Measuring Crime Reporting and Incidence: Method and Application to #MeToo

Germain Gauthier

CREST - Ecole Polytechnique

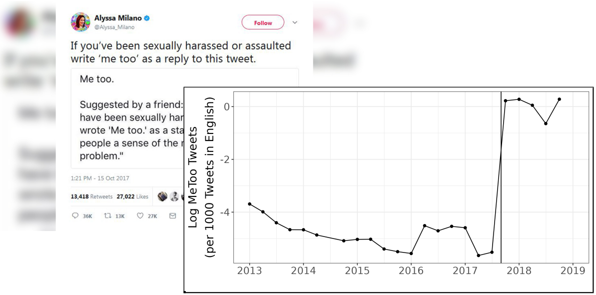
October 31, 2022

The Me Too Movement

In October 2017, the Me Too movement led millions of women worldwide to protest against sexual violence.

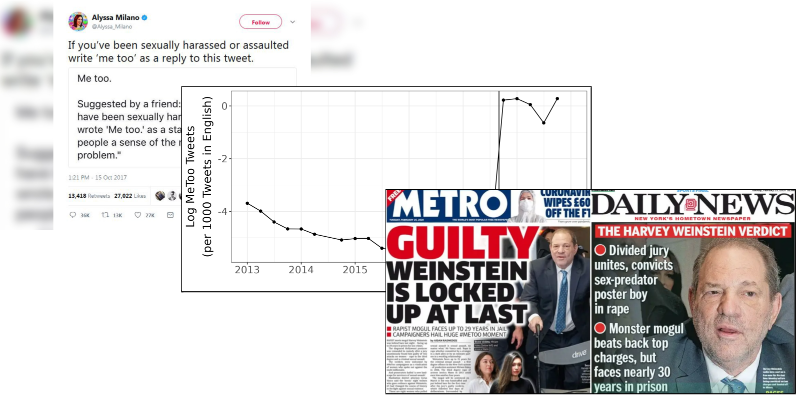
The Me Too Movement

In October 2017, the Me Too movement led millions of women worldwide to protest against sexual violence.



The Me Too Movement

In October 2017, the Me Too movement led millions of women worldwide to protest against sexual violence.



The Me Too Movement

- The Me Too movement is arguably the largest public awareness campaign against sexual violence in history.
- Enthusiastic commentators portrayed the movement as a game-changer for women's rights.
- Others, more skeptical, raised concerns about false allegations, backlash effects, and socioeconomic and racial divides.

Has the Me Too movement been successful in combatting sex criminality?

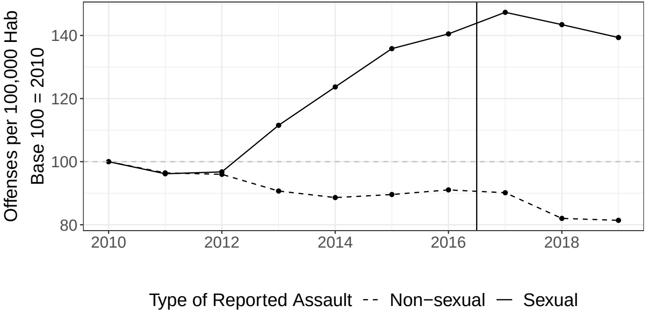
Empirical Challenge

In the United States, national surveys estimate:

$\approx$ 300,000 victims of sexual violence per year.

$\approx$ 70% do not report to the police.

$\rightarrow$ Reported crimes reflect variations in crime incidence *and* reporting (Quêtelet, 1831).



This paper proposes a new method...

- I decompose a time-series of reported crimes into variations in victim reporting *and* crime incidence.
- More than half of sex crimes are reported with a **delay**.
- (Modified) mixed proportional hazards **duration model**.
- Maximum likelihood estimation, but non-trivial likelihood because of **double-truncation** in the data.
- Validation via Monte Carlo simulations.

This paper proposes a new method...

- I decompose a time-series of reported crimes into variations in victim reporting *and* crime incidence.
- More than half of sex crimes are reported with a **delay**.
- (Modified) mixed proportional hazards **duration model**.
- Maximum likelihood estimation, but non-trivial likelihood because of **double-truncation** in the data.
- Validation via Monte Carlo simulations.

This paper proposes a new method...

- I decompose a time-series of reported crimes into variations in victim reporting *and* crime incidence.
- More than half of sex crimes are reported with a **delay**.
- (Modified) mixed proportional hazards **duration model**.
- Maximum likelihood estimation, but non-trivial likelihood because of **double-truncation** in the data.
- Validation via Monte Carlo simulations.

This paper proposes a new method...

- I decompose a time-series of reported crimes into variations in victim reporting *and* crime incidence.
- More than half of sex crimes are reported with a **delay**.
- (Modified) mixed proportional hazards **duration model**.
- Maximum likelihood estimation, but non-trivial likelihood because of **double-truncation** in the data.
- Validation via Monte Carlo simulations.

This paper proposes a new method...

- I decompose a time-series of reported crimes into variations in victim reporting *and* crime incidence.
- More than half of sex crimes are reported with a **delay**.
- (Modified) mixed proportional hazards **duration model**.
- Maximum likelihood estimation, but non-trivial likelihood because of **double-truncation** in the data.
- Validation via Monte Carlo simulations.

... with an application to #MeToo.

- I use incident-level police records for New York City, Los Angeles, Seattle, and Cincinnati, between 2010 and 2020.
- I estimate **trends** in sex crime reporting and incidence.
- I assess #MeToo's impact on sex crime **reporting**...

  I compare the length of delayed reports to the police before/after #MeToo (conditional on controls and time-invariant unobservables).
- ... and on sex crime **incidence**.

  I compare estimated sex crimes with reported non-sexual crimes in a difference-in-differences setup.

... with an application to #MeToo.

- I use incident-level police records for New York City, Los Angeles, Seattle, and Cincinnati, between 2010 and 2020.
- I estimate **trends** in sex crime reporting and incidence.
- I assess #MeToo's impact on sex crime **reporting**...

  I compare the length of delayed reports to the police before/after #MeToo (conditional on controls and time-invariant unobservables).
- ... and on sex crime **incidence**.

  I compare estimated sex crimes with reported non-sexual crimes in a difference-in-differences setup.

... with an application to #MeToo.

- I use incident-level police records for New York City, Los Angeles, Seattle, and Cincinnati, between 2010 and 2020.
- I estimate **trends** in sex crime reporting and incidence.
- I assess #MeToo's impact on sex crime **reporting**...

  I compare the length of delayed reports to the police before/after #MeToo (conditional on controls and time-invariant unobservables).
- ... and on sex crime **incidence**.

  I compare estimated sex crimes with reported non-sexual crimes in a difference-in-differences setup.

... with an application to #MeToo.

- I use incident-level police records for New York City, Los Angeles, Seattle, and Cincinnati, between 2010 and 2020.
- I estimate **trends** in sex crime reporting and incidence.
- I assess #MeToo's impact on sex crime **reporting**...

  I compare the length of delayed reports to the police before/after #MeToo (conditional on controls and time-invariant unobservables).
- ... and on sex crime **incidence**.

  I compare estimated sex crimes with reported non-sexual crimes in a difference-in-differences setup.

Preview of Results

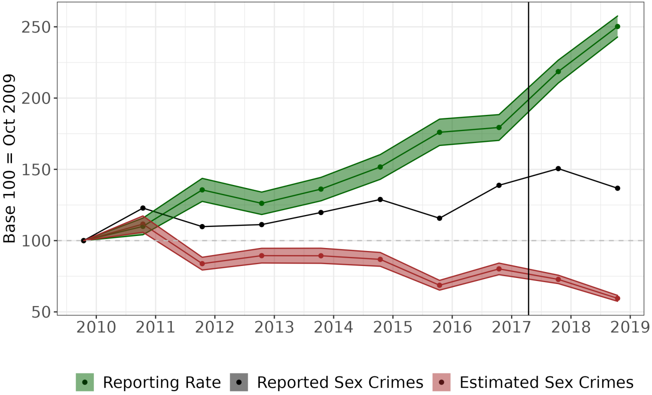


Figure: Trends in Sex Crime Incidence and Reporting

Preview of Results

- Crime reporting and incidence partly **cancel each other out** in the time series of reported crimes.
  - Corroborated by alternative proxies.
- #MeToo **increased** the reporting rate (+25%).
  - Juveniles, racial minorities, and victims of old crimes and misdemeanors are particularly responsive to the intervention.
  - Robust to alternative specifications for the baseline hazard, unobserved heterogeneity, and polynomial time-trends.
  - No effect for placebo dates and negligible effects for placebo crimes.
- #MeToo had a large **deterrent effect** ($\approx$ -20%).
  - Robust to alternative counterfactual models.
  - No effect for placebo dates and placebo crimes.

