Oil Prices and Unemployment in Canadian Provinces

Alexander Ondrus

2/27/2020

Question

Originally, my question was:

> How does the monthly average West Texas Intermediate Crude spot price correlate to the unemployment rate in each province in Canada?

Now, I have expanded my question to:

> What is the impact of various energy prices on employment in regions across Canada generally and in Alberta specifically?

Data

Original Chart

The data that I used for my original question is:

- Federal Reserve Bank of St. Louis, Spot Crude Oil Price: West Texas Intermediate (WTI) [WTISPLC], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/WTISPLC, February 27, 2020
- Statistics Canada. Table 14-10-0287-03 Labour force characteristics by province, monthly, seasonally adjusted

Table 14-10-0287-03 is a nearly 1 gigabyte CSV containing many different fields over many different levels of aggregation, so I subset it for the following:

- All age groups ("15 years and over")
- Estimates of the unemployment rate only
- Both sexes
- Seasonally adjusted data

In the future, I may want to adjust these parameters, but for the purposes of uploading this to GitHub I will only post the subset .rds file. The function that I used to generate the file is defined below.

```
filter_unemployment_table <- function(){
  unemployment <- read_csv("14100287.csv") %>%
  filter(GEO != "Canada",
         `Age group` == "15 years and over",
         Statistics == "Estimate",
         `Labour force characteristics` == "Unemployment rate",
         Sex == "Both sexes", `Data type` == "Seasonally adjusted") %>%
  select(REF_DATE, GEO, VALUE)

unemployment$Year <- substr(unemployment$REF_DATE, 1, 4) %>% as.numeric()
```

```
unemployment$Month <- substr(unemployment$REF_DATE, 6, 7) %>% as.numeric()
unemployment$REF_DATE <- NULL
saveRDS(unemployment, file = "unemployment_by_province_monthly.RDS")
}
```

The code chunk below relies on the RDS file produced by the function defined above. If one wanted to introduce a date offset, it could easily be done by adjusting the date column in the `oil` data frame below. Note that I reduce the unemployment by a factor of 100 to make the '`scales::percent` argument work well below.

```
unemployment <- readRDS("unemployment_by_province_monthly.RDS")

oil <- read_csv("WTISPLC.csv")
oil$Month <- month(oil$DATE)
oil$Year <- year(oil$DATE)

unemployment_and_oil <- left_join(unemployment, oil, by = c("Year", "Month")) %>%
  mutate(VALUE = VALUE / 100) %>%
  drop_na()

colnames(unemployment_and_oil) <- c("Province", "Unemployment",
                                    "Year", "Month", "Date", "Oil")
```

I calculate simple correlation coefficients for each province. A more sophisticated analysis will consider the effects of autocorrelation of the respective time series for each variable. The `x` and `y` columns are set manually for a nice positioning of the R^2 values on the final charts.

```
u_o_rsq <- unemployment_and_oil %>%
  group_by(Province) %>%
  summarise(rsq = cor(Oil, Unemployment)^2)

u_o_rsq$labels <- paste("R^2 ==", round(u_o_rsq$rsq, 2))
u_o_rsq$x <- 100
u_o_rsq$y <- 0.20

kable(u_o_rsq) %>% kable_styling(full_width = FALSE)
```

Province	rsq	labels	x	y
Alberta	0.1341367	R^2 == 0.13	100	0.2
British Columbia	0.2877269	R^2 == 0.29	100	0.2
Manitoba	0.2284668	R^2 == 0.23	100	0.2
New Brunswick	0.3603982	R^2 == 0.36	100	0.2
Newfoundland and Labrador	0.4145582	R^2 == 0.41	100	0.2
Nova Scotia	0.3097419	R^2 == 0.31	100	0.2
Ontario	0.0076244	R^2 == 0.01	100	0.2
Prince Edward Island	0.1872470	R^2 == 0.19	100	0.2
Quebec	0.3526427	R^2 == 0.35	100	0.2
Saskatchewan	0.2466897	R^2 == 0.25	100	0.2

