

Domestic Violence in Chicago, 2001 - 2023

A Comparison to Violent and Property Crime Trends

Caroline Davidson

Table of contents

Motivation	1
Data Prep	1
Exploratory Analysis - Graphing Trends Over Time for the City of Chicago	3
Regression Analysis	7
Appendix	13

Motivation

This document uses the csv files generated by the accompanying Python script that contain a monthly count of reported crimes by type (DV, property crime, violent crime) and area (community areas or police districts) for 2001 - 2023. The objective of this analysis is to see whether domestic violence trends over time more closely follow trends in property crime or trends in violence crime, and whether this changed at all during the Covid-19 pandemic. This file includes exploratory plotting followed by panel regressions with year, month, and area fixed effects.

Data Prep

The csv files created by the Python script have data on reported domestic violence crimes, reported property crimes, and reported violent crimes for each month from January 2001 through December 2023. One file has this data at the community area level and the other at the police district level. While community area is perhaps the most commonly used geographic aggregation level used to study patterns in Chicago, police district may be more relevant if districts vary in their ability to prevent or respond to crime. Police district may also be more relevant if districts differ in the level of trust that residents feel in the police and subsequently their likelihood to report crimes when they occur.

The data are in long format, so I begin by pivoting into a wide format for ease of graphing and regression analysis. I also create variables for the percentage change in reported crime by type at the city level. Note that analysis of percentage changes at the community area or district level is made difficult by a large number of zero count values.

I also create a covid dummy variable that is equal to 1 for months beginning in March 2020, when the Covid-19 pandemic hit Chicago and the first school closures and other social distancing measures were introduced.

```
comm_areas_monthly <- read_csv(
  "Data/regression ready aggregations/monthly_comm_area_2001_2023.csv",
  col_select = -1
)

comm_areas_monthly_wide <- pivot_wider(comm_areas_monthly,
  names_from = reg_crime_group,
  values_from = n_reported
) %>%

rename(
  property_crime = "property crime",
  violent_crime = "violent crime"
) %>%

mutate(date = make_datetime(year = year, month = month, day = 1)) %>%
group_by(community_area) %>%
mutate(
  perc_change_property = (
    (property_crime - lag(property_crime)) / lag(property_crime)) * 100,
  perc_change_violent = (
    (violent_crime - lag(violent_crime)) / lag(violent_crime)) * 100
) %>%
ungroup()

# will consider Covid-19 period to start in March 2020 (when schools first closed in IL)
comm_areas_monthly_wide$covid <- as.integer(
  comm_areas_monthly_wide$date >= as.Date("2020-03-01")
)

# these manual interaction terms will be helpful for regressions below
comm_areas_monthly_wide$property_crime_covid <- (
  comm_areas_monthly_wide$property_crime * comm_areas_monthly_wide$covid
)

comm_areas_monthly_wide$violent_crime_covid <- (
  comm_areas_monthly_wide$violent_crime * comm_areas_monthly_wide$covid
)
```

```

)

districts_monthly <- read_csv(
  "Data/regression ready aggregations/monthly_district_2001_2023.csv",
  col_select = -1
)

districts_monthly_wide <- pivot_wider(districts_monthly,
                                     names_from = reg_crime_group,
                                     values_from = n_reported
                                   ) %>%

  rename(
    property_crime = "property crime",
    violent_crime = "violent crime"
  ) %>%
  mutate(date = make_datetime(year = year, month = month, day = 1))

districts_monthly_wide$covid <- as.integer(
  districts_monthly_wide$date >= as.Date("2020-03-01")
)
districts_monthly_wide$property_crime_covid <- (
  districts_monthly_wide$property_crime * districts_monthly_wide$covid
)
districts_monthly_wide$violent_crime_covid <- (
  districts_monthly_wide$violent_crime * districts_monthly_wide$covid
)

