Causal Inference II

Mixtape Session

MIXTAPE SESSIONS

Roadmap

Two Contemporary Examples

- Facebook and Mental Health
- Generative AI and Worker Productivity

Concluding remarks

Bringing them together

- Now we will look at a couple of studies using differential timing to understand both the studies (as a nice break), but also how they approached what we covered
    1. Braghieri, Levy and Makarin (2022), "Social Media and Mental Health", *American Economic Review*, 112(11): 3660-3693
    2. Brynjolfsson, Li and Raymond (2023), "Generative AI at Work", *NBER Working Paper 31161*
- I also love these papers because what I hope to do is suggest why TWFE seems problematic in the first, but weirdly, not the second

POSTED JAN 31, 2024 AT 10:32 AM EST

0 Comments (0 New)

ADI ROBERTSON

Mark Zuckerberg is trying to reset the conversation on social media's mental health effects. His opening testimony emphasizes that "the existing body of scientific work has not shown a causal link between using social media and young people having worse mental health outcomes." (Which isn't necessarily wrong, partly because causal links are very hard to prove — the overall body of research is complicated and seems to suggest social media has a variety of possible effects.)

All the news from Congress' Big Tech child safety hearing

ADI ROBERTSON JAN 31

Mental health and Social Media

- Unclear what he means; he may mean there is no experimental evidence
- Very difficult to imagine a randomized experiment being realistic or ethical
- Once the claim out there is that it is harmful, Institutional Review Boards likely wouldn't approve it
- Herein is a reason we so often focus on quasi-experimental approaches – for many policy questions, it's the only way

Reviewing the contribution

- Premiere journals are looking for important research questions, high quality data and appropriate research designs (if causal)
- My observations are about the paper's contribution
  1. Social media platforms impact on youth mental health is a major policy question and difficult to answer (see Zuckerberg testifying before Congress about it)
  2. Meticulous data collection with ingenious linkages (but which are increasingly common)
  3. Quasi-experimental research design using differential timing difference-in-differences
  4. Data visualization from event studies (appropriately specified) are compelling

Overview

- Facebook (“theFacebook”) initially targeted different American universities from 2004 to 2006 *at different points in time*
- They found an online data source that allowed them to pin point precisely when a university was “treated” with theFacebook
- They then linked that data with a longrunning health survey of college students (both before and after) in a very clever way
- Estimated the effect of a new social media platform’s presence at a university on student revealed mental health problems

Five elements of a strong DiD

1. **Bite**: Nothing. They cannot really show much here. No data on Facebook usage. They had to rely on anecdote and Facebook as a "first mover", but there had been Friendster and MySpace so this does weaken the paper maybe. More intent-to-treat
2. **Main Results**: Very strong evidence, mostly expressed using rich survey data and questions transformed into z-scores (standard deviations)
3. **Falsifications**: None. Authors do not perform falsifications. Remember Miller, Johnson and Wherry looking at Medicaid's effect on Medicare eligible population? There isn't anything like that here.
4. **Event studies**: Extremely compelling evidence and robustness across a half dozen different models
5. **Mechanism**: Speculative but let's see what you think

Main Specification: TWFE

$$Y_{icgt} = \alpha_g + \delta_t + \beta \times Facebook_{gt} + X_i \times \gamma + X_c \times \psi + \varepsilon_{icgt} \quad (1)$$

- Authors in 2022 *American Economic Review* made a TWFE specification their main model.
- Why given they will also estimate the robust methods?
- One of them told me that when they submitted, DiD literacy was much lower than it is now, and they could not take for granted people would know this material – so they present the new work as "robustness"
- Does not mean that this is true now, but it's something to remember – know your audience, anticipate their expectations and write accordingly
- Dan and Mark, what's your thoughts as editors about this?

Data on Facebook

- When does Facebook appear at a school?
  - → Facebook only publishes a fraction of that information
  - → They came up with a workaround
- The Wayback Machine has been taking almost daily photographs of every website since the Internet's beginning – including the frontpage of "TheFacebook"
- This is a very useful website for you to know about – we also used it in a recent publication of mine (Cunningham, DeAngelo and Tripp 2023)

INTERNET ARCHIVE WaybackMachine

DONATE

Explore more than 294 billion web pages saved over time

thefacebook.com

Calendar · Collections · Changes · Summary · Site Map · URLs

Saved **7,057 times** between February 12, 2004 and February 16, 2024.

2001 2002 2003 **2004** 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2

JAN

1 2 3
4 5 6 7 8 9 10
11 12 13 14 15 16 17
18 19 20 21 22 23 24
25 26 27 28 29 30 31

FEB

1 2 3 4 5 6 7
8 9 10 11 12 13 14
15 16 17 18 19 20 21
22 23 24 25 26 27 28
29

MAR

1 2 3 4 5 6
7 8 9 10 11 12 13
14 15 16 17 18 19 20
21 22 23 24 25 26 27
28 29 30 31

APR

1 2 3
4 5 6 7 8 9 10
11 12 13 14 15 16 17
18 19 20 21 22 23 24
25 26 27 28 29 30

MAY

1
2 3 4 5 6 7 8
9 10 11 12 13 14 15
16 17 18 19 20 21 22
23 24 25 26 27 28 29
30 31

JUN

1 2 3 4 5
6 7 8 9 10 11 12
13 14 15 16 17 18 19
20 21 22 23 24 25 26
27 28 29 30

JUL

1 2 3
4 5 6 7 8 9 10
11 12 13 14 15 16 17
18 19 20 21 22 23 24
25 26 27 28 29 30 31

AUG

1 2 3 4 5 6 7
8 9 10 11 12 13 14
15 16 17 18 19 20 21
22 23 24 25 26 27 28
29 30 31

SEP

1 2 3 4

OCT

1 2

NOV

1 2 3 4 5 6

DEC

1 2 3 4

Differential Timing?