Plots

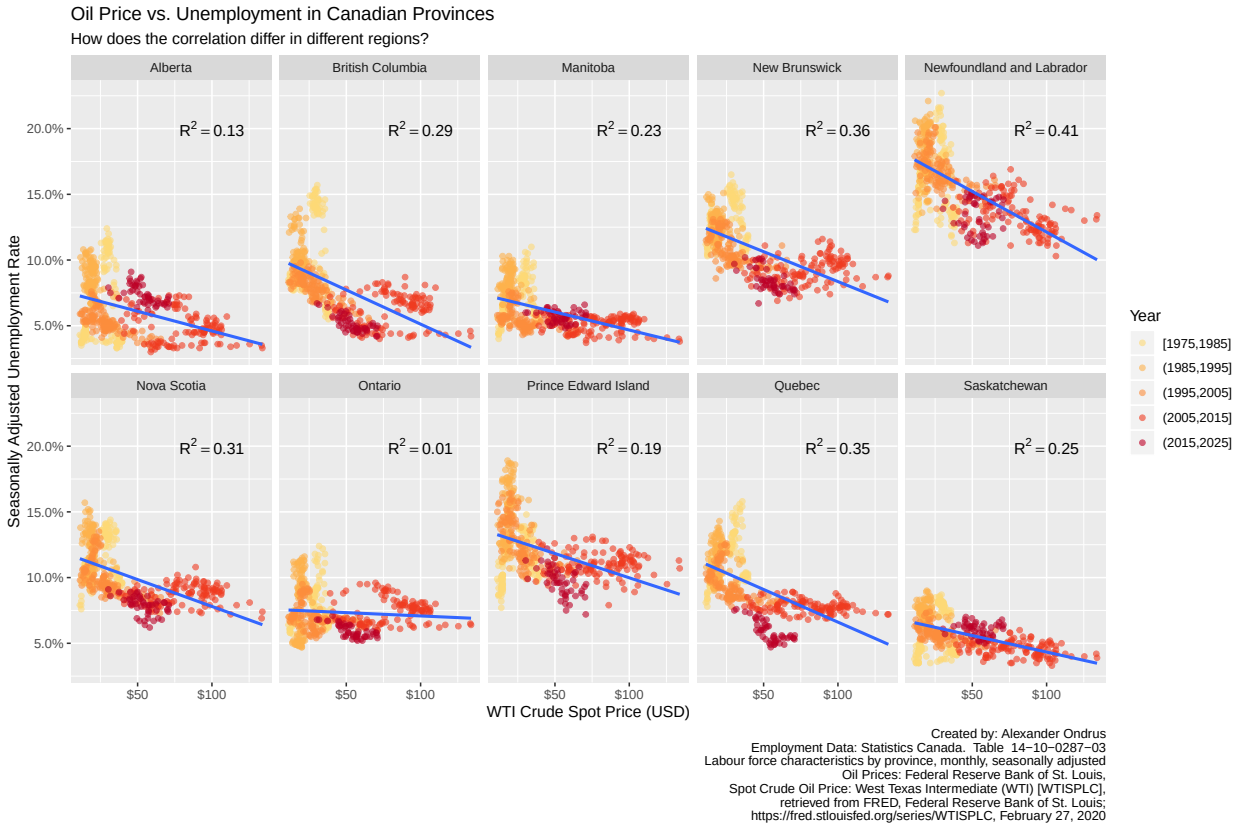
Original

To reduce the problem of lighter colours, I use a custom palette which makes one more colour than categories but drops the lightest colour.

```
custom_palette <- brewer.pal(6, "YlOrRd")[2:6]

p <- ggplot(unemployment_and_oil, aes(x = Oil, y = Unemployment)) +
    geom_point(aes(colour = cut_width(Year, 10)),
               alpha = 0.6) +
    geom_smooth(method = "lm", se = FALSE) +
    facet_wrap (~ Province, nrow = 2) +
    scale_colour_manual(values = custom_palette) +
    geom_text(data = u_o_rsq, aes(x = x, y = y, label = labels), parse = TRUE) +
    labs(title = "Oil Price vs. Unemployment in Canadian Provinces",
         subtitle = "How does the correlation differ in different regions?",
         x = "WTI Crude Spot Price (USD)",
         y = "Seasonally Adjusted Unemployment Rate",
         colour = "Year",
         caption = "Created by: Alexander Ondrus\nEmployment Data: Statistics Canada.  Table  14-10-028
  scale_x_continuous(labels = scales::dollar) +
  scale_y_continuous(labels = scales::percent)

plot(p)
```



Second Version

Thanks to Jens von Bergmann for the idea of using `shape = 21`. This will allow me to use a lower value of alpha (to compensate for over-plotting) while still maintaining the visibility of lighter colours.

```
p1 <- ggplot(unemployment_and_oil, aes(x = Oil, y = Unemployment)) +
    geom_point(aes(fill = cut_width(Year, 10)),
               alpha = 0.6,
               shape = 21,
               colour = "white") +
    geom_smooth(method = "lm", se = FALSE) +
    facet_wrap (~ Province, nrow = 2) +
    scale_fill_manual(values = custom_palette) +
    geom_text(data = u_o_rsq, aes(x = x, y = y, label = labels), parse = TRUE) +
    labs(title = "Oil Price vs. Unemployment in Canadian Provinces",
         subtitle = "How does the correlation differ in different regions?",
         x = "WTI Crude Spot Price (USD)",
         y = "Seasonally Adjusted Unemployment Rate",
         fill = "Year",
         caption = "Created by: Alexander Ondrus\nEmployment Data: Statistics Canada.  Table  14-10-028
  scale_x_continuous(labels = scales::dollar) +
  scale_y_continuous(labels = scales::percent)

plot(p1)
```

Outputting the Plot

I include the `ggsave()` command below as it contains the parameters that I find work well on either mobile or desktop browsers (in twitter at least).

```
ggsave("Oil_vs_Unemployment.jpeg", plot = p, width = 30, height = 20, units = "cm", dpi = "retina")
```

Next Steps

1. Investigate the statistical significance of the difference in R^2-values for each of the different provinces.
2. Examine the impacts of autocorrelation on the significance of the differences in values.
3. Examine the impacts of natural gas and interprovincial migration.
4. For Alberta specifically, look at the WCS price versus unemployment as opposed to WTI.
5. For Alberta specifically, look at the correlation between *production* and unemployment as opposed to *price* (then do the same for *employment* as opposed to *unemployment*).
6. For all provinces, examine the correlations that exist for *employment* as opposed to *unemployment*.

New Data and Ideas

Data Sources Some data sources that I had not considered but were mentioned to me by various people on twitter are:

- EIA
- Government of Alberta Economic Dashboard - Oil Production
- Government of Alberta Economic Dashboard - Oil Prices
- Sproule Price Forecast
- Interprovincial Employment in Canada, 2002 to 2011

Ideas Thanks to Steven Klaiber-Noble and (several) others who pointed out that I had not considered the impacts of autocorrelation or whether the variables were cointegrated. Steven pointed me to a well-written blog post on the topic as well.

Rob Johnston also mentioned that I could integrate the impacts of natural gas prices on employment.

Here is a resource on variance of the error term (*heteroskedasticity*) that contains some useful R code as well as theoretical background. (Thanks to Brenton Kenkel)