

Regression for Econ 468 Project

Aviel Fradkine

2022-12-01

To Do:

- Check if there was a change in coefficients between different time periods (e.g. before NAFTA in 1996)
- Check the approach done by other literature, how our approach fits with it, and what we can say about their approaches and our approaches given the regression we've run

Load in the datasets we'll be using. The datasets draw on 67 unique stations, with daily weather data from 1991-2020. See the file "getweatherdata.Rmd" for the code that generated the weather data used to create these indexes. The raw daily weather data used for the creation of the indexes is found in the file "sample_weather.csv".

```
setwd("C:/Users/Aviel/Desktop/Econ468project/regression_inputs/")

monthly_precipitation <- read.csv("monthly_precipitation.csv")
monthly_mean_25 <- read.csv("monthly_over_25.csv")
monthly_max_25 <- read.csv("monthly_max_25.csv")
monthly_mean_30 <- read.csv("monthly_over_30.csv")
monthly_max_30 <- read.csv("monthly_max_30.csv")
monthly_mean_temp <- read.csv("monthly_mean_temp.csv")
fppi <- read.csv("fppi1991to2021prelim.csv")
cpi_diff <- read_csv("cpi_diff_monthly.csv")
```

Clean dataset format and construct our df

```
fppi$date <- monthly_mean_25$month_year

fppi_crops <- fppi %>%
  select(`date`, `Total.crops`) %>%
  rename(month_year = date)

reg_data <- data.frame(fppi_crops$month_year,
  fppi_crops$Total.crops,
  cpi_diff$cpi_diff_monthly,
  monthly_mean_25$total_station_days_over_25,
  monthly_max_25$total_station_max_days_over_25,
  monthly_mean_30$total_station_days_over_30,
  monthly_max_30$total_station_max_days_over_30,
  monthly_mean_temp$month_mean_temp,
  monthly_precipitation$monthly_precip,
  monthly_precipitation$avg_precip)
```

```

colnames(reg_data) <- c('month_year',
                        'fppi_crops',
                        'cpi_diff',
                        'st_days_mean_over_25',
                        'st_days_max_over_25',
                        'st_days_mean_over_30',
                        'st_days_max_over_30',
                        'monthly_mean_temp',
                        'monthly_precipitation_total',
                        'monthly_precipitation_average')

# add monthly difference of fppi and add a lag for fppi
fppi_diffs <- c(0.3, diff(reg_data$fppi_crops)) # length 359
reg_data <- reg_data %>%
  mutate(diff_fppi = fppi_diffs) %>%
  mutate(lag_st_days_max_30 = lag(st_days_max_over_30, n = 3)) %>%
  mutate(lag_st_days_max_25 = lag(st_days_max_over_25, n = 3)) %>%
  mutate(lag_monthly_mean_temp = lag(monthly_mean_temp, n = 3))

# add monthly difference of fppi as percent
reg_data <- reg_data %>%
  mutate(diff_fppi_per = round(fppi_diffs/fppi_crops, 3))

write.csv(reg_data,
          "C:\\Users\\Aviel\\Desktop\\Econ468project\\regression_inputs\\reg_data.csv")