- The Wayback Machine only lets you see a website at different points in time (more or less the universe though)
- But how are they going to take a website to support a difference-in-differences design?
- Look at what was on the front page

Go

[thefacebook]

login register about

Email:

Password:

login register

Welcome to Thefacebook!

[Welcome to Thefacebook]

Thefacebook is an online directory that connects people through social networks at colleges.

We have opened up Thefacebook for popular consumption at **Harvard**, **Columbia**, **Stanford**, **Yale**, **Cornell**, **Dartmouth**, **UPenn**, **MIT**, and now **BU** and **NYU**.

Your facebook is limited to your own college or university.

You can use Thefacebook to:

- Search for people at your school
- Find out who is in your classes
- Look up your friends' friends
- See a visualization of your social network

To get started, click below to register. If you have already registered, you can log in.

Register Login

about contact faq terms privacy

a Mark Zuckerberg production

[thefacebook]

login register about

Email:

Password:

login register

Welcome to Thefacebook!

[Welcome to Thefacebook]

Thefacebook is an online directory that connects people through social networks at colleges.

We have opened up Thefacebook for popular consumption at:

Amherst • BC • Berkeley • Bowdoin • Brown • BU • Bucknell
Caltech • **Carnegie Mellon** • Chicago • Columbia • Cornell • Dartmouth
Duke • Emory • Florida • **Florida State** • Georgetown • **Georgia** • **GWU** • Hamilton
Harvard • Illinois • **Johns Hopkins** • **Maryland** • Michigan
Michigan State • Middlebury • MIT • Northeastern • Northwestern • **Notre Dame**
NYU • Oberlin • Penn • Princeton • Rice • Rochester • **South Florida**
Stanford • Swarthmore • **Syracuse** • Tufts • Tulane • UCDavis
UCF • UCLA • UCSB • UCSD • UNC • USC • UVA • **Vanderbilt** • WashU
Wellesley • Wesleyan • Williams • Yale

Your facebook is limited to your own college or university.

You can use Thefacebook to:
- Search for people at your school
- Find out who is in your classes
- Look up your friends' friends
- See a visualization of your social network

To get started, click below to register. If you have already registered, you can log in.

Register Login

a Mark Zuckerberg production

Go MAR APR MAY
28
2004 2005 2006

[thefacebook]

login register about faq

Email:

Password:

login register

Welcome to Thefacebook!

[Welcome to Thefacebook]

Thefacebook is an online directory that connects people through social networks at colleges.

We have recently opened up Thefacebook at the following schools:

Abilene Christian • Agnes Scott • Albright • Allegheny
Anderson • Angelo State • Arcadia • Arts • Azusa Pacific
Beloit • Boise State • C. Arkansas • C. Missouri • Cal. Lutheran
Campbell • Canisius • Capital • Carthage • Christian Brothers
Clark Atlanta • Cleveland State • Columbus State • CSCC
CUNY City • Delta State • DeSales • Eckerd • Endicott • ETSU
Evansville • Frostburg • Guilford • Gustavus • Hampden-Sydney
Hartwick • Hendrix • Illinois Tech • IPFW • Jacksonville
John Jay • Kettering • Lake Forest • Lamar • Liberty • Lock Haven
McDaniel • Messiah • Millsaps • MSOE • Murray State • N. Georgia
N. Kentucky • NJIT • Nova • NYIT • Otterbein • Philadelphia
RIC • S. Alabama • S.E. Louisiana • Saginaw Valley • Salem State
Shepherd • St. Cloud • St. Rose • Sweet Briar • Tarleton
TN Chattanooga • TN Tech • UMass Lowell • Valencia • W. Florida
W. Oregon • Widener • WNEC • Wofford • Yeshiva

For a complete list of supported schools, click here.

Your facebook is limited to your own college or university.

You can use Thefacebook to:

- Search for people at your school
- Find out who is in your classes
- Look up your friends' friends
- See a visualization of your social network

To get started, click below to register. If you have already registered, you can log in.

Register Login

a Mark Zuckerberg production

Timing Dates

- They went through three years, day by day, of daily screenshots on Wayback machine to find when a school appeared on the front page
- The first time MIT or University of Mississippi appears on the front page, the authors mark that as the date when the school got Facebook
- But now they need information on mental health outcomes
- They find it with an old long running repeated cross section survey of college students

NCHA survey by ACHA

”Our second main data source consists of more than 430,000 responses to the NCHA survey, a survey administered to college students on a semi-annual basis by the American College Health Association (ACHA). The NCHA survey was developed in 1998 by a team of college health professionals with the purpose of obtaining information from college students about their mental and physical health. Specifically, the NCHA survey inquires about demographics, physical health, ***mental health****, alcohol and drug use, sexual behaviors, and perceptions of these behaviors among one’s peers.”*

No evidence of bite

> *The NCHA survey does not include any questions on social media use; therefore, it is not possible for us to determine whether a particular survey respondent had a Facebook account.*

This is probably the problem in any study in which your treatment is more or less the first of its kind – most likely the standard surveys have not yet incorporated the questions into their surveys

Linking Facebook data with NCHA data

In order to protect the privacy of the institutions that participate in the NCHA survey while still allowing us to carry out the analysis, the ACHA kindly agreed to provide us with a customized dataset that includes a variable indicating the semester in which Facebook was rolled out at each college. Specifically, the ACHA adopted the following procedure: (i) merge our dataset containing the Facebook introduction dates to the NCHA dataset; (ii) add a variable listing the semester in which Facebook was rolled out at each college;15 (iii) strip away any information that could allow us to identify colleges (including the specific date in which Facebook was introduced at each college).

