
COLLEGIO CARLO ALBERTO
2025
2x2 = (Ẏk,t1 − Ẏk,t-1) −
− (Ẏu,t1 − Ẏut-1)

Roadmap

Background

- The origin of the robust diff-in-diff papers identifying pathologies in TWFE was Borusyak and Jaravel (2016) working paper
- Both problems with static and dynamic specifications were discussed, and the identification of the “already treated” as controls
- Paper remained in working paper until 2021 when Jan Speiss was brought on; the team developed a new estimator
- We will focus primarily on the estimator, to avoid redundancies

My Outline

1. Discussion of their interpretation of “basic” DiD assumptions
2. Critique of TWFE OLS when strong assumptions don’t hold
3. Introduction of new assumptions
4. Robust efficient imputation estimator

ATT parameter

Unlike CS and SA, BJS has a unit-level treatment effect parameter as its target. This is then aggregated to the ATT, but point is it starts lower.

$$\tau_w = \sum_{it \in \Omega_1} w_{it} \tau_{it} = w_1' \tau$$

Note the weights – they need not add up to one. Weights could be $\frac{1}{N}$ for all $it \in \Omega_1$. We have a number of options.

Standard TWFE Assumptions

- Like SA and dCDH, they both ”reverse engineer” TWFE and they ”forward engineer” a new estimator.
- First, the reverse engineering of TWFE which is unbiased under three main assumptions without covariates:
  1. Parallel trends – but it will have a fixed effects specification, not ”four averages and three subtractions”
  2. No anticipation
  3. Homogenous treatment effects
- After we discuss this, we will discuss their new estimator

TWFE Assumption 1: Parallel trends

Assumption 1: Parallel trends. There exists fixed parameters α_i and β_t in a Y^0 model:

$$Y_{it}^0 = \alpha_i + \beta_t + \varepsilon_{it}$$

with

$$E[\varepsilon_{it}] = 0$$

for all $it \in \Omega$.

TWFE Assumption 1: Parallel trends

- Only imposes restrictions on Y^0, not treatment effects themselves.
- This is the same data generating process as in our `baker.do` simulation
- It is not a parallel trends assumption expressed as "four averages and three subtractions" – this is parallel trends in *all pre-treatment periods*, not just from baseline to post
- Notice that this is a twoway fixed effects model of Y_{it}^0 – which will be important later

TWFE Assumption 2: No anticipation

- No anticipation rules out anticipatory behavior that would cause treatment effects to materialize even before the treatment occurred:

$$Y_{it} = Y_{it}(0)$$

  for all $it \in \Omega_0$.
- Notice how as an assumption, it literally imposes $\tau = 0$ for all pre-treatment periods.

TWFE Assumption 3: Homogenous treatment effects

- Up until now, these are the common diff-in-diff assumptions – parallel trends and no anticipation – but recall for unbiasedness, the canonical TWFE assumption also needs homogenous treatment effects
- Parallel trends and no anticipation are both restrictions on Y^0 – parallel trends in the future, no anticipation in the past – but homogenous treatment effects is Y^1 relative to Y^0:
- If we can assume something like homogenous treatment effects, parallel trends in all periods of your data, and NA, then you get unbiased estimate of a variance weighted ATT and smaller confidence intervals though new work suggests there are better methods (Chen, Sant'Anna and Chie 2025)

Negative weighting and violations of Assumption 3

- Goodman-Bacon (2021) like result – if you estimate TWFE, have PT and NA, and have a single dummy (”static specification”), then the TWFE coefficient has *negative weights*
- And weights may not sum to 1 either
- Hence the canonical TWFE estimator used in diff-in-diff with differential timing is not a ”properly weighted average” of treatment effects, as some of the weights are negative (and maybe flip the sign)
- Some find their illustration helpful, which I’ll show now.

Simple illustration

Table: TWFE dynamics

$E(y_{it})$	$i = A$	$i = B$
t=1	α_A	α_B
t=2	$\alpha_A + \beta_2 + \delta_{A2}$	$\alpha_B + \beta_2$
t=3	$\alpha_A + \beta_3 + \delta_{A3}$	$\alpha_B + \beta_3 + \delta_{B3}$
Event date	$E_i = 2$	$E_i = 3$

Static: $\delta = \delta_{A2} + \frac{1}{2}\delta_{B3} - \frac{1}{2}\delta_{A3}$.

Notice the negative weight on the furthest lag when dynamic and zero if constant. Similar to Bacon's negative weight on ΔATT_U.

Short-run bias of TWFE

- TWFE OLS has a severe short-run bias because the larger the effects in the long-run, the smaller the static TWFE coefficient will be
- It's caused by "forbidden comparisons" (late to early treated) in differential timing setups
- Forbidden comparisons create downward bias on long-run effects with treatment effect heterogeneity, *but not with treatment effect homogeneity*
- It's a violation of assumption 3 – heterogenous treatment effects

Modifications of general model

Modification of A1 to A1':

$$Y_{it}(0) = A'_{it}\lambda_i + X'_{it}\delta + \varepsilon_{it}$$

Assumption 4 is introduced (homoskedastic errors). This is key, because they will be building an "efficient estimator" with BLUE like OLS properties.

Using A1' to A4, we get the "efficient estimator" which is for all linear unbiased estimates of δ_W, the unique efficient estimator $\widehat{\delta^*_W}$ can be obtained with 3 steps

Role of the untreated observations

"The idea is to estimate the model of Y_{it}^0 using the untreated observations and extrapolate it to impute Y_{it}^0 for treated observations" (Borusyak, Jaravel and Speiss 2024)

"At some level, all methods for causal inference can be viewed as imputation methods, although some more explicitly than others." – Imbens and Rubin (2015)

Steps

Their solution is to estimate the Y_{it}^0 model using the control group only, similar to Heckman, Ichimura and Todd (1997) with covariates (though here it's with fixed effects)

1. Estimate expected $\widehat{Y^0}_{it}$ for all untreated observations using TWFE (i.e., $t_i < E_i$)
2. Impute $\widehat{Y^0}_{it}$ using those estimated fixed effects for the treatment group in the treated period
3. Subtract $\widehat{\delta}_{it} = Y_{it}^1 - \widehat{Y}_{it}^0$ to get individual treatment effects
4. Aggregate the estimated treatment effects into an ATT (e.g.,

$$\widehat{\delta}_W = \sum_{it} w_{it} \widehat{\delta}_{it}$$

Why is this working?

- Think back to that original statement of the PT assumption – you're modeling $Y(0)_{it}$.
- That is, without treatment – so the potential outcomes do not depend on any treatment effect
- Hence where we get treatment heterogeneity
- We obtain consistent estimates of the fixed effects which are then used to extrapolate to the counterfactual units for all $Y(0)_{it \in \Omega_1}$
- I think this is a very cool trick personally, and as it is still OLS, it's computationally fast and flexible to unit-trends, triple diff, covariates and so forth (though remember what we said about covariates)

Comparisons to other estimators

Table 3: Efficiency and Bias of Alternative Estimators

Horizon	Estimator	Baseline simulation		More pre-periods	Heterosk. residuals	AR(1) residuals	Anticipation effects
		Variance (1)	Coverage (2)	Variance (3)	Variance (4)	Variance (5)	Bias (6)
$h = 0$	Imputation	0.0099	0.942	0.0080	0.0347	0.0072	-0.0569
	DCDH	0.0140	0.938	0.0140	0.0526	0.0070	-0.0915
	SA	0.0115	0.938	0.0115	0.0404	0.0066	-0.0753
$h = 1$	Imputation	0.0145	0.936	0.0111	0.0532	0.0143	-0.0719
	DCDH	0.0185	0.948	0.0185	0.0703	0.0151	-0.0972
	SA	0.0177	0.948	0.0177	0.0643	0.0165	-0.0812
$h = 2$	Imputation	0.0222	0.956	0.0161	0.0813	0.0240	-0.0886
	DCDH	0.0262	0.958	0.0262	0.0952	0.0257	-0.1020
	SA	0.0317	0.950	0.0317	0.1108	0.0341	-0.0850
$h = 3$	Imputation	0.0366	0.928	0.0255	0.1379	0.0394	-0.1101
	DCDH	0.0422	0.930	0.0422	0.1488	0.0446	-0.1087
	SA	0.0479	0.952	0.0479	0.1659	0.0543	-0.0932
$h = 4$	Imputation	0.0800	0.942	0.0546	0.3197	0.0773	-0.1487
	DCDH	0.0932	0.950	0.0932	0.3263	0.0903	-0.1265
	SA	0.0932	0.954	0.0932	0.3263	0.0903	-0.1265

Notes: See Section 4.6 for a detailed description of the data-generating processes and reported statistics.

