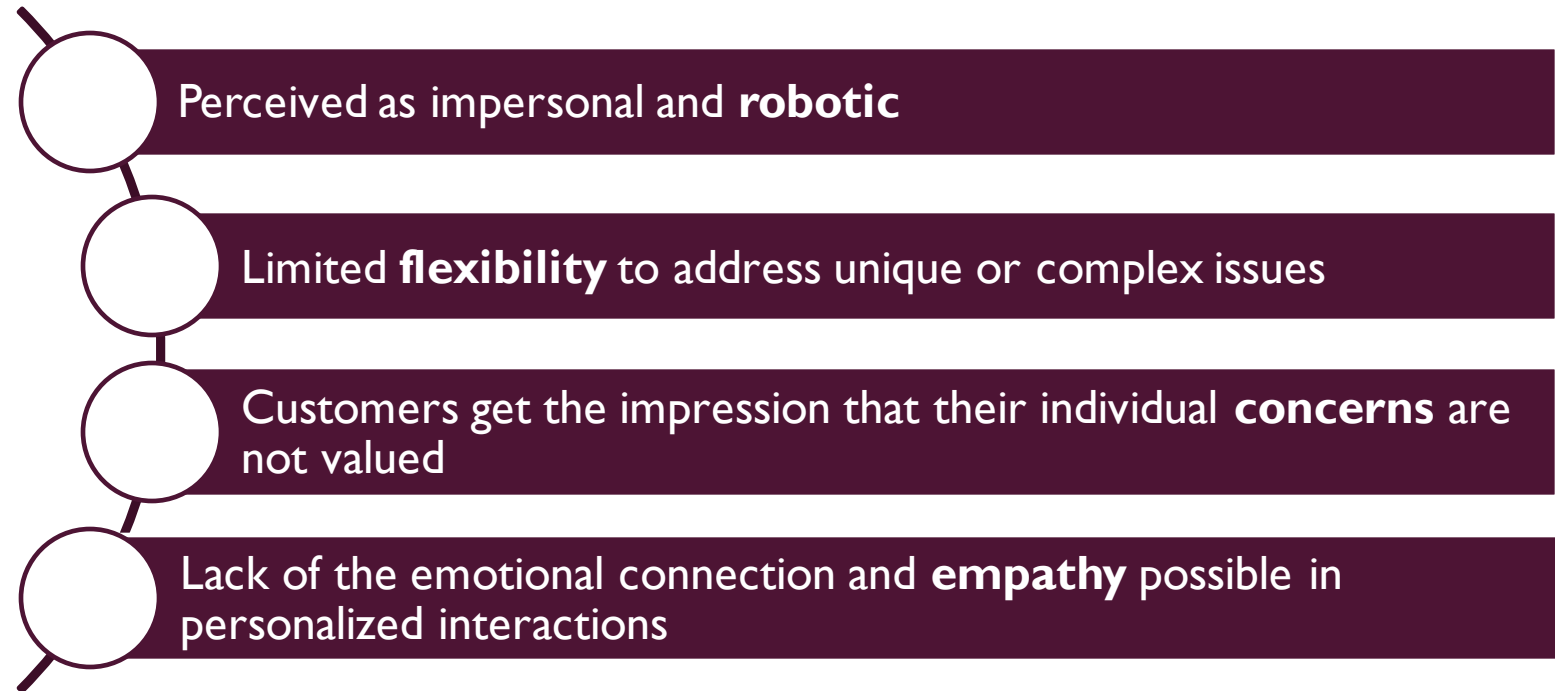


Satisfied customers are more likely to become repeat customers and advocates for the brand, potentially leading to increased **sales**.

# THEORETICAL FRAMEWORK

## EXPECTATIONS

### STANDARDIZATION **DOWNSIDES**

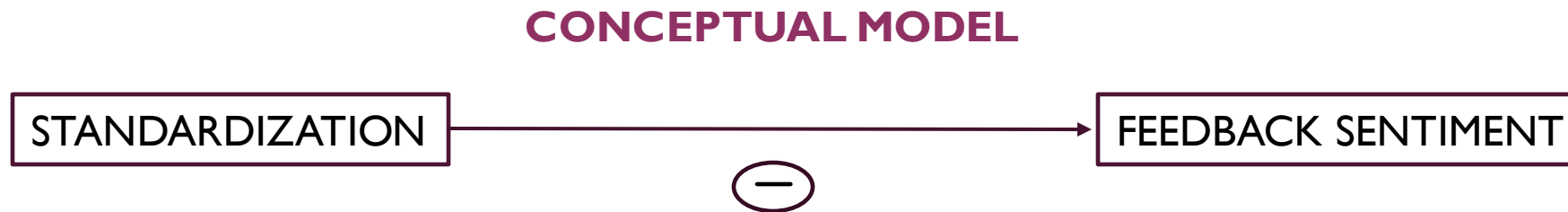


### PREVIOUS RESEARCH

- Personalization has a **positive impact** on business processes and customer satisfaction
- However, personalization might be a **cost** and **overpersonalization** may be counterproductive