```

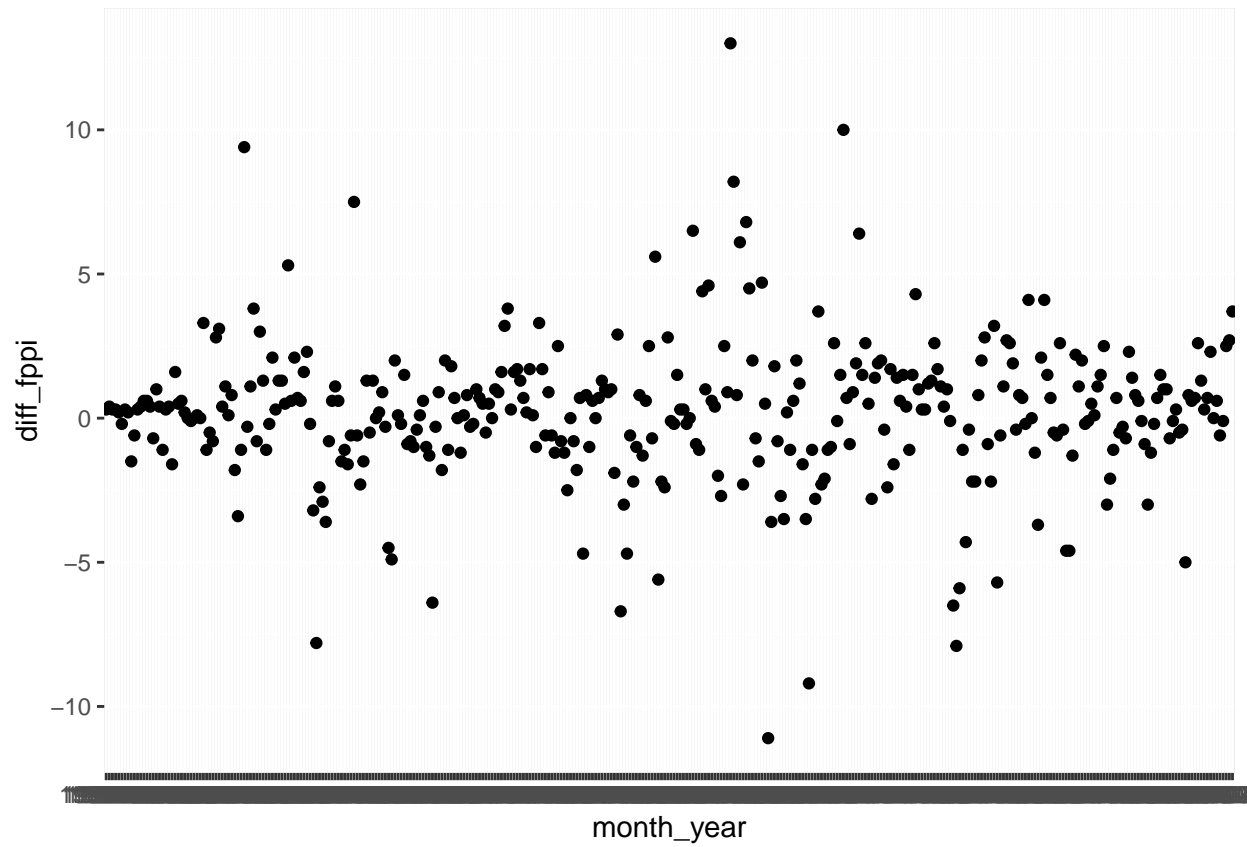
Data visualization

Simple scatterplot of diff_fppi and months

```

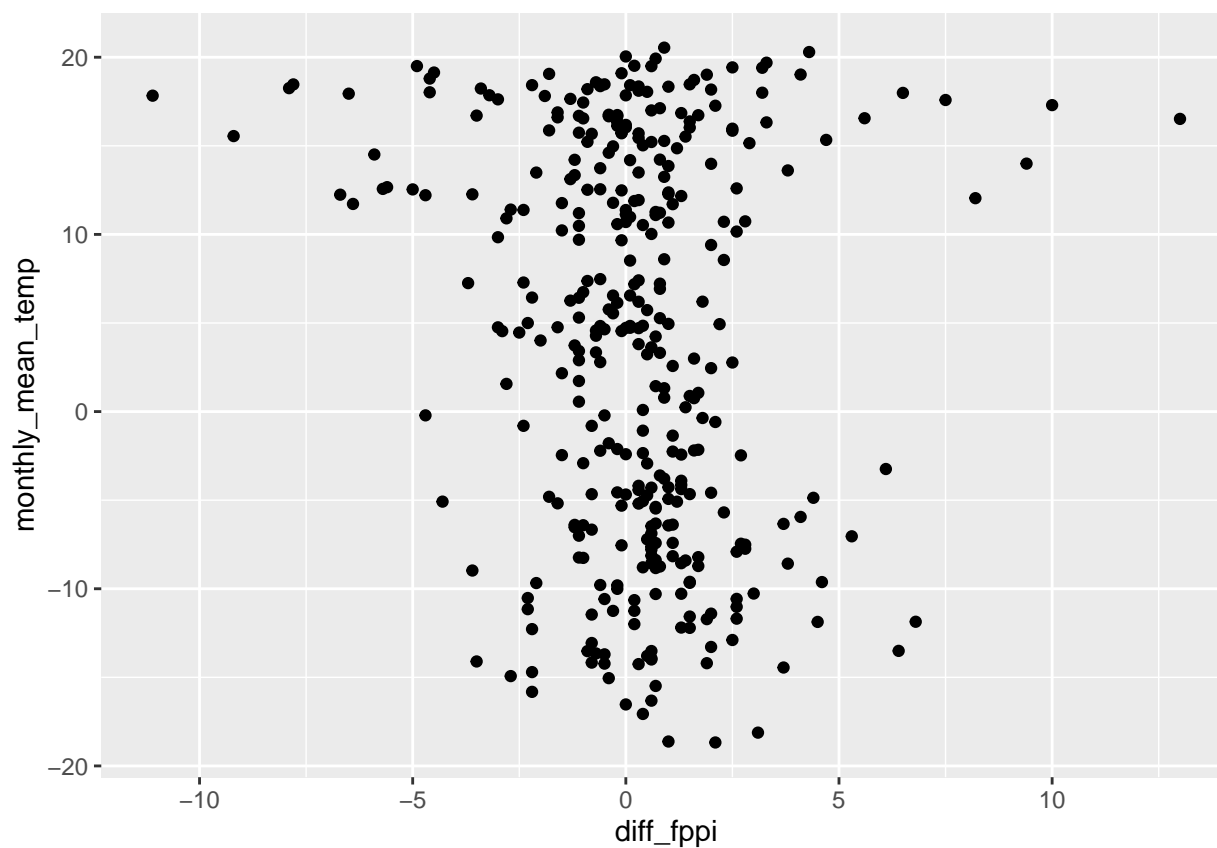
diff_fppi <- ggplot(reg_data, aes(x = month_year,
                                  y = diff_fppi)) +
  geom_point()
diff_fppi