Returning to the minimum wage

- Clemens and Strain (2021) implemented the BJS imputation estimator to estimate the effect of the minimum wage (post Great Recession) on employment
- One comment abt the following graphics: BJS procedure does not have a “base” period in the same sense as the regression models do because it is not contrasting each period relative to some omitted group
- Since it is imputing counterfactuals, we can calculate each period’s effect

BJS Results

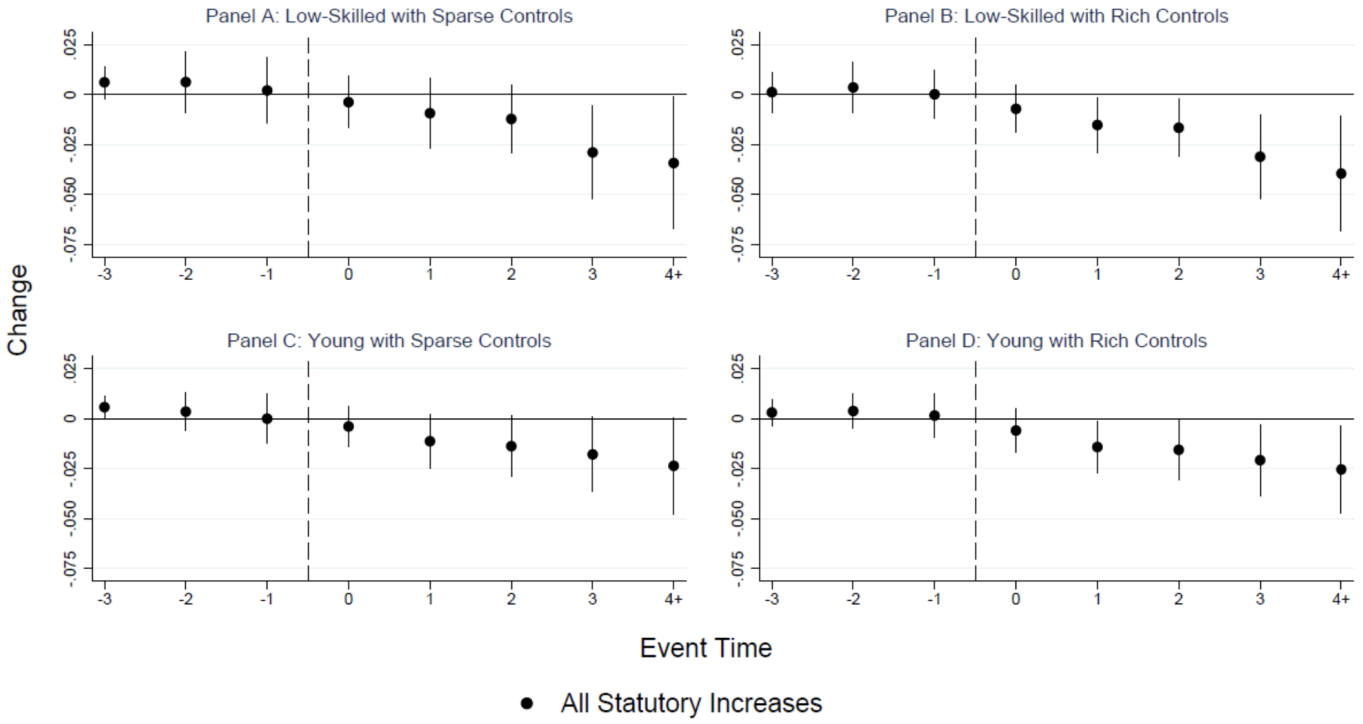


Figure 11. Event Studies of Changes in Employment Following Statutory Minimum Wage Increases Using the BJS Imputation Estimator: This figure displays coefficients obtained using the imputation estimator proposed by Borusyak, Jaravel and Spiess (2021) (BJS). For the BJS estimator, we code the first treatment year as the year in which a state's first statutory minimum wage increase took effect. Note that this appears graphically as "year 0" in the BJS figures, but corresponds with year 1 in the stacked event study figures. Panels A and B plot coefficients for low-skilled individuals defined as individuals ages 16–25 without a completed high school education. Panels C and D plot coefficients for young individuals defined as all individuals ages 16–21. The samples are from the ACS. Regressions with "sparse controls" include state and year fixed effects, as well as the log of annual average *per capita* income and the annual average state house price index used in our main regressions. Regressions with "rich controls" include all controls in the base controls regressions plus the three-year lag of log *per capita* income and the house price index, as well as a dummy variable for each education group and age. Error bars denote 95 percent confidence intervals around each estimated coefficient. Standard errors are clustered by state.

BJS Results

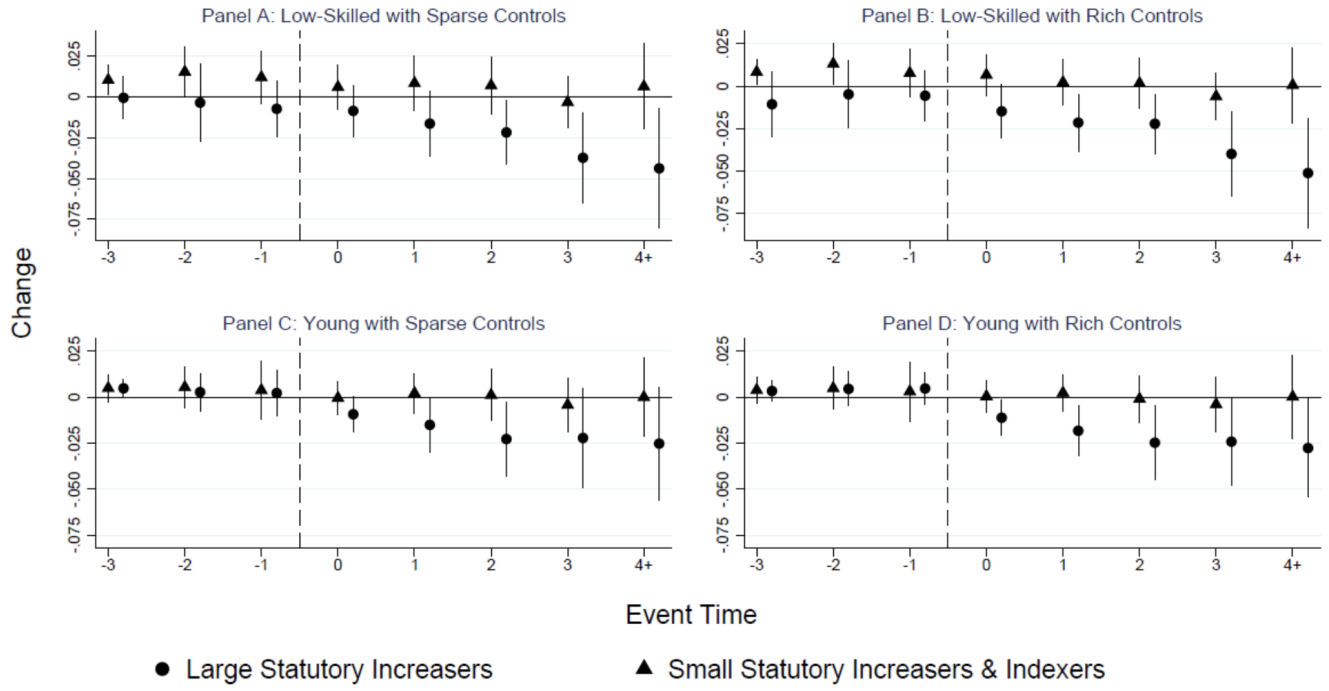


Figure 12. Event Studies of Changes in Employment Following Large and Small Statutory Minimum Wage Increases Using the BJS Imputation Estimator: This figure displays coefficients obtained using the imputation estimator proposed by Borusyak, Jaravel and Spiess (2021) (BJS). For the BJS estimator, we code the first treatment year as the year in which a state's first statutory minimum wage increase took effect. Note that this appears graphically as "year 0" in the BJS figures but corresponds with year 1 in the stacked event study figures. We compare estimates for large vs. small increases as defined in the main text. Panels A and B plot coefficients for low-skilled individuals defined as individuals ages 16–25 without a completed high school education. Panels C and D plot coefficients for young individuals defined as all individuals ages 16–21. The samples are from the ACS. Regressions with "sparse controls" include state and year fixed effects, as well as the log of annual average *per capita* income and the annual average state house price index used in our main regressions. Regressions with "rich controls" include all controls in the base controls regressions plus the three-year lag of log *per capita* income and the house price index, as well as a dummy variable for each education group and age. Error bars denote 95 percent confidence intervals around each estimated coefficient. Standard errors are clustered by state.

Comments abt the minimum wage study

- Elasticity of employment with respect to minimum wage is -0.124 and -0.082 for those without high school and the young, respectively
- Differences by size of minimum wage increase:
  - → Large increases (around $2.90): own-wage elasticity is -1.01 for 16-25yo with less than HS and -0.41 for 16 to 21yo (large effects)
  - → Small increases (around $1.90): own-wage elasticity is 0.46 (i.e., no employment effects)
  - → Inflation-index increases (around $0.90): own-wage elasticity is 0.16 (no effect) and -0.17 (no effect)

Two Stage DiD

> *“It seems natural that TWFE should identify the ATT” – Gardner (2021)*

It just seems like TWFE with a DiD will estimate the ATT with weights that we’ll find intuitive. Was this just a conjecture and was never true? Why isn’t this working?

Two Stage DiD

- Why does TWFE fail under differential timing? Violates strict exogeneity under heterogeneity
- The logic of the failure suggests an obvious, but previously unknown, solution which is the 2SDiD
- I'll explain 2SDiD, focus on the parallel trends implications, and show we can get a consistent and unbiased estimate of group and relative time fixed effects
- If you can get consistent and unbiased estimates of group and relative time fixed effects, then you can delete them and run normal analysis

Background

- By now, we all agree that TWFE just doesn't handle heterogeneity under differential timing very well
- We've seen in the Goodman-Bacon decomposition why – it's caused by TWFE implicitly calculating late to early 2x2s, which are a source of bias
- But some of you are coming straight from a panel econometrics course that maybe didn't use potential outcomes notation
- Isn't strict exogeneity enough for consistent estimates? What then does strict exogeneity have to do with heterogeneity and differential timing?

High level discussion

- TWFE identifies the ATT when the heterogeneous effects are distributed equally across all groups and periods, but since that is a knife-edge situation, it is likely that TWFE will not in our applications meet this special scenario
- In the two group case, that is what happens though which is why TWFE worked fine there
- Metaphorically, the two group case that we always used to pin our intuition of what DiD was doing was the exception not the rule
- Goodman-Bacon (2021) shows the problem is caused by late-to-early comparisons; Gardner (2021) will show that the problem is misspecification
- Think of these as different perspectives on the same problem

Model misspecification

> *"Misspecified DiD regression models project heterogenous treatment effects onto group and period fixed effects rather than the treatment status itself"*

Spoiler: This analysis of the problem suggests solution – why don't we remove those?