# THEORETICAL FRAMEWORK

## BASE MODEL



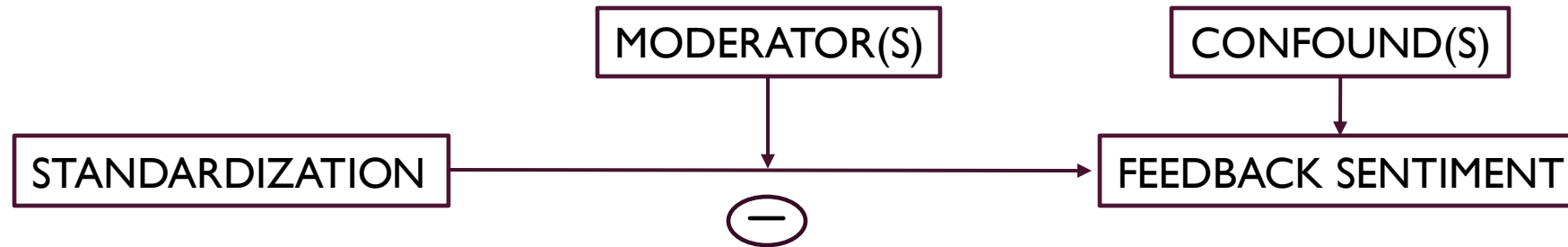
## EMPIRICAL MODEL

$$feedback\_sentiment = \beta_0 + \beta_1 stadardization$$

# THEORETICAL FRAMEWORK

## EXTENDED MODEL

### CONCEPTUAL MODEL



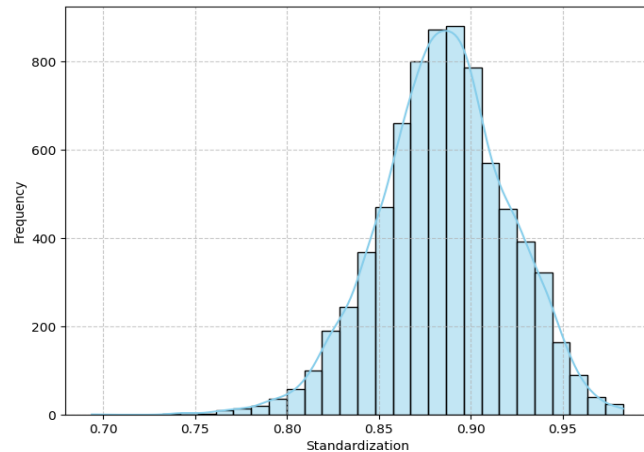
### EMPIRICAL MODEL

$$feedback\_sentiment = \beta_0 + \beta_1 stadardization + \beta_{2..K} confounders$$

$$feedback\_sentiment = \beta_0 + \beta_1 stadardization + \beta_{2..K} confounders + \beta_{K+1..M} stadardization \cdot moderators$$

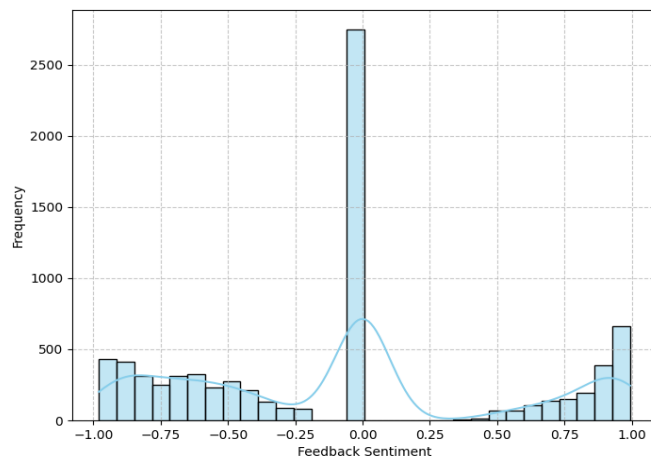
# DATA

## IV AND DV VISUALIZATION



### STANDARDIZATION

- The variable is continuous and ranges from **0 to 1**, although only values greater than 0.7 are observed.
- Standardization seems to have a distribution recalling a **normal** with mean around 0.88.



### FEEDBACK SENTIMENT

- The variable ranges from **-1 to 1**, with -1 a negative feedback sentiment a 1 a positive one. A feedback sentiment of 0 is neutral.
- The majority of the feedbacks are **neutral**, as underlined by the distribution.
- The rest of the feedback sentiments tend to be very **polarized**, with either very low or very high values.



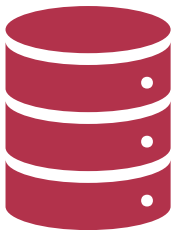
# DATA

## EDA AND PREPROCESSING



### EXPLORATORY DATA ANALYSIS

- The data contains **7,975** observations for **141** distinct companies
- The data exhibits panel data structure over the industries, as well as other dimensions
- The dataframe contains categorical variables such as: `has_dedicated_support_company`, `firm_texted_last` and `MarketCap`
- In particular we ordinally encoded the `MarketCap` variable to preserve the ordinal nature of the feature