```

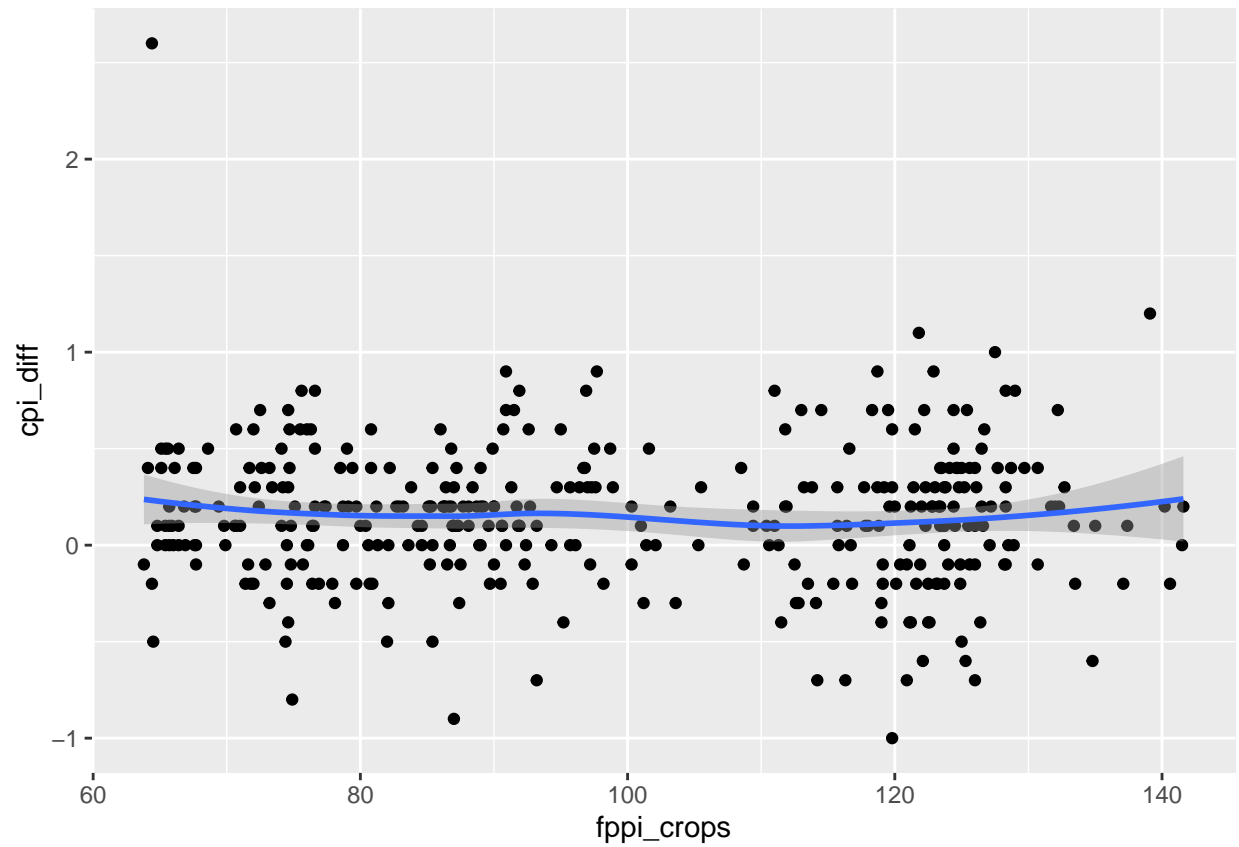


Relationship between fppi and cpi The CPI and fppi have a weird relationship because the fppi is indexed to 2007 whereas the inflation is a monthly inflation

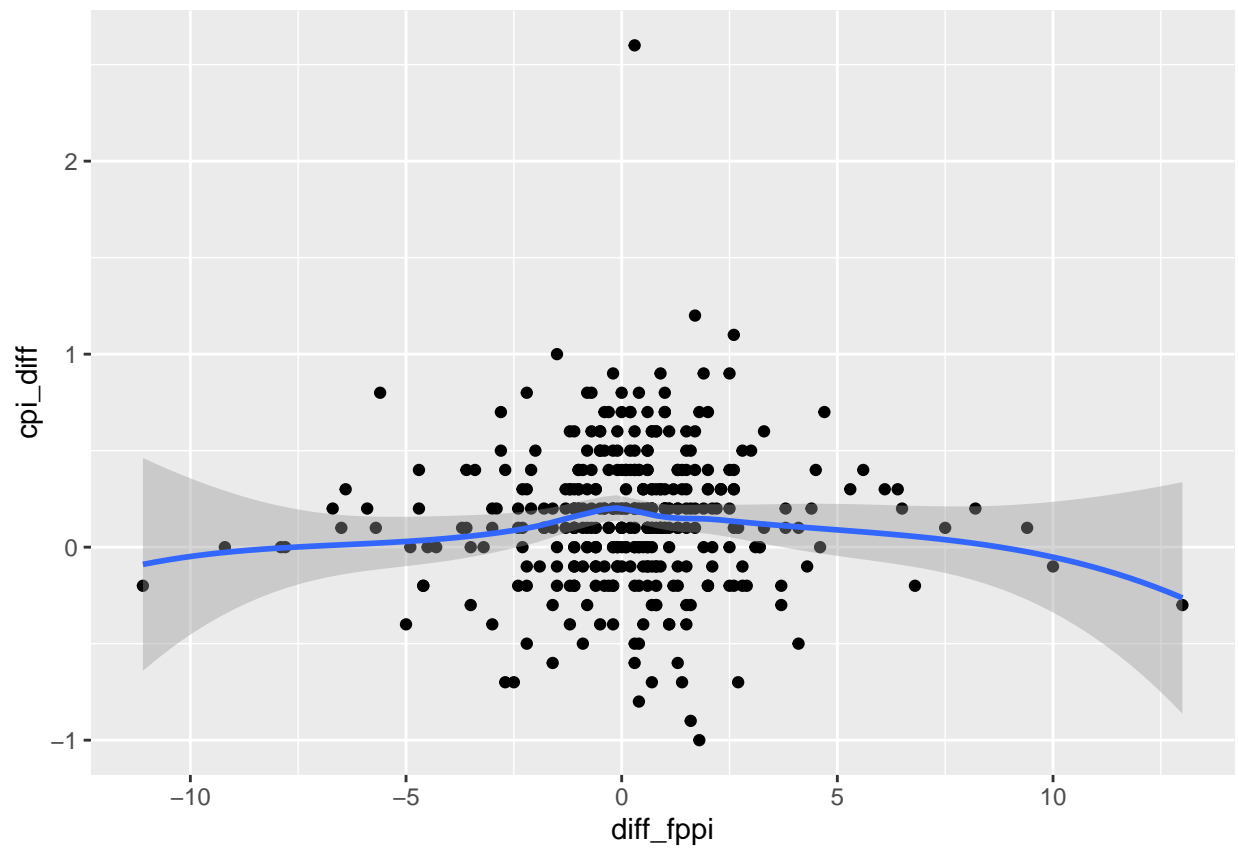
```
fppi_mean_temp_plot <- ggplot(reg_data) +  
  aes(x = diff_fppi,  
      y = monthly_mean_temp) +  
  geom_point()  
fppi_mean_temp_plot
```



```
fppi_cpi_plot <- ggplot(reg_data) +  
  aes(x = fppi_crops,  
      y = cpi_diff) +  
  geom_point() +  
  geom_smooth()  
fppi_cpi_plot
```



```
diffs_plot <- ggplot(reg_data) +  
  aes(x = `diff_fppi`,  
      y = `cpi_diff`)+  
  geom_point()+  
  geom_smooth()  
diffs_plot
```