2SDiD

- First stage – estimate the group and relative time fixed effects using only the $D = 0$ observations
- Second stage – using predicted values based off those fixed effect coefficients, run your model off the transformed outcome
- Get the standard errors right just like 2SLS by taking the first stage into account (uses GMM)

More high level

- The second step recovers the average difference in outcomes between treated and untreated units after removing group and period fixed effects
- Strong parallel trends assumption compared to CS and SA, but unclear if this is a big deal in general

Notation

i: panel units
t: calendar time – think of real dates
$g \in \{0, 1, \ldots, G\}$ – groups
$p \in \{0, 1, \ldots, P\}$ – relative time or “periods”

Periods are successive. Group 0 – never treated. Group 1 – treated in period 1, 2, and on. Group 2 – treated in period 2, etc.

Parameters

$$\beta_{gp} = E\left[Y^1_{gpit} - Y^0_{gpit} | g, p\right]$$

It's a group-time ATT but expressed in a more traditional econometric notation that you could easily find in Wooldridge or some such

Modeling basics

Under parallel trends, mean outcomes will satisfy the following equation

$$E\left[Y_{gpit}|g,p,D_{gp}\right] = \lambda_g + \gamma_p + \beta_{gp}D_{gp}$$

In two-group, group and period effects are eliminated with dummies because TWFE uses dummies to demean across multiple dimensions. Then TWFE identifies ATT. But this does not hold when average effects vary across group and period. There are many ways to express a treatment effect's across group and time, but Gardner presented it as a weighted average of the coefficients for only that group-period situation:

$$E\left(\beta_{gp}|D_{gp}=1\right) = E\left(Y^1_{gpit} - Y^0_{gpit}|D_{gp}=1\right)$$

Strict exogeneity violation

Rewriting the above we get:

$$
\begin{aligned}
E\left[Y_{gpit}|g,p,D_{gp}\right] &= \lambda_g + \gamma_p + E\left[\beta_{gp}|D_{gp}=1\right]D_{gp} \\
&\quad \left[\beta_{gp} - E(\beta_{gp}|D_{gp}=1)\right]D_{gp}
\end{aligned}
$$

The problem is there's this weird new error term and it isn't mean zero under heterogenous treatment effects spread across group and period. Unlike the two group case, the coefficient on D_{gp} from TWFE doesn't identify the average $E(\beta_{gp}|D_{gp}=1)$

So let's see Gardner's solution, but note – his solution was suggested by the problem itself. Gardner is thoughtful and observant.

DiD regression estimand

- So if TWFE isn't recovering $E(\beta_{gp}|D_{gp}=1)$, then what is it recovering?
- He shows that under PT, the coefficient on D_{gp} is:

$$\beta^* = \sum_{g=1}^{G}\sum_{p=g}^{P} w_{gp}\beta_{gp}$$

- So then – what are the weights w_{gp}? They are variance weights

Estimation

$$Y_{gpit} = \lambda_g + \gamma_p + \beta D_{gp} + \varepsilon_{gpit}$$

This specification assumes a conditional expectation function that is linear in group, period and treatment status. But when the model is misspecified, it will attribute some of the heterogeneity impacts of the treatment to group and period fixed effects. The longer the treatment, the greater $\overline{D}$ is, the more that group's treatment effects will be absorbed by group fixed effects. When misspecified, TWFE doesn't recover $E[\beta|D = 1]$.

Statistical issues

- Common support: “as long as there are untreated and treated observations for each group and period, λ_g and γ_p are identified from the subpopulation of untreated groups and periods.”
- Identification: “the overall group $\times$ period ATT is identified from a comparison of mean outcomes between treated and untreated groups after removing group and period effects.”

Estimation: First stage

First stage:

$$Y_{gpit} = \lambda_g + \gamma_p + \varepsilon_{gpit}$$

using only $D_{gp} = 0$, retaining the fixed effects. Collect the $\widehat{\lambda_g}$ and $\widehat{\gamma_p}$.

Estimation: Second stage

Second stage:

$$\begin{aligned} \widehat{y}_{gpit} &= y_{gpit} - \widehat{\lambda_g} - \widehat{\gamma_p} \\ \widehat{y}_{gpit} &= \alpha + \beta D_{gp} + \psi_{gpit} \end{aligned}$$

Why does this work? Parallel trends assumption implies:

$$E(y_{gpit}|g,p,D_{gp}) - \lambda_g - \gamma_p = E\Big[\beta_{gp}|D_{gp}=1\Big]D_{gp} + \Big[\beta_{gp} - E(\beta_{gp}|D_{gp}=1)\Big]D_{gp}$$

But because

$$E\Big\{[\beta_{gp} - E(\beta_{gp}|D_{gp}=1)]D_{gp}|D_{gp}\Big\} = 0$$

Estimand

Then this procedure will identify $E(\beta_{gp}|D_{gp}=1)$. Consistency and unbiasedness proofs.

This is $E(\beta_{gp}|D_{gp}=1) = \sum^{G}\sum^{P}\beta_{gp}P(g,p|D_{gp}=1)$. It will tend to put more weight, by definition, on groups earlier into their treatment. But this isn't the same as the negative weighting that BJS say occurs oof the long lags. It just means there are more of them.

Event studies are:

$$y_{gpit} = \lambda_g + \gamma_p + \sum_{r=-R}^{P}\beta_r D_{rgp} + \varepsilon_{gpit}$$

Just change the second stage with the transformed outcome.

Inference

- Standard errors are wrong on the second stage because the dependent variable uses estimates obtained from the first stage.
- The asymptotic distribution of the second stage can be obtained by interpreting the two-stage procedure as a joint GMM

Roadmap

Bringing them together

- An advanced area is when areas adopt policies at different points in time – called differential timing
- Very popular, but methodologically somewhat more complex than the one we've reviewed
- I'm going to just walk you through their findings, and if you're interested in learning more about this, then you can attend my workshop for $1 in a few weeks!

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 – especially once the claim out there is that it is harmful, Institutional Review Boards likely wouldn't approve it
- Quasi-experimental evidence can step in to answer important questions like this
- Braghieri, Levy and Makarin (2022), "Social Media and Mental Health", *American Economic Review*, 112(11): 3660-3693

Overview of design and data

- Authors take advantage of a clever quirk in Facebook (then “theFacebook”) targeted different universities from 2004 to 2006
- 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

DID in Court

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
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**: Very weak in my opinion

DiD in Court

- So in many ways the strength of the project lies in a few areas:
  1. Important question – social media and youth mental health problems is a major policy question (see Zuckerberg testifying before Congress about it)
  2. Excellent research design – difference-in-differences
  3. Meticulous data collection
  4. Data visualization is compelling
- And it publishes in the premiere journal in economics, which I think shows that the research question and high quality data combined with research design can lift a paper

TWFE

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

This is a version of the regression model we looked at called ”twoway fixed effects”. Somewhat complicated to dig into, so I will just say that they use it plus some other methods that are appropriate when you have several difference-in-differences events. But the focus is on β

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"
- Guess what was on the front page of TheFacebook …

DONATE

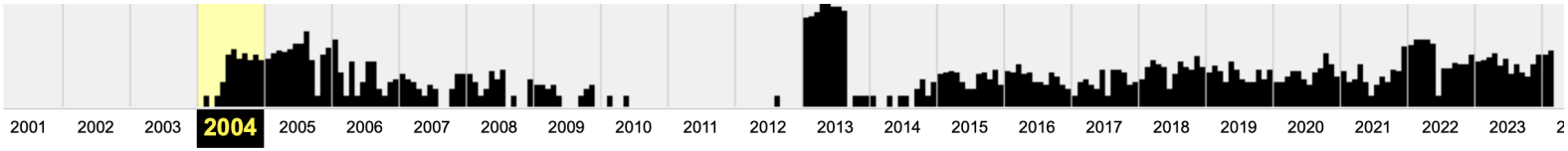
INTERNET ARCHIVE
WayBackMachine

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.



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**	

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.

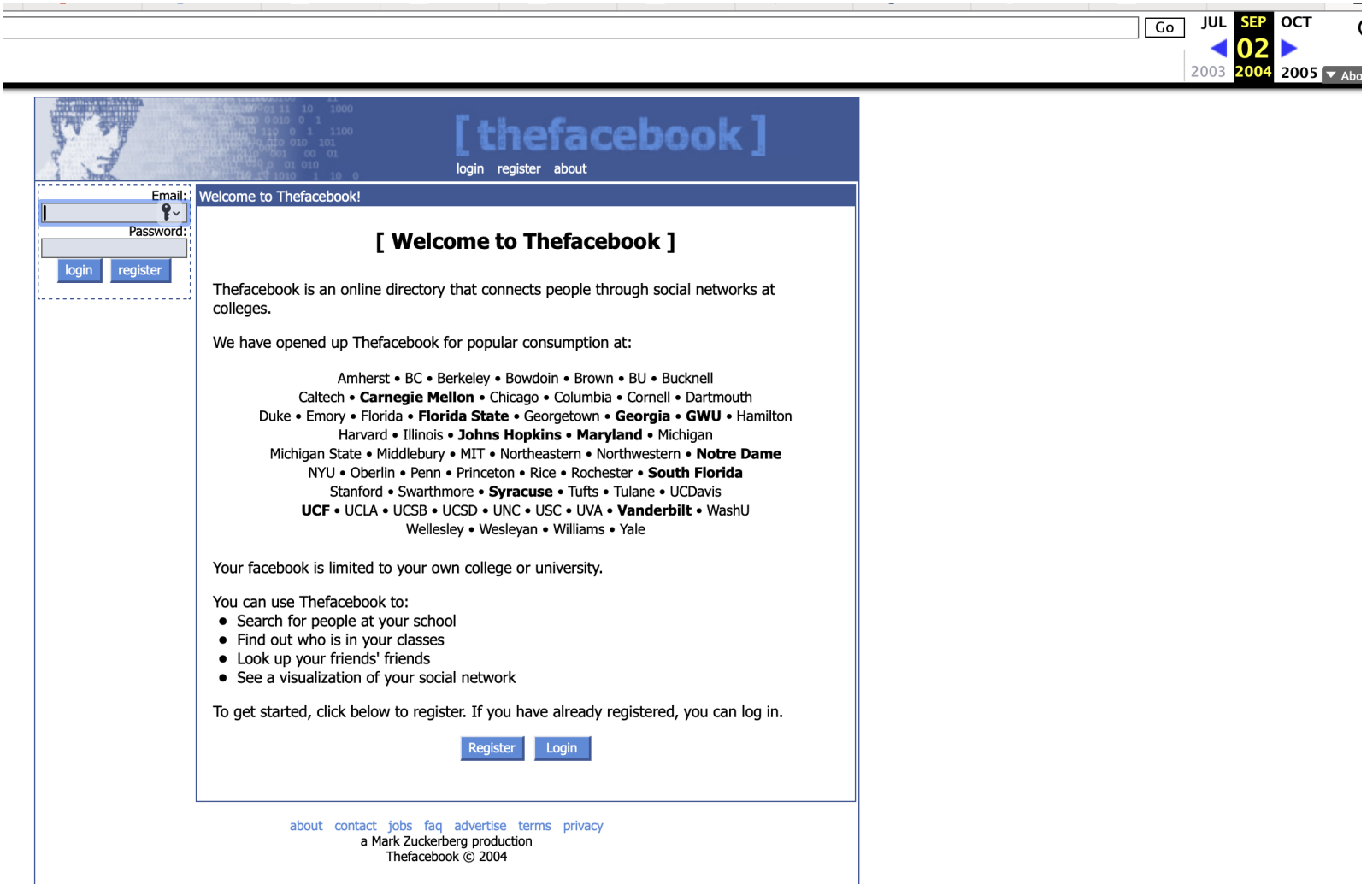
You can use Thefacebook to:

