Divorce during COVID-19 Pandemic: Google knows best?

Analysis of Divorce-related Google Queries during COVID-19 Pandemic and Nowcasting of Crude Divorce Rates in the United States.

Sansu Mathis

Institut National d'Études Démographiques

Séminaire café, January 2021

Table of Contents

1. Introduction & Motivation
2. Data
3. Analysis
4. Results
5. Limitations & Extensions
6. Conclusion

Table of Contents

1. Introduction & Motivation
2. Data
3. Analysis
4. Results
5. Limitations & Extensions
6. Conclusion

Divorce amid COVID-19 pandemic

- *China's Divorce Spike Is a Warning to Rest of Locked-Down World*, Bloomberg, March 2020, Sheridan Prasso
  https://www.bloomberg.com/news/articles/2020-03-31/
  divorces-spike-in-china-after-coronavirus-quarantines
- *US Divorce Rates Soar During COVID-19 Crisis*, Legal Templates, July 2020
  https:
  //legaltemplates.net/resources/personal-family/divorce-rates-covid-19/

Divorce amid COVID-19 pandemic

- COVID-19 pandemic: a global shock
  - traumatic event (life-threatening)
  - binding measures
  - uncertainty & stress
- Various sanitary measures (state level)
- Potential ambiguous effects on divorce-related search interest and on divorce rates

Divorce amid COVID-19 pandemic

- Current lack of systematic data at the aggregate level (delay to measure demographic phenomena)
- Nowcasting attempt (actual crude divorce rates)
- Seminal works:
  - Ginsberg, Mohebbi, Patel, Brammer, Smolinski, and Brilliant (2009)
  - Choi and Varian (2012)
  - Billari, D'Amuri, and Marcucci (2016)

Table of Contents

1. Introduction & Motivation
2. Data
3. Analysis
4. Results
5. Limitations & Extensions
6. Conclusion

Google Trends Data

Main characteristics:

- Internet data: important volume, immediate availability, anonymity
- Normalization procedure (Google): Search Interest Index (Relative Search Volume in the literature)
- Outlier problem (among other limitations)

Extracted Data

- January 2009 - October 2020
- Queries (and not topics)
- Keyword validity (Mellon (2014): face validity and content validity)
- Group of selected keywords: divorce, legal procedures, administrative features, financial traits, divorce questioning and a control set

Extracted Data

- Normalization procedure:
  Let $\tilde{R}_{st}$ be the observation provided by Google Trends for the variable $\tilde{R}$ in state s at date t. Let $\overline{\tilde{R}_s}$ be the mean of the series for the variable $\tilde{R}$ in state s. The normalized variable of interest is noted R_{st}.

$$R_{st} = \frac{\tilde{R}_{st}}{\overline{\tilde{R}_s}} * 100$$

- Outlier problem: subtraction command

Google Trends Data: «divorce» query

Figure 1: Search interest associated to the keyword «divorce»

Ground Data

- Divorce data: National Vital Statistics System, Center for Disease Control and Prevention (until 2018, not all states)
- Unemployment data: U.S. Bureau of Labor Statistics
- Population data: U.S. Census Bureau and Weldon Cooper Center, Demographics Research Group (University of Virginia)
- COVID-19 data: Johns Hopkins University Center for Systems Science and Engineering

