

# Empirical Economics with R

## 4 Difference-in-Difference Estimation and Estimating the Impact of Search Engine Marketing

Uni Ulm

Prof. Dr. Sebastian Kranz

WiSe 20/21

### Search Engine Marketing

2 / 22

- Internet Advertising is a huge business. Google's advertising revenue has grown from 0.07 billion USD in 2001 and 28.24 billion USD in 2010 to 134.81 USD in 2019. See here for a graph: <https://www.statista.com/statistics/266249/advertising-revenue-of-google/>
- The largest share of Google's advertisement revenue comes from search engine marketing (SEM), also called paid search advertisement. These are advertisements linked to particular keywords and shown when users enter particular search terms.
- For many firms an important marketing question is whether and which forms of search engine marketing are profitable.
- Usually firms track clicks on advertising and analyze what sales such customers generate. However...
  - A customer that clicks on an ad, possibly would have reached the website and made the purchase even without the ad.
  - Possibly, a customer finds the shop and product through the ad, but makes the purchase later by directly visiting the website once more (and can't be tracked because he turned off cookies).

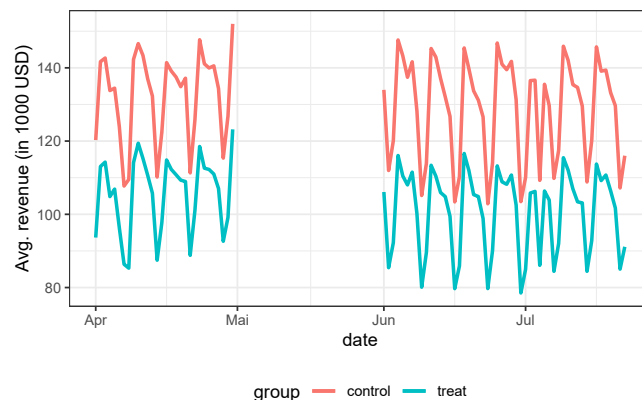
# Controlled Experiments to Study Revenue Impact of SEM

3 / 22

- Blake, Nosko and Tadelis (2014) convinced eBay to run controlled experiments to study the revenue impact of search engine marketing.
- The original data is not available, but Matt Taddy published a modified version on this Github page where the data was scaled and translated so that eBay's original revenues remain unknown.
- eBay used intensively search engine marketing by bidding on different key words in Google's adwords.
- For 8 weeks following May 22th 2012, eBay stopped search engine marketing in a treatment group of 65 of 210 "designated market areas" (DMA) in the USA.
  - A DMA is a metropolitan area like New York, or Los Angeles
  - Google allows advertisement to be regionally targeted. It guesses a user's DMA with information like the IP address.
- eBay tracked their revenues in each DMA using the shipping address of customers.

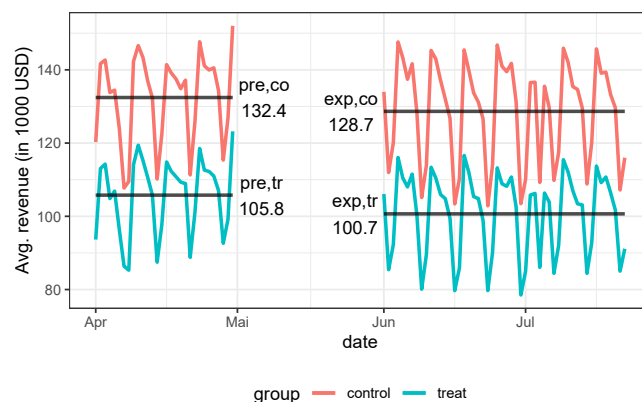
4 / 22

- The plot shows average revenues per DMA in the treatment group (ads turned off) and the control group (ads stayed on) on a daily basis before and during the experiment (June + July).
- The experiment was not perfectly randomized, e.g. ads were not allowed to be turned off in some of the largest market areas.
- We already see in the pre-experimental phase, larger revenues per DMA in the control group than in the treatment group.



5 / 22

- I now added to the plot the average revenues in the pre-experimental (pre) and experimental (exp) phases for the treatment and control group.
- Just taking these 4 mean revenues, what would be a good estimator of the treatment effect? (Here we mean by the treatment effect the average revenue impact per DMA of turning search engine marketing off.)