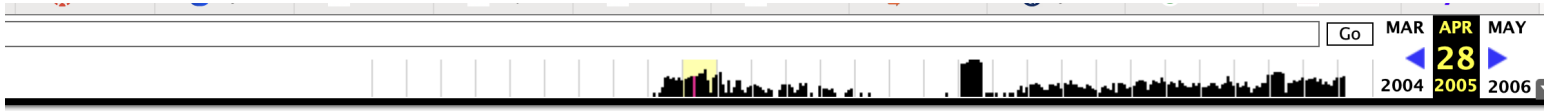
- Search for people at your school
- Find out who is in your classes
- Look up your friends' friends
- See a visualization of your social network

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 © 2004


Go
JUL SEP OCT
02
2003 2004 2005
[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
about contact jobs faq advertise terms privacy
a Mark Zuckerberg production
Thefacebook © 2004



[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

Timing Dates

- They went through three years of daily screenshots on Wayback machine to find when a school appeared on the front page
- The first time Agnes Scott, or Covenant, 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.

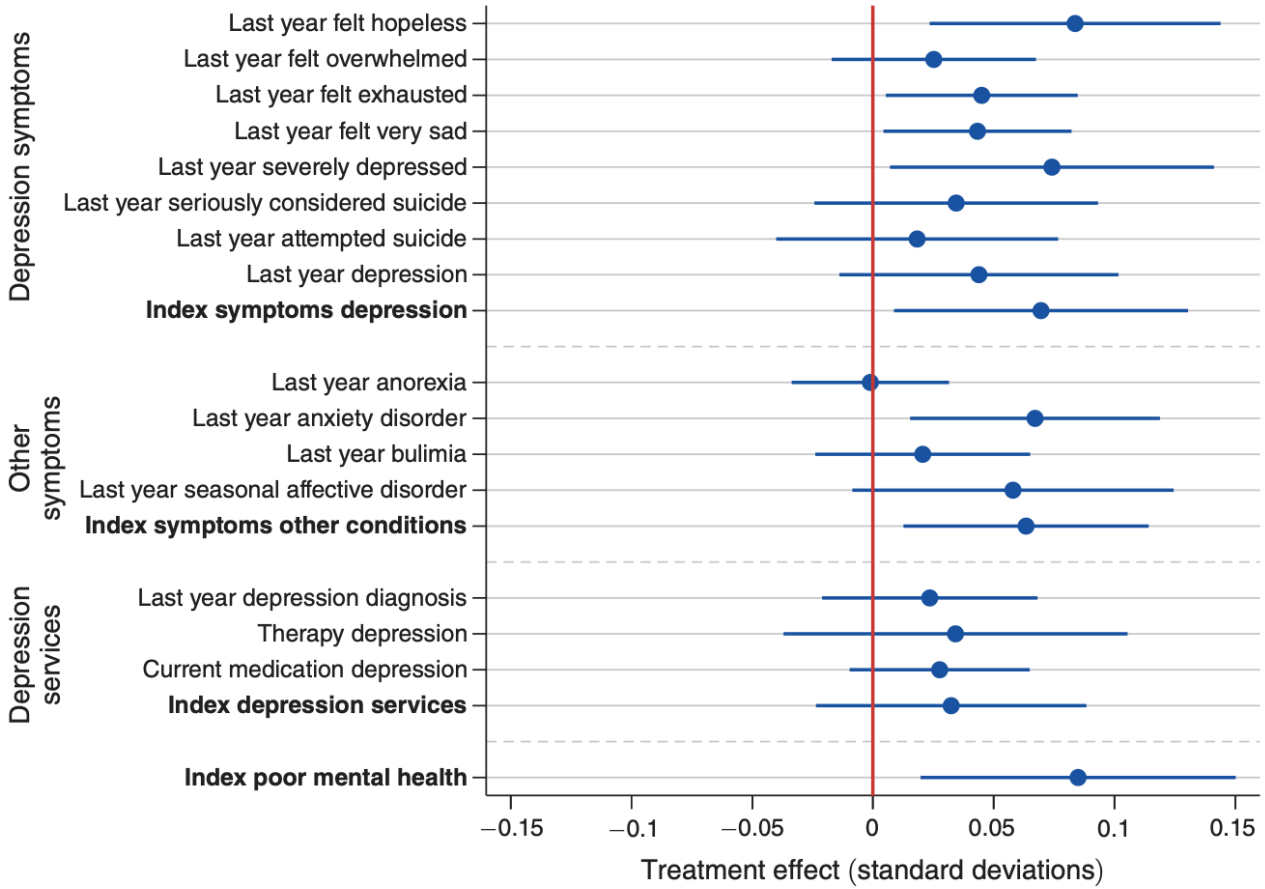


FIGURE 1. EFFECTS OF THE INTRODUCTION OF FACEBOOK ON STUDENT MENTAL HEALTH

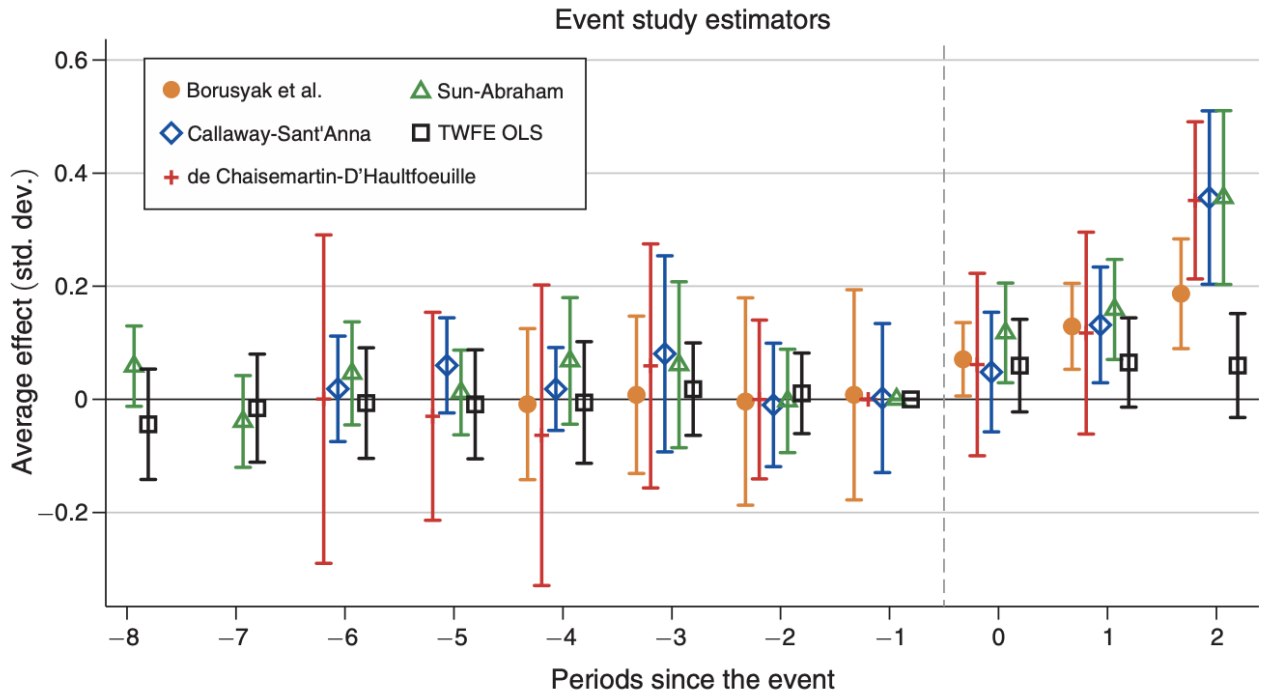


FIGURE 2. EFFECTS OF FACEBOOK ON THE INDEX OF POOR MENTAL HEALTH BASED ON DISTANCE TO/FROM FACEBOOK INTRODUCTION

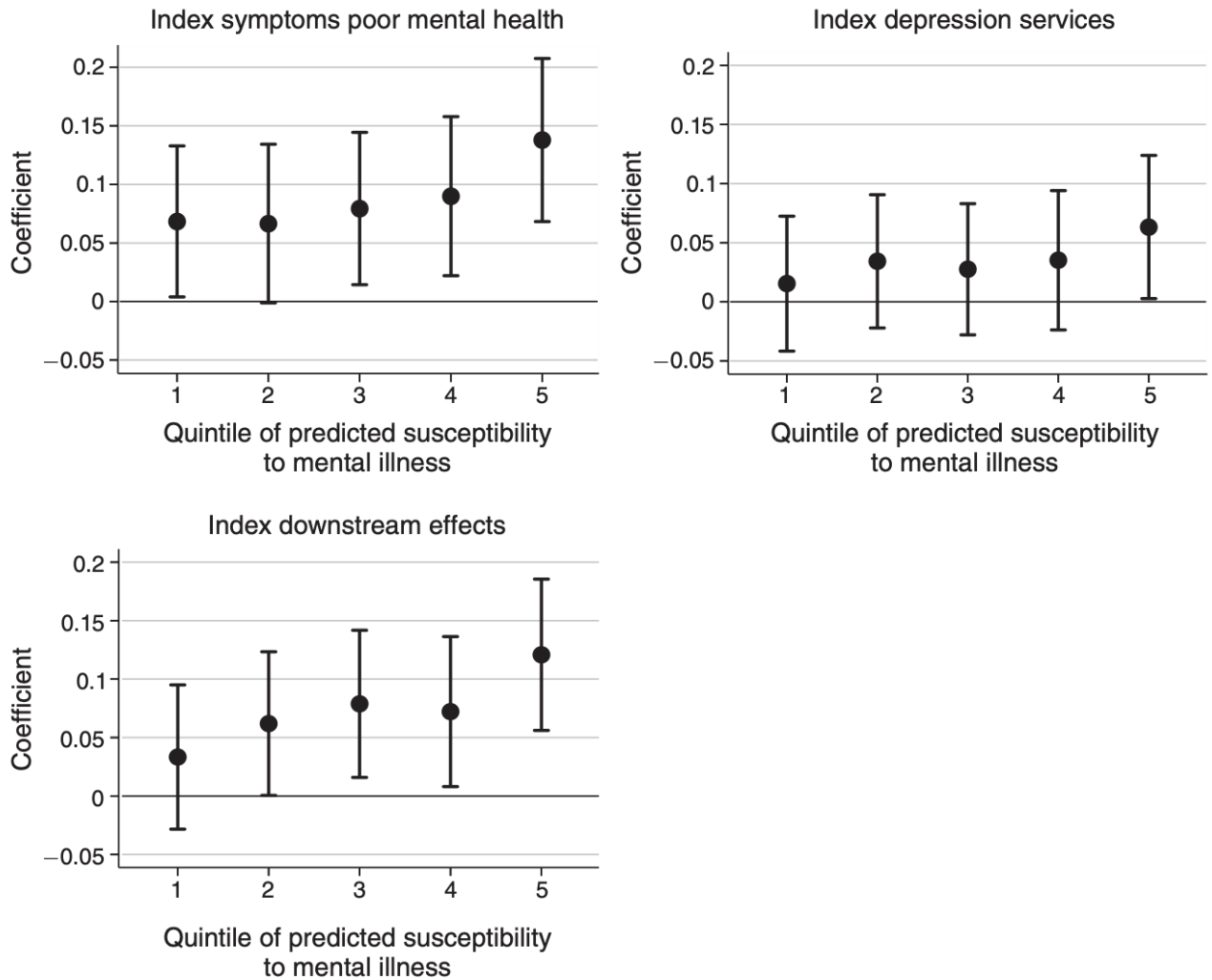


FIGURE 3. HETEROGENEOUS EFFECTS BY PREDICTED SUSCEPTIBILITY TO MENTAL ILLNESS

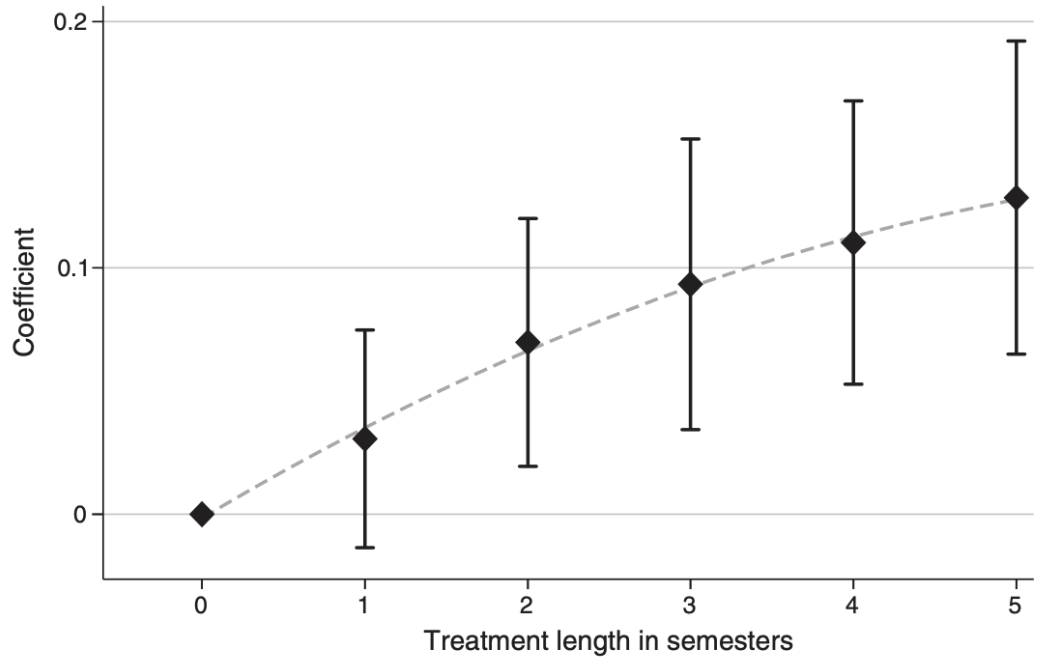


FIGURE 4. EFFECT ON POOR MENTAL HEALTH BY LENGTH OF EXPOSURE TO FACEBOOK

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.

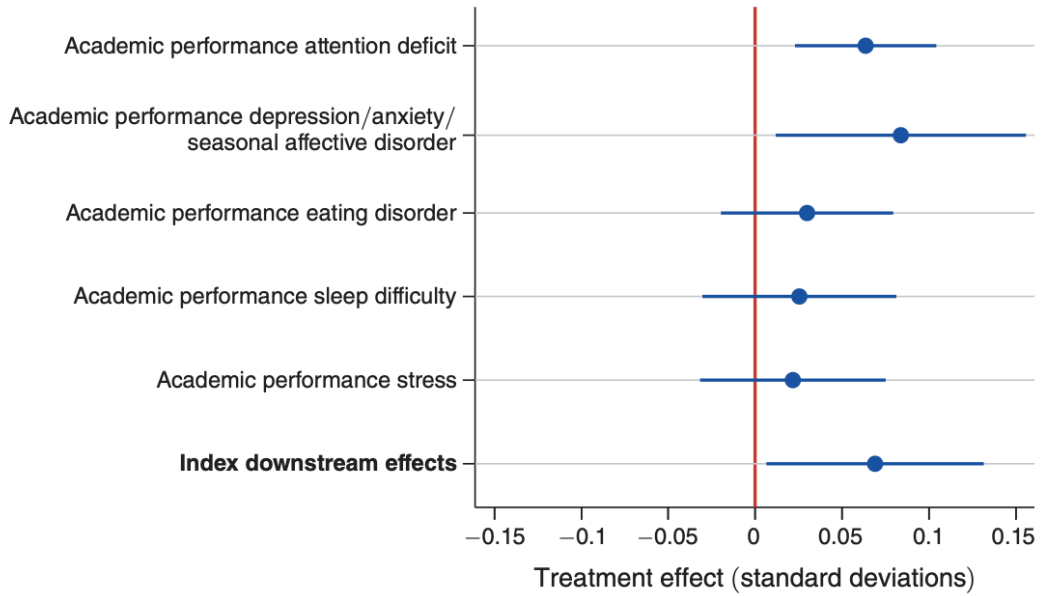


FIGURE 5. DOWNSTREAM EFFECTS ON ACADEMIC PERFORMANCE

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.