Basic facts about early and late adopters

- Colleges in earlier Facebook expansion groups are more selective in terms of test scores, larger, more likely to be on the East Coast, and have more residential undergraduate programs than colleges in later Facebook expansion groups.
- Colleges in earlier Facebook expansion groups enroll students from relatively more advantaged economic backgrounds.
- Students in colleges that received Facebook relatively earlier have worse baseline mental health outcomes than students attending colleges in later Facebook expansion groups.

Measurement

- The survey data is very rich with a lot of questions about mental health with different scales
- They create their own combinations of these questions into aggregate indices – “index of poor mental health” where higher numbers mean worse mental health
- Each outcome survey question is normalized into what is called a “z-score” which is interpreted as a fraction of a standard deviation
- Estimates are ATT parameters – average effect of Facebook on students at schools that got Facebook

Table 1—Baseline Results: Index of Poor Mental Health

	Index of poor mental health			
	(1)	(2)	(3)	(4)
Post-Facebook introduction	0.137	0.124	0.085	0.077
	(0.040)	(0.022)	(0.033)	(0.032)
Observations	374,805	359,827	359,827	359,827
Survey-wave fixed effects	✓	✓	✓	✓
Facebook-expansion-group fixed effects	✓	✓		
Controls		✓	✓	✓
College fixed effects			✓	✓
FB-expansion-group linear time trends				✓

Notes: This table explores the effect of the introduction of Facebook at a college on student mental health. Specifically, it presents estimates of coefficient β from equation (1) with our index of poor mental health as the outcome variable. The index is standardized so that, in the preperiod, it has a mean of zero and a standard deviation of one. Column 1 estimates equation (1) without including controls; column 2 estimates equation (1) including controls; column 3, our preferred specification, replaces Facebook-expansion-group fixed effects with college fixed effects; column 4 includes linear time trends estimated at the Facebook-expansion-group level. Our controls consist of age, age squared, gender, indicators for year in school (freshman, sophomore, junior, senior), indicators for race (White, Black, Hispanic, Asian, Indian, and other), and an indicator for international student. Column 2 also includes indicators for geographic region of college (Northeast, Midwest, West, South); such indicators are omitted in columns 3 and 4 because they are collinear with the college fixed effects. For a detailed description of the outcome, treatment, and control variables, see online Appendix Table A.31. Standard errors in parentheses are clustered at the college level.

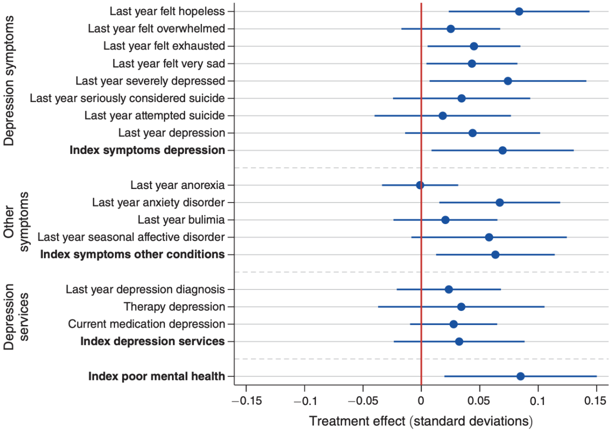


Figure 1. Effects of the Introduction of Facebook on Student Mental Health

Probably Caused the Acceptance

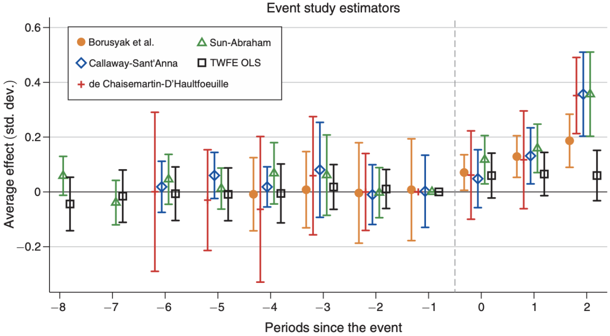


FIGURE 2. EFFECTS OF FACEBOOK ON THE INDEX OF POOR MENTAL HEALTH BASED ON DISTANCE TO/FROM FACEBOOK INTRODUCTION

Heterogeneity Analysis and "Smell Test"