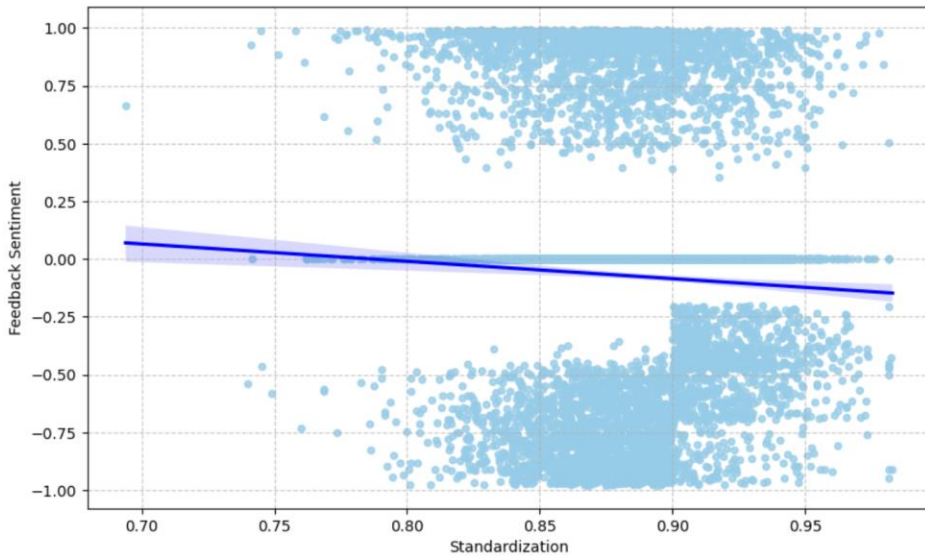


### PREPROCESSING

- There are **376 duplicates** and we have decided to **drop** them
- Some other features such as `offensive_company_count`, with almost **zero variability** across the dataset, have been **dropped** from start.
- **Categorical** columns such as `MarketCap` have been **encoded** with a ordinal encoder so that the ordinal nature of the data was not lost
- As good practice requires we also centered all variables we plan to use as moderators and our main independent variable 'standardization'

# BASE MODEL

## OLS



## RESULTS

- We run a **basic OLS** regression and use it as a base model to compare with our further analysis
- The coefficient of standardization is **negative**, which can also be inferred from the graph as the fitting line slopes downward
- Specifically, an increase in one unit of standardization is associated to a 0.75 decrease in one unit of feedback sentiment
- The regressor is **significant** at any conventional level of significance, as the p-value is virtually zero.
- The R-squared is 0.0019 so our model barely explains the variation of feedback sentiment

## CONSIDERATIONS

- The analysis cannot be considered conclusive. First, we are assuming that the **OLS assumptions** hold. Second, we cannot infer about **causality**. Third, we are not exploiting the **panel data** structure of our dataset.
- We proceed to **expand** the model

```
. regress feedback_sentiment standardization
```

Source	SS	df	MS	Number of obs	=	7,599
Model	5.33649823	1	5.33649823	F(1, 7597)	=	14.57
Residual	2781.8276	7,597	.36617449	Prob > F	=	0.0001
Total	2787.1641	7,598	.366828653	R-squared	=	0.0019
				Adj R-squared	=	0.0018
				Root MSE	=	.60512

feedback_sent~t	Coefficient	Std. err.	t	P> t	[95% conf. interval]
standardization	-.751226	.1967825	-3.82	0.000	-1.136974 - .3654779
_cons	.5908896	.1743584	3.39	0.001	.249099 .9326801

# CONFOUNDS

## CONSIDERATIONS

	feedback_s~t	standardiz~n	sentiment_~n
feedback_s~t	1.0000		
standardiz~n	-0.0438	1.0000	
sentiment_~n	0.2034	-0.0205	1.0000

### Sentiment\_score\_company\_mean

The **overall sentiment** of the company's tweets could positively influence customer feedback sentiment. At the same time, in order to keep positive tones, companies require a higher degree of personalization, i.e., less standardization.

### Response\_time\_company\_mean\_s

**Faster response** times is likely to be responsible of higher customer satisfaction, i.e., better feedback sentiment. At the same time, faster responses might require a higher degree of standardization.

	feedback_s~t	standardiz~n	response_t~s
feedback_s~t	1.0000		
standardiz~n	-0.0438	1.0000	
response_t~s	0.0014	-0.0104	1.0000

	feedback_s~t	standardiz~n	word_~y_mean
feedback_s~t	1.0000		
standardiz~n	-0.0438	1.0000	
word_~y_mean	-0.1027	0.2000	1.0000

### Word\_count\_company\_mean

Verbose responses may be perceived as less engaging and customers might prefer **concise** and to-the-point interactions. At the same time, standardized responses may include predefined scripts or templates, so the higher the word count the more likely the response to be standardized.

# CONFOUNDS

## OLS EXTENSION

```
. reg c_feedback_sentiment c_standardization ///
> word_count_company_mean response_time_company_mean_s sentiment_score_company_mean
```

Source	SS	df	MS	Number of obs	=	7,556
Model	139.601539	4	34.9003848	F(4, 7551)	=	100.03
Residual	2634.45606	7,551	.348888368	Prob > F	=	0.0000
Total	2774.0576	7,555	.367181681	R-squared	=	0.0503
				Adj R-squared	=	0.0498
				Root MSE	=	.59067

c_feedback_sentiment	Coefficient	Std. err.	t	P> t	[95% conf. interval]
c_standardization	-.3740846	.1967497	-1.90	0.057	-.7597688 .0115995
word_count_company_mean	-.004461	.0005788	-7.71	0.000	-.0055956 -.0033264
response_time_company_mean_s	.014861	.0098014	1.52	0.130	-.0043524 .0340744
sentiment_score_company_mean	.2944037	.0167449	17.58	0.000	.2615791 .3272283
_cons	.0758868	.0212824	3.57	0.000	.0341673 .1176063

## RESULTS

- The coefficient of standardization is **still negative**, but has notably decreased in absolute value
- This is evidence of the fact that the **confounders** have **absorbed** some of the variation in feedback sentiment that was attributed to standardization
- Overall, the model is significant as evidenced by the F-test of **joint significance**
- The R-squared remarkably higher with respect to our base OLS model meaning that the model explains the variation in feedback sentiment better

## CONSIDERATIONS

- The effect of standardization might depend on some other variables, so we must take into consideration the effect of **moderators** and **mediators**
- The data exhibits a panel structure, that we might exploit to capture the **latent heterogeneity** which otherwise could act as a further confounder