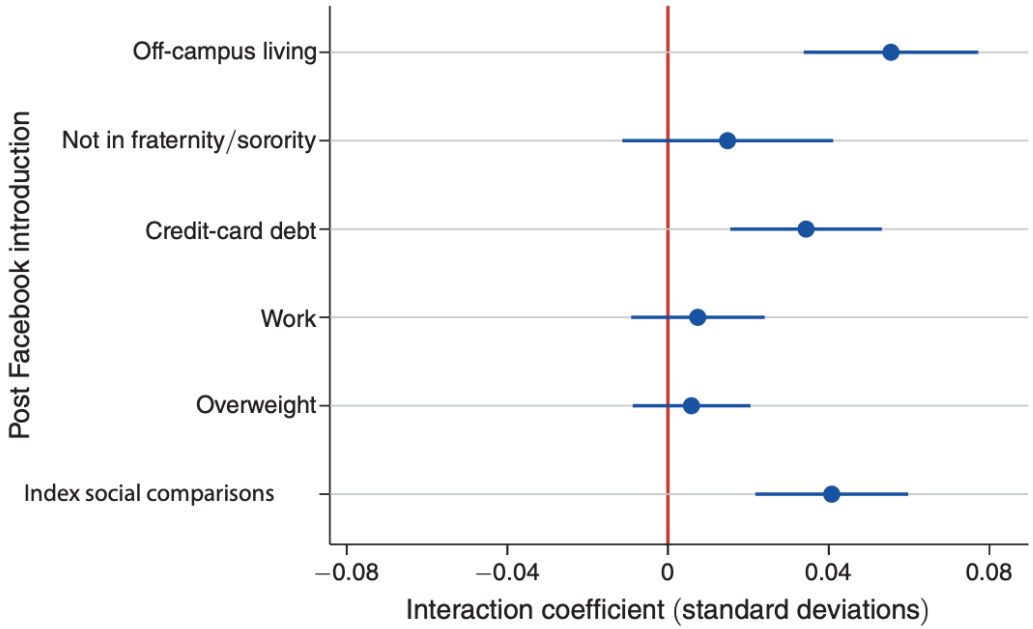


FIGURE 6. HETEROGENEOUS EFFECTS AS EVIDENCE OF UNFAVORABLE SOCIAL COMPARISONS

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

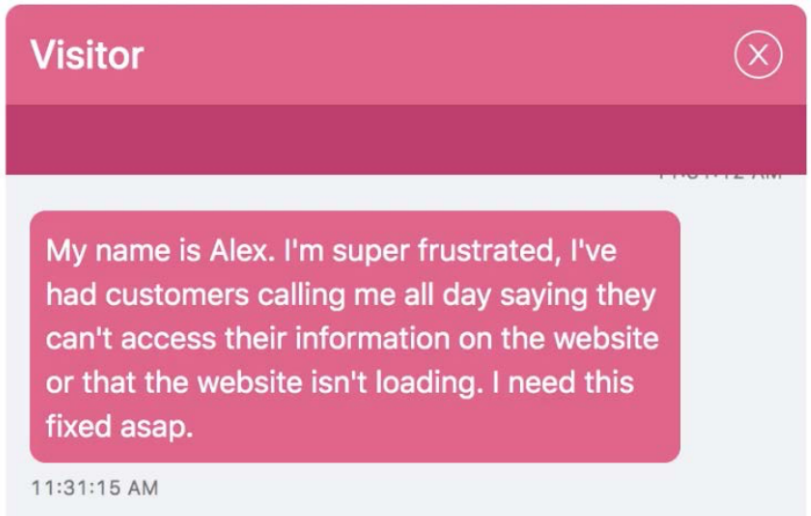
Outcomes and Pictures

- 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

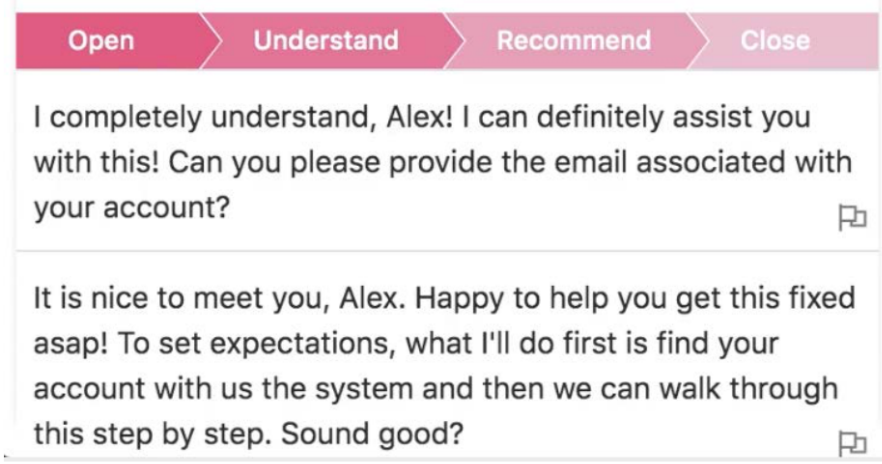
Example of ChatBot

Figure 1: Sample AI Output

A. Sample Customer Issue

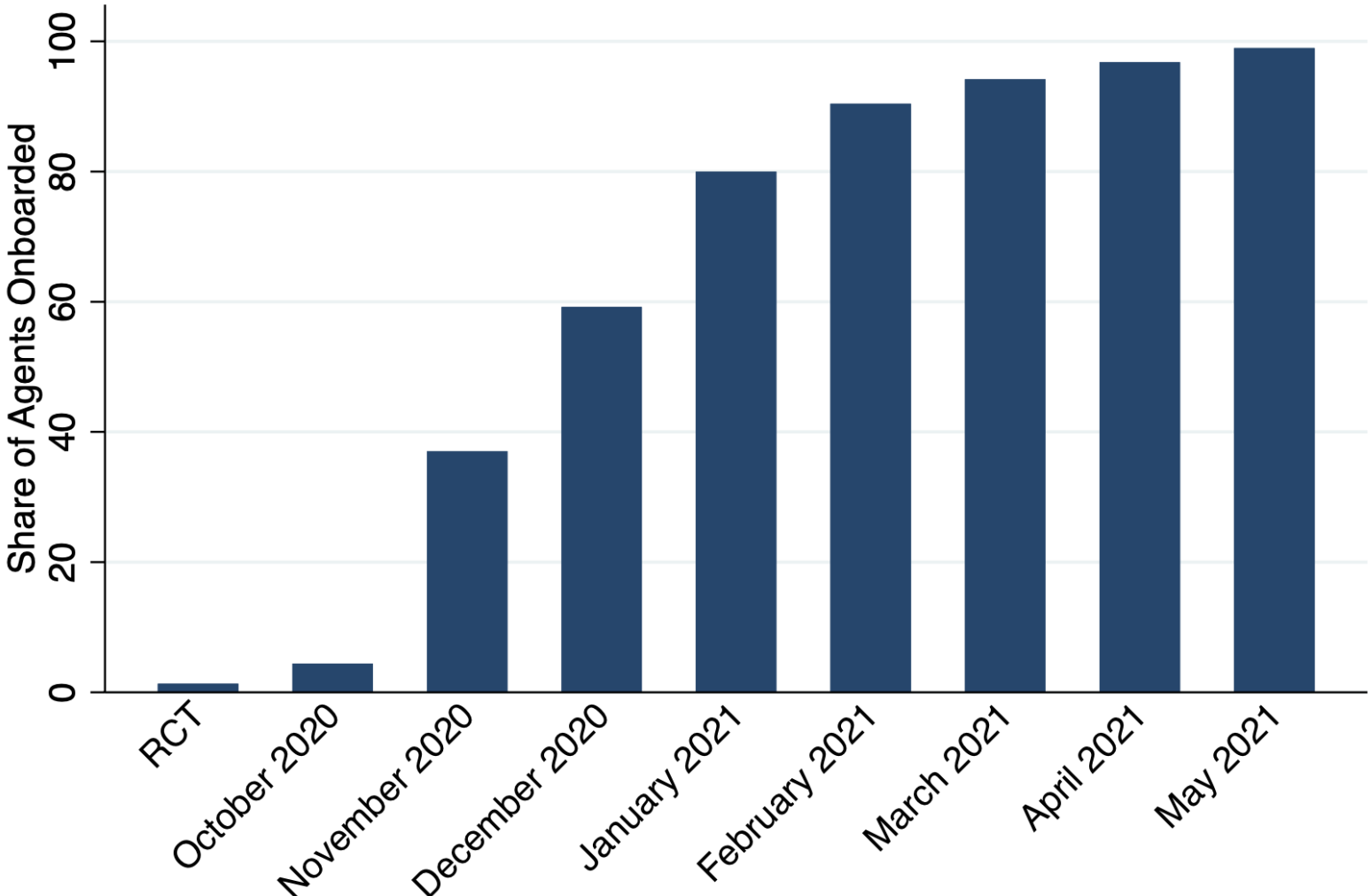


B. Sample AI-generated Suggested Response



Rollout

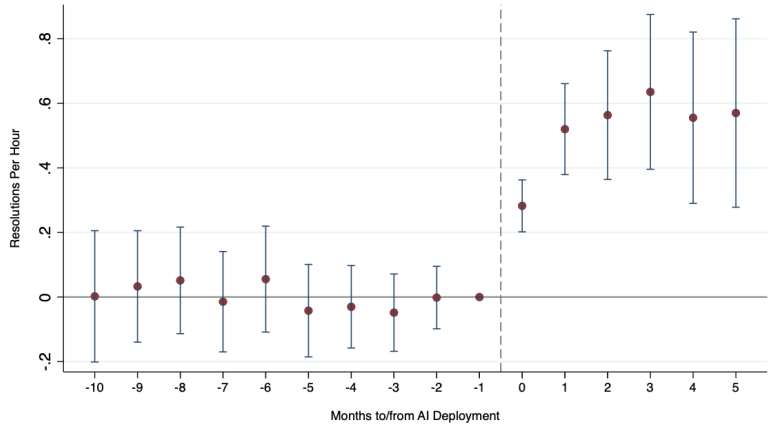
FIGURE 2: DEPLOYMENT TIMELINE



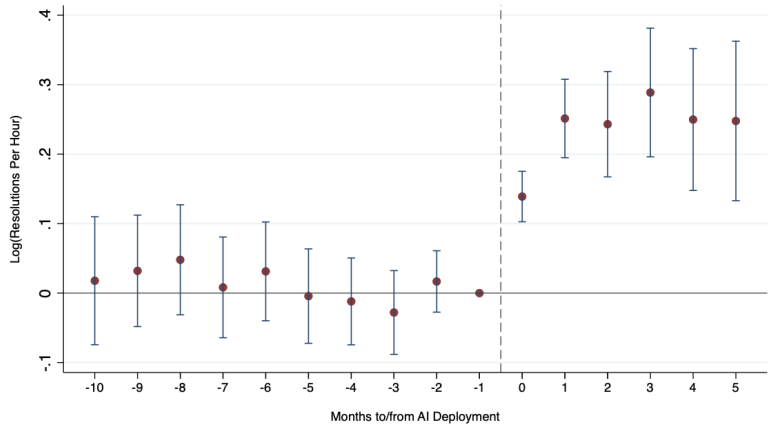
Resolutions of Customer Problems