Preview of Results

- Crime reporting and incidence partly **cancel each other out** in the time series of reported crimes.
  - Corroborated by alternative proxies.
- #MeToo **increased** the reporting rate (+25%).
  - Juveniles, racial minorities, and victims of old crimes and misdemeanors are particularly responsive to the intervention.
  - Robust to alternative specifications for the baseline hazard, unobserved heterogeneity, and polynomial time-trends.
  - No effect for placebo dates and negligible effects for placebo crimes.
- #MeToo had a large **deterrent effect** ($\approx$ -20%).
  - Robust to alternative counterfactual models.
  - No effect for placebo dates and placebo crimes.

Preview of Results

- Crime reporting and incidence partly **cancel each other out** in the time series of reported crimes.
  - Corroborated by alternative proxies.
- #MeToo **increased** the reporting rate (+25%).
  - Juveniles, racial minorities, and victims of old crimes and misdemeanors are particularly responsive to the intervention.
  - Robust to alternative specifications for the baseline hazard, unobserved heterogeneity, and polynomial time-trends.
  - No effect for placebo dates and negligible effects for placebo crimes.
- #MeToo had a large **deterrent effect** ($\approx$ -20%).
  - Robust to alternative counterfactual models.
  - No effect for placebo dates and placebo crimes.

Methodological Contributions

- **Measurement of Crime**
  - Quêtelet (1831); Coleman and Moynihan (1996); Levitt (1998); Aizer (2010); Stephens-Davidowitz (2013); Bellégo and Drouard (2019)

  → Underreporting is a real threat to empirical studies on crime.

  → New model to disentangle crime incidence and reporting from police data.

- **Double-truncation in Survival Analysis**
  - Van den Berg (2001); Abbring and Van den Berg (2003); Dörre and Emura (2019); Vakulenko-Lagun et al. (2019); Rennert and Xie (2018); Mandel et al. (2018)

  → Mixed proportional hazards model under double-truncation.

Methodological Contributions

- **Measurement of Crime**
  - Quêtelet (1831); Coleman and Moynihan (1996); Levitt (1998); Aizer (2010); Stephens-Davidowitz (2013); Bellégo and Drouard (2019)

→ Underreporting is a real threat to empirical studies on crime.

→ New model to disentangle crime incidence and reporting from police data.

- **Double-truncation in Survival Analysis**
  - Van den Berg (2001); Abbring and Van den Berg (2003); Dörre and Emura (2019); Vakulenko-Lagun et al. (2019); Rennert and Xie (2018); Mandel et al. (2018)

→ Mixed proportional hazards model under double-truncation.

Substantive Contributions

- **Sexual Violence**
  - Iyer et al. (2012); Miller and Segal (2019); Bottan and Perez-Truglia (2015); Mathur et al. (2019); Sahay (2021); McDougal et al. (2021); Levy and Mattsson (2021)

→ New insights on the Me Too movement.

→ Public awareness campaigns can work.

- **Crime Deterrence**
  - Becker (1968); Nagin (2013); Chalfin and McCrary (2017); Doleac (2019); Hay and Shleifer (1998); Berkowitz et al. (2003); Benabou and Tirole (2011); Young (2015); Akerlof and Yellen (1994); Dyck et al. (2010); Acemoglu and Jackson (2017)

→ Empirical evidence of a deterrent effect through increased reporting.

→ Norms-based interventions can enforce socially desirable behaviors when the legal system fails to do so on its own.

Substantive Contributions

- **Sexual Violence**
  - Iyer et al. (2012); Miller and Segal (2019); Bottan and Perez-Truglia (2015); Mathur et al. (2019); Sahay (2021); McDougal et al. (2021); Levy and Mattsson (2021)

  → New insights on the Me Too movement.

  → Public awareness campaigns can work.

- **Crime Deterrence**
  - Becker (1968); Nagin (2013); Chalfin and McCrary (2017); Doleac (2019); Hay and Shleifer (1998); Berkowitz et al. (2003); Benabou and Tirole (2011); Young (2015); Akerlof and Yellen (1994); Dyck et al. (2010); Acemoglu and Jackson (2017)

  → Empirical evidence of a deterrent effect through increased reporting.

  → Norms-based interventions can enforce socially desirable behaviors when the legal system fails to do so on its own.

Table of Contents

Data

I collect data between 2010 and 2020.

- **City-level police datasets** ▸ Additional Details
  - New York City, Los Angeles, Seattle, and Cincinnati
  - Incident-level complaint data
  - Offender-level arrest data (for NYC and LA)
- **FBI national consolidated databases**
  - Uniform Crime Reporting Program
  - Supplementary Homicide Reports
- **National Crime Victimization Survey (NCVS)**
- **Tweets on sex crimes**
  - Tweets for the 1st of each month
- **Google queries on sex crimes**

Setup

- Denote τ_1 and τ_2 respectively the first and last calendar date of data collection.
- Number of crimes committed in period t, reported between τ_1 and τ_2 (observed): R_{t,τ_1,τ_2}
- Binary treatment implemented in period $t^* \in [\tau_1, \tau_2]$ (observed): D_t
- Number of crimes committed in period t (unobserved): C_t

Objective: Estimate the effect of D_t on C_t.

Classic Omitted Variable Bias

- Assume **no delayed reports** and denote r_t the reporting rate.
- Then, we have

$$R_{t,\tau_1,\tau_2} = r_t \times C_t.$$

- Linear regressions face an omitted variable bias:

$$\log(C_t) = \beta_0 + \beta_1 D_t + \underbrace{u_t}_{\varepsilon_t - \log\left(r_t\right)}.$$

But delayed reports are the norm, not the exception.

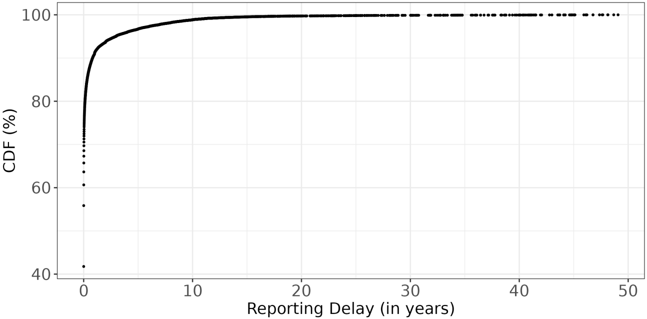


Figure: Distribution of Reporting Delays for Sex Crimes

But delayed reports are the norm, not the exception.

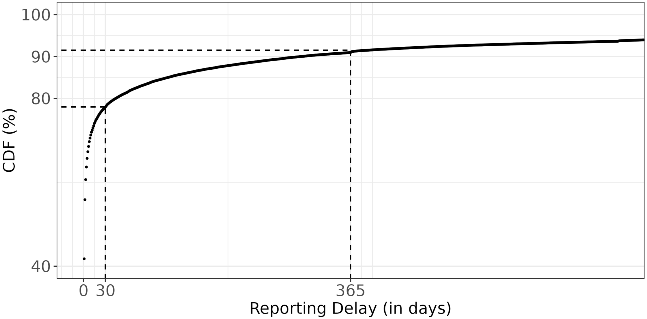


Figure: Distribution of Reporting Delays for Sex Crimes

Factoring in Delayed Reports

- If reports are indexed by the **date of the incident** t, we have

$$\mathbb{E}\Big[R_{t,\tau_1,\tau_2}\Big] = p_t \times \mathbb{E}\Big[C_t\Big],$$

where:

- $p_t = F\Big(max(\tau_2 - t, 0) \mid \chi_t\Big) - F\Big(max(\tau_1 - t, 0) \mid \chi_t\Big)$ is the probability of reporting a crime committed at date t within the study period.
- F is the cumulative distribution function of delays Y.
- χ_t is the history of interventions until time t.

Solution: Estimate the distribution of Y to compute p_t.

▸ Additional issues with delayed reports.

