

Internet search data shows increasing interest in online education during the COVID-19 pandemic, with females showing greater increases

Supplementary Information

A. Extended Methods and Model Results

In this section, we provide more details of the models and results.

The dependent variable in our models is the search intensity, i.e., it is the proportions of the search volume of online-education-related queries out of all searches. Hence the dependent variable is an estimated proportion with sampling variance. Therefore, when fitting the models, we weigh the observations by the inverse of the standard errors associated with each observed dependent variable, as in Saxonhouse (1976) and Hornstein (2013) (1, 2).

We summarize the results of Eq. 1 in the upper panel of **Table A-1**. The coefficients α^T are positive and significant for $T = 1$ and 2, which show significantly higher search intensity of online education during the lockdown phase and the “new normal” phase. Comparing the coefficient α^T with the pre-pandemic average (i.e., on or before 03/08/2020), we find that searches of online education increased by about 52.7% (95% CI: 46.9 to 58.4%) in the lockdown phase, and 52.9% (95% CI: 45.8 to 60.0%) in the “new normal” phase.

We summarize the results of Eq. 2 in the lower panel of **Table A-1**. The coefficients γ^T are positive and significant for $T = 1$ and 2, which show that the increased searches of online education among females was significantly higher compared to males, during the lockdown phase ($\gamma^1 = 11.01$, $p < 0.01$) and the “new normal” phase ($\gamma^2 = 11.51$, $p < 0.01$). Since we use males as the baseline, the coefficient α^T ($T = 1, 2$) captures the increased searches among males during the lockdown and “new normal” phase respectively. Calculating the ratio of γ^T and α^T , we find that compared to males, the increased search intensity was 124.1% (95% CI: 78.1 to 193.7%) and 131.4% (95% CI: 74.9 to 226.5%) higher among females during the lockdown and “new normal” phases respectively.

We summarize the estimated $\gamma^{T,A}$ of Eq. 3 in **Table A-2**. The coefficients $\gamma^{T,A}$ are positive and significant when $T = 1$ and 2 among the age group $A=0,1,2,3$ (i.e., below or equal to 24, 25~34, 35~44, and 45~64 years old). It shows that during the lockdown phase and the “new normal” phase, among all the four age groups below 64, there was significantly greater increase in search intensity of online education among females than males.

We also fit Eq. 1 and Eq. 2 for the search intensity of economic concerns (which includes the searches for queries such as “unemployment”, “layoff” and so on), to provide support for the potential mechanisms related to economic concerns, as described in the Discussion: Potential Explanations section. The results for the search intensity of economic concerns are summarized in the right panel of **Table A-1**. From the result of Eq. 1, the coefficients α^T are positive and significant for $T = 1$ and 2, which show that searches of economic concerns increased significantly during the lockdown and “new normal” phases. From the result of Eq. 2, the coefficients γ^T are positive and significant for $T = 1$ and 2, which show that the increased search intensity of economic concerns among females was significantly higher compared to males, during the lockdown phase, ($\gamma^1 = 15.46$, $p < 0.01$) and the “new normal” phase ($\gamma^2 = 10.80$, $p < 0.01$).

Table A-1. Changes of search intensity during the pandemic

Overall search intensity during the progress of the pandemic		
	Online Education	Economic Concerns
Lockdown Phase α^1	14.28 (0.79) ***	21.20 (0.87) ***
“New Normal” Phase α^2	14.33 (0.98) ***	19.69 (1.10) ***

R-square	0.864	0.699
----------	-------	-------

Differential effect among females and males

	Online Education	Economic Concerns
Lockdown Phase γ^1	11.01 (1.45) ***	15.46 (1.53) ***
"New Normal" Phase γ^2	11.51 (1.79) ***	10.80 (1.93) ***
R-square	0.887	0.768

Note. **Upper panel:** the coefficients α^T are positive and significant for $T = 1$ and 2, which show significantly higher search intensity of online education and economic concerns during the lockdown phase ($T=1$) and the "new normal" phase ($T=2$). **Lower panel:** the coefficients γ^T are positive and significant for $T = 1$ and 2, which shows that the increased searches among females were significantly higher compared to males, during the lockdown phase and the "new normal" phase. *** $p < .01$, ** $p < .05$, * $p < .10$.

