A couple of months ago [I blogged about consumer spending on vehicle fuel by income](http://freerangestats.info/blog/2018/07/01/petrol-spend). The impetus for that post was the introduction on 1 July 2018 of an 11.5 cent per litre [“regional fuel tax” in Auckland](https://www.transport.govt.nz/land/regional-fuel-tax/).

One vehement critic of the levy, largely on the grounds of its impact on the poor, has been [Sam Warburton](https://nzinitiative.org.nz/about-us/our-people/sam-warburton/) (@Economissive on Twitter) of the New Zealand Initiative. Since early May 2018 (ie seven weeks before the fuel levy began), Sam has been collecting fuel prices from [pricewatch.co.nz](http://pricewatch.co.nz/), and he recently [made the data available](https://twitter.com/Economissive/status/1028861304306425856).

One of the issues of controversy about a levy like this is whether it will lead to “price spreading” – fuel companies absorbing some of the extra tax in Auckland and increasing prices in other regions. A relatively small number of firms make retail pricing decisions about fuel in New Zealand so it’s plausible that imperfect competition is making this possible. I had a look at the data to see if the intervention of the fuel levy in New Zealand’s biggest city can be seen to impact on fuel pricing in the rest of the country.

To cut to the chase, this graphic exemplifies my approach and results:

We see that after the spike at the time the tax was introduced, fuel prices in other regions have converged somewhat on Auckland’s prices (particularly when considering the relative change happening before the tax). The impact of the tax is still clearly felt much more strongly in Auckland than anywhere else (as of course would be expected – the question at issue is whether *anywhere* else would be impacted at all). More on that later.

**The data**

First, let’s look at the data Sam’s collected. The Twitter thread linked to above provides an Excel workbook on Dropbox. There’s a worksheet for each region (which are defined similarly, but not identically, to New Zealand’s official Regions) as well as one describing the source. For each region we have data on fuel prices for combinations of Company, Date, fuel type (eg diesel, 91 octane, etc) and Region. If we plot all the data other than liquid petroleum gas (which has particularly sparse observations), it looks like this:

“Companies” have been sorted in order of increasing average price for that graphic, but fairly crudely (ie not taking into account different mixes of fuel type by Company).

We can see we have more data for 91 octane petrol than the other types. For the rest of this post I’ll be focusing on just 91 octane.

Here’s the R code to tidy up the data to this point and draw the graphic. It assumes you’ve manually downloaded the Excel workbook to your working folder (I’m currently working with severely restricted internet, so couldn’t experiment in automating that process.)

library(tidyverse)

library(scales)

library(openxlsx)

library(nlme)

# Import data:

sn <- getSheetNames("fuel price data.xlsx")

sn <- sn[sn != "Source"]

fuel\_orig <- list()

for(i in 1:length(sn)){

tmp <- read.xlsx("fuel price data.xlsx", sheet = sn[i], cols = 1:7,

detectDates = TRUE, na.strings = c("NA", "n/a"))

tmp[ , "region"] <- sn[i]

fuel\_orig[[i]] <- tmp

}

# Combine into a single data frame

fuel\_df <- do.call("rbind", fuel\_orig)

# some useful extra information:

south\_island <- c("Canterbury", "Nelson", "Otago", "Southland", "West Coast")

big\_four <- c("CALTEX", "Z ENERGY", "BP", "MOBIL")

# Make long, thin, tidy version:

fuel\_tidy <- fuel\_df %>%

select(-LPG) %>%

gather(fueltype, value, -Company, -Date, -region) %>%

filter(!is.na(value)) %>%

mutate(island = ifelse(region %in% south\_island, "South", "North"),

company\_type = ifelse(Company %in% big\_four, "Big Four", "Smaller")) %>%

mutate(region = fct\_reorder(region, as.numeric(as.factor(island))),

Company = fct\_reorder(Company, value))

# Overview graphic:

fuel\_tidy %>%

ggplot(aes(x = Date, y = value, colour = Company)) +

facet\_grid(fueltype~region, scales = "free\_y") +

geom\_point(size = 0.8) +

scale\_y\_continuous("Price per litre at the pump", label = dollar) +

labs(x = "Date in 2018",

caption = "Source: pricewatch.co.nz, collated by @Economissive") +

ggtitle("Petrol prices in New Zealand over several months in mid 2018") +

guides(colour = guide\_legend(override.aes = list(size=5))) +

scale\_colour\_brewer(palette = "Set1")

**Regional comparisons with Auckland**

I tried a couple of different ways of comparing prices in individual regions with those in Auckland. I think this graphic is probably the most informative and straightforward:

The grey line in the background of each facet represents Auckland’s price; the shaded blue rectangle is the post-tax period (ie 1 July 2018 and onwards). The grey shaded area shows the difference between the given region’s price and that of Auckland.