```

Exploratory Analysis - Graphing Trends Over Time for the City of Chicago

Before doing my community area and district level panel regressions, I aggregate the data to the city level to plot how reported crimes have trended over time city-wide.

```

city_monthly <- comm_areas_monthly %>%
  group_by(year, month, reg_crime_group) %>%
  summarize(reported_crimes = sum(n_reported)) %>%
  mutate(date = make_datetime(year = year, month = month, day = 1)) %>%
  ungroup() %>%
  group_by(reg_crime_group) %>%
  mutate(perc_change = (
    (reported_crimes - lag(reported_crimes)) / lag(reported_crimes)) * 100
  )

```

```

    ) %>%
ungroup()

graph_colors <- c(
  "DV" = "#3e0c99",
  "property crime" = "#2a63d4",
  "violent crime" = "#940503")

crime_type_labels <- c(
  "DV" = "Domestic Violence",
  "property crime" = "Property Crime (non DV, index crime)",
  "violent crime" = "Violent Crime (non DV, index crime)"
)

# adding a vertical line at March 13, 2020 - date Governor Pritzker
# first closed schools in Illinois, first lock-down type measure
# will use to mark onset of the Covid-19 pandemic in Illinois

crime_counts_graph <- ggplot(filter(city_monthly),
  aes(
    x = as.Date(date),
    y = reported_crimes,
    color = fct_reorder2(
      reg_crime_group, as.Date(date), reported_crimes
    )
  )
) +
  geom_line(key_glyph = "timeseries") +
  scale_color_manual(values = graph_colors) +
  geom_vline(xintercept = as.Date("2020-03-13"), linetype = "dashed") +
  labs(
    title = "Reported Crimes in Chicago by Month and Type",
    y = "Number of Monthly Reported Crimes",
    x = NULL,
    color = "type of crime"
  ) +
  scale_x_date(
    breaks = seq(as.Date("2002-01-01"), as.Date("2025-01-01"), by = "5 years"),
    date_labels = "%Y"
  ) +

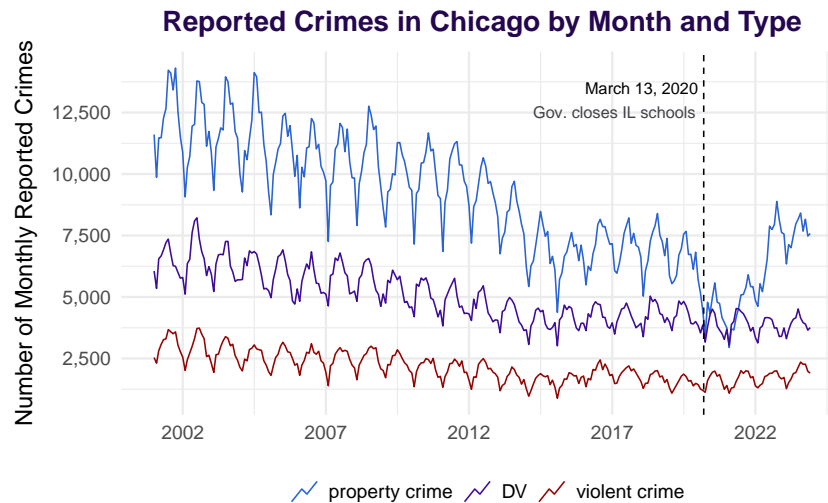
```

```

scale_y_continuous(
  breaks = seq(from = 2500, to = 12500, by = 2500),
  labels = label_comma()
) +
guides(color = guide_legend(title = "Crime Type")) +
theme(
  legend.position = "bottom",
  legend.title = element_blank(),
  axis.title.y = element_text(margin = margin(r = 10)),
  plot.title = element_text(
    hjust = 0.5, vjust = 1.5, face = "bold", color = "#24074f"
  ),
  plot.background = element_rect(fill = "white",
    color = "white")
) +
annotate(
  geom = "text",
  x = as.Date("2020-03-13"), y = 13500,
  label = "March 13, 2020", hjust = 1.05
) +
annotate(
  geom = "text",
  x = as.Date("2020-03-13"), y = 12500,
  label = "Gov. closes IL schools",
  hjust = 1.05, color = "#434047"
)

crime_counts_graph

```



```
ggsave("images/crimes_by_month_type.png", crime_counts_graph)
```

We can see from this graph that all 3 types of crimes have clear annual cycles and have generally been (slowly) trending downwards since 2001. There is a marked pattern in property crime after the onset of the Covid-19 pandemic, with an initial decrease followed by a rapid rise in 2022. Any changes in the pattern for domestic violence and violent crimes after the onset of the Covid-19 pandemic are much more subtle.