Table A-2. Gender difference in search intensity of online education among each age group

$\gamma^{T,A}$	Lockdown Phase (T=1)	"New Normal" Phase (T=2)
24 yr or below (A=0)	21.25 (3.38) ***	23.77 (4.26) ***
25 to 34 yr (A=1)	19.89 (2.88) ***	19.83 (3.53) ***
35 to 49 yr (A=2)	11.81 (2.26) ***	12.44 (2.79) ***
50 to 64 yr (A=3)	8.67 (1.97) ***	9.75 (2.44) ***
65 yr or above (A=4)	2.10 (2.28)	1.45 (2.80)
R-square	0.940	

Note. The table demonstrates the coefficient $\gamma^{T,A}$ in Eq. 3, which shows that there was a greater increase in search intensity of online education among females than males, among all age groups except people of 65 yr or above, during the lockdown phase and the "new normal" phase. *** $p < .01$, ** $p < .05$, * $p < .10$.

B. Query selection

In this section, we provide more details of the query selection.

As illustrated in the Measures section, we represent search intensity of online education as the share of a set of online-education-related queries out of all searches. Specifically, we collect searches of queries that contain keywords related to online education, including names of top online learning websites/platforms (<https://www.bestcollegereviews.org/50-top-online-learning-sites/> and <https://www.learnworlds.com/online-learning-platforms/>), popular online courses (<https://coursemethod.com/popular-online-courses.html>), and common terms related to online education and learning (<https://www.uni-prep.com/online-education/understanding-online-learning-terms/>). We eliminate the queries that contain keywords related to kid education, including "kid", "pre-k", "k-12" and names of kid online education programs (<https://codakid.com/11-best-online-education-programs-for-kids/>).

We also obtain search intensity of economic concerns (i.e., the share of queries containing keywords "unemployment", "layoff", "furlough", "economic relief", "stimulus check", "pay cut", "salary freeze" and "hiring freeze") to explore potential mechanisms.

C. Robustness checks

In Eq. 2, we may include fixed effect of each week instead of the effect of phase indicators ($Phase_t^1$ and $Phase_t^2$), as follows.

$$y_{i,j,t} = \mu + (\gamma^1 Phase_t^1 Female_j + \gamma^2 Phase_t^2 Female_j) + \sum_k \alpha_k Week_t^k + \sum_{A=1}^4 \omega_A Age_i^A + \rho Female_j + e_{i,j,t},$$

where k indexes each week and the dummy variable $Week_t^k = 1$ if $t = k$. We summarize the results in **Table B-1**, which yield similar conclusions.

Second, for DID analysis, a key assumption is that the observed effect did not occur before the pandemic. Following a widely used approach in the literature (3, 4), we split the key covariate into a series of time indicators to reveal effects over time. Specifically, we replace the interaction of phrase indicators and female indicator in Eq. 2 with the interaction of week indicators and female indicator, as follows.

$$y_{i,j,t} = \mu + \sum_k \gamma_k Week_t^k Female_j + \sum_k \alpha_k Week_t^k + \sum_{A=1}^4 \omega_A Age_i^A + \rho Female_j + e_{i,j,t}.$$

We use the week of March 2 – 8, 2020 as baseline, and plot the estimated coefficient γ_k in **Fig. B-1**.

The figures show that the effects (i.e., females show greater increase in searches of online education than males) did not occur before the pandemic, and started to occur immediately after the week of March 11. The results for searches of economic concerns show similar trends.

Third, we rule out seasonality effect by conducting similar analysis using data of the previous year (October 2018 – July 2019). Results of regression Eq. 1 and Eq. 2 are summarized in **Table B-2**, which show no significant increase in search intensity of online education or economic concerns in 2019, and females did not show higher increased search intensity of online education or economic concerns than males during March 2019 to July 2019. **Fig. 1** also confirms that the effect observed in 2020 during the pandemic did not occur in the same season in 2019.

Table B-1. Robustness check with week fixed effect in the DID model

Differential effect among females and males		
	Online Education	Economic Concerns
Lockdown Phase γ^1	11.09 (1.32) ***	15.26 (1.17) ***
“New Normal” Phase γ^2	11.45 (1.63) ***	10.78 (1.48) ***
R-square	0.915	0.877

Note. In this robustness check, we include fixed effect of each week instead of the effect of phase indicators. The coefficients γ^T are positive and significant for $T = 1$ and 2, which show that the increased searches of online education and economic concerns among females were significantly higher compared to males. Hence the model with week fixed effect yields similar conclusions to Eq. 2. *** $p < .01$, ** $p < .05$, * $p < .10$.