Crude Divorce Rates Series

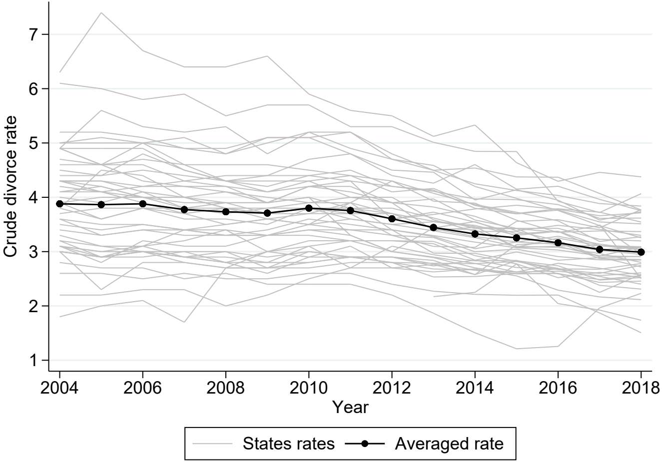

Figure 2: States crude divorce rates and averaged crude divorce rate

Table of Contents

1. Introduction & Motivation
2. Data
3. Analysis
4. Results
5. Limitations & Extensions
6. Conclusion

Theoretical Model

- Google Trends data: proxy of aggregate interest for the divorce issue
- Because of growing tensions among couples due to confinement measures[1,2], relative search volumes for divorce-related queries might increase during the lockdown and remain at a higher level than usual for a short period afterwards because of the postponement of divorce filings.
- The increase in unemployment[3] and in leisure time with the spouse may lead to a decrease in relative search volumes for divorce-related queries. Moreover the rise in Internet use during lockdown periods[4] might also induce a mechanical decrease.
- Ambiguous expectations.

---

[1] Biroli, Bosworth, Giusta, Girolamo, Jaworska, and Vollen, 2020.

[2] Leslie and Wilson, 2020.

[3] Amato and Beattie, 2011.

[4] Feldmann, Gasser, Lichtblau, Pujol, Poese, Dietzel, Wagner, Wichtlhuber, Tapiador, Vallina-Rodriguez, Hohlfeld, and Smaragdakis, 2020.

Lockdown and Pandemic «Treatments»

- Disentanglement of the lockdown effect and of the pandemic effect
- Variation regarding confinement measures (implementation and temporality)
- Spatio-temporal variation for the impact of the pandemic
- Threshold: not too low (no variation) and not too high (pandemic definition) $\longrightarrow$ 3,000 new confirmed cases per month and per million inhabitants of the state

Pandemic Threshold

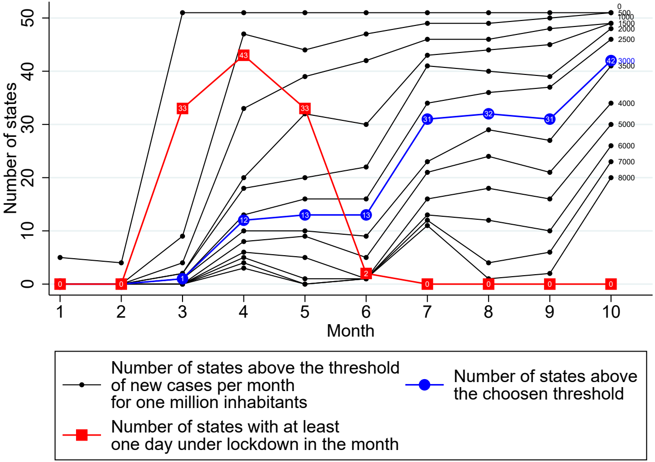

Figure 3: Number of states under lockdown and exceeding a threshold of new confirmed cases proportional to state population during the year

Event Study Analyses

- Influential works:
  - Brodeur, Clark, Fleche, and Powdthavee (2020)
  - Berger, Ferrari, Leturcq, Panico, and Solaz (2020)

,

- Events considered:
  - Statewide lockdown (SAHO-SIPO)
  - Exceeding of the pandemic threshold for states hit by the «first wave»
  - Exceeding of the pandemic threshold for states hit by the «second wave»

Event Study Analyses

- Let us denote t the number of months before or after the event. t is equal to 0 for the first month where we observe the corresponding event. R_{mys} is the value of the variable R in the state s for the month m and the year y. The month fixed-effects, year fixed-effects and state fixed-effects are respectively denoted by α_m, α_y and α_s.

$$R_{myst} = \beta_0 + \sum_{\tau=t_{pre}}^{t_{post}} \beta_\tau 1\{t = \tau\} + \alpha_m + \alpha_y + \alpha_s \qquad (1)$$

- Dynamic response relatively to the event

- Disentanglement of the lockdown and pandemic effects

Triple-difference (DDD) analysis

- Let us denote L and P be respectively the lockdown treatment and pandemic treatment. The tripe-difference model is written as follows:

$$R_{mys} = \beta_0 + \beta_1 L + \beta_2 P + \beta_3 L * P + \alpha_m + \alpha_y + \alpha_s + \epsilon_{mys} \tag{2}$$