Table of Contents

The Structure of (Delayed) Crime Reports

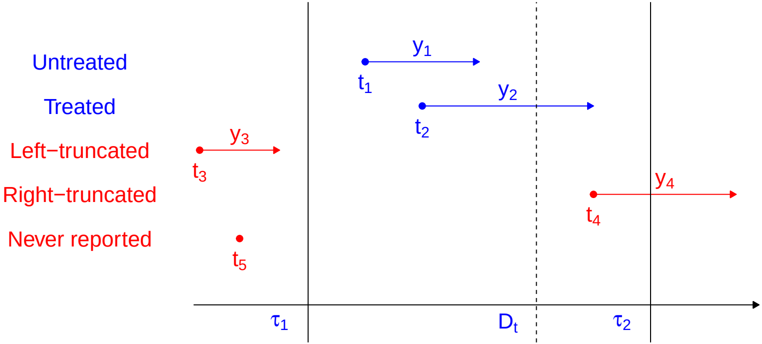


Figure: The Structure of (Delayed) Crime Reports

Notes: Graphical representation of delays in crime reporting. Blue incidents are observed. Red incidents are unobserved.

Hazard of Reporting of Plaintiffs

- **Mixed proportional hazards model** (Van den Berg, 2001)

$$h_{it}^{(p)}\left(y_i \mid \gamma_i,\ x_{ity}\right) = h_0(y_i)\exp(\beta' x_{ity})\gamma_i,$$

where:

- $h_0 : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is the baseline hazard function.
- $x_{ity} \in \mathbb{R}^d$ is a vector of observed covariates (e.g., #MeToo).
- $(\beta') \in \mathbb{R}^{d+1}$ is the vector of regression coefficients.
- $\gamma_i \in \mathbb{R}$ is an individual-specific frailty term.

Hazard of Reporting of Plaintiffs

- **Mixed proportional hazards model** (Van den Berg, 2001)

$$h_{it}^{(p)}\left(y_i \mid \gamma_i,\ x_{ity}\right) = \underbrace{h_0(y_i)}_{\text{baseline hazard}} \exp(\beta' x_{ity})\gamma_i,$$

where:

- $h_0 : \mathbb{R}^+ \to \mathbb{R}^+$ is the baseline hazard function.
- $x_{ity} \in \mathbb{R}^d$ is a vector of observed covariates (e.g., #MeToo).
- $(\beta') \in \mathbb{R}^{d+1}$ is the vector of regression coefficients.
- $\gamma_i \in \mathbb{R}$ is an individual-specific frailty term.

Hazard of Reporting of Plaintiffs

- **Mixed proportional hazards model** (Van den Berg, 2001)

$$h_{it}^{(p)}\big(y_i \mid \gamma_i,\ x_{ity}\big) = h_0(y_i) \underbrace{\exp(\beta' x_{ity})}_{\text{effects of covariates}} \gamma_i,$$

where:

- $h_0 : \mathbb{R}^+ \to \mathbb{R}^+$ is the baseline hazard function.
- $x_{ity} \in \mathbb{R}^d$ is a vector of observed covariates (e.g., #MeToo).
- $(\beta') \in \mathbb{R}^{d+1}$ is the vector of regression coefficients.
- $\gamma_i \in \mathbb{R}$ is an individual-specific frailty term.

Hazard of Reporting of Plaintiffs

- **Mixed proportional hazards model** (Van den Berg, 2001)

$$h_{it}^{(p)}\left(y_i \mid \gamma_i,\ x_{ity}\right) = h_0(y_i)\exp(\beta' x_{ity}) \underbrace{\gamma_i}_{\text{unobserved heterogeneity}},$$

  where:
  - $h_0 : \mathbb{R}^+ \to \mathbb{R}^+$ is the baseline hazard function.
  - $x_{ity} \in \mathbb{R}^d$ is a vector of observed covariates (e.g., #MeToo).
  - $(\beta') \in \mathbb{R}^{d+1}$ is the vector of regression coefficients.
  - $\gamma_i \in \mathbb{R}$ is an individual-specific frailty term.
- But what about victims who do not report to the police?

Hazard of Reporting of Victims

- I modify the MPH model to account for **never-reporters**:

$$h_{it}^{(v)}\left(y_i \,|\gamma_i,\ x_{ity}\right) = f_0(y_i)\exp\left(\alpha + \beta' x_{ity}\right)\gamma_i,$$

where:
  - f_0 is a density function.
  - $\alpha \in \mathbb{R}$ is the share of never-reporters at baseline and is not identified.

Hazard of Reporting of Victims

- I modify the MPH model to account for **never-reporters**:

$$h_{it}^{(v)}\left(y_i \mid \gamma_i,\ x_{ity}\right) = \underbrace{f_0(y_i)}_{\substack{\text{baseline hazard}\\ \text{is a density}}} \exp\left(\alpha + \beta' x_{ity}\right)\gamma_i,$$

where:

- f_0 is a density function.
- $\alpha \in \mathbb{R}$ is the share of never-reporters at baseline and is not identified.

Hazard of Reporting of Victims

- I modify the MPH model to account for **never-reporters**:

$$h_{it}^{(v)}\left(y_i \,|\gamma_i,\; x_{ity}\right) = f_0(y_i)\underbrace{\exp\left(\alpha + \beta' x_{ity}\right)}_{\substack{\text{plugin parameter}\\ \text{for never-reporters}}}\gamma_i,$$

where:

- f_0 is a density function.
- $\alpha \in \mathbb{R}$ is the share of never-reporters at baseline and is not identified.

Hazard of Reporting of Victims

- I modify the MPH model to account for **never-reporters**:

$$h_{it}^{(v)}\left(y_i \,|\gamma_i,\; x_{ity}\right) = f_0(y_i)\exp\left(\alpha + \beta' x_{ity}\right)\gamma_i,$$

where:

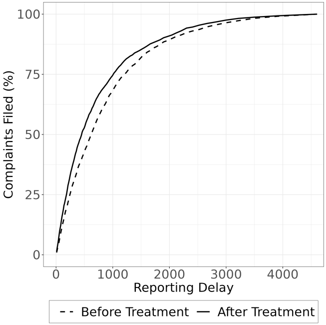
  - f_0 is a density function.
  - $\alpha \in \mathbb{R}$ is the share of never-reporters at baseline and is not identified.

- In the simplest case of no heterogeneity, the **reporting rate** is

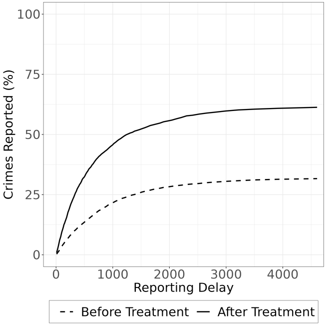
$$\lim_{y\to\infty} F^{(v)}(y) = 1 - \exp\Big(-\exp(\alpha)\Big).$$

▸ How restrictive is the model?

Intuition



α such that $F^{(v)}(+\infty)=30\%$



Plaintiffs

- Observed in police datasets
- Lower bound on treatment effects

Victims

- Estimates of the reporting rate
- Unbiased treatment effects

Estimation

- Maximum likelihood estimation where observations are **weighted** by the inverse of their **sampling probability**.
- Under double-truncation, the density of each data point (u_i, y_i) is

$$P(U = u_i,\ Y = y_i \,|\, U \leq Y \leq U + d),$$

where:

  - $d = \tau_2 - \tau_1$ is the length of the study period.
  - $U = \max(0, \tau_1 - T)$ is the left-truncation time.
  - T is the date of the incident.

- If we assume ***conditional* independence** between Y and T, the likelihood eventually simplifies to

$$L(\Theta) = \prod_{i=1}^{n} \frac{f(y_i)}{\displaystyle\int_{u_i}^{u_i+d} f(y)dy}.$$

Validation via Monte Carlo Simulations

- I simulate realistic time series of reported crimes to the police.
- Three interventions affect crime reporting *and* incidence.

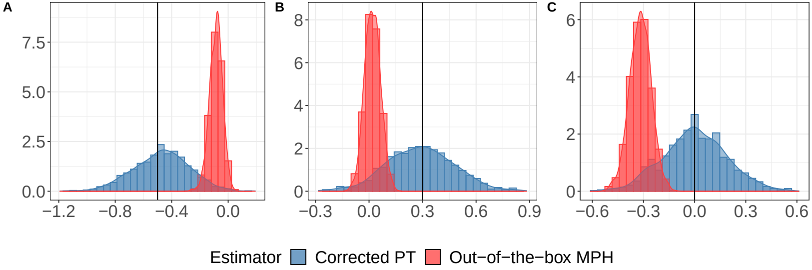


Figure: Density of Estimates for the Effects of the Interventions

Notes: I simulate 1000 datasets. The solid vertical line is the “true” treatment effect.