FIGURE 4: EVENT STUDIES, RESOLUTIONS PER HOUR

A. RESOLUTIONS PER HOUR



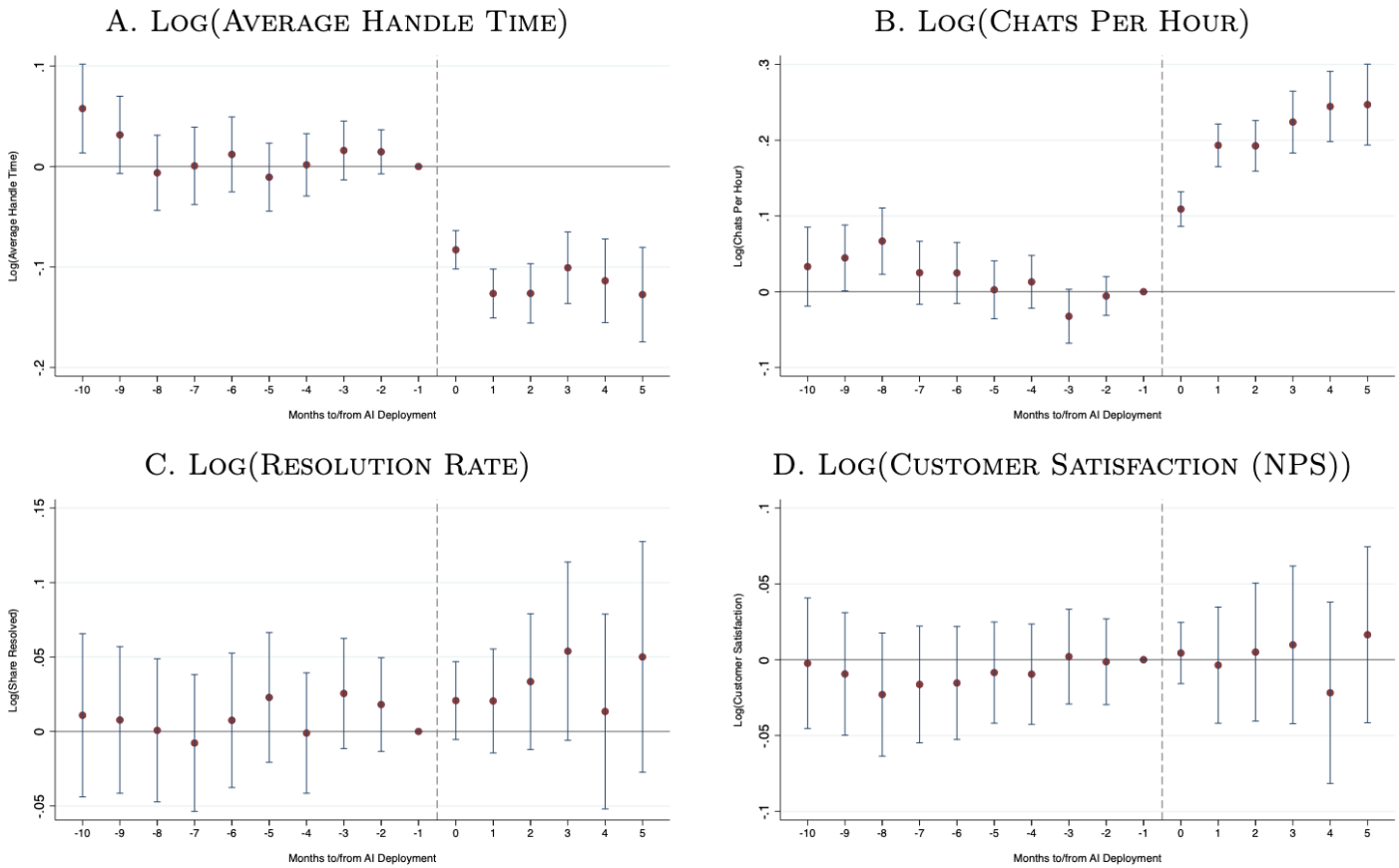
B. LOG(RESOLUTIONS PER HOUR)



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.

Additional Outcomes

FIGURE 5: EVENT STUDIES, ADDITIONAL OUTCOMES

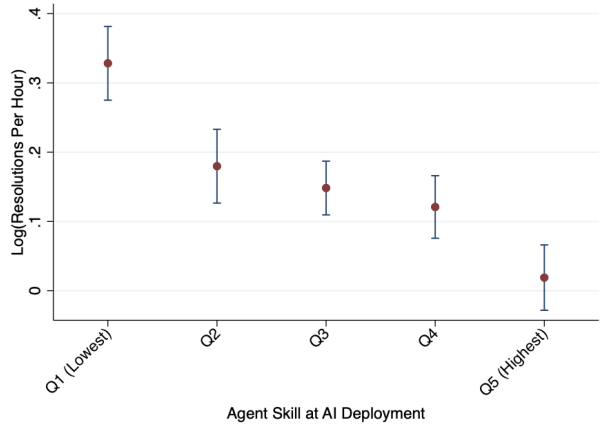


NOTES: These figures plot the coefficients and 95 percent confidence interval from event study regressions of AI model deployment using the Sun and Abraham (2021) interaction weighted estimator. See text for additional details. Panel A plots the average handle time or the average duration of each technical 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.