- In a well randomized experiment, we could just look at the experimental period and compute the difference between the average outcome in the treatment and experimental group:

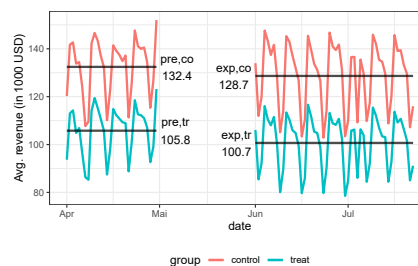
$$\bar{y}_{exp,tr} - \bar{y}_{exp,co} = 100.7 - 128.7 = -28$$

This approach would suggest that turning off search engine marketing reduces daily revenues by 28 thousand USD on average per DMA.

- However, we found already previous to the experiment substantially lower daily revenues in the DMAs that were later treated:

$$\bar{y}_{pre,tr} - \bar{y}_{pre,co} = 105.8 - 132.4 = -26.6$$

So attributing the complete difference of -28 in the experimental period to the causal effect of the treatment just seems plainly wrong here.



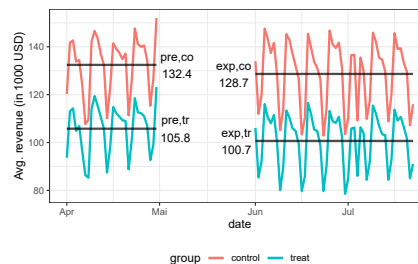
## Difference-in-Difference Estimator

- The difference-in-difference (short *diff-in-diff* or *DiD*) estimator for the causal effect of the treatment computes how the difference between the treatment and control group has changed in the experimental period compared to the pre-experimental period:

$$\begin{aligned} \text{DiD} &= (\bar{y}_{exp,tr} - \bar{y}_{exp,co}) - (\bar{y}_{pre,tr} - \bar{y}_{pre,co}) \\ &= (100.7 - 128.7) - (105.8 - 132.4) \\ &= -28 - (-26.6) = -1.4 \end{aligned}$$

This is a very intuitive natural estimator that under comparatively weak conditions consistently estimates the causal effect of the treatment.

- Here we would thus estimate that turning off search engine marketing reduces eBay's average daily revenues per DMA by 1.4 thousand USD. That is only a small share of the average pre-experimental revenues.



## Alternative derivation of the DiD Estimator

8 / 22

- There is an alternative way how we can think about the DiD estimator.
- We first look how the average daily revenues changed in the treatment group during the experiment compared to before:

$$\bar{y}_{exp,tr} - \bar{y}_{pre,tr} = 100.7 - 105.8 = -5.1$$

This might suggest that turning off advertisement reduced daily revenues by 5.1 thousands USD.

- However, there could be seasonal effects or time trends that would have changed revenues even absent the experiment. To get an estimate of these time effects, we can look how revenues have changed over time in the control group:

$$\bar{y}_{exp,co} - \bar{y}_{pre,co} = 128.7 - 132.4 = -3.7$$

- The DiD estimator "corrects" the change in the treatment group by this change in the control group:

$$\begin{aligned} \text{DiD} &= (\bar{y}_{exp,tr} - \bar{y}_{pre,tr}) - (\bar{y}_{exp,co} - \bar{y}_{pre,co}) \\ &= (10.7 - 105.8) - (128.7 - 132.4) \\ &= -5.1 - (-3.7) = -1.4 \end{aligned}$$

We get exactly the same DiD estimator as before using this alternative route.

## Estimating DiD via linear regression

9 / 22

- Typically, one computes the DiD estimator by running a linear regression. This e.g. allows to conveniently get standard errors of the estimated causal effect.
- Here is a sample of 6 rows from the data set:

	date	dma	rev	treat	exp
p	group	period			
1	2012-04-01	500	76.7	1	
0	treat	pre			
2	2012-04-01	501	2096.2	0	
0	control	pre			
3	2012-04-01	502	35.0	0	
0	control	pre			

	date	dma	rev	treat	exp
t	exp	group	period		
10501	2012-06-21	500	63.0		
1	1	treat	exp		
10502	2012-06-21	501	2220.1		
0	1	control	exp		
10503	2012-06-21	502	27.4		
0	1	control	exp		

- `date` denotes the day. We will index the day with  $t$  in our regression equations.
- `dma` is the designated market area, we will index it with  $i$  in our regression equations.
- `rev` denoted by  $rev_{i,t}$  in our equation are the revenues of DMA  $i$  on day  $t$  in 1000 USD.
- `treat` denoted by  $treat_i$  is a dummy variable that is 1 if the DMA  $i$  is part of the treatment group and 0 if it is part of the control group. Note that  $treat_i$  does not depend on the day  $t$ . The variable `group` encodes the same information as `text`.
- `exp` denoted by  $exp_t$  is a dummy variable that is 1 if day  $t$  is in the experimental period and 0 if the day is previous to the start of the experiment. The variable `period` encodes the same information as `text`.