- Pandemic «treatment»: definition of subgroups $\longrightarrow$ months impacted by the pandemic (for each state)
- Differencing strategy on the pandemic «treatment»: sort of double difference-in-difference
- Avoid an endogeneity problem since the pandemic influences both the dependent variable (the search rates) and the independent variable (the lockdown)
- Adequate identification if enough variation between states for the treatments

Common Trend Assumption: «child support» query

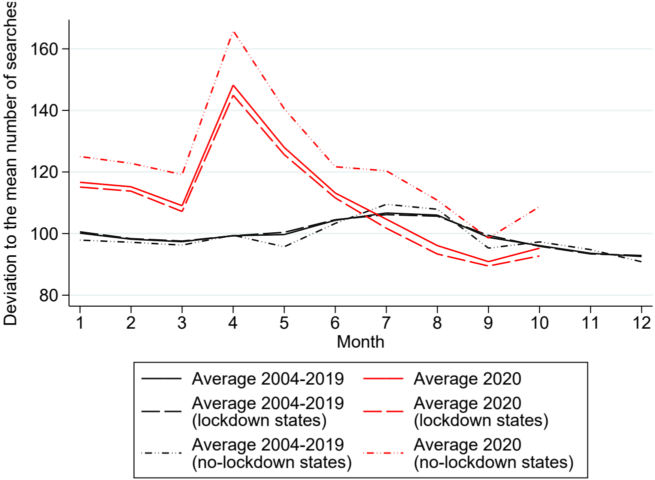

Figure 4: Searches associated to the combined keywords «child support» and averaged time series for the states with and without a lockdown

Other Treatment & Robustness Checks

Treatment: DDD analysis with heterogeneous effects for states of different cohorts (first or second wave).

Robustness checks:

- Sampling period: since 2004
- Regression without population weight (state interpretation)
- Addition of trends (linear and linear-quadratic)
- Control variables: number of COVID-19 cases and deaths
- Placebo treatments to replicate the analysis in absence of events (for years 2018 and 2019)

Static Panel Data Model

- Hypothesis: correlation between Internet behaviours and marital behaviours
- Nowcasting attempt
- Aggregation issue (with monthly lags allowed):

$$\begin{cases} R^a_{syl} = \dfrac{1}{12} \times \displaystyle\sum_{t=l-11}^{l} R_{syt} & \text{for the years 2004 to 2019} \\ R^a_{syl} = \dfrac{1}{10} \times \displaystyle\sum_{t=l-9}^{l} R_{syt} & \text{for the year 2020} \end{cases}$$

Static Panel Data Model

- Influential works:
    - Billari, D'Amuri, and Marcucci (2016)
    - Wilde, Chen, and Lohmann (2020)

- Cross-validation method: Mean Squared Predicted Error (MSPE) on each 2-year window (iterative process to avoid overfitting)

$$MSPE := \frac{1}{n} \sum_{i=1}^{n} \left(D_i - \hat{D}_i \right)^2$$

- Data-driven approach limited by the keyword validity step at the beginning (*cf.* Google Flu Trends)

Static Panel Data Model

- Let us denote D_{sy} the crude divorce rate for the state s in year y and U^{l}_{sy} the annual average for the monthly unemployment rates of state s in year y with a lag l. X is a set of control variables and W is the set of selected keywords after the cross-validation exploration.

$$D_{sy} = \sum_{w \in W} \beta_w R_{syw} + \beta_U U^{l}_{sy} + \beta_X X_{sy} + \alpha_s + \epsilon_{sy} \tag{3}$$

- Selected variables: «divorce», «divorce law » with a 7-month lag, «divorce» with a 11-month lag, annual average of the monthly unemployment rates