Table of Contents

Plaintiffs report faster to the police after #MeToo.

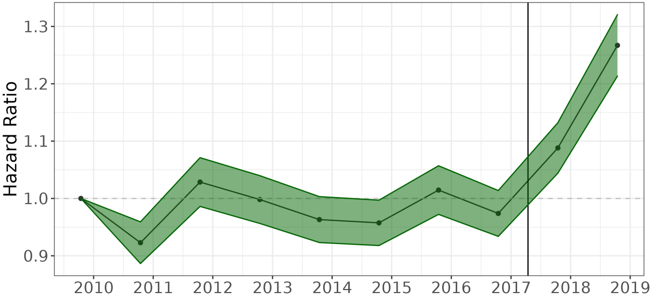


Figure: Yearly Hazard Ratios – Lower Bounds

▸ Specification ▸ Naive Estimator ▸ By Quarter ▸ Robustness ▸ Heterogeneity

This implies that reported sex crimes hide massive trends...

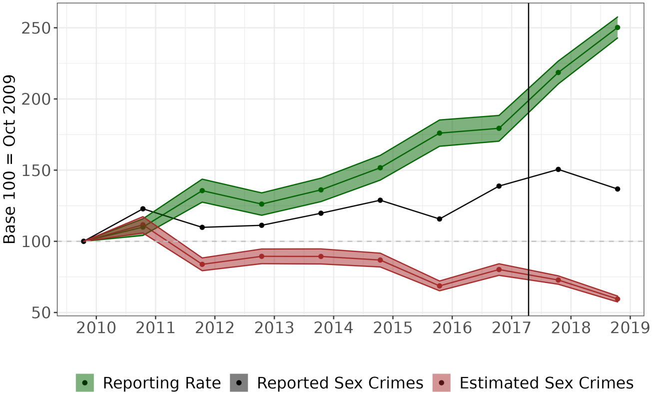


Figure: Trends in Sex Crime Incidence and Reporting

▸ Specification ▸ Nominal values ▸ Alternative alphas ▸ Unfounded allegations ▸ Time-dependent Effects

▸ Delayed Reports ▸ Survey Estimates ▸ Homicides ▸ Twitter ▸ Google

... and that #MeToo had a large deterrent effect.

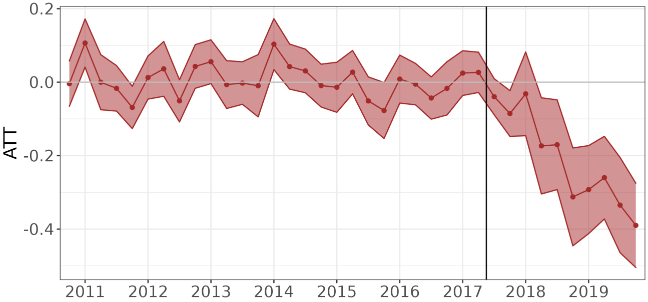


Figure: Interactive Fixed Effects Model – ATTs

▸ Specification ▸ DID ▸ Matrix Completion ▸ Robustness ▸ Probability of arrest

Table of Contents

Concluding Remarks

- The considerable uncertainty surrounding (sex) crime incidence raises many empirical issues.
- I leveraged delayed reports to disentangle crime incidence and reporting in police data.
- The reporting, incidence, and public discourse regarding sexual violence had been changing for over a decade in the United States.
- #MeToo reinforced these broader trends.
- The movement's long-run effects on institutions remain unknown and represent an opportunity for future research.

Measuring Crime Reporting and Incidence: Method and Application to #MeToo

Germain Gauthier

CREST - Ecole Polytechnique

October 31, 2022

Table of Contents

Sex crimes are often reported with a delay.

Characteristic	Sex Crime[1]	Murder[1]	Assault[1]	Robbery[1]	Burglary[1]
N	110,591	7,478	1,239,729	295,097	536,312
Report Type					
Delayed	58%	12%	21%	17%	54%
Direct	42%	88%	79%	83%	46%
Time to Report (days)					
Mean	197.19	105.47	4.40	2.68	6.05
Median	2.00	1.00	1.00	1.00	2.00
SD	857.99	948.55	57.12	52.73	62.07

[1] N; %

Figure: Delayed Reporting by Crime Type

Go Back

Sex crimes are largely gender-specific, targeted at minorities and juveniles.

Characteristic	Sex Crime[1]	Murder[1]	Assault[1]	Robbery[1]	Burglary[1]
N	110,591	7,478	1,239,729	295,097	536,312
Victim Sex					
Female	87%	17%	53%	30%	46%
Male	13%	83%	47%	70%	54%
Unknown	5,709	297	110,790	36,111	125,031
Victim Age					
Adult	57%	92%	91%	85%	95%
Juvenile	43%	7.5%	9.0%	15%	4.5%
Unknown	2,381	102	15,952	10,089	19,145
Victim Race					
White	22%	9.9%	16%	23%	42%
Black	40%	67%	48%	40%	37%
Hispanic	38%	23%	36%	37%	20%
Unknown	43,216	2,858	571,717	144,585	413,992

[1] N; %

Figure: Victim Characteristics by Crime Type

Offenders are mainly adult males belonging to a minority.

Characteristic	Sex Crime[1]	Murder[1]	Assault[1]	Robbery[1]	Burglary[1]
N	110,591	7,478	1,239,729	295,097	536,312
Supect Sex					
Female	8.0%	7.8%	25%	6.4%	8.7%
Male	92%	92%	75%	94%	91%
Unknown	57,595	4,641	730,803	143,505	498,513
Suspect Age					
Adult	97%	98%	98%	97%	100%
Juvenile	3.4%	2.3%	2.2%	3.2%	0.3%
Unknown	8,485	735	94,134	44,055	46,622
Suspect Race					
White	14%	7.7%	11%	5.5%	16%
Black	49%	65%	57%	73%	61%
Hispanic	37%	27%	32%	21%	23%
Unknown	66,307	4,764	782,366	152,000	502,594

[1] N; %

Figure: Suspect Characteristics by Crime Type

Go Back

Delays vary across socio-demographic groups.

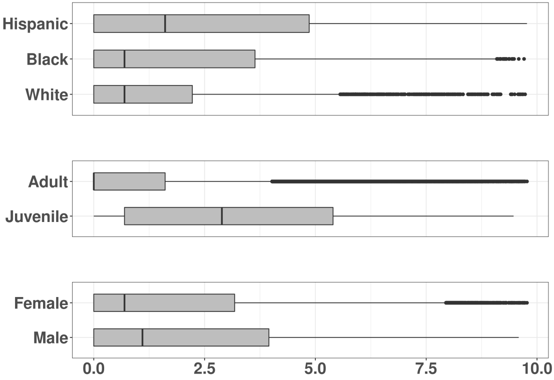


Figure: Delayed Reports and Victim Characteristics (log-scale)

Go Back

Reported sex crimes increased over time.

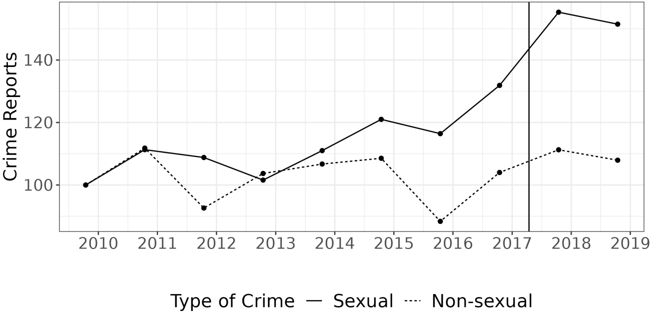


Figure: Sexual and Non-sexual Crime Reports

Go Back

Delayed reports suggest the depletion of a stock of unreported crimes.