## Estimating DiD via linear regression

10 / 22

- We can estimate the DiD estimator as the OLS estimator  $\hat{\beta}_1$  of the following regression:

$$rev_{i,t} = \beta_0 + \beta_1 treat_i \cdot exp_t + \beta_2 treat_i + \beta_3 exp_t + u_{i,t}$$

- The interaction term  $treat_i \cdot exp_t$  is 1 only for observations that were actually treated, i.e. if the DMA  $i$  is part of the treatment group and the day  $t$  is in the experimental period.  $\beta_1$  shall thus denote the causal effect of being treated (search engine marketing turned off) on a DMA's daily revenues.
- The regression results are on the right. We estimate a causal effect of turning off SEM of  $\hat{\beta}_1 = -1.3$ .
  - While we computed in the previous slides DiD estimator of  $-1.4$ , the differences are just due to rounding errors. Absent rounding errors both estimators are the same.

	Dependent variable:
	rev
treat_exp	-1.3 (7.7)
treat	-26.7*** (6.2)
exp	-3.8 (4.4)
Constant	132.4*** (3.5)
Observations	17,220
Note:	*p<0.1; **p<0.05; ***p<0.01

## Showing that $\hat{\beta}_1$ is the DiD estimator

- Let us show that  $\hat{\beta}_1$  in our regression is indeed the DiD estimator. Consider our estimated regression:

$$\begin{aligned} \hat{r}ev_{i,t} &= \hat{\beta}_0 + \hat{\beta}_1 treat_i \cdot exp_t + \hat{\beta}_2 treat_i + \hat{\beta}_3 exp_t \\ &= 132.4 - 1.3 \cdot treat_i \cdot exp_t - 26.7 \cdot treat_i - 3.8 \cdot exp_t \end{aligned}$$

We will link the estimated coefficients with the group means, we have used for the DiD calculation earlier:

$$\begin{aligned} \bar{y}_{pre,co} &= 132.4 & \bar{y}_{exp,co} &= 128.7 \\ \bar{y}_{pre,tr} &= 105.8 & \bar{y}_{exp,tr} &= 100.7 \end{aligned}$$

- Which revenue would we predict for an observation in the control group in the pre-experimental period? From the group means it would be  $\bar{y}_{pre,co} = 132.4$ . In our regression, we would have  $treat_i = exp_t = 0$  and thus our prediction would be just the constant  $\hat{\beta}_0 = 132.4$ . We have

$$\hat{\beta}_0 = \bar{y}_{pre,co}$$

- Now let us predict revenues for an observation in the treatment group in the pre-experimental period, i.e. we have  $treat_i = 1$  and  $exp_t = 0$ . Our revenue prediction (correcting rounding errors) would be

$$\hat{\beta}_0 + \hat{\beta}_2 = 132.4 - 26.7 = 105.8 = \bar{y}_{pre,tr}$$

Rearranging and substituting  $\hat{\beta}_0 = \bar{y}_{pre,co}$ , we find

$$\hat{\beta}_2 = \bar{y}_{pre,tr} - \bar{y}_{pre,co}$$

- If we want to predict revenues for a control group observation in the experimental period, i.e.  $treat_i = 0$  and  $exp_t = 1$ , our revenue prediction (correcting rounding errors) would be

$$\hat{\beta}_0 + \hat{\beta}_3 = 132.4 - 3.8 = 128.7 = \bar{y}_{exp,co}$$

Rearranging we find

$$\hat{\beta}_3 = \bar{y}_{exp,co} - \bar{y}_{pre,co}$$

- Now finally, let us predict the revenues for a treatment group member in the experimental period, i.e. we have  $treat_i = exp_t = treat_i \cdot exp_t = 1$ . Our revenue prediction (correcting rounding errors) would be

$$\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3 = 132.4 - 1.3 - 26.7 - 3.8 = 100.7 = \bar{y}_{exp,tr}$$