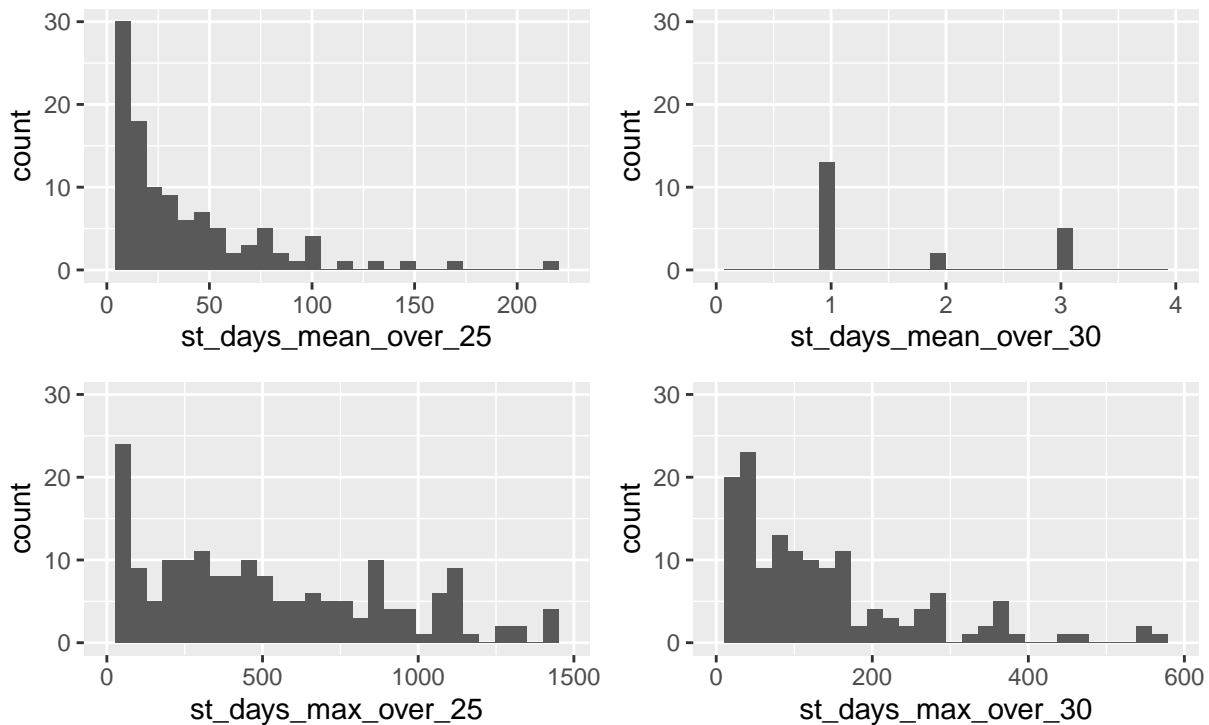


Density plot of weather indexes

```
mean.25.plot <- ggplot(reg_data) +  
  geom_histogram(aes(x = st_days_mean_over_25)) +  
  xlim(0, max(reg_data$st_days_mean_over_25)) +  
  ylim(0, 30)  
mean.30.plot <- ggplot(reg_data) +  
  geom_histogram(aes(x = st_days_mean_over_30)) +  
  xlim(0, max(reg_data$st_days_mean_over_30)) +  
  ylim(0, 30)  
max.25.plot <- ggplot(reg_data) +  
  geom_histogram(aes(x = st_days_max_over_25)) +  
  xlim(0, max(reg_data$st_days_max_over_25)) +  
  ylim(0, 30)  
max.30.plot <- ggplot(reg_data) +  
  geom_histogram(aes(x = st_days_max_over_30)) +  
  xlim(0, max(reg_data$st_days_max_over_30)) +  
  ylim(0, 30)  
mean.25.plot + mean.30.plot + max.25.plot + max.30.plot +  
  plot_annotation(  
    title = "Histograms of weather indexes",  
    subtitle = "Each month has max value of 2010 or 2077"  
  )
```