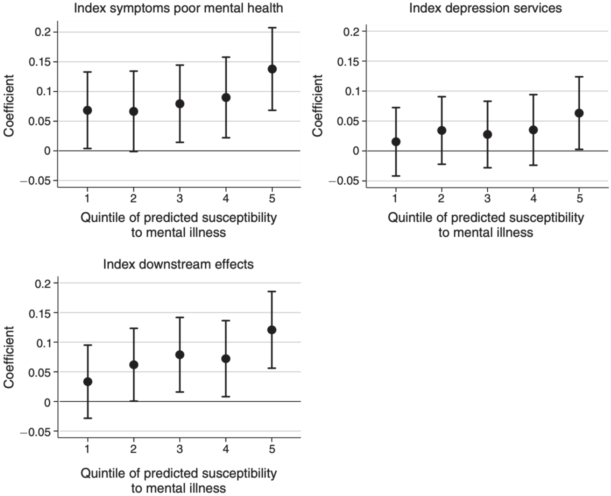


FIGURE 3. HETEROGENEOUS EFFECTS BY PREDICTED SUSCEPTIBILITY TO MENTAL ILLNESS

Exposure Worsens

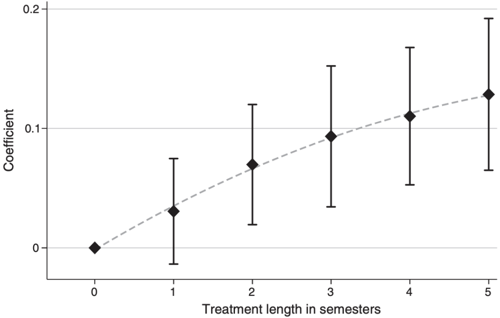


Figure 4. Effect on Poor Mental Health by Length of Exposure to Facebook

Notes: This figure explores the effects of length of exposure to Facebook on our index of poor mental health by presenting estimates of equation (4). The index is standardized so that, in the preperiod, it has a mean of zero and a standard deviation of one. The dashed curve is the quadratic curve of best fit. Our controls consist of age, age squared, gender, indicators for year in school (freshman, sophomore, junior, senior), indicators for race (White, Black, Hispanic, Asian, Indian, and other), and an indicator for international student. Students who entered college in 2006 might have been exposed to Facebook already in high school, because, starting in September 2005, college students with Facebook access could invite high school students to join the platform. Such students are excluded from the regression. For a detailed description of the outcome, treatment, and control variables, see online Appendix Table A.31. The bars represent 95 percent confidence intervals. Standard errors are clustered at the college level.

Academic Effects

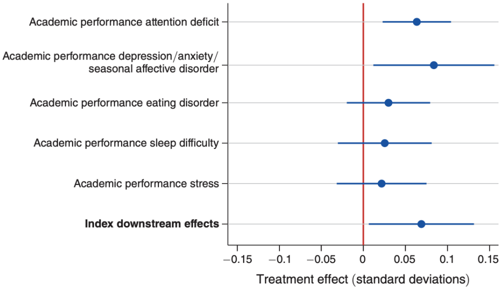


FIGURE 5. DOWNSTREAM EFFECTS ON ACADEMIC PERFORMANCE

Notes: This figure explores downstream effects of the introduction of Facebook on the students' academic performance. It presents estimates of coefficient β from equation (1) using our preferred specification, including survey-wave fixed effects, college fixed effects, and controls. The outcome variables are answers to questions inquiring as to whether various mental health conditions affected the students' academic performance and our index of downstream effects. All outcomes are standardized so that, in the preperiod, they have a mean of zero and a standard deviation of one. For a detailed description of the outcome, treatment, and control variables, see online Appendix Table A.31. The bars represent 95 percent confidence intervals. Standard errors are clustered at the college level.

Mechanisms

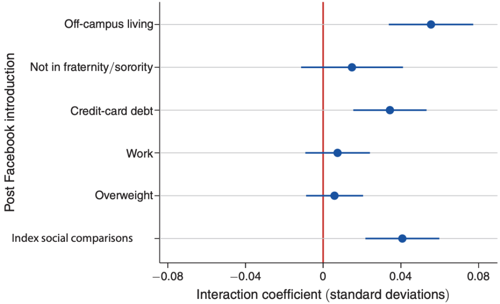


FIGURE 6. HETEROGENEOUS EFFECTS AS EVIDENCE OF UNFAVORABLE SOCIAL COMPARISONS

Notes: This figure explores the mechanisms behind the effects of Facebook on mental health. It presents estimates from a version of equation (1) in which our treatment indicator is interacted with a set of indicators for belonging to a certain subpopulation of students. The outcome variable is our overall index of poor mental health. The estimates are obtained using our preferred specification, namely the one including survey-wave fixed effects, college fixed effects, and controls. For a detailed description of the outcome, treatment, interaction, and control variables, see online Appendix Table A.31. The bars represent 95 percent confidence intervals. Standard errors are clustered at the college level.

Questions and Comments

- Mark Zuckerberg quote: “the existing body of scientific work has not shown a causal link between using social media and young people having worse mental health outcomes” – What is your reaction to his claim?
- What is your reaction to this study’s evidence?
- Which parts of this study do you think is more memorable and more convincing and why?
- How might you replicate this study yourself?