# MODERATORS

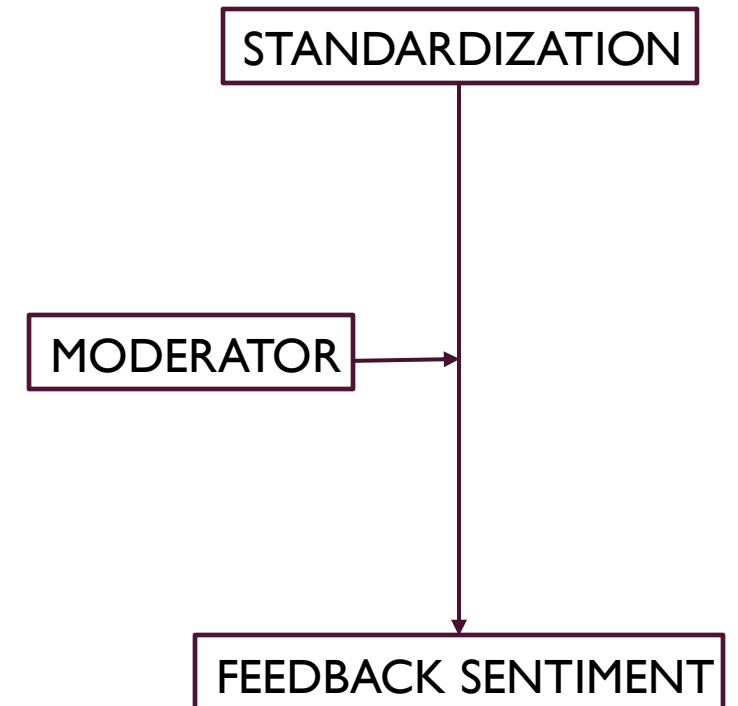
## CONSIDERATIONS

### MESSAGE UNIQUENESS

- Our hypothesis is that if a customer posts a very unique tweet, he will expect a personalized response from the company, otherwise she will be disappointed.
- As message **uniqueness increases**, we expect that more **standardized** responses have a **worse** effect on feedback.

### MESSAGE SENTIMENT

- Our hypothesis is that the effect of standardization is moderated by the emotional state of the customer at the beginning of the interaction.
- An **enthusiastic** customer will not require an immediate response, but values more a **human interaction** and acknowledgment.
- On the other hand, an **animated** customer will probably require a **timely response**, fast and standardized is the best option.



# MODERATORS

## RESULTS

- Once we account for possible moderator and keep the confounds, the overall regression is still **significant**.
- The effect of standardization on feedback is now explained by the following equation

$$0.92 - 35.5\text{message\_uniqueness} - 2.6\text{message\_sentiment}$$

- In general, the interaction with **uniqueness** **dominates** the effect on feedback.
- In fact, no matter the sentiment, answering in a **standardized** manner to a unique complaint will be **detrimental** to customer feedback