We can see a lot of regional variation here, and an interesting pattern with three (or maybe even all) of the South Island regions experiencing price declines in June then picking up in July. Of course, the 11.5 cent increase in price in Auckland is very obvious in the grey lines and shading. Later on I’ll be using time series intervention analysis on this data; this is an approach commonly used in evaluating the impact of evaluations. If we were only after the direct impact, there would be no need to do any statistical tests beyond this graphic above; the big spike in prices hits you between the eyes, and there is no doubt about the discontinuity in Auckland’s prices on 1 July! The question, of course, is how sustained that impact is, and whether it bled into secondary impacts in other regions.

Here’s a second graphic that tries to visually simplify what’s going on, by calculating a single line of the ratio of prices in each region to those in Auckland. I think what it gains in visual simplicity (less lines and shading) it loses in clear interpretability. In particular, it’s not possible to tell from this graphic what changes in the graphic come from changes in Auckland, and which come from changes in the comparison region. That’s not a deal-breaker for using a graphic like this, but it does strongly suggest we should also include the first one, with the rawer average prices per region plainly shown without transformation, for context.

Note that even after the introduction of the Auckland-specific fuel tax, 91 octane petrol in several regions still costs more than in Auckland. The regions with prices higher than Auckland in the most recent data in the collection are West Coast, Otago, Nelson, Canterbury and Wellington.

Here’s the R code for those two graphics:

# construct two convenient summaries of the data, different ways of comparing regions to Auckland:

p91 <- fuel\_tidy %>%

filter(fueltype == "91") %>%

group\_by(region, island, Date) %>%

summarise(value = mean(value, tr = 0.2)) %>%

ungroup()

p91\_rel <- p91 %>%

group\_by(Date) %>%

mutate(Auckland = value[region == "Auckland"]) %>%

filter(! region %in% c("Auckland", "Wairarapa")) %>%

mutate(perc\_of\_auck = value / Auckland)

# Plot showing original price data

ggplot() +

# annoying trick necessary here to draw a semi-transparent background rectangle:

geom\_rect(data = data.frame("hello world"),

xmin = as.Date("2018-07-01"), xmax = Inf, ymin = -Inf, ymax = Inf, fill = "blue", alpha = 0.1) +

# now draw the actual data:

geom\_ribbon(data = p91\_rel, aes(x = Date, ymin = Auckland, ymax = value), fill = "grey", alpha = 0.5) +

geom\_line(data = p91\_rel, aes(x = Date, y = Auckland), colour = "grey50") +

geom\_line(data = p91\_rel, aes(x= Date, y = value, colour = island), size = 1.2) +

facet\_wrap(~region, ncol = 3) +

scale\_y\_continuous("Price of 91 octane petrol compared to in Auckland\n", label = dollar) +

labs(x = "2018; grey line shows Auckland",

caption = "Source: pricewatch.co.nz, collated by @Economissive")

# ratio of Auckland prices to others

ggplot() +

geom\_hline(yintercept = 1, colour = "grey50") +

geom\_rect(data = data.frame("hello world"),

xmin = as.Date("2018-07-01"), xmax = Inf, ymin = -Inf, ymax = Inf, fill = "blue", alpha = 0.1) +

geom\_line(data = p91\_rel, aes(x= Date, y = perc\_of\_auck, colour = island)) +

facet\_wrap(~region, ncol = 3) +

scale\_y\_continuous("Price of 91 octane petrol as a percentage of in Auckland\n", label = percent) +

labs(x = "2018",

caption = "Source: pricewatch.co.nz, collated by @Economissive")

**Modelling the impact of an intervention over time**

When it came to directly addressing our question of interest regarding price spreading, I opted to group all non-Auckland regions together and compare average prices there with those in Auckland. There are better ways of modelling this that make full use of the granular data available (mostly involving mixed effects models, and more complex ways of representing the trend over time than linearly; and they would certainly take into account weighting from the spread in population over regions) but they come with big costs in complexity that I don’t have time for right now. Plus, the difference-of-averages method struck me as the easiest way to interpret and communicate, not to mention think about, the question of whether prices were converging back towards eachother after the initial shock of the addition of the tax. This leads me to the graphic I showed earlier in this post:

The linear regression lines shown in that graphic are a simplified version of the formal statistical model we want to fit and use to test our hypothesis. We’re looking for evidence that the slope of the post-tax line is materially less than the slope of the pre-tax line; in other words, is the gap in pricing between Auckland and other regions declining after the initial 11.5 cent shock of the tax.