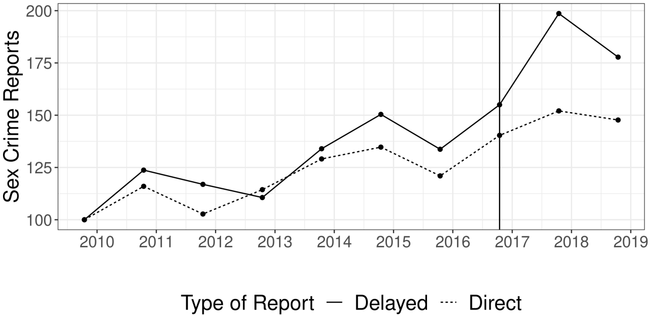


Figure: Delayed vs Direct Sex Crime Reports

Go Back

Sex-related Homicides

Sex-related homicides of women suggest a decrease in sex crime incidence.

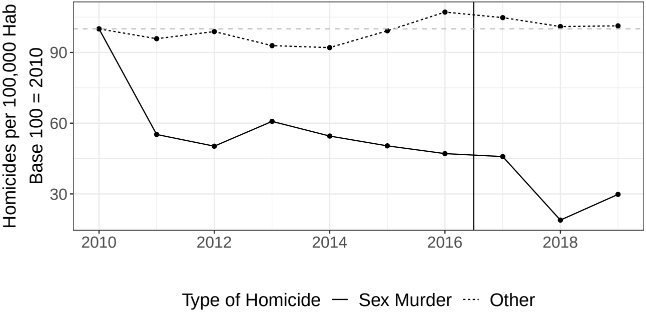


Figure: Trends in Homicides

Source: Supplementary Homicide Reports

Go Back

Survey-Based Estimates

National surveys suggest a large backlash effect after #MeToo.

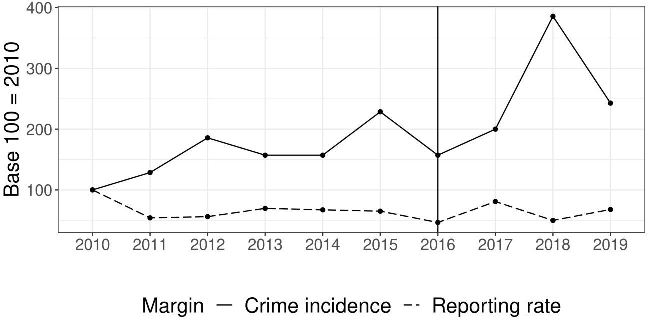


Figure: Survey Estimates of Sexual Assault Reporting and Incidence

Source: National Crime Victimization Survey (NCVS)

Go Back

#MeToo on Twitter

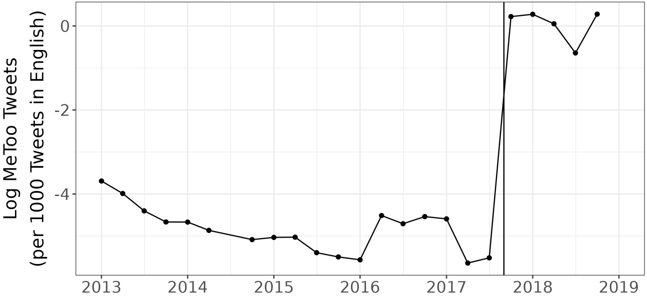


Figure: #MeToo Tweets

Go Back

Sexual Violence on Twitter

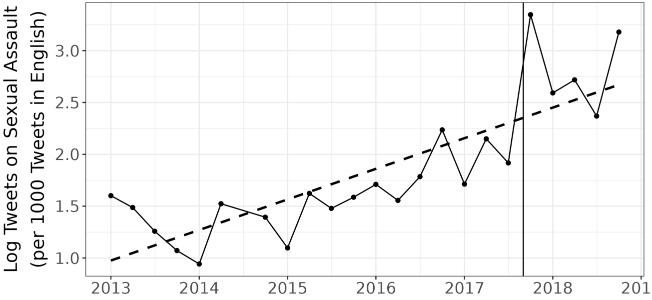


Figure: Tweets on Sexual Violence

Go Back

#MeToo Google Queries

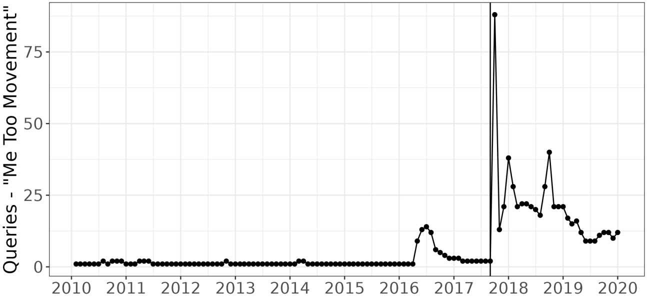


Figure: "Me Too Movement" Google Queries

Go Back

Sexual Assault Google Queries

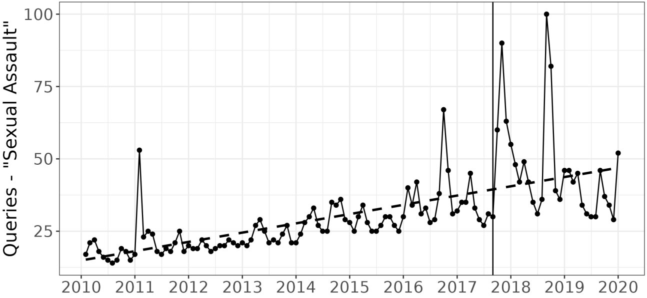


Figure: “Sexual Assault” Google Queries

Go Back

Arrest Rates

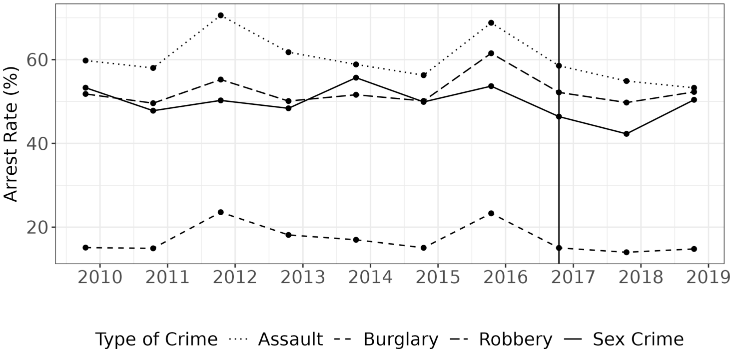


Figure: Arrest Rates for NYC and LA (2010 – 2020)

Go Back

Table of Contents

Theoretical Interpretation of the Duration Model

- If a crime is committed, a victim **j** is exposed to **K** (i.i.d.) potential decisive arguments to voice out:

$$K \sim \text{Pois}\Big(\theta_t\Big).$$

- The time for each argument to trigger a report to the police is drawn from a distribution **F**.
- The survival function of victims is

$$S^{(v)}(y_i) = \exp\Big(-\theta_t \cdot F(y_i)\Big) \xrightarrow[Y_i \to \infty]{} \exp\Big(-\theta_t\Big)$$

Go Back

Theoretical Interpretation of θ_t – Formal Proof

$$S_{Vt}(y) = P(Y > y)$$

$$= P(N = 0) + P(W_1 > y \cap ... \cap W_N > y \cap N \geq 1)$$

$$= e^{-\theta_t} + \sum_{N=1}^{\infty} (1 - F(y))^N e^{-\theta_t} \frac{\theta_t^N}{N!}$$

$$= \sum_{N=0}^{\infty} (1 - F(y))^N e^{-\theta_t} \frac{\theta_t^N}{N!}$$

$$= e^{-\theta_t F(y)}$$

Go Back

How restrictive is the model?

- Consider the mixture cure model:

$$S^{(v)}(y_i \mid x_i) = 1 - p(x_i) + p(x_i) \cdot S(y_i \mid x_i)$$

- In some extreme cases, the hazard of reporting of plaintiffs provides no information on the share of never-reporters. For example:

$$S_{it}^{(v)}(y_i \mid x_i) = 1 - p + p \cdot S(y_i \mid x_i).$$

Go Back

Police crime trends are sensitive to delayed reports.

Figure: Sex Felonies Under Different Data Processing Approaches

Source: City-level datasets

Go Back

Police crime trends are sensitive to delayed reports.