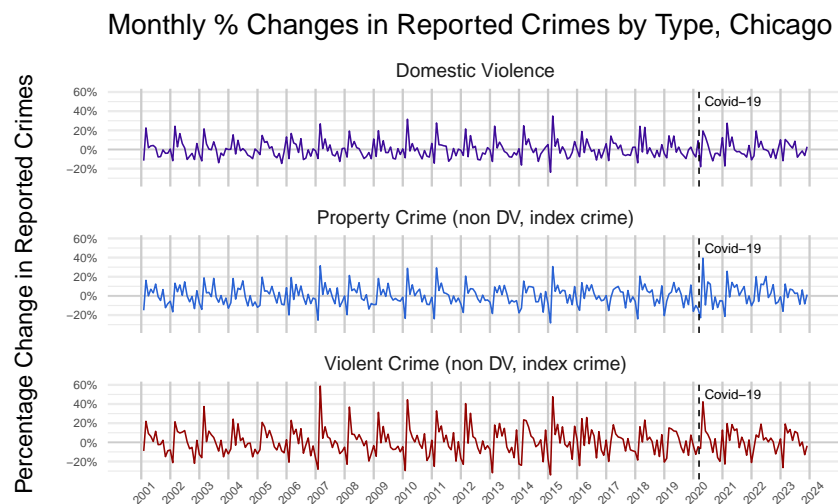
```
percent_changes_by_type_graph <- ggplot(city_monthly, aes(
  x = as.Date(date),
  y = perc_change/100,
  color = reg_crime_group)
) +
geom_hline(aes(yintercept = 0), color = "gray80") +
geom_line() +
geom_vline(xintercept = as.Date("2020-03-13"), linetype = "dashed") +
facet_wrap(
  vars(reg_crime_group), ncol = 1,
  labeller = labeller(reg_crime_group = crime_type_labels)
) +
scale_color_manual(values = graph_colors) +
labs(
  title = "Monthly % Changes in Reported Crimes by Type, Chicago",
  x = NULL,
  y = "Percentage Change in Reported Crimes"
) +
```

```

theme(legend.position = "none") +
theme(
  axis.title.y = element_text(margin = margin(r = 20)),
  axis.title.x = element_text(margin = margin(t = 10)),
  axis.text.x = element_text(size = 8, angle = 45),
  axis.text.y = element_text(size = 8),
  panel.grid.major.x = element_line(color = "gray80"),
  panel.grid.minor.x = element_blank(),
  plot.background = element_rect(fill = "white",
                                color = "white")
) +
scale_x_date(
  breaks = seq(as.Date("2001-01-01"), as.Date("2024-01-01"), by = "1 year"),
  date_labels = "%Y",
  limits = c(as.Date("2001-01-01"), as.Date("2024-01-01"))
) +
scale_y_continuous(labels = percent) +
  annotate(
    geom = "text",
    x = as.Date("2020-03-13"), y = 0.5,
    label = "Covid-19", hjust = -0.1, size = 3
  )

```

percent_changes_by_type_graph



```
ggsave("images/percent_changes_graph.png", percent_changes_by_type_graph)
```

While the graph of crime counts showed that property crime had the highest levels and most extreme variability in counts, when looking at monthly percent changes (rather than counts) we see that violent crime actually appears to be the most variable of the three, with the highest peaks and lowest troughs.