Working Paper on ChatBot

NBER WORKING PAPER SERIES

GENERATIVE AI AT WORK

Erik Brynjolfsson
Danielle Li
Lindsey R. Raymond

Working Paper 31161
http://www.nber.org/papers/w31161

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2023

Chatbots and Workers

- An unnamed firm released gradually a generative AI-based conversational assistant chatbot to its 5,179 customer support agents
- These chatbots provided assistance in handling complaints
- Very stressful job as the only time customers reached out was when they were very upset
- It isn't a randomized experiment so they're going to estimate the effect of the adoption of the chatbot using difference-in-differences

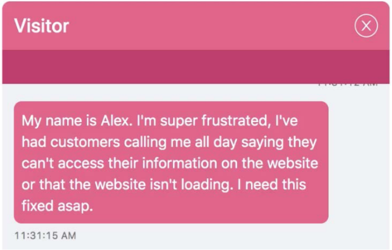
Outcomes and Pictures

- Main focus is on various measures of customer support agents handling of calls, which is the proxy for their productivity
- But they also focus on high and low skill workers (heterogeneity like before)
- Authors are going to present evidence almost entirely using event study graphs
- They also present regression tables, but the event study graphs are very powerful

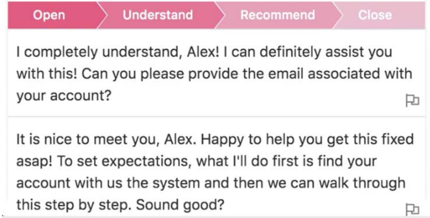
Example of ChatBot

Figure 1: Sample AI Output

A. Sample Customer Issue

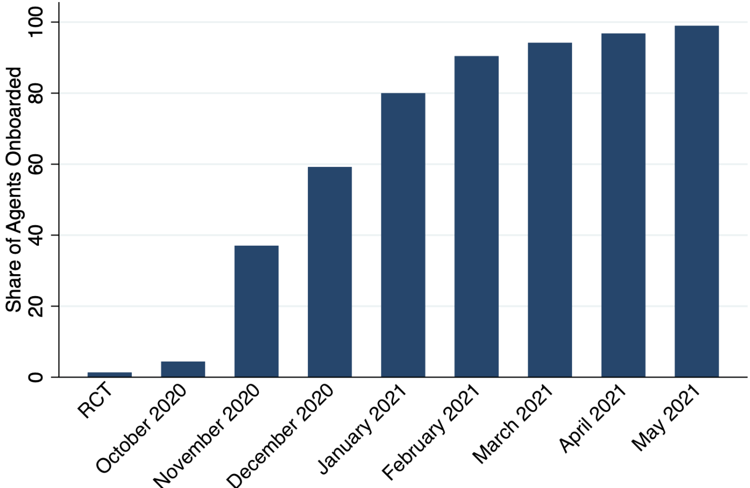


B. Sample AI-generated Suggested Response



Rollout

FIGURE 2: DEPLOYMENT TIMELINE



All the diff-in-diffs vs Sun and Abraham

Figure A.2: Event Studies, Resolutions Per Hour

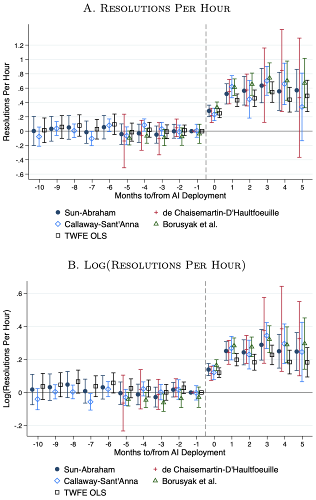
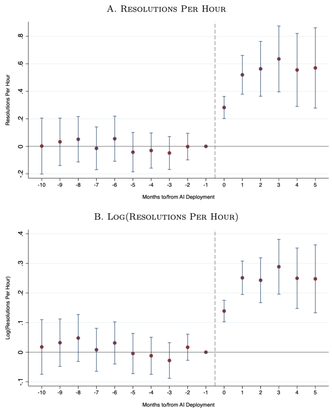


Figure 4: Event Studies, Resolutions Per Hour



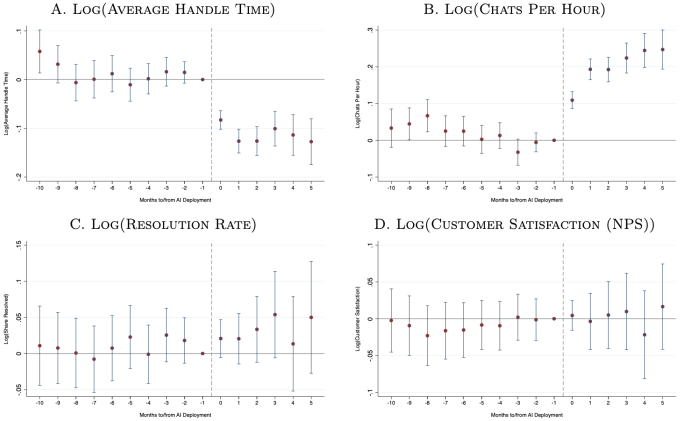
Notes: These figures plot the coefficients and 95 percent confidence interval from event study regressions of AI model deployment using the Sun and Abraham (2021) interaction weighted estimator. See text for additional details. Panel A plots the resolutions per hour and Panel B plots the natural log of the measure. All specifications include agent and chat year-month, location, agent tenure and company fixed effects. Robust standard errors are clustered at the agent level.