Table B-2. Search intensity patterns in the previous year (October 2018 – July 2019)

Overall search intensity		
	Online Education	Economic Concerns
Lockdown Phase α^1	-0.44 (0.52)	-1.02 (0.15) ***
“New Normal” Phase α^2	-0.61 (0.63)	-0.67 (0.18) ***
R-square	0.858	0.583

Differential effects among females and male		
Lockdown Phase γ^1	-0.25 (1.03)	-0.09 (0.30)
"New Normal" Phase γ^2	-0.22 (1.26)	-0.07 (0.36)
R-square	0.858	0.583

Note. In this robustness check, we rule out seasonality effect by conducting similar analysis (Eq. 1 and Eq. 2) using data of the previous year (October 2018 – July 2019). The coefficients α^T in the upper panel are the results of Eq. 1, and the coefficients γ^T in the lower panel are the results of Eq. 2. They show that in the previous year, the search intensity of online education or economic concerns did not show significant increase, and females did not show higher increased search intensity of online education or economic concerns than males during March 2019 to July 2019. *** $p < .01$, ** $p < .05$, * $p < .10$.

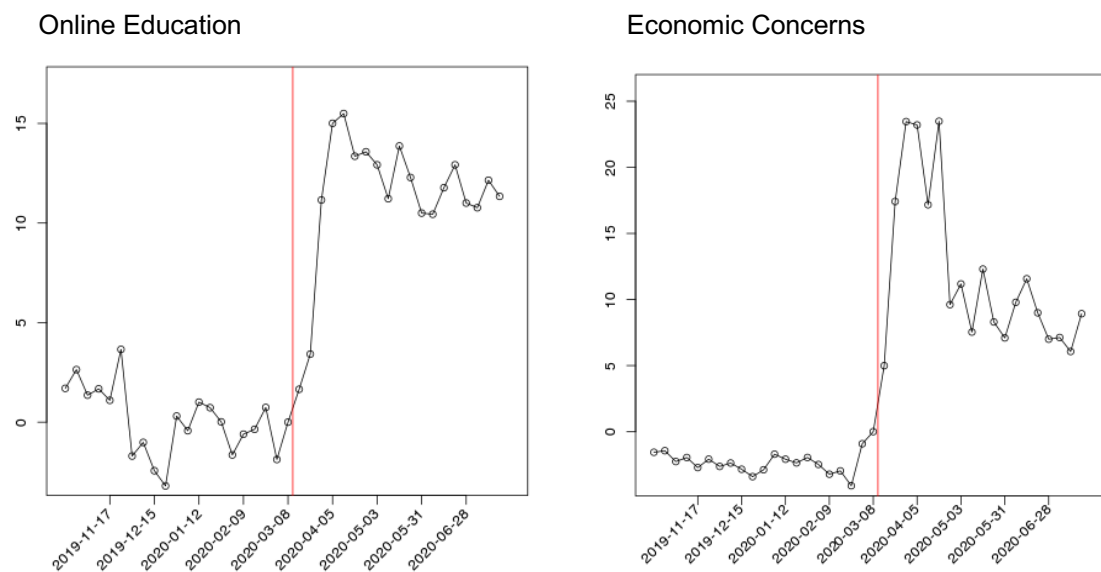


Figure B-1. DID assumption check for online education (left) and economic concerns (right). The figures depict the estimated coefficient of the interaction of week indicator and female indicator, which show that the effects (i.e., females show greater increase in searches of online education or economic concerns than males) did not occur before the pandemic, and started to occur immediately after the week of March 11. The two vertical red lines indicate the date of March 11, 2020, and June 1, 2020.

References

1. G. Saxonhouse, Estimated Parameters as Dependent Variables. *Am. Econ. Rev.* **66** (1976).
2. A. S. Hornstein, M. Zhao, Corporate capital budgeting decisions and information sharing. *J. Econ. Manag. Strateg.* **20** (2011).
3. J. D. Angrist, J. S. Pischke, *Mostly harmless econometrics: An empiricist's companion* (2008) <https://doi.org/10.1111/j.1475-4932.2011.00742>.
4. P. Wang, G. Xiong, J. Yang, Frontiers: Asymmetric effects of recreational cannabis legalization. *Mark. Sci.* **38** (2019).