Regression Analysis

I now conduct panel regressions with domestic violence count as my dependent variable, property crime and violent crime counts as independent variables, and fixed effects for:

- month: to account for the (average) annual cycle revealed by the graphs
- year: to account for long-term trends / yearly shocks that affect DV in all areas equally
- area (community area or police district): to account for area-specific factors that might influence the level of DV

Including these fixed effects allows us to make within-area comparisons, i.e. to compare community areas to themselves over time, reducing issues of bias that could creep in when making cross-area comparisons. Specifically, these fixed effects eliminate bias from: area-specific characteristics that do not vary over time; year-specific shocks that affect all areas equally; average seasonal patterns.

Since I'm including year fixed effects, I expect them to soak up most of the Covid effect, so I will exclude a Covid dummy variable from these regression and just include (previously) manually coded interaction terms. The results do not change substantively when a Covid dummy variable is included, as shown in the appendix.

```
reg_comm_areas <- felm(
  DV ~ property_crime + violent_crime + property_crime_covid + violent_crime_covid |
  month + year + community_area | # fixed effects
  0 | # no IV
  community_area, # where to cluster SEs
  data = comm_areas_monthly_wide
)

reg_districts <- felm(
  DV ~ property_crime + violent_crime + property_crime_covid + violent_crime_covid |
  month + year + district | # fixed effects
  0 | # no IV
  district, # where to cluster SEs
  data = districts_monthly_wide
)
```



```

fe_lines <- list(
  c("Month FEs:", "X", "X"),
  c("Year FEs:", "X", "X"),
  c("Community Area FEs:", "X", ""),
  c("Police District FEs:", "", "X")
)

table_regs <- stargazer(reg_comm_areas, reg_districts,
  type = "latex",
  header = FALSE,
  title = "Regression with Year, Month, Area Fixed Effects",
  df = FALSE,
  add.lines = fe_lines)

```

The regression coefficients on `property_crime` and `violent_crime` capture the statistical association between these two types of crime and domestic violence (the dependent variable) for the period before March 2020. To get the statistical association between these types of crime and domestic violence during/post Covid, sum those coefficients with their respective covid interaction terms. For example:

- In the pre-Covid period, one additional reported property crime was statistically associated with an additional 0.062 reported DV crimes at the community area level on average, holding all else (i.e. violent crime) constant.
- After the onset of the Covid-19 pandemic, one additional property crime was statistically associated with an additional 0.104 ($0.062 + 0.042 = 0.104$) reported DV crimes at the community area level on average, holding all else (i.e. violent crime) constant.

Both regressions, at the community area level and the police district level, show that the level of domestic violence is more strongly statistically associated with violent crime than with property crime in the pre-Covid era, as seen by the larger coefficients on `violent_crime` than `property_crime`. This remains true in the Covid period, however the magnitude of the difference (in associations with property crime vs. violent crime) is smaller due to the negative coefficient on `violent_crime_covid`.

All coefficients are statistically significant at the 1% level except for the coefficient on `property_crime` in the regression at the police district level. Note that Chicago has 77 community areas but only 22 police districts, so the district-level regression has considerably fewer observations and we could expect it to be less precise. Indeed, the standard errors on coefficients in this regression are slightly larger than in the community area level regression, with the exception of the `property_crime` coefficient, so the lack of statistical significance for that variable does not appear to arise from a decrease in precision.

Table 1: Regression with Year, Month, Area Fixed Effects

	<i>Dependent variable:</i>	
	DV	
	(1)	(2)
property_crime	0.062*** (0.022)	−0.005 (0.018)
violent_crime	0.832*** (0.056)	0.838*** (0.091)
property_crime_covid	0.042*** (0.015)	0.076*** (0.022)
violent_crime_covid	−0.555*** (0.062)	−0.456*** (0.080)
Month FEs:	X	X
Year FEs:	X	X
Community Area FEs:	X	
Police District FEs:		X
Observations	21,252	6,072
R ²	0.956	0.950
Adjusted R ²	0.956	0.950
Residual Std. Error	14.182	27.642
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01	

```

# differently formatted regression output table to save as image
comm_areas_table <- tbl_regression(
  reg_comm_areas,
  estimate_fun = function(x) style_number(x, digits = 3)
) %>%
  add_significance_stars()

districts_table <- tbl_regression(
  reg_districts,
  estimate_fun = function(x) style_number(x, digits = 3)
) %>%
  add_significance_stars()

combined_table <- tbl_merge(
  list(comm_areas_table, districts_table),
  tab_spanner = c(
    "**Community Area Level**", "**Police District Level**"
  )
) %>%
  modify_caption(
    "**Panel Regression Results with Year, Month, Area Fixed Effects**"
  ) %>%
  modify_header(label = "**Variable**")

combined_table