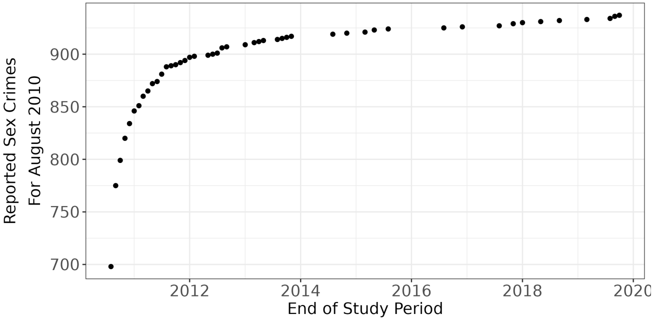


Figure: Crime counts for the same date vary depending on the study period.

Source: City-level datasets

Go Back

Table of Contents

#MeToo Effects on Crime Reporting – Lower Bounds

- I compare the distribution of reporting delays across years.
- The main specification is

$$h_{itc}^{(p)}(y) = h_0(y)\exp\left(\delta_c + \sum_{k=\text{Oct.15,2010}}^{\text{Oct.15,2019}} \beta_k \mathbb{1}(t+y \geq k)\right)\gamma_i,$$

where:

- h_0 is a piecewise constant exponential hazard with breaks on days 1, 30, 90, 180, and 365.
- δ_c is a city fixed effect.
- β_k is a year fixed effect.
- γ_i is a random effect for unobserved heterogeneity.

Go Back

Double-truncation leads to heavily biased estimates.

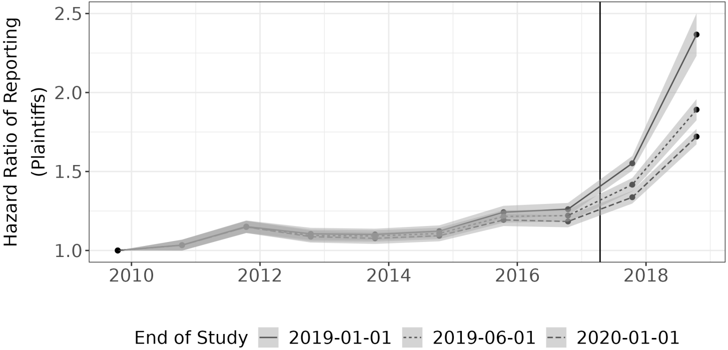


Figure: Naive Mixed Proportional Hazards (MPH) Estimates

Go Back

Estimates By Quarter

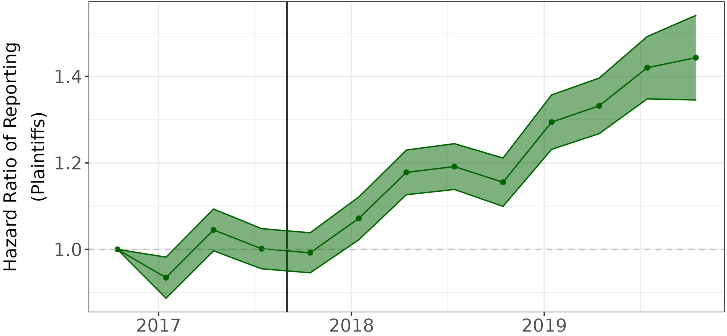


Figure: Quarterly Hazards Ratios – Lower Bounds

Go Back

#MeToo Effects on Crime Reporting – Lower Bounds

- I compare the distribution of reporting delays before and after #MeToo.
- The main specification is

$$h^{(p)}_{itc}(y) = h_0(y) \exp\left(\phi \mathrm{MeToo}_{ity} + \zeta X_i + \delta_c\right)\gamma_i,$$

where:

- h_0 is a piecewise constant exponential hazard.
- MeToo_{ity} is dummy for sex crimes after #MeToo.
- X_i are incident-level controls.
- δ_c is a city fixed effect.
- γ_i is time-invariant unobserved heterogeneity.

Go Back

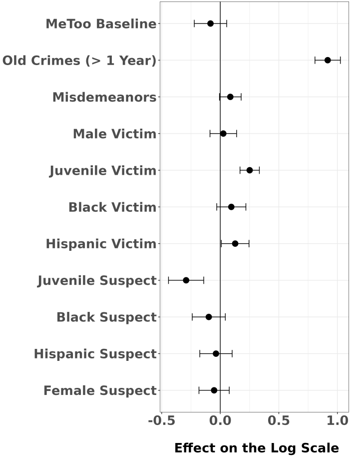
#MeToo Effects on Reporting – Lower Bounds

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Plaintiff Hazard							**Sex Crimes**						**Assaults**	**Robberies**	**Burglaries**
#MeToo: October 15, 2017	0.154	0.169	0.145	0.13	0.10	0.224	0.170	0.166	0.166	0.135	0.196	0.196	0.0018	0.03	0.05
	(0.009)	(0.011)	(0.02)	(0.012)	(0.017)	(0.012)	(0.011)	(0.011)	(0.011)	(0.013)	(0.020)	(0.019)	(0.011)	(0.011)	(0.011)
Baseline Hazard															
Day 0	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Day 1+	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Day 30+	X	X	X	X	X		X	X	X	X	X	X	X		
Day 90+	X	X	X	X	X			X	X	X	X	X	X	X	X
Day 180+	X	X	X	X	X				X	X	X	X	X	X	X
Day 365+	X	X	X	X	X					X	X	X	X	X	X
Controls															
Crime Category												X			
Victim Characteristics												X			
Suspect Characteristics												X			
City Fixed Effects	X	X	X	X	X	X	X	X	X	X			X	X	X
Time-Trends															
Linear				X	X										
Quadratic					X										
Unobserved Heterogeneity		Gamma	Discrete	Gamma	Gamma	Gamma	Gamma	Gamma	Gamma	Gamma	Gamma	Gamma	Gamma	Gamma	Gamma
τ_1	2010	2010	2010	2010	2010	2010	2010	2010	2010	2014	2010	2010	2010	2010	2010
τ_2	2019	2019	2019	2019	2019	2019	2019	2019	2019	2019	2019	2019	2019	2019	2019
Cities	All	All	All	All	All	All	All	All	All	All	NYC	NYC	All	All	All
N Observations	111869	111869	111869	111869	111869	111869	111869	111869	111869	72730	32442	32442	50000	50000	50000
Algorithm	BFGS	BFGS	rgenoud	BFGS	BFGS	BFGS	BFGS	BFGS	BFGS	BFGS	BFGS	BFGS	BFGS	BFGS	BFGS
Log-Likelihood	-419895	-416769.5	-413414	-419889	-419886	-420521	-417059	-416848	-416811	-173196	-121129	-119837	-78038	-68458	-108183

Table 1: Main Results and Robustness

Go Back

#MeToo Effects on Reporting – Heterogeneity



Notes: The baseline is a white adult woman accusing a white adult man. The heterogeneity analysis is restricted to New York City.

Go Back

Trends in Sex Crime Reporting *and* Incidence

- I decompose a time series of reported sex crimes into two margins of crime reporting and incidence.
- The main specification is

$$h_{itc}^{(v)}(y) = f_0(y)\exp\left(\alpha + \delta_c + \sum_{k=\text{Oct.15,2010}}^{\text{Oct.15,2019}} \beta_k \mathbb{1}(t+y \geq k)\right),$$

where:

- f_0 is a density. Its hazard is a piecewise constant exponential hazard with breaks on days 1, 30, 90, 180, and 365.
- α = 70% of never-reporters at baseline (i.e., October 2009).
- δ_c is a city fixed effect.
- β_k is a year fixed effect.