Histograms of weather indexes

Each month has max value of 2010 or 2077



Regressions

```
setwd("C:/Users/Aviel/Desktop/Econ468project/regression_inputs/")
reg_data <- read_csv("reg_data.csv")
```

Regression #1 - Diff in fppi on monthly mean temperature

```
reg.diff.fppi <- lm(diff_fppi ~ monthly_mean_temp, reg_data)
summary(reg.diff.fppi)
```

Call:

```
lm(formula = diff_fppi ~ monthly_mean_temp, data = reg_data)
```

Residuals:

Min	1Q	Median	3Q	Max
-10.8646	-1.1125	0.0518	1.0184	13.1945

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.32083	0.13982	2.295	0.02234 *
monthly_mean_temp	-0.03119	0.01190	-2.622	0.00911 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.512 on 358 degrees of freedom

Multiple R-squared: 0.01884, Adjusted R-squared: 0.0161

F-statistic: 6.875 on 1 and 358 DF, p-value: 0.009115

Regression #2 - Diff in fppi on mean temperature

```
reg.diff.mean.30 <- lm(diff_fppi ~ st_days_mean_over_30, reg_data)
reg.diff.mean.25 <- lm(diff_fppi ~ st_days_mean_over_25, reg_data)
summary(reg.diff.mean.25)
```

Call:

```
lm(formula = diff_fppi ~ st_days_mean_over_25, data = reg_data)
```

Residuals:

Min	1Q	Median	3Q	Max
-11.4223	-1.0797	0.1273	1.1457	12.6941

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.154271	0.144241	1.070	0.286
st_days_mean_over_25	0.004099	0.004617	0.888	0.375

Residual standard error: 2.533 on 358 degrees of freedom

Multiple R-squared: 0.002197, Adjusted R-squared: -0.0005899

F-statistic: 0.7884 on 1 and 358 DF, p-value: 0.3752


```
summary(reg.diff.mean.30)
```

Call:

```
lm(formula = diff_fppi ~ st_days_mean_over_30, data = reg_data)
```

Residuals:

Min	1Q	Median	3Q	Max
-11.2607	-1.0607	0.1393	1.1393	12.8393

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.1607	0.1364	1.179	0.239
st_days_mean_over_30	0.3783	0.2614	1.447	0.149

Residual standard error: 2.528 on 358 degrees of freedom

Multiple R-squared: 0.005816, Adjusted R-squared: 0.003039

F-statistic: 2.094 on 1 and 358 DF, p-value: 0.1487

Regression #3 - Diff in fppi on max temperature

```
reg.diff.max.30 <- lm(diff_fppi ~ st_days_max_over_30, reg_data)
reg.diff.max.25 <- lm(diff_fppi ~ st_days_max_over_25, reg_data)
summary(reg.diff.max.25)
```

Call:

```
lm(formula = diff_fppi ~ st_days_max_over_25, data = reg_data)
```

Residuals:

Min	1Q	Median	3Q	Max
-11.0245	-1.1044	0.0956	1.0221	12.9814

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.3044392	0.1617886	1.882	0.0607 .
st_days_max_over_25	-0.0003953	0.0003559	-1.111	0.2674

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.531 on 358 degrees of freedom

Multiple R-squared: 0.003434, Adjusted R-squared: 0.0006508

F-statistic: 1.234 on 1 and 358 DF, p-value: 0.2674

```
summary(reg.diff.max.30)
```

Call:

```
lm(formula = diff_fppi ~ st_days_max_over_30, data = reg_data)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-11.1665	-1.0872	0.0944	1.0604	12.8716

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.2395542	0.1509839	1.587	0.113
st_days_max_over_30	-0.0006657	0.0012734	-0.523	0.601

Residual standard error: 2.535 on 358 degrees of freedom
Multiple R-squared: 0.0007627, Adjusted R-squared: -0.002028
F-statistic: 0.2733 on 1 and 358 DF, p-value: 0.6015

Regression #4 - Diff in fppi on days max over 30c, diff in cpi, monthly precipitation