```

Table 2: **Panel Regression Results with Year, Month, Area Fixed Effects**

Variable	Beta	SE	Beta	SE
property_crime	0.062**	0.022	-0.005	0.018
violent_crime	0.832***	0.056	0.838***	0.091
property_crime_covid	0.042**	0.015	0.076**	0.022
violent_crime_covid	-0.555***	0.062	-0.456***	0.080

```

combined_table %>%
  as_gt() %>%
  gt::gtsave(filename = "images/combined_table.png")

```

Plots of coefficient estimates and their confidence intervals can be helpful in visually communicating these results to audiences less comfortable with reading regression output.

```

# Community Areas
comm_areas_plot <- plot(comm_areas_table)
ggsave("images/temp/comm_areas_coefficients_plot.png", comm_areas_plot)

comm_areas_image <- image_read("images/temp/comm_areas_coefficients_plot.png")
# crop image to remove repeated variable names
comm_areas_image_cropped <- image_crop(comm_areas_image,
                                         geometry = geometry_area(width = 1800,
                                                                    height = 2100,
                                                                    x_off = 500,
                                                                    y_off = 0))

image_write(comm_areas_image_cropped, "images/comm_areas_coeffs_cropped.png")

# Police Districts
districts_plot <- plot(districts_table)
ggsave("images/temp/districts_coefficients_plot.png", districts_plot)

districts_image <- image_read("images/temp/districts_coefficients_plot.png")
districts_image_cropped <- image_crop(districts_image,
                                       geometry = geometry_area(width = 1800,
                                                                  height = 2100,
                                                                  x_off = 500,
                                                                  y_off = 0))

image_write(districts_image_cropped, "images/districts_coeffs_cropped.png")

```

Appendix

```
reg_comm_areas_a <- felm(  
  DV ~ property_crime + violent_crime + property_crime*covid + violent_crime*covid |  
  month + year + community_area | # fixed effects  
  0 | # no IV  
  community_area, # where to cluster SEs  
  data = comm_areas_monthly_wide  
)  
  
reg_districts_a <- felm(  
  DV ~ property_crime + violent_crime + property_crime*covid + violent_crime*covid |  
  month + year + district | # fixed effects  
  0 | # no IV  
  district, # where to cluster SEs  
  data = districts_monthly_wide  
)  
  
fe_lines <- list(  
  c("Month FEs:", "X", "X"),  
  c("Year FEs:", "X", "X"),  
  c("Community Area FEs:", "X", ""),  
  c("Police District FEs:", "", "X")  
)  
  
table_regs_a <- stargazer(reg_comm_areas_a, reg_districts_a,  
  type = "latex",  
  header = FALSE,  
  title = "Regression with Year, Month, Area Fixed Effects",  
  df = FALSE,  
  add.lines = fe_lines)
```

Table 3: Regression with Year, Month, Area Fixed Effects

	<i>Dependent variable:</i>	
	DV	
	(1)	(2)
property_crime	0.062*** (0.022)	−0.005 (0.018)
violent_crime	0.833*** (0.056)	0.838*** (0.091)
covid	6.619*** (1.817)	1.223 (5.587)
property_crime:covid	0.041*** (0.015)	0.075*** (0.023)
violent_crime:covid	−0.562*** (0.063)	−0.458*** (0.083)
Month FEs:	X	X
Year FEs:	X	X
Community Area FEs:	X	
Police District FEs:		X
Observations	21,252	6,072
R ²	0.956	0.950
Adjusted R ²	0.956	0.950
Residual Std. Error	14.174	27.644
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01	