Comments

- They showed what is becoming standard “all the diff in diffs”
- But then they stick with Sun and Abraham for all analysis (all the diff in diffs is an appendix figure)
- They also produce a table in the appendix of their simple ATTs for everything
- Contrast this with the Facebook paper which showed TWFE *even though TWFE was biased*

Additional Outcomes

Figure 5: Event Studies, Additional Outcomes

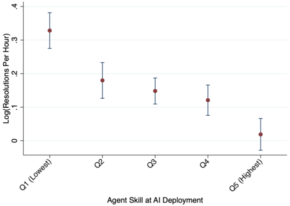


Notes: These figures plot the coefficients and 95 percent confidence interval from event study regressions of AI model deployment using the Sun and Abraham (2021) interaction weighted estimator. See text for additional details. Panel A plots the average handle time or the average duration of each technical support chat. Panel B plots the number of chats an agent completes per hour, incorporating multitasking. Panel C plots the resolution rate, the share of chats successfully resolved, and Panel D plots net promoter score, which is an average of surveyed customer satisfaction. All specifications include agent and chat year-month, location, agent tenure and company fixed effects. Robust standard errors are clustered at the agent level.

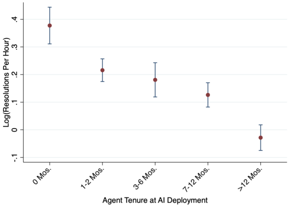
Heterogeneity by Skill

Figure 6: Heterogeneity of AI Impact, by Skill and Tenure

A. Impact of AI on Resolutions Per Hour, by Skill at Deployment

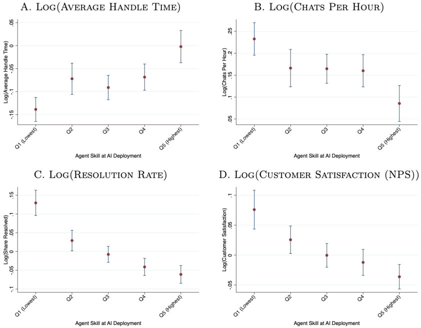


B. Impact of AI on Resolutions Per Hour, by Tenure at Deployment



Heterogeneity by Skill

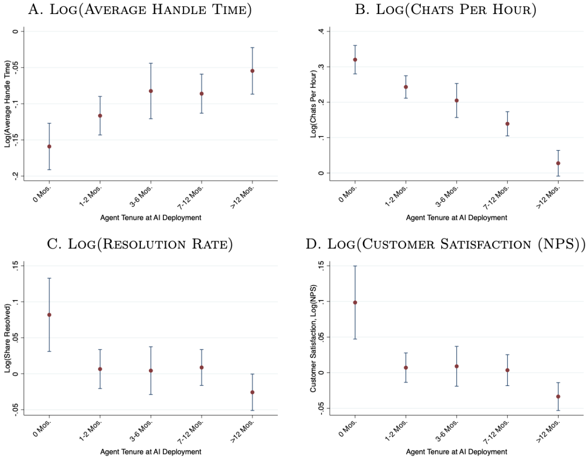
Figure 7: Heterogeneity of AI Impact by pre-AI Worker Skill, Additional Outcomes



Notes: These figures plot the impacts of AI model deployment on four measures of productivity and performance, by pre-deployment worker skill. Agent skill is calculated as the agent's trailing three month average of performance on average handle time, call resolution, and customer satisfaction, the three metrics our firm uses for agent performance. Within each month and company, agents are grouped into quintiles, with the most productive agents within each firm in quintile 5 and the least productive in quintile 1. Panel A plots the average handle time or the average duration of each technical support chat. Panel B graphs chats per hour, or the number of calls an agent can handle per hour. Panel C plots the resolution rate, and Panel D plots net promoter score, an average of surveyed customer satisfaction. All specifications include agent and chat year-month, location, and company fixed effects and standard errors are clustered at the agent level.

Heterogeneity by Worker Tenure

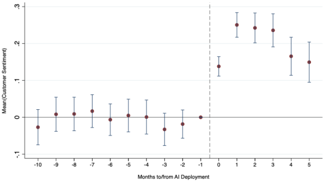
Figure 8: Heterogeneity of AI Impact by pre-AI Worker Tenure, Additional Outcomes



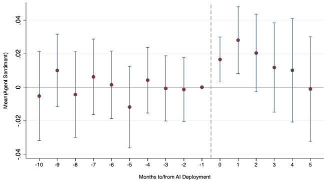
Notes: These figures plot the impacts of AI model deployment on measures of productivity and performance by pre-AI worker tenure, defined as the number of months an agent has been employed when they receive access to the AI model. Panel A plots the average handle time or the average duration of each technical support chat. Panel B

Sentiment

C. Customer Sentiment, Event Study



D. Agent Sentiment, Event Study



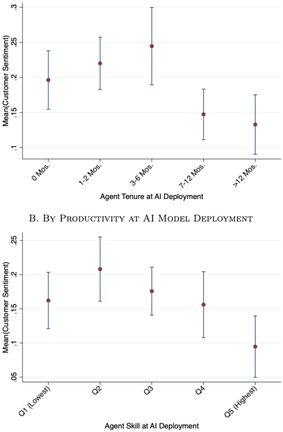
Notes: Each panel of this figure plots the impact of AI model deployment on conversational sentiment. Panel A shows average customer sentiments. Panel B shows average agent sentiments. Panel C plots the event study of AI model deployment on customer sentiment and Panel D plots the corresponding estimate for agent sentiment. Sentiment is measured using SiEBERT, a fine-tuned checkpoint of a RoBERTA, an English language transformer model. All data come from the firm's internal software systems.