We can rearrange this for  $\hat{\beta}_1$  as follows:

$$\begin{aligned}
 \hat{\beta}_1 &= \bar{y}_{exp,tr} - \hat{\beta}_0 - \hat{\beta}_2 - \hat{\beta}_3 \\
 &= \bar{y}_{exp,tr} - \bar{y}_{pre,co} - (\bar{y}_{pre,tr} - \bar{y}_{pre,co}) - (\bar{y}_{exp,co} - \bar{y}_{pre,co}) \\
 &= (\bar{y}_{exp,tr} - \bar{y}_{exp,co}) - (\bar{y}_{pre,tr} - \bar{y}_{pre,co})
 \end{aligned}$$

This is exactly the formula for the DiD estimator.

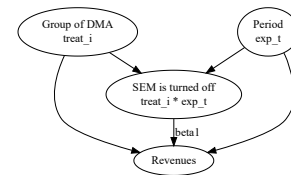
## Alternative view on the DiD regression.

12 / 22

- Consider our DiD regression

$$rev_{i,t} = \beta_0 + \beta_1 treat_i \cdot exp_t + \beta_2 treat_i + \beta_3 exp_t + u_{i,t}$$

We can view  $treat_i$  and  $exp_t$  as control variables that pick up confounders. Consider the graph on the right.



- We saw that even before the experiment the DMAs in the control and treatment group had different revenues. We control for this problem by adding the group dummy  $treat_i$ .
- By adding the  $exp_t$  period dummy, we control for seasonal effects or time trends that would cause difference between the experimental and previous period even absent the treatment.
- If we would just estimate a short regression

$$rev_{i,t} = \beta_0 + \beta_1 treat_i \cdot exp_t + \varepsilon_{i,t}$$

those factors would be part of the error term  $\varepsilon_{i,t}$ . Then our variable of interest  $treat_i \cdot exp_t$  would be endogenous and  $\hat{\beta}_1$  would be inconsistent.

## Controlling with DMA fixed effects

13 / 22

- Instead of controlling with a dummy  $treat_i$  that indicates whether the DMA was in the treatment group, we could also control by adding separate dummies for each DMA.
- For example, if there were only 3 DMAs (501, 502, and 503) we could generate dummy variables `dma501`, `dma502` and `dma503` for our regression as follows (the first dummy will be left-out if we add a constant to the regression):

	date	dma	dma501	dma502	dma503
1	2012-04-01	501	1	0	0
2	2012-04-01	502	0	1	0
3	2012-04-01	503	0	0	1
4	2012-04-02	501	1	0	0
5	2012-04-02	502	0	1	0

- In our data set, we have 210 different DMA, implying a lot of dummies. Such dummies for categories with many levels are often called *fixed effects*.
  - The expression *fixed effect* comes from the *panel data* literature. A panel data set is a data set like ours where we observe different units  $i$  (here DMA) over several time periods  $t$  (here several days).
- Using the standard OLS computation method for a model with so many dummies can be quite memory and time consuming. (One has to invert a matrix with  $K + 1$  rows and columns, or apply some similar complex matrix decomposition methods, where  $K$  is the number of explanatory variables including all dummies. The time to invert a matrix roughly increases proportionally to  $K^3$ .)
- However, there are efficient computational tricks to control for fixed effects. In R we can use the function `fe1m` from the package `lfe`.



## Fixed Effects Regressions

14 / 22

- We can estimate our regression with DMA fixed effects in R as follows:

```
library(lfe)
felm(rev ~ treat_exp + exp | dma, data=dat)
```

- We could additionally replace the  $exp_t$  dummy by day fixed effects:

```
felm(rev ~ treat_exp | dma+date, data=dat)
```

- On the right, you see the regression results beside the original DiD regression. We estimate the same treatment effect  $\hat{\beta}_1 = -1.33$  in all 3 specifications.
  - Often a DiD estimator is implemented by such fixed effects regression in empirical studies.
- Interestingly, the standard errors of  $\hat{\beta}_1$  are lower in the fixed effects regression. This is not necessarily the case but can happen.