- Control variables: sex ratio and median age

Table of Contents

1. Introduction & Motivation
2. Data
3. Analysis
4. Results
5. Limitations & Extensions
6. Conclusion

Event Study Model: «divorce» query relatively to the lockdown

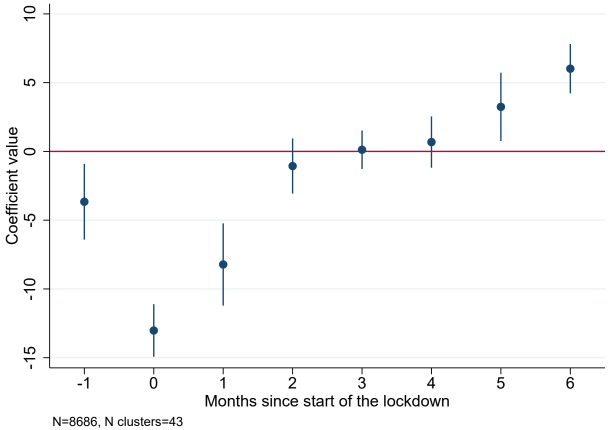

Figure 5: Event study analysis for the query «divorce» relatively to the lockdown

Event Study Model: «child support» query relatively to the lockdown

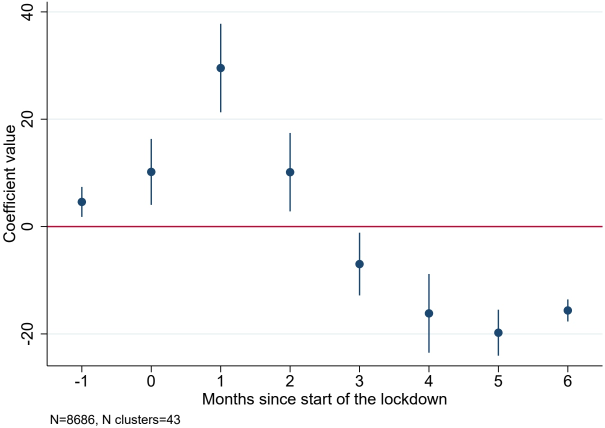

Figure 6: Event study analysis for the query «child support» relatively to the lockdown

Event Study Models

	«lockdown»	«divorce»	«divorce law»	«child support»	«alimony»
-1	975.3*	-3.659*	-1.190	4.593**	-0.741
	(404.5)	(1.364)	(2.023)	(1.386)	(5.015)
0	3121.5***	-13.02***	-12.30***	10.18**	-22.93***
	(305.6)	(0.948)	(1.892)	(3.045)	(2.498)
1	1098.0***	-8.222***	-12.44***	29.54***	-22.59***
	(99.83)	(1.482)	(1.429)	(4.087)	(3.753)
2	704.6***	-1.064	-7.969***	10.13**	-12.19**
	(76.06)	(0.994)	(1.876)	(3.621)	(3.868)
3	269.4***	0.119	-6.090**	-6.987*	-1.076
	(33.51)	(0.694)	(2.073)	(2.895)	(3.963)
4	167.9**	0.678	-4.351**	-16.17***	-8.101*
	(49.84)	(0.925)	(1.571)	(3.636)	(3.328)
5	71.80***	3.238*	-2.375	-19.78***	-8.573**
	(16.57)	(1.230)	(1.447)	(2.123)	(3.146)
6	64.56***	6.016***	-3.263	-15.63***	1.961
	(14.41)	(0.888)	(2.314)	(1.019)	(3.453)
Cons.	7.599	102.5***	111.8***	78.61***	108.4***
	(6.623)	(3.024)	(3.923)	(3.348)	(4.307)
R^2	0.757	0.568	0.663	0.331	0.133
Obs.	6106	6106	5822	6106	6106

Standard errors in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1: Event Study on GT Indicators relatively to the lockdown

Event Study Results

- Important lockdown effect: transitory with repercussions in the mid-term
- Stock-flux interpretation: search rates render the evolution of the amount of information shared in the society $\longrightarrow$ catch-up effect (postponement) for instance
- Information effect: novelty and media coverage of the lockdown $\longrightarrow$ shift the rest
- Overwhelming financial stress: cost of divorce perspective