Sentiment

Figure A.7: Heterogeneity in Customer Sentiment

A. By Tenure at AI Model Deployment

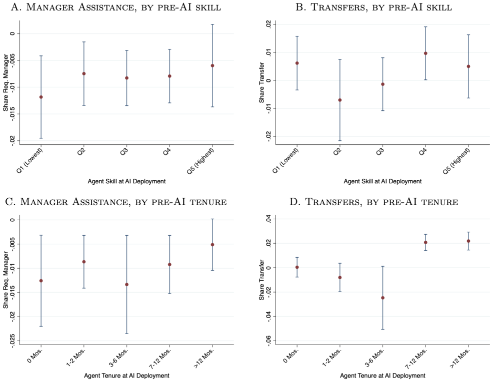
B. By Productivity at AI Model Deployment



Notes: Each panel of this figure plots the impact of AI model deployment on the mean sentiment per conversation. Sentiment refers to the emotion or attitude expressed in the text of the customer chat and ranges from −1 to 1 where −1 indicates very negative sentiment and 1 indicates very positive sentiment. Panel A plots the effects of AI model deployment on customer sentiment by agent tenure when AI deployed and Panel B plots the impacts by agent ex-ante productivity. All data come from the firm's internal software systems. Average sentiment is measured using SiEBERT, a fine-tuned checkpoint of a RoBERTA, an english language transformer model.

Manager Assistance

FIGURE A.8: ESCALATION AND TRANSFERS, HETEROGENEITY BY WORKER TENURE AND SKILL



NOTES: Panels A and C show the effects of AI on customer requests for manager assistance, by pre-AI agent skill and in by pre-AI agent tenure. Panels B and D show the impacts on transfers by pre-AI agent skill and pre-AI agent tenure. All robust standard errors are clustered at the agent location level. All data come from the firm's internal software systems.

Outcomes

- Across many dimensions, worker productivity rose
- And the productivity increases were higher for the least skilled workers – just like we had seen in the experiment
- They suggest that generative AI “reallocated experience” to the least experienced workers making them essentially appear as though they had been their awhile
- Findings suggest that it improves customer sentiment, reduces requests for managerial intervention, and improves employee retention
- Still unclear how generalizeable this is, and what impact we should see on overall aggregate employment as this was AI assisted, not AI alone

Facebook vs AI Event Study

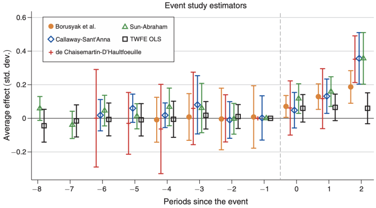
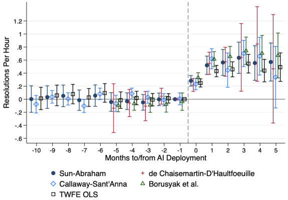


FIGURE 2. EFFECTS OF FACEBOOK ON THE INDEX OF POOR MENTAL HEALTH BASED ON DISTANCE TO/FROM FACEBOOK INTRODUCTION

FIGURE A.2: EVENT STUDIES, RESOLUTIONS PER HOUR

A. RESOLUTIONS PER HOUR



- Facebook TWFE estimate is biased downward; AI TWFE is unbiased
- Recall what TWFE needs:
    1. SUTVA (no interference)
    2. No Anticipation (baseline is untreated)
    3. Parallel Trends
    4. Homogeneous Treatment Profiles: All cohorts have the same dynamic treatment effects

Sun and Abraham decomposition

$$
\begin{aligned}
\mu_g &= \underbrace{\sum_{l \in g} \sum_{e} w^g_{e,l} (E[Y_{i,e+l} - Y^\infty_{i,0} | E_i = e] - E[Y^\infty_{i,e+l} - Y^\infty_{i,0}])}_{\text{Targets}} \\
&+ \underbrace{\sum_{g' \neq g} \sum_{l \in g'} \sum_{e} w^g_{e,l} (E[Y_{i,e+l} - Y^\infty_{i,0} | E_i = e] - E[Y^\infty_{i,e+l} - Y^\infty_{i,0}])}_{\text{Contamination from other leads and lags}} \\
&+ \underbrace{\sum_{l \in g^{excl}} \sum_{e} w^g_{e,l} (E[Y_{i,e+l} - Y^\infty_{i,0} | E_i = e] - E[Y^\infty_{i,e+l} - Y^\infty_{i,0}])}_{\text{Contamination from dropped periods}}
\end{aligned}
$$

Any relative time indicator (e.g., $g = -2$) is equal to the sum of three (weighted) types of difference-in-differences calculations

Sun and Abraham decomposition

$$
\begin{aligned}
\mu_g \quad = \quad & \underbrace{\sum_{l \in g} \sum_{e} w^g_{e,l} CATT_{e,l}}_{\text{Desirable}} \\
& + \underbrace{\sum_{g' \neq g, g' \in G} \sum_{l' \in g'} \sum_{e} w^g_{e,l'} CATT_{e,l'}}_{\text{Bias from other specified bins}} \\
& + \underbrace{\sum_{l' \in g^{excl}} \sum_{e} w^g_{e,l'} CATT_{e,l'}}_{\text{Bias from dropped relative time indicators}}
\end{aligned}
$$