	Dependent variable:		
	rev		
	OLS	felm	
	(1)	(2)	(3)
treat_exp	-1.33 (7.73)	-1.33 (1.03)	-1.33 (0.94)
treat	-26.66*** (6.16)		
exp	-3.78 (4.40)	-3.78*** (0.59)	
DMA fixed effects?	No	Yes	Yes
Day fixed effects?	No	No	Yes
Observations	17,220	17,220	17,220

Note:

\*p<0.1; \*\*p<0.05;  
\*\*\*p<0.01

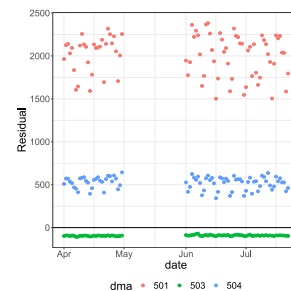
## Excursion: Cluster-Robust Standard Errors

15 / 22

- Consider our original DiD regression without fixed effects:

$$rev_{i,t} = \beta_0 + \beta_1 treat_i \cdot exp_t + \beta_2 treat_i + \beta_3 exp_t + u_{i,t}$$

- R's default formula for the standard errors of the coefficients  $\hat{\beta}_k$  only is correct if the error terms  $u_{i,t}$  are identically and independently distributed (iid) from each other.
- The plot on the right shows the residuals  $\hat{u}_{i,t}$  from our regression for 3 different DMA indicated by color. If the  $u_{i,t}$  were iid then the residuals of each DMA should be randomly be distributed with mean 0.
- However, we see that the residuals of each DMA tend to keep similar values over all days, they are strongly correlated with each other. The default formula for standard errors does not apply in such a situation.
- One remedy is to compute so called *cluster-robust standard errors* clustered on the DMA. They allow the error terms for the same DMA to be correlated with each other.



## Excursion: Cluster-Robust Standard Errors

16 / 22

- The `fe1m` command in the `lfe` package also allows to conveniently compute such cluster-robust standard errors in R as follows:

```
fe1m(rev ~ treat_exp + treat
+ exp|0|0|dma, data=dat)
```

- We see on the right that interestingly the cluster-robust standard errors for  $\hat{\beta}_1$  are smaller than the default standard errors. (This can sometimes happen.)
- We can also compute cluster-robust standard errors for DMA together with a fixed effect for DMA. But there we find no big difference between the standard errors.
- Remark: There is active econometric research that examines situations in which clustered-standard errors don't work well yet and how to improve them. E.g. the package `clubSandwich` contains some modern formulas for clustered standard errors.

	Dependent variable:			
	rev			
	(1)	(2)	(3)	(4)
treat_exp	-1.33 (7.73)	-1.33 (1.09)	-1.33 (0.94)	-1.33 (1.09)
treat	-26.66*** (6.16)	-26.66 (27.70)		
exp	-3.78 (4.40)	-3.78*** (0.68)		
DMA fixed effects?	No	No	Yes	Yes
Day fixed effects?	No	No	Yes	Yes
Cluster robust s.e.?	No	Yes	No	Yes
Observations	17,220	17,220	17,220	17,220
Note:	*p<0.1; **p<0.05; ***p<0.01			

## Discussion: Effects of Search Engine Marketing

17 / 22

- Using our rule-of-thumb, we find in our fixed-effect specification with cluster-robust standard errors the following 95% confidence interval for the causal effect of turning-off SEM on revenues:

$$[\hat{\beta}_1 - 2 \cdot se(\hat{\beta}_1); \hat{\beta}_1 + 2 \cdot se(\hat{\beta}_1)] =$$

$$[-1.33 - 2 \cdot 1.09; -1.33 + 2 \cdot 1.09] = [-3.51; 0.85]$$

Given that average daily revenues per DMA in the pre-experiment period were 123.8 thousand USD, this means we estimate that turning off SEM reduces revenues by just  $1.33/123.8 = 1.07\%$  and we are 95% confident that it does not reduce revenues by not more than  $3.51/123.8 = 2.8\%$  of the average pre-experiment revenues.

- As noted in the beginning, these are not the true numbers for eBay, because the publicly available data set has been shifted and scaled by some unknown factors.
- Using the original data, Blake et. al. (2014) also find quite small effects of SEM on eBays revenues. In particular, they find that the costs of SEM exceeded the estimated revenue gains from it, so that SEM had a negative return on investment for eBay in the experimental period.