Trends in Nominal Values

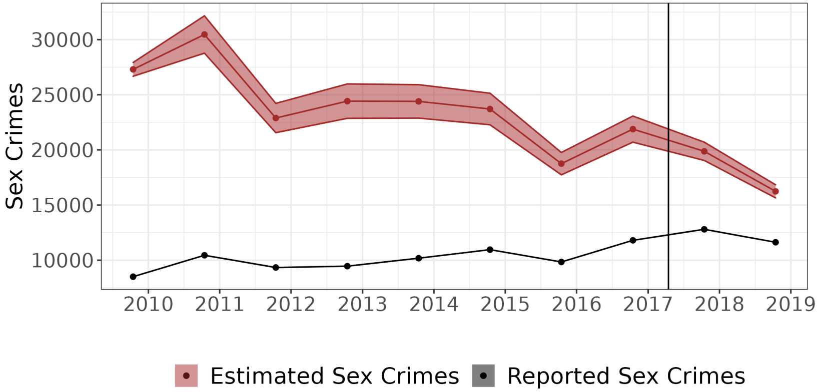


Figure: Estimated and Reported Sex Crimes

Go Back

Sensitivity to the Value of α

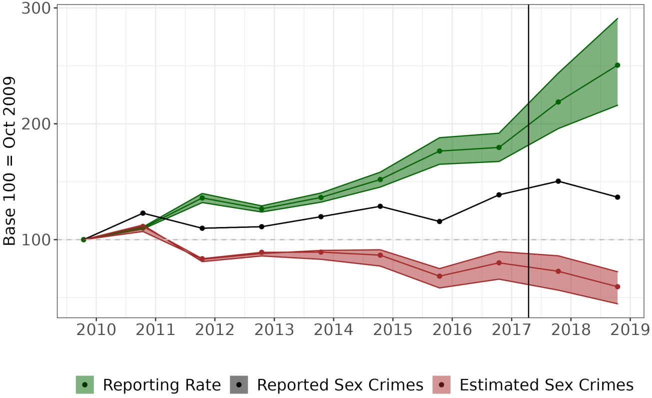


Figure: Sex Crime Reporting for Alternative Parameter Values

Go Back

Sensitivity to the Value of α

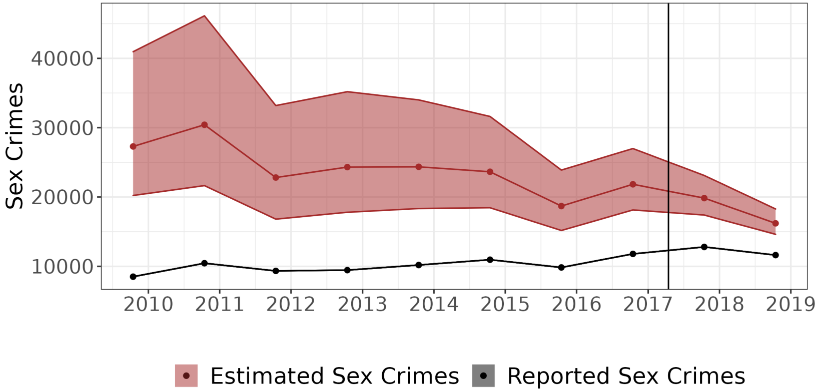


Figure: Sex Crime Reporting for Alternative Parameter Values

Go Back

Sensitivity to the Value of α

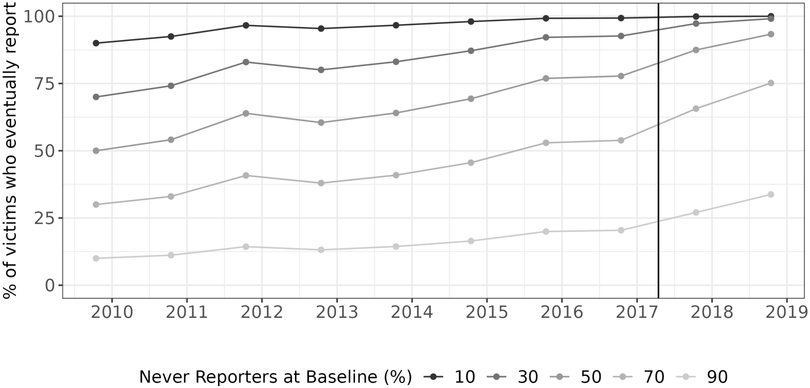


Figure: Sex Crime Reporting for Alternative Parameter Values

Go Back

Sensitivity to the Value of α

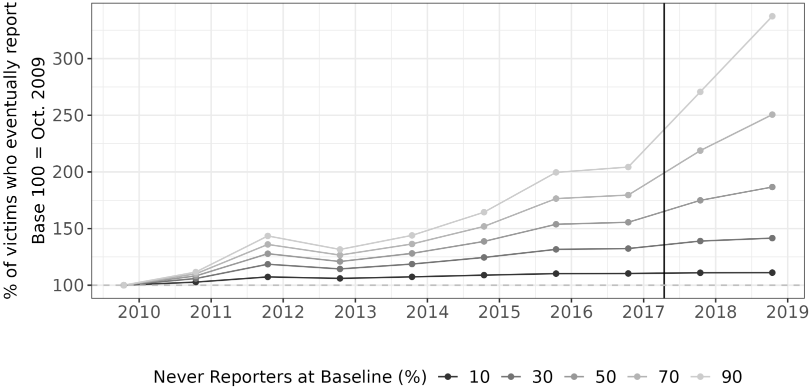


Figure: Sex Crime Reporting for Alternative Parameter Values

Go Back

Sex Crime Reporting and Incidence – Los Angeles

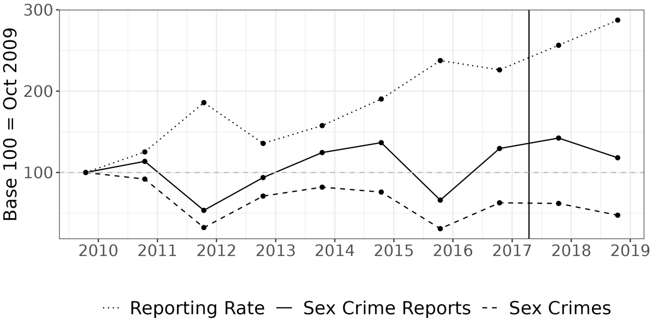


Figure: Trends – Full Sample

Go Back

Conservative Estimates – Los Angeles

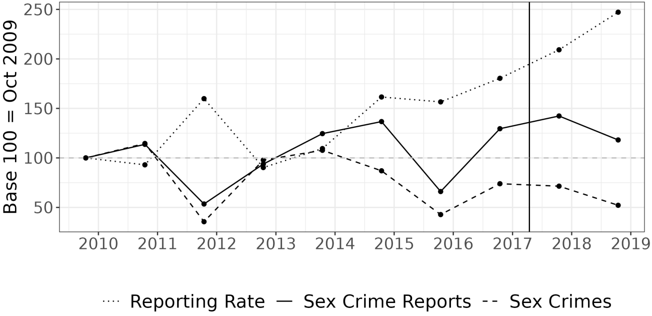


Figure: Trends – Complaints that lead to an arrest.

Go Back

Test for Time-Dependent Effects

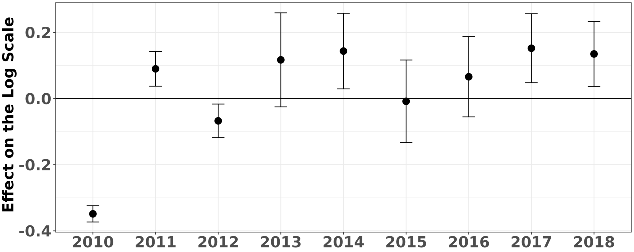


Figure: Additional Marginal Yearly Effect for Old Crimes (365+ days)

Robustness to Time-Dependent Effects

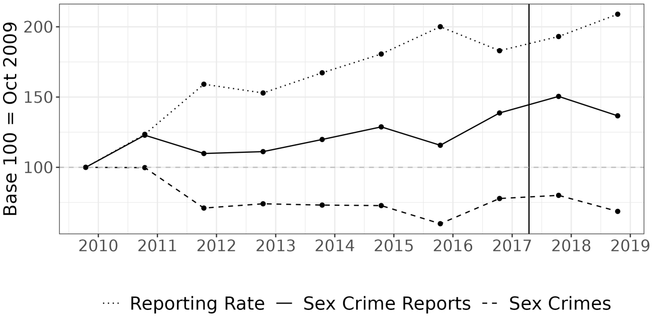


Figure: Trends Under Time-Dependent Effects

Go Back

#MeToo Effects on Crime

- I compare sex crimes to reported non-sexual crimes in an extended DID framework.
- The main specification is

$$\log(\text{Crimes})_{it} = \beta \text{MeToo}_{it} + \delta_i + \delta_t + \lambda_i' f_t + \varepsilon_{it},$$

where:

  - MeToo_{it} is dummy for sex crimes after #MeToo.
  - δ_i is a crime by city fixed effect.
  - δ_t is a time fixed effect.
  - f_t is a vector of factors and their associated loadings λ_i.
  - ε_{itc} is the error term.

- Additional assumption for causal inference:
  - The reporting rate of non-sexual crimes is uncorrelated to #MeToo.