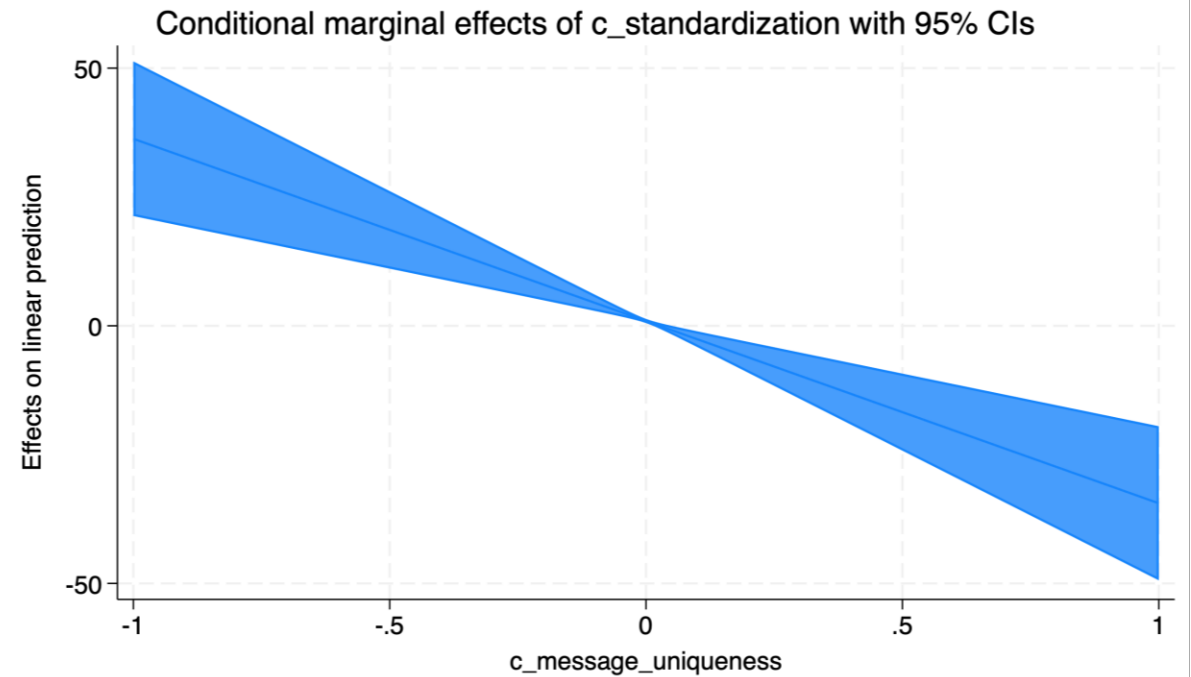
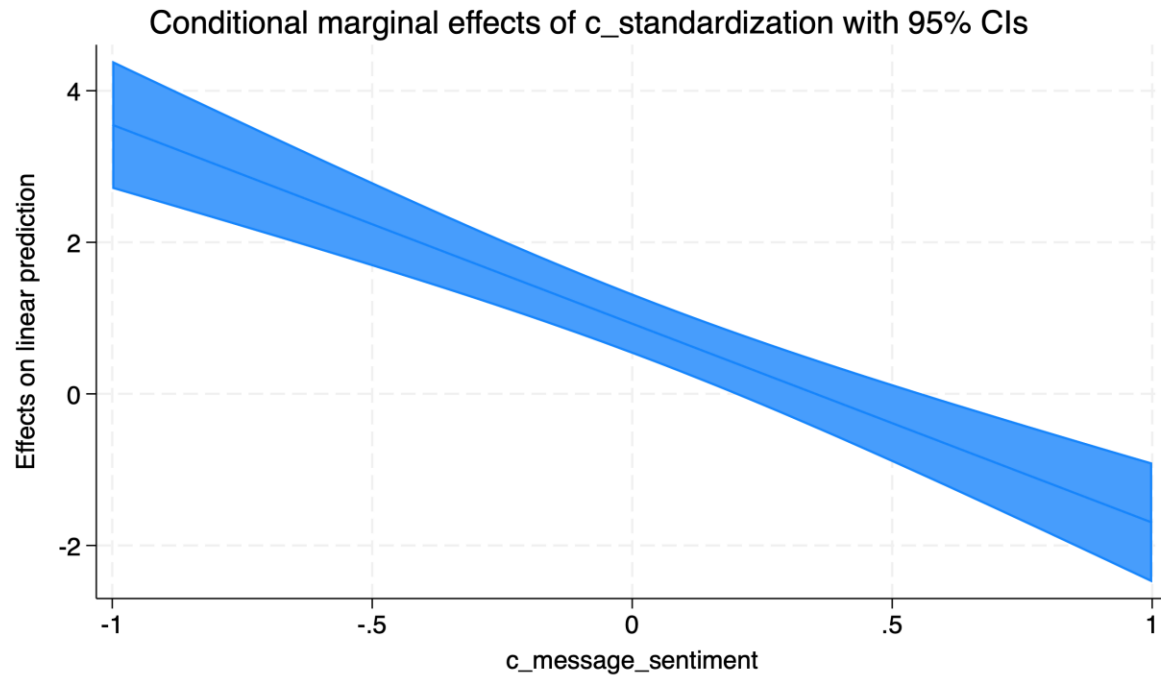
Source	SS	df	MS	Number of obs	=	7,556
Model	300.92301	8	37.6153762	F(8, 7547)	=	114.79
Residual	2473.13463	7,547	.327697712	Prob > F	=	0.0000
				R-squared	=	0.1085
				Adj R-squared	=	0.1075
Total	2774.05764	7,555	.367181687	Root MSE	=	.57245

feedback_sentiment	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
c_standardization	.9271755	.20324	4.56	0.000	.5287684	1.325582
word_count_company_mean	-.0030769	.0005669	-5.43	0.000	-.0041881	-.0019656
response_time_company_mean_s	.0100037	.0095072	1.05	0.293	-.0086331	.0286405
sentiment_score_company_mean	.2014112	.0168102	11.98	0.000	.1684584	.2343639
c_standardization	0	(omitted)				
c_message_uniqueness	.8129048	.3037155	2.68	0.007	.2175378	1.408272
c.c_standardization#c.c_message_uniqueness	-35.36919	7.634125	-4.63	0.000	-50.3342	-20.40418
c_standardization	0	(omitted)				
c_message_sentiment	.2446054	.0152064	16.09	0.000	.2147966	.2744141
c.c_standardization#c.c_message_sentiment	-2.624812	.3647896	-7.20	0.000	-3.339901	-1.909723
_cons	-.040847	.0208325	-1.96	0.050	-.0816845	-.9.47e-06

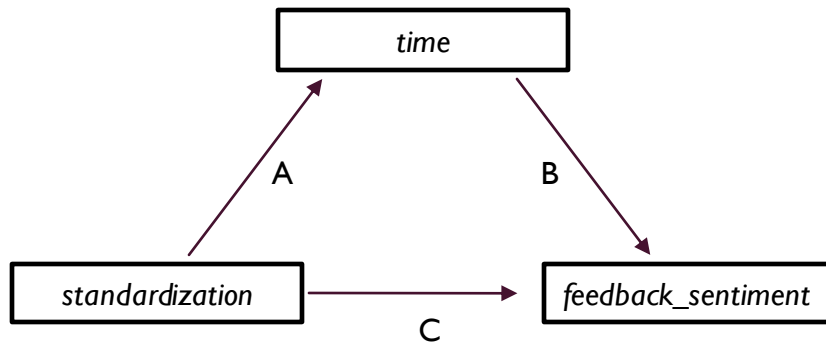
# MODERATORS

## MARGIN PLOTS



# TIME MEDIATION CONSIDERATIONS

Our theory is that **Standardization** has a different impact on final feedback depending on the initial **sentiment** of the customer, **mediated** by the **time** needed to respond such that:



Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6), 1173–1182.