## Discussion: Effects of Search Engine Marketing 2

18 / 22

- So can we conclude from the study by Blake et. al. (2014) that search engine marketing does generally not pay off? No. Below are some counter arguments.
- The experiment only estimates the short term effects of turning off SEM during the 8 weeks period. But advertisement also has long term effects, e.g. by increasing brand awareness. Also filling the advertisement slots in search engines can make market entry by potential competitors harder. Unfortunately, such long term effects are very hard to measure.
- Possibly, eBays SEM strategy was not yet optimal. Possibly, bidding on other keywords, or targeting different customer groups may be more effective. To compare different SEM strategies, one should run new experiments.
- eBay is a large well-known firm. This means probably many customers directly search on the eBay site. Also eBay may show up quite early in the unpaid "organic" search results on Google. This suggests that for eBay SEM may not be as important as for other smaller firms.

## Discussion: Effects of Search Engine Marketing 3

19 / 22

- Coviello, Gneezy and Goette (2017) also study the effect of SEM and summarize their findings as follows:

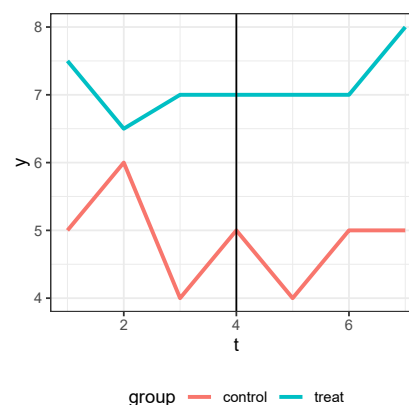
*Abstract: Companies spend billions of dollars online for paid links to branded search terms. Measuring the effectiveness of this marketing spending is hard. Blake, Nosko and Tadelis (2015) ran an experiment with eBay, showing that when the company suspended paid search, most of the traffic still ended up on its website. Can findings from one of the largest companies in the world be generalized? We conducted a similar experiment with Edmunds.com, arguably a more representative company, and found starkly different results. More than half of the paid traffic is lost when we shut off paid-links search. These results suggest money spent on search-engine marketing may be more effective than previously documented.*

- If you want to evaluate the impact of SEM marketing for your company, the important point is that you should run experiments. Successful tech companies run 1000nds of experiments per year. (You can e.g. Google "How many experiments does Amazon run".)

## The parallel trends assumption for DiD estimation

20 / 22

- Let us come back to DiD estimation. Consider the data set on the right, where an fictitious experimental treatment started in period  $t = 4$ . Is it plausible that here a DiD estimator consistently estimates the causal effect of the treatment?
- No. The problem is that already in the pre-experimental phase, the treatment and control group had very different time trends with respect to the dependent variable  $y$ . This means we don't know whether how much of the change in the differences between the two lines in the experimental phase is due to the experiment and how much just arises because even absent the experiment the time trends would have continued to differ.
- A crucial assumption for the DiD estimator is the so called *parallel trends assumption*. It means that if no experiment would take place, the difference between the average outcome in the 'treatment' and 'control' group should be constant over time.
- One requirement of the parallel trend assumption is that in the pre-experimental period the curves for the treatment and control group run parallel. This is typically checked graphically. This is always to some degree subjective. The curves will seldom be exactly parallel, but should look parallel enough.
  - Most economists would agree that in the figure on the right, the parallel trend assumption seems strongly violated.



## Discussion: Difference in Difference Estimation

21 / 22

- DiD estimation is a powerful method to estimate causal effects if we observe data before and during a treatment intervention.
- DiD can e.g. be used in experiments that are not perfectly randomized.
- DiD has also been often used in examining policy changes. You will study in the RTutor problem set of this chapter the causal effect of an increase in minimum wages on employment using a DiD approach.
- In non-experimental settings, often the main problem is to find a suitable control group, so that the parallel trends assumption is not violated.

## References

22 / 22

- Blake, Thomas, Chris Nosko, and Steven Tadelis. 2015. "Consumer heterogeneity and paid search effectiveness: A large-scale field experiment." *Econometrica* 83, no. 1: 155-174.
- Coviello, Lorenzo, Uri Gneezy, and Lorenz Goette. 2017. "A large-scale field experiment to evaluate the effectiveness of paid search advertising."