Parallel trends turns each diff-in-diff into a cohort specific ATT (i.e., CATT)

Sun and Abraham decomposition

$$\begin{aligned}\mu_g \quad &= \quad \underbrace{\sum_{l \in g} \sum_e w^g_{e,l} CATT_{e,l}}_{\text{Desirable}} \\ &+ \underbrace{\sum_{g' \neq g, g' \in G} \sum_{l' \in g'} \sum_e w^g_{e,l'} CATT_{e,l'}}_{\text{Bias from other specified bins}} \\ &+ \underbrace{\sum_{l' \in g^{excl}} \sum_e w^g_{e,l'} CATT_{e,l'}}_{\text{Bias from dropped relative time indicators}}\end{aligned}$$

No Anticipation causes third row to be zero; causes all $g < 0$ to be zero too

Sun and Abraham decomposition

$$\begin{aligned} \mu_g &= \sum_{l \in g} w_l^g ATT_l \\ &+ \sum_{g' \neq g} \sum_{l' \in g'} w_{l'}^g ATT_{l'} \\ &+ \sum_{l' \in g^{excl}} w_{l'}^g ATT_{l'} \end{aligned}$$

Treatment effect homogeneity profiles cause all $CATT_{e,l}$ to be the same; therefore ATT_l and not dependent on the timing

What will cause second row to vanish?

$$\begin{aligned}
\mu_g &= \sum_{l \in g} w_l^g ATT_l \\
&\quad + \sum_{g' \neq g} \sum_{l' \in g'} w_{l'}^g ATT_{l'} \\
&\quad + \sum_{l' \in g^{excl}} w_{l'}^g ATT_{l'}
\end{aligned}$$

Second row weights sum to 0 (for each l bin) meaning some are negative. But if they're all the same for each l bin, then they cancel out. So question is why would they ever be the same?

Two ways TWFE becomes unbiased

1. Not a lot of heterogeneity. Maybe in your application there isn't a lot of treatment effect heterogeneity and so trivially there are no unique $CATT_{e,l}$
2. **Randomization of the treatment**. If the treatment rolled out to different groups *randomly*, then it implies:

$$
\begin{aligned}
E[Y^0|D=1] &= E[Y^0|D=0] = E[Y^0] \\
E[Y^1|D=1] &= E[Y^1|D=0] = E[Y^1] \\
E[\delta|D=1] &= E[\delta|D=0] = E[\delta]
\end{aligned}
$$

3. Mean covariates and potential outcomes are distributed equally for all groups under randomization, in other words, which causes treatment effect homogeneity profile, and therefore TWFE is unbiased
4. Diff-in-diff does not require this – it's more than we need – but if the rollout was random, then there are good reasons to use TWFE as it's

Facebook vs Generative AI Rollout

- We know from the paper that early adopters of Facebook had *worse* baseline mental health; this could mean that the treatment effect profiles will be different from later adopters, hence why TWFE was biased
- But is the firm targeting the best or worst customer service agents and giving them chatbots earlier than later?
- That's the question – is the selection into treatment based on the dynamic profiles? Then TWFE is likely biased, but that is not guaranteed to be the case
- And homogenous treatment effects also, regardless of the profiles, also addresses this

So why not then TWFE?

- Why not just assume randomized roll outs?
- You don’t assume randomization – it either was a literal, physical randomized roll out or it was not
- And assuming homogenous treatment profiles is to make a strong statement about some underlying production functions *about which we know very little beforehand*
- Assuming constant treatment effects is a strong assumption and unnecessary

Roadmap

Two Contemporary Examples

- Facebook and Mental Health
- Generative AI and Worker Productivity

Concluding remarks

DiD vs ATT

- We learned that difference-in-differences was just four averages and three subtractions
- But it was also a specific regression specification
- We saw that difference-in-differences could be used to estimate average treatment effects
- But the DiD equation is distinct from the ATT parameter we care about

Parallel Trends

- DiD only was equal to the ATT if the parallel trends assumption was true
- But it's not verifiable so it's a difficult assumption
- Parallel trends is not something a statistical model fixes – it's something a control group fixes
- Some comparison groups will satisfy parallel trends, but some won't
- Neither of the following test the actual parallel trends assumption, but ultimately you have to take a stand on the trend and these are tests that you were justified

Evidence for parallel trends

1. Event study graphics are placebos on the *pre-treatment period*
   - → Plot coefficients and 95% confidence intervals
   - → Check if pre-trends are zero
   - → Consider bounding if you think that's not likely ("honest diff-in-diff" by Rambachan and Roth 2024)
2. Falsifications are placebos on the *post-treatment period*
   - → Assume some confounder and find a nearly identical group that should be affected by that confounder (but not eligible for the treatment)
   - → Run your diff-in-diff on that "nearly identical" group
   - → Recall Miller, Johnson and Wherry using the 65 and older group as a falsification – shouldn't they be affected by the confounder?

Five Types of Evidence

1. Show bite – first order effects
2. Main results – What's your study about?
3. Event study graphs – This will be your main results and your evidence of parallel trends keeping in mind pre-trends and parallel trends are technically distinct
4. Falsifications – If you can find falsifications, use them
5. Mechanisms – can you find any explanation?

Imputation methods

- But what if parallel trends really isn't realistic – what then? At minimum, then diff-in-diff is not a solution
- We will review that next tomorrow when Kyle discusses imputation methods, including his own work
- Thank you!