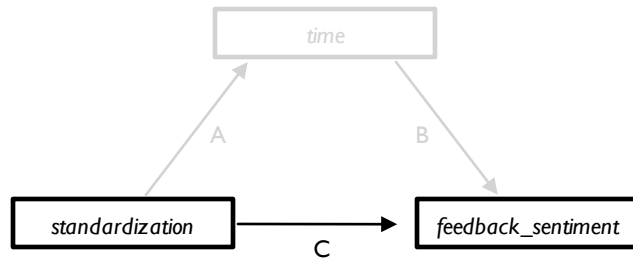
- A **happy customer** with a high message sentiment score probably does not need help, instead said customer is looking for a **human acknowledgment** of his comment. Then it is reasonable to conclude **high standardization** will have a **negative impact** on feedback if message sentiment is high. We deem this type of customer more standardization than time sensitive.
- An **upset customer** on the other hand, has the necessity of a **timely response** capable of solving the issue. Non standardized responses usually require more time. We conclude that an **untimely** non standardized response has a **worse** effect compared to a timely standardized response. We deem this type of customer more time than standardization sensitive.

To test this theory we use the Causal Steps Approach developed by Baron and Kenny.

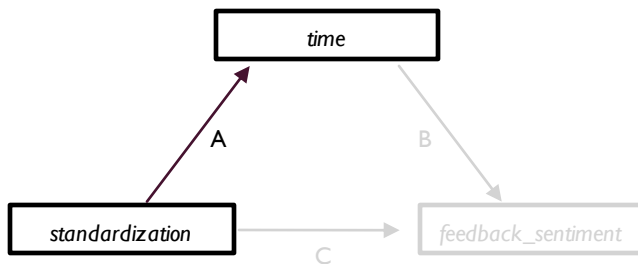


# TIME MEDIATION

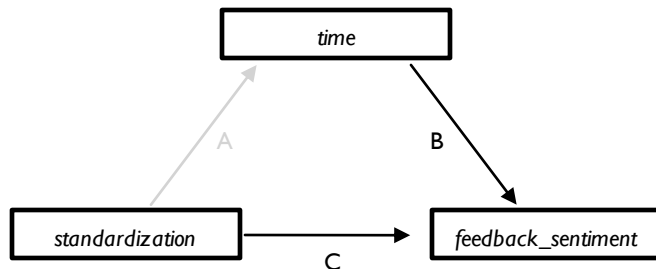
## CAUSAL STEPS



c_feedback_sentiment	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
c_standardization	-.3863685	.1959751	-1.97	0.049	-.7705338	-.0022031
word_count_company_mean	-.004315	.0005738	-7.52	0.000	-.0054399	-.0031901
sentiment_score_company_mean	.294248	.0166854	17.64	0.000	.2615401	.3269559
_cons	.0728322	.0211893	3.44	0.001	.0312953	.1143691



response_time_company_mean_s	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
c_standardization	-.3126923	.2311577	-1.35	0.176	-.7658223	.1404376
c_message_uniqueness	-.4705212	.3535017	-1.33	0.183	-1.163478	.2224353
focal_user_followers_count_s	-.0005386	.0004196	-1.28	0.199	-.0013611	.0002838
like_count_focal_user_first_s	.3838718	.3929391	0.98	0.329	-.3863923	1.154136
_cons	.1552954	.0078854	19.69	0.000	.139838	.1707528

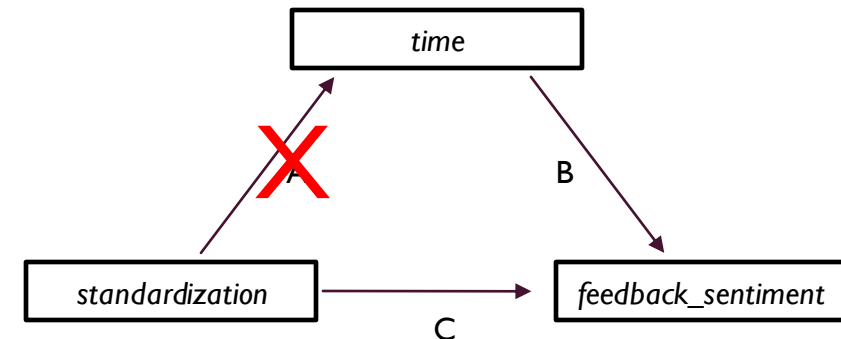


c_feedback_sentiment	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
c_standardization	-.3740846	.1967497	-1.90	0.057	-.7597688	.0115995
word_count_company_mean	-.004461	.0005788	-7.71	0.000	-.0055956	-.0033264
response_time_company_mean_s	.014861	.0098014	1.52	0.130	-.0043524	.0340744
sentiment_score_company_mean	.2944037	.0167449	17.58	0.000	.2615791	.3272283
_cons	.0758868	.0212824	3.57	0.000	.0341673	.1176063