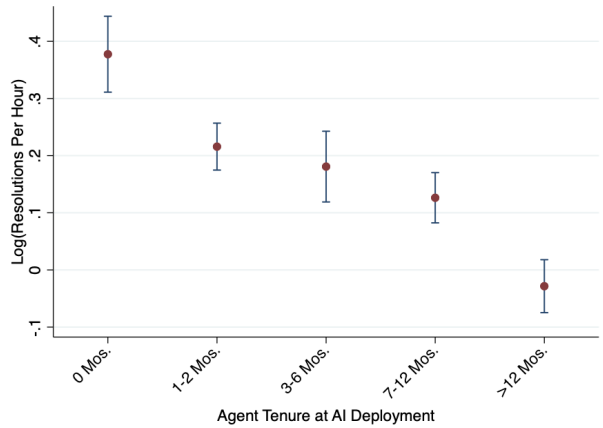
Heterogeneity by Skill

FIGURE 6: HETEROGENEITY OF AI IMPACT, BY SKILL AND TENURE

A. IMPACT OF AI ON RESOLUTIONS PER HOUR, BY SKILL AT DEPLOYMENT

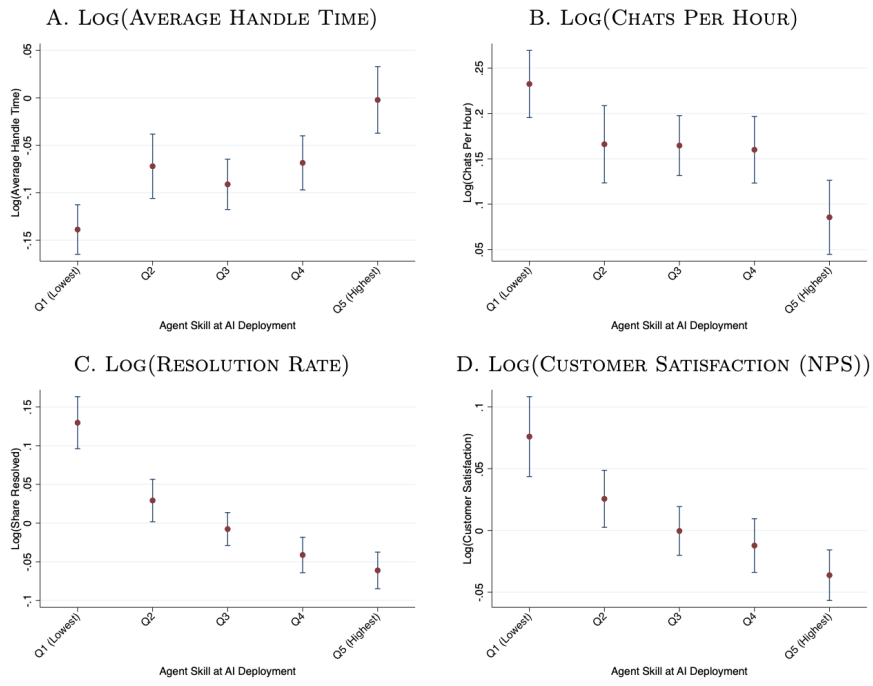


B. IMPACT OF AI ON RESOLUTIONS PER HOUR, BY TENURE AT DEPLOYMENT



Heterogeneity by Skill

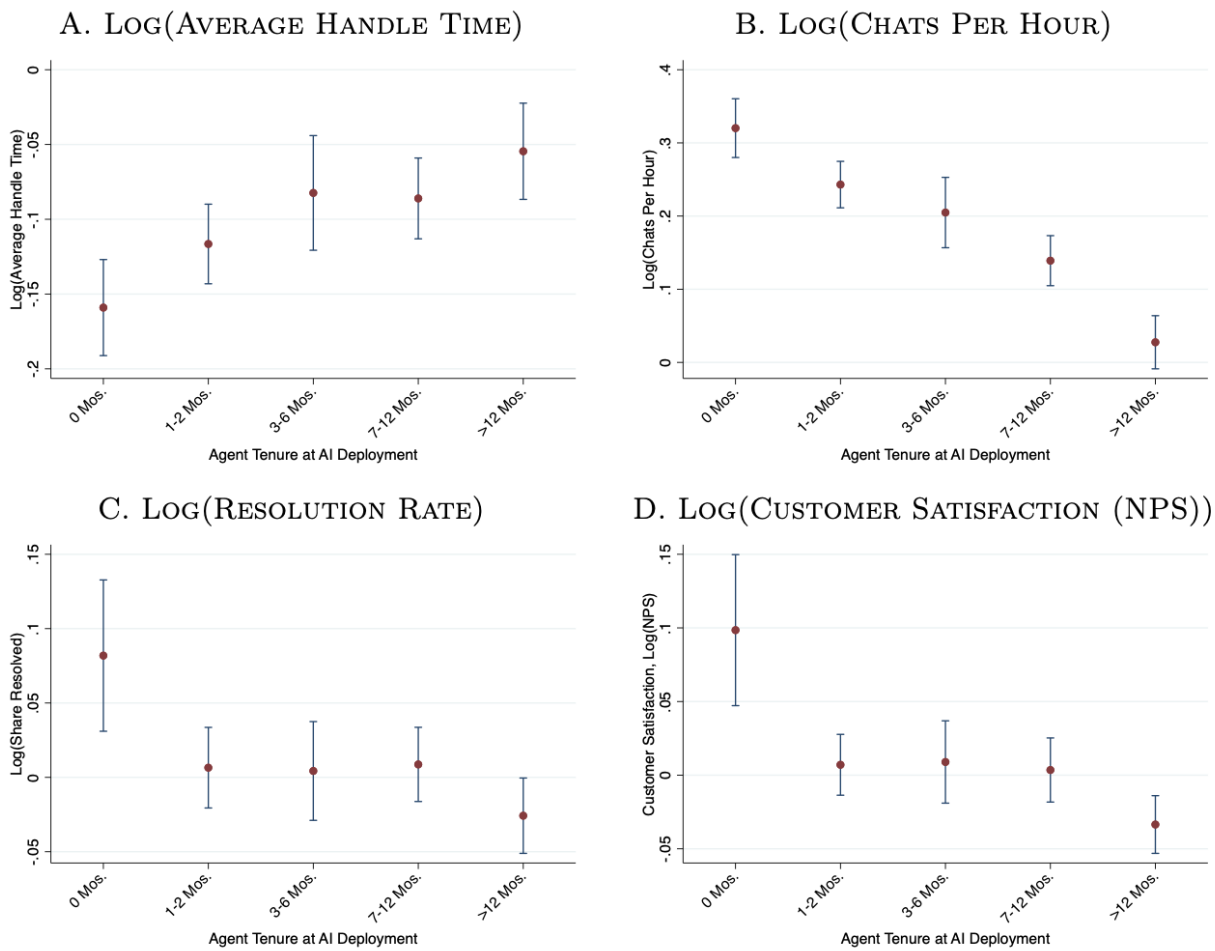
FIGURE 7: HETEROGENEITY OF AI IMPACT BY PRE-AI WORKER SKILL, ADDITIONAL OUTCOMES



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

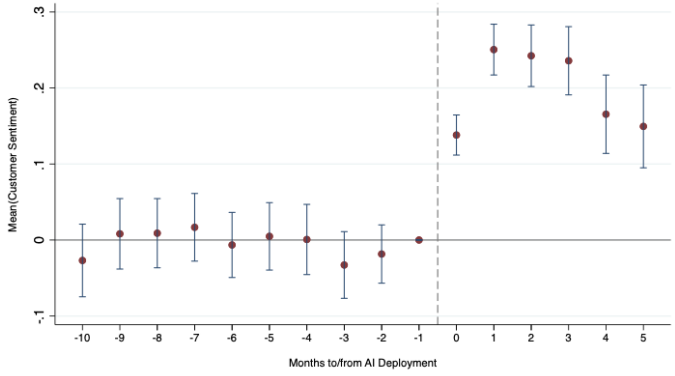
FIGURE 8: HETEROGENEITY OF AI IMPACT BY PRE-AI WORKER TENURE, ADDITIONAL OUTCOMES



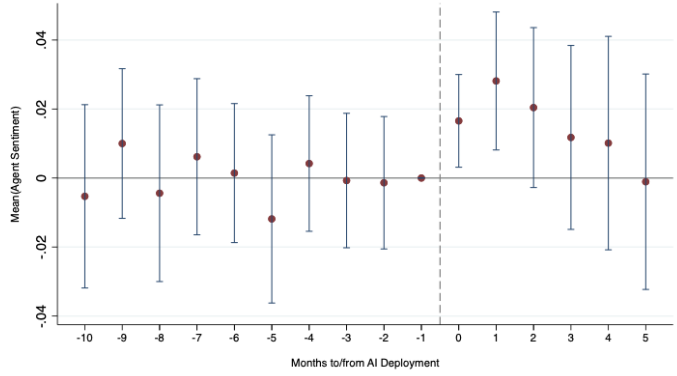
NOTES: These figures plot the impacts of AI model deployment on measures of productivity and performance by pre-AI worker tenure, defined as the number of months an agent has been employed when they receive access to the AI model. Panel A plots the average handle time or the average duration of each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

Sentiment

C. Customer Sentiment, Event Study



D. Agent Sentiment, Event Study

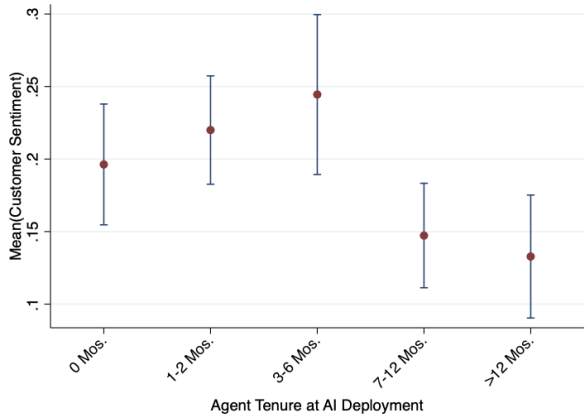


Notes: Each panel of this figure plots the impact of AI model deployment on conversational sentiment. Panel A shows average customer sentiments. Panel B shows average agent sentiments. Panel C plots the event study of AI model deployment on customer sentiment and Panel D plots the corresponding estimate 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.

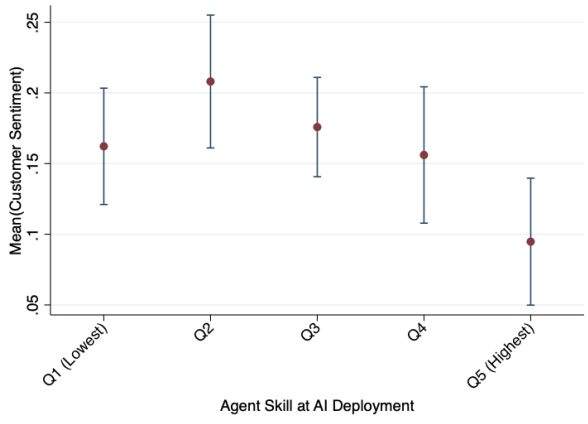
Sentiment

FIGURE A.7: HETEROGENEITY IN CUSTOMER SENTIMENT

A. BY TENURE AT AI MODEL DEPLOYMENT



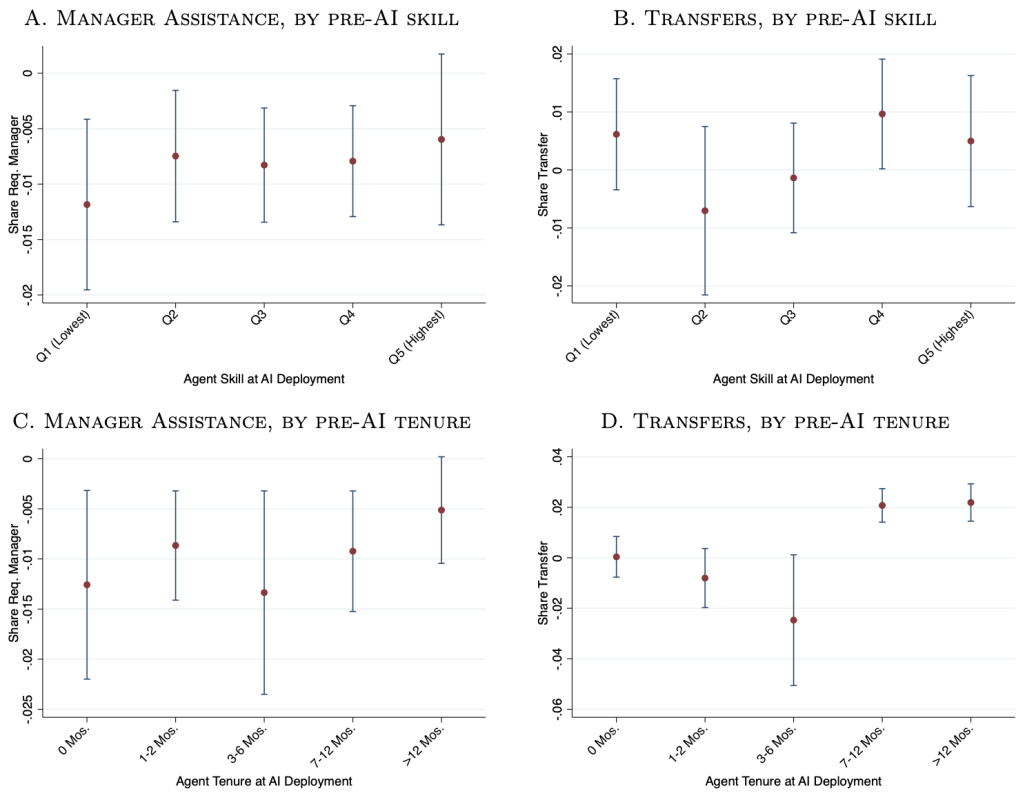
B. BY PRODUCTIVITY AT AI MODEL DEPLOYMENT



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

Roadmap

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

Evidence for parallel trends

1. Event study graphics – plot coefficients and confidence intervals to check if pre-trends are zero so that post-treatment coefficients are compelling evidence for causal effects
2. Falsifications – if possible, rule out competing hypotheses using falsifications (e.g., 65+ year olds can't be on Medicaid as they're already on Medicare)

Roadmap

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?

Synthetic control

- But what if parallel trends really isn't realistic – what then?
- Then you may need to create your own control group that follows the same approximately trajectory as your treatment group pre-treatment
- A method by Abadie and Gardazebal (2003) and follow up papers worked out a method for this called synthetic control
- We will review that next

Temporary page!

LaTeX was unable to guess the total number of pages correctly.
was some unprocessed data that should have been added to
page this extra page has been added to receive it.
If you rerun the document (without altering it) this surplus pag
away, because LaTeX now knows how many pages to expect fo
document.