DDD Models

	«lockdown»	«divorce»	«divorce law»	«child support»
Lock.	1263.1***	-8.960***	-9.958***	35.60***
	(195.6)	(0.978)	(1.270)	(6.135)
Pand.	-381.0***	3.476***	-0.439	-10.76***
	(97.54)	(0.832)	(1.306)	(2.980)
Lock. * Pand.	-15.87	-3.573	-4.703	-10.95
	(225.4)	(2.398)	(3.005)	(11.09)
Cons.	19.45**	102.5***	122.0***	79.14***
	(7.109)	(2.927)	(3.754)	(3.271)
R^2	0.456	0.558	0.652	0.313
Obs.	7242	7242	6958	7242

Standard errors in parentheses — * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2: DDD on GT Indicators (1)

DDD Models

	«alimony»	«divorce how»	«divorce papers»	«divorce court»
Lock.	-23.43***	-14.13***	-23.43***	-10.09***
	(2.995)	(1.947)	(3.579)	(2.133)
Pand.	-1.977	-0.456	-6.136	5.209**
	(2.585)	(1.180)	(3.624)	(1.662)
Lock. * Pand.	5.185	-2.265	-3.416	-18.74***
	(6.312)	(3.797)	(5.765)	(4.753)
Cons.	109.8***	90.26***	116.5***	95.16***
	(4.316)	(3.314)	(6.768)	(3.985)
R^2	0.120	0.188	0.138	0.145
Obs.	7242	7242	6532	7100

Standard errors in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: DDD on GT Indicators (2)

DDD Results

- Confirmation of a lockdown effect on divorce-related queries
- Early reactions possibilities $\longrightarrow$ mostly focused on financial issues
- Overall search volume problem
- Global reaction to the events (and not especially state reactions)

Panel Prediction

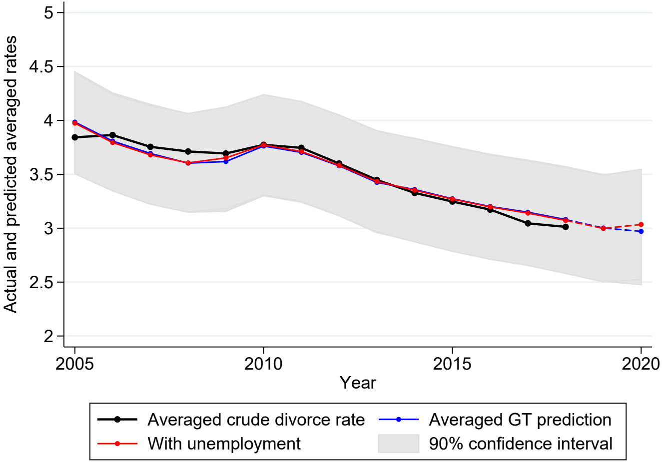

Figure 7: Nowcasting of the average crude divorce rates in the United States

Panel Results

- Not causal inferences
- Predictions higher than the trend of the series
- GT indicators push downward the predicted points while unemployment rates push them upwards
- Link between Internet behaviours and marital behaviours in period of crisis?

Table of Contents

1 Introduction & Motivation

2 Data

3 Analysis

4 Results

5 Limitations & Extensions

6 Conclusion

Google Trends Data Limitations

- Sparsity issues:
  - Threshold of absolute number of queries
  - Series of missing values
- Representativeness problem
  - Distribution of Internet use among demographic groups
  - Google is only one search engine
  - Time specificity of keyword validity
- Volume problems
  - Relative Search Volumes when overall volume remains unknown
  - Outliers
  - Media coverage bias
- Google black box and commercial algorithm (Lazer, Kennedy, King, and Vespignani (2014))

Extensions

- Refined divorce rate (number of divorces per 1,000 women married to men)
- Accounting for the stringency of measures *e.g.* lockdown stringency
- Adopt a double event-study approach?
- Handling mixed-frequency data (MIDAS regression)
- Handling infra-seasonality thanks to this type of data

Table of Contents

1 Introduction & Motivation

2 Data

3 Analysis

4 Results

5 Limitations & Extensions

6 Conclusion

Conclusion

- Google Trends data: interesting but limited avenues
- Lockdown effect on divorce-related queries
- Small rise of crude divorce rates?
- Necessity of a more robust framework to draw conclusions for demographic issues
- COVID-19 pandemic unfinished: long-term consequences

Thanks!