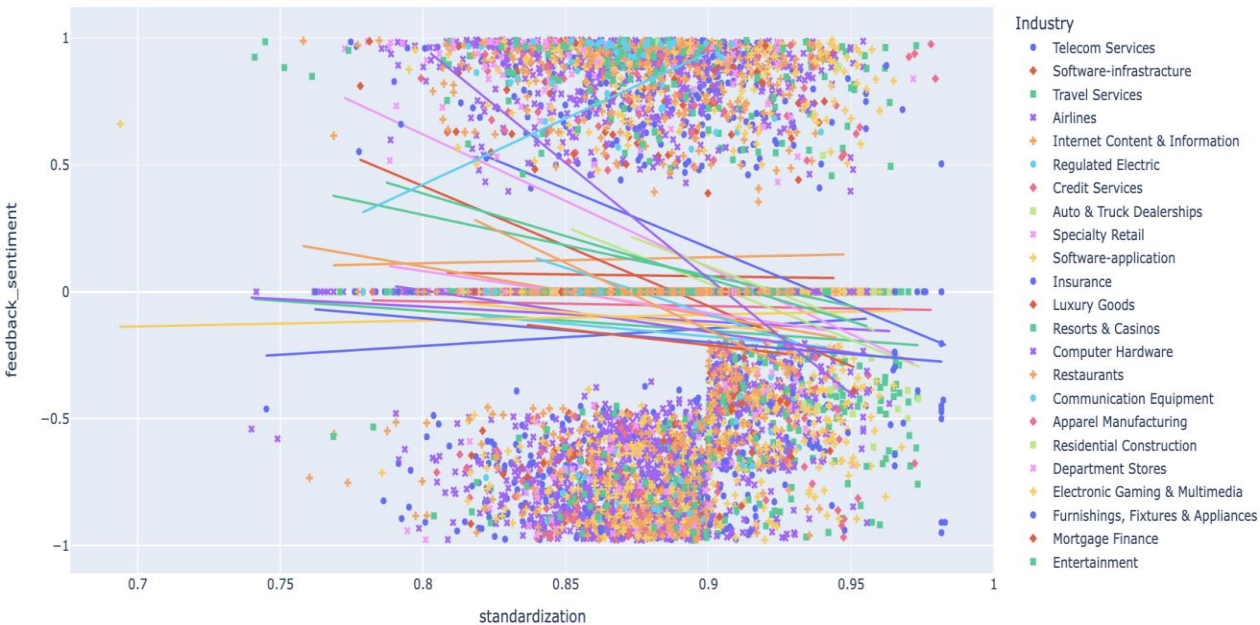
# TIME MEDIATION

## RESULTS

- The results clearly **contradict** our theory. In fact, the regression of time on standardization is **not significant**, other factors must determine the observed interaction between sentiment and standardization.
- A possible alternative explanation could be:
  - **Happy** customer only cares about **human acknowledgement** of positive tweet (as before)
  - **Angry** customer has **no point** in leaving a bad feedback to a **standardized** customer support account, which won't take the anger into consideration



# PANEL DATA VISUALIZATION



w~y_mean	overall	33.81369	12.06864	3	69
	between	6.967812	18.10569	51.35764	
	within	10.41411	.4753428	78.39682	

sentim~n	overall	.2483399	.4067576	-.9503567	.9934992
	between	.1693004	-.0519069	.8733534	
	within	.395503	-1.055161	1.091964	

- A visual representation of best fitting line of feedback sentiment on standardization, per industry
- A visual inspection suggests that there is a **downward sloping** trend shared among the majority of industries
- Only a **few** industries such as Travel Services seem to be associated to a **positive** relation between standardization and customer feedback
- Looking at our aggregated data we explore the between and the within variability, trying to understand if the fixed effect model is needed
- Some variables show a high degree of **within variability**, such as user tweet count or the mean word count of the company.
- We proceed to **implement FE** and make use of statistical tests for significance

# PANEL DATA CONSIDERATIONS

## LATENT HETEROGENEITY

- Customer **demographics**: different industries may attract distinct customer demographics, which could influence the relationship between standardization and feedback sentiment.
- **Technological** maturity: industries may differ in their technological maturity and adoption rates and potentially influence customer reaction to the firm behaviour on twitter
- Social media **engagement**: industries may have different approaches to social media engagement (e.g. younger industries make of social engagement their strength, while older industries care less)
- Product or service **characteristics**: each industry might offer products or services with unique characteristics, some of which require more personalized responses

## FIXED EFFECTS SOLUTION

- Exploit panel data structure to account for possible latent **heterogeneity** among different **industries**.
- The model does this by introducing a dummy variable for each of the industries, hence the name LSDV (least squares dummy variables)
- We can also obtain the same results by running the regression in group mean deviations

# PANEL DATA

## FIXED EFFECTS

Fixed-effects (within) regression

Group variable: industry\_id

Number of obs = 7,556

Number of groups = 26

R-squared:

Within = 0.0979

Between = 0.5323

Overall = 0.1078

Obs per group:

min = 1

avg = 290.6

max = 2,252

F(8, 7522) = 102.06

corr(u\_i, Xb) = 0.0919

Prob > F = 0.0000

feedback_sentiment	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
c_standardization	1.013023	.2096214	4.83	0.000	.6021068	1.42394
word_count_company_mean	-.0039921	.0006453	-6.19	0.000	-.0052571	-.0027271
response_time_company_mean	.0049571	.0097058	0.51	0.610	-.014069	.0239832
sentiment_score_company_mean	.1925094	.0172298	11.17	0.000	.1587341	.2262846
c_standardization	0 (omitted)					
c_message_uniqueness	.6516508	.3065604	2.13	0.034	.0507067	1.252595
c.c_standardization#						
c.c_message_uniqueness	-36.10204	7.680431	-4.70	0.000	-51.15783	-21.04625
c_standardization	0 (omitted)					
c_message_sentiment	.2223918	.0153673	14.47	0.000	.1922675	.2525161
c.c_standardization#						
c.c_message_sentiment	-2.623222	.3675278	-7.14	0.000	-3.343679	-1.902764
_cons	-.0070245	.0237442	-0.30	0.767	-.0535699	.0395208
sigma_u	.19476735					
sigma_e	.56759538					
rho	.10534419	(fraction of variance due to u_i)				

F test that all u\_i=0: F(25, 7522) = 6.18

Prob > F = 0.0000

- Our previous results remain consistent under the FE model
- The **fixed effect** model proves to be statistically **significant**: indeed, the **F-test** that jointly tests the relevance of the model dummies is associated to a p-value of zero