Go Back

#MeToo Effect on Crime – Alternative Specifications

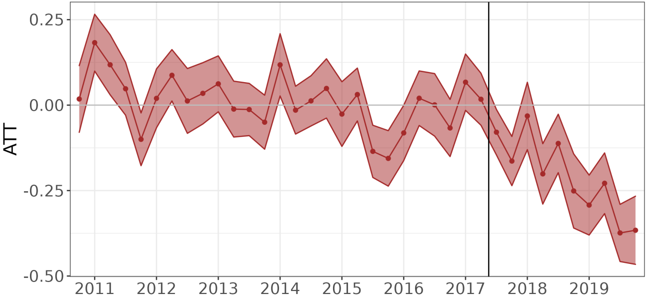


Figure: Difference-in-differences – ATTs

Go Back

#MeToo Effect on Crime – Alternative Specifications

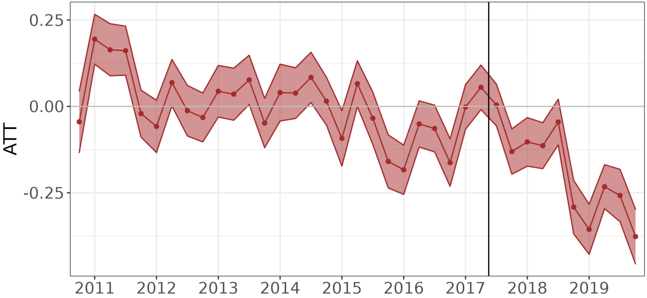


Figure: Matrix Completion Method – ATTs

Go Back

#MeToo Effect on Crime

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Treated Group	**Estimated Sex Crimes**							**Reported Sex Crimes**		**Murders**	**Assaults**	**Robberies**	**Burglaries**
								(Incident Date)	**(Report Date)**				
#MeToo: October 15, 2017	-0.23	-0.23	-0.21	-0.21	-0.19	-0.27	-0.26	0.20	0.32	0.11	0.17	-0.16	-0.13
	(0.02)	(0.03)	(0.015)	(0.016)	(0.02)	(0.02)	(0.02)	(0.105)	(0.10)	(0.15)	(0.12)	(0.19)	(0.17)
Model	DID	IFE	MC	DID	DID	DID	DID	DID	DID	DID	DID	DID	DID
Fixed Effects													
City Fixed Effects	X	X	X	X	X	X	X	X	X	X	X	X	X
Time Fixed Effects	X	X	X	X	X	X	X	X	X	X	X	X	X
Crime Fixed Effects	X	X	X	X	X	X	X	X	X	X	X	X	X
Control Groups													
Murders	X	X	X		X	X	X	X	X		X	X	X
Assaults	X	X	X	X		X	X	X	X	X		X	X
Robberies	X	X	X	X	X		X	X	X	X	X		X
Burglaries	X	X	X	X	X	X		X	X	X	X	X	
Standard Errors	Bootstrap	Bootstrap	Bootstrap	Bootstrap	Bootstrap	Bootstrap	Bootstrap	Clustered	Clustered	Clustered	Clustered	Clustered	Clustered
N Observations	740	740	740	592	592	592	592	740	740	592	592	592	592

Table 2: Main Results and Robustness

Go Back

The probability of arrest increased over time.

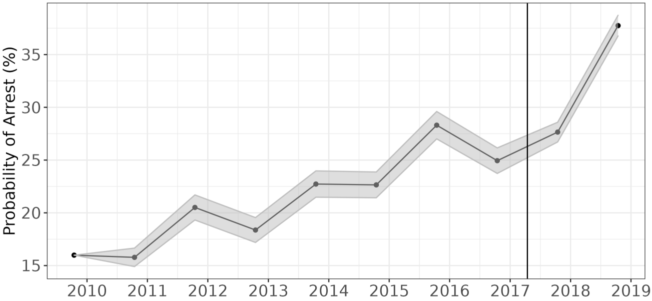


Figure: Probabilities of Arrest for Sex Offenders

Notes: Probabilities of arrest are computed for New York City and Los Angeles.

◂ Go Back

Abbring, J. H. and Van den Berg, G. J. (2003). The nonparametric identification of treatment effects in duration models. *Econometrica*, 71(5):1491–1517.

Acemoglu, D. and Jackson, M. O. (2017). Social norms and the enforcement of laws. *Journal of the European Economic Association*, 15(2):245–295.

Aizer, A. (2010). The gender wage gap and domestic violence. *American Economic Review*, 100(4):1847–59.

Akerlof, G. and Yellen, J. L. (1994). *Gang behavior, law enforcement, and community values*. Canadian Institute for Advanced Research Washington, DC.

Becker, G. S. (1968). Crime and punishment: An economic approach. In *The economic dimensions of crime*, pages 13–68. Springer.

Bellégo, C. and Drouard, J. (2019). Does it pay to fight crime? evidence from the pacification of slums in rio de janeiro.

Benabou, R. and Tirole, J. (2011). Laws and norms. Technical report, National Bureau of Economic Research.

Berkowitz, D., Pistor, K., and Richard, J.-F. (2003). Economic development, legality, and the transplant effect. *European economic review*, 47(1):165–195.

Bottan, N. L. and Perez-Truglia, R. (2015). Losing my religion: The effects of religious scandals on religious participation and charitable giving. *Journal of Public Economics*, 129:106–119.

Chalfin, A. and McCrary, J. (2017). Criminal deterrence: A review of the literature. *Journal of Economic Literature*, 55(1):5–48.

Coleman, C. and Moynihan, J. (1996). *Understanding crime data: Haunted by the dark figure*, volume 120. Open University Press Buckingham.

Doleac, J. L. (2019). Encouraging desistance from crime. Technical report, Working paper.

Dörre, A. and Emura, T. (2019). *Analysis of Doubly Truncated Data: An Introduction*. Springer.

Dyck, A., Morse, A., and Zingales, L. (2010). Who blows the whistle on corporate fraud? *The journal of finance*, 65(6):2213–2253.

Hay, J. R. and Shleifer, A. (1998). Private enforcement of public laws: A theory of legal reform. *The American Economic Review*, 88(2):398–403.

Iyer, L., Mani, A., Mishra, P., and Topalova, P. (2012). The power of political voice: women's political representation and crime in india. *American Economic Journal: Applied Economics*, 4(4):165–93.

Levitt, S. D. (1998). Why do increased arrest rates appear to reduce crime: deterrence, incapacitation, or measurement error? *Economic inquiry*, 36(3):353–372.

Levy, R. and Mattsson, M. (2021). The effects of social movements: Evidence from# metoo. *Available at SSRN 3496903*.

Mandel, M., de Uña-Álvarez, J., Simon, D. K., and Betensky, R. A. (2018). Inverse probability weighted cox regression for doubly truncated data. *Biometrics*, 74(2):481–487.

Mathur, A., Munasib, A., Roy, D., Bhatnagar, A., et al. (2019). Sparking the# metoo revolution in india: The'nirbhaya'case in delhi. Technical report, American Enterprise Institute.

McDougal, L., Krumholz, S., Bhan, N., Bharadwaj, P., and Raj, A. (2021). Releasing the tide: how has a shock to the acceptability of gender-based sexual violence affected rape reporting to police in india? *Journal of interpersonal violence*, 36(11-12):NP5921–NP5943.

Miller, A. R. and Segal, C. (2019). Do female officers improve law enforcement quality? effects on crime reporting and domestic violence. *The Review of Economic Studies*, 86(5):2220–2247.

Nagin, D. S. (2013). Deterrence: A review of the evidence by a criminologist for economists. *Annu. Rev. Econ.*, 5(1):83–105.

Quêtelet, A. (1831). *Research on the Propensity for Crime at Different Ages*.

Rennert, L. and Xie, S. X. (2018). Cox regression model with doubly truncated data. *Biometrics*, 74(2):725–733.

Sahay, A. (2021). The silenced women.

Stephens-Davidowitz, S. (2013). Unreported victims of an economic downturn. *Unpublished paper, Harvard University, Department of Economics, Cambridge, MA*.

Vakulenko-Lagun, B., Mandel, M., and Betensky, R. A. (2019). Inverse probability weighting methods for cox regression with right-truncated data. *Biometrics*.

Van den Berg, G. J. (2001). Duration models: specification, identification and multiple durations. In *Handbook of econometrics*, volume 5, pages 3381–3460. Elsevier.

Young, H. P. (2015). The evolution of social norms. 7(1):359–387.