```
reg.diff.ovb <- lm(diff_fppi ~ st_days_max_over_30 +
                    monthly_precipitation_average +
                    cpi_diff,
                    reg_data)
summary(reg.diff.ovb)
```

Call:

```
lm(formula = diff_fppi ~ st_days_max_over_30 + monthly_precipitation_average +
    cpi_diff, data = reg_data)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-11.1684	-1.0882	0.1029	1.0089	12.9594

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.3889085	0.3074413	1.265	0.207
st_days_max_over_30	-0.0002854	0.0013812	-0.207	0.836
monthly_precipitation_average	-0.0044076	0.0066200	-0.666	0.506
cpi_diff	0.1321216	0.3716487	0.356	0.722

Residual standard error: 2.54 on 356 degrees of freedom
Multiple R-squared: 0.002463, Adjusted R-squared: -0.005943
F-statistic: 0.293 on 3 and 356 DF, p-value: 0.8305

Appendix

Regression of diff in fppi on lagged weather max over 30

```
reg.lag.30 <- lm(diff_fppi ~ lag_st_days_max_30, reg_data)
summary(reg.lag.30)
```

Call:

```
lm(formula = diff_fppi ~ lag_st_days_max_30, data = reg_data)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-11.3657	-1.0669	0.1172	1.1197	12.7616

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.270602	0.152290	1.777	0.0764 .
lag_st_days_max_30	-0.001237	0.001279	-0.967	0.3341

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.543 on 355 degrees of freedom

(3 observations deleted due to missingness)

Multiple R-squared: 0.002628, Adjusted R-squared: -0.0001813

F-statistic: 0.9355 on 1 and 355 DF, p-value: 0.3341

Regression of diff in fppi on lagged weather max over 25

```
reg.lag.25 <- lm(diff_fppi ~ lag_st_days_max_25, reg_data)
summary(reg.lag.25)
```

Call:

```
lm(formula = diff_fppi ~ lag_st_days_max_25, data = reg_data)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-11.3787	-1.0402	0.0598	1.1598	12.7790

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.3401668	0.1632079	2.084	0.0379 *
lag_st_days_max_25	-0.0005343	0.0003575	-1.494	0.1360

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.538 on 355 degrees of freedom

(3 observations deleted due to missingness)

Multiple R-squared: 0.006251, Adjusted R-squared: 0.003452

F-statistic: 2.233 on 1 and 355 DF, p-value: 0.136

Regression of diff in fppi on lagged weather max

```
reg.lag.mean <- lm(diff_fppi ~ lag_monthly_mean_temp, reg_data)
summary(reg.lag.mean)
```

Call:

```
lm(formula = diff_fppi ~ lag_monthly_mean_temp, data = reg_data)
```

Residuals:

Min	1Q	Median	3Q	Max
-11.1892	-1.0770	-0.0169	1.1137	12.9314

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.26911	0.14207	1.894	0.059 .
lag_monthly_mean_temp	-0.01759	0.01204	-1.461	0.145

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.539 on 355 degrees of freedom

(3 observations deleted due to missingness)

Multiple R-squared: 0.005973, Adjusted R-squared: 0.003173

F-statistic: 2.133 on 1 and 355 DF, p-value: 0.145

Display results in nice table

```
stargazer(reg.lag.30, reg.lag.25, reg.lag.mean,
  type = "latex",
  title = "Lagged regressions",
  column.labels = c("Lagged days max over 30",
                    "Lagged days max over 25",
                    "Lagged monthly mean temp"),
  header = T,
  font.size = "small")
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Thu, Dec 01, 2022 - 1:54:26 PM

Table 1: Lagged regressions

	<i>Dependent variable:</i>		
	diff_fppi		
	Lagged days max over 30	Lagged days max over 25	Lagged monthly mean temp
	(1)	(2)	(3)
lag_st_days_max_30	−0.001 (0.001)		
lag_st_days_max_25		−0.001 (0.0004)	
lag_monthly_mean_temp			−0.018 (0.012)
Constant	0.271* (0.152)	0.340** (0.163)	0.269* (0.142)
Observations	357	357	357
R ²	0.003	0.006	0.006
Adjusted R ²	−0.0002	0.003	0.003
Residual Std. Error (df = 355)	2.543	2.538	2.539
F Statistic (df = 1; 355)	0.935	2.233	2.133

Note:

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