# PANEL DATA

## RANDOM EFFECTS

```
. hausman fe_wmods re_wmods, sigmaless
```

	—— Coefficients ——		(b-B) Difference	sqrt(diag(V_b-V_B)) Std. err.
	(b) fe_wmods	(B) re_wmods		
c_standard~n	1.013023	.9271755	.0858478	.0577241
word_~y_mean	-.0039921	-.0030769	-.0009152	.000317
response_t~s	.0049571	.0100037	-.0050466	.0023111
sentiment_~n	.1925094	.2014112	-.0089018	.0043651
c_message_~s	.6516508	.8129048	-.161254	.057391
c.				
c_standard~n#				
c.				
c_message_~s	-36.10204	-35.36919	-.7328516	1.301236
c_message_~t	.2223918	.2446054	-.0222136	.0029708
c.				
c_standard~n#				
c.				
c_message_~t	-2.623222	-2.624812	.0015906	.0652087

b = Consistent under H0 and Ha; obtained from xtreg.

B = Inconsistent under Ha, efficient under H0; obtained from xtreg.

Test of H0: Difference in coefficients not systematic

```
chi2(8) = (b-B)'[(V_b-V_B)^(-1)](b-B)
= 92.86
```

```
Prob > chi2 = 0.0000
```

- For the sake of completeness, we run a **Hausman test** to confirm the correctness of the model specification.
- In particular, we test the FE against the RE.
- The p-value is virtually zero
- We reject the null under which both the estimators are **consistent**, in favor of the alternative hypothesis where **only the FE** is.

# FEATURE AUGMENTATION

## ADDRESSING BY NAME

- When interacting with someone, it makes sense to address him/her using their **first name**, it kind of gives a **personal** and **human-like** flavour to the **conversation**
- We wanted to test whether there is evidence or not of this phenomenon also in twitter customer support conversations
- We had the **partial** texts from the first **tweet** posted by the company, thus we classified each tweet as containing a personal name (1) or not (0)
- In order to do so, we used the Stanford Named Entity Recognition (NER), [here](#) the release.

Performance benchmark [F1 score – higher is better]

	English	Spanish	German	Portuguese
list	59.1	62.3	54.9	45.2
stanford	94.9	91.1	93.6	...
stanza	82.2	98.4	87.6	...
flair	95.8	96.9	91.3	62.9
spacy	86.1	87.0	74.1	42.0
pii-tools	92.3	90.6	89.3	85.9

# FEATURE AUGMENTATION

## NLP ON TWEETS

- The classification took **longer** than what we expected, but we managed to obtain the tagged tweets.
- We then added the binary feature «**addressed\_with\_name**» to our dataset and run the regression



		✓ 802m 37.5s
...	0	[(Hey, 0), (Tracy, PERSON), (, 0), (we, 0), (...]
	1	[(Hey, 0), (Tracy, PERSON), (, 0), (we, 0), (...]
	2	[(Hey, 0), (Mike, PERSON), (!, 0), (I, 0), (wo...]
	3	[(We, 0), (dislike, 0), (spam, 0), (as, 0), (m...]
	4	[(Hello, 0), (Simon, PERSON), (, 0), (we, 0), ...]
		...



# FEATURE AUGMENTATION RESULTS

- As expected, addressing customers with their name has a positive impact on the intercept, resulting in a 0.02 upshift of the curve
- However, the confidence interval contains 0, suggesting the variable is not significant at a statistical level

feedback_sentiment	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
c_standardization	1.040512	.2104722	4.94	0.000	.6279272	1.453096
word_count_company_mean	-.0041046	.00065	-6.31	0.000	-.0053788	-.0028305
response_time_company_mean_s	.004393	.009713	0.45	0.651	-.0146471	.0234332
sentiment_score_company_mean	.1911903	.0172529	11.08	0.000	.1573699	.2250108
c_standardization	0 (omitted)					
c_message_uniqueness	.6594171	.3065858	2.15	0.032	.0584232	1.260411
c.c_standardization#c.c_message_uniqueness	-36.28611	7.680942	-4.72	0.000	-51.34291	-21.22932
c_standardization	0 (omitted)					
c_message_sentiment	.2221738	.015367	14.46	0.000	.1920502	.2522974
c.c_standardization#c.c_message_sentiment	-2.635172	.367595	-7.17	0.000	-3.355761	-1.914583
addressed_with_name	.021039	.0145969	1.44	0.150	-.007575	.0496531
_cons	-.0092803	.0237941	-0.39	0.697	-.0559233	.0373627
sigma_u	.19462199					
sigma_e	.56755474					
rho	.10521703	(fraction of variance due to u_i)				
F test that all u_i=0: F(25, 7521) = 6.16				Prob > F = 0.0000		

# CONCLUSION

## Comprehensive variables matter

- Consideration of message uniqueness, response time, sentiment score, and word count refines the impact of standardization on sentiment

## Standardization vs Sentiment

- Increased standardization leads to less positive customer sentiment, affirming the preference for human-like interactions.

## Must account for uniqueness

- This is the attribute that matters more when judging the effect of standardization

## Time doesn't mediate

- Baron and Kenny's approach rejected the theory of time mediating standardization-feedback sentiment interaction

## Panel data structure is important

- Different industries might have peculiarities that influence the relationship of standardization and customer feedback, so we must account for the panel structure.

## Strategic insight

- Balancing standardization for efficiency with personalized interactions is crucial for a company's success in customer. Personalization is an investment, not a cost.