I defined this as a simple linear model, but fit it using generalized least squares with time series residuals (auto-regressive moving average of order (1, 1)). This is straightforward to specify and fit using the gls function in Pinheiro, Bates et al’s nlme package, but there are other ways of doing it too.

This results in the following:

|  |  |
| --- | --- |
|  | |
|  | *Dependent variable:* |
|  |  |
|  | value |
|  | |
| Date | 0.001\*\* |
|  | (0.0003) |
|  |  |
| post\_tax | 19.159\*\* |
|  | (7.804) |
|  |  |
| Date:post\_tax | -0.001\*\* |
|  | (0.0004) |
|  |  |
| Constant | -10.470\*\* |
|  | (4.946) |
|  |  |
|  | |
| Observations | 93 |
| Log Likelihood | 330.591 |
| Akaike Inf. Crit. | -647.182 |
| Bayesian Inf. Crit. | -629.761 |
|  | |
| *Note:* | \*p<0.1; \*\*p<0.05; \*\*\*p<0.01 |

…which simply confirms what is visually obvious, that there is indeed statistically significant evidence of the slope changing direction downwards after the tax is introduced. In other words, we do have evidence consistent with some degree of “spreading” taking place. After the initial clean shock of the introduction of the tax, prices in Auckland and in the rest of the country are indeed converging somewhat; although nowhere near as much as the full cost of the tax.

This effect holds whether I use all of New Zealand as the comparison point or just the South Island (which has less competition in fuel retailers) or just the North Island, although in the latter case the effect is not as strong (as can be seen in the graphic). It also doesn’t seem to matter whether we use all available prices, or just those of the “big four” companies that are present in all regions.

We can’t say for sure the effect comes from introducing the tax from just looking at the numbers. Drawing that conclusion would require carefully considering any other possible causality options. For example, one driver of the pattern we’re seeing is clearly that prices in Canterbury, Nelson and Otago stopped declining and started rising slightly in July. What are other plausible causes of that pattern? Understanding and considering such alternative theories would need more knowledge of the fuel market in New Zealand than I have, so I’ll leave it to others to debate that. All I can safely conclude is what I wrote above:

after the spike caused by the tax, fuel prices in Auckland and in the rest of the country are converging somewhat (although much less than the full cost of the tax), and plausibly this is because of companies’ price adjustments down in Auckland and up elsewhere to spread the cost of the tax over a broader base.

Here’s the code for that final graphic and statistical modelling;

# Data on the difference between Auckland's average price and those in other areas:

diff\_data <- fuel\_tidy %>%

filter(fueltype == "91" & company\_type == "Big Four") %>%

group\_by(Date) %>%

summarise(auck\_v\_rest =

mean(value[region == "Auckland"]) -

mean(value[region != "Auckland"]),

auck\_v\_si =

mean(value[region == "Auckland"]) -

mean(value[island == "South"]),

auck\_v\_ni =

mean(value[region == "Auckland"]) -

mean(value[island == "North" & region != "Auckland"]),

) %>%

mutate(post\_tax = as.integer(Date >= as.Date("2018-07-01"))) %>%

gather(comparison, value, -Date, -post\_tax) %>%

mutate(comparison = case\_when(

comparison == "auck\_v\_si" ~ "Compared to South Island",

comparison == "auck\_v\_ni" ~ "Compared to rest of North island",

comparison == "auck\_v\_rest" ~ "Compared to all NZ except Auckland"))

# Graphic:

ggplot(diff\_data, aes(x = Date, y = value)) +

facet\_wrap(~comparison, ncol = 3) +

geom\_line() +

geom\_smooth(aes(group = post\_tax), method = "lm") +

scale\_y\_continuous("Average price of 91 octane petrol in Auckland\nminus average price in comparison area",

label = dollar) +

labs(x = "Date in 2018\nAverage prices have not been weighted by population or sales",

caption = "Source: pricewatch.co.nz, collated by @Economissive") +

ggtitle("Fuel prices in Auckland compared to three other comparison areas",

"Restricted to prices from BP, Caltex, Mobil and Z Energy")

# Modelling:

# first, make a convenient subset of the data (useful later in various tests and diagnostics):

D <- subset(diff\_data, comparison == "Compared to all NZ except Auckland")

# Fit model, taking care to specify time series residuals, which aren't as useful for inference

# as i.i.d. residuals and hence lead to more conservative inference:

model <- gls(value ~ Date \* post\_tax,

data = subset(diff\_data, comparison == "Compared to all NZ except Auckland"),

cor = corARMA(p = 1, q = 1))

# print-friendly summary of coefficieints

stargazer::stargazer(model, type = "html")

# more comprehensive summary (not shown in blog):

summary(model)

**Observations per region**

A topic of interest in fuel pricing debate in New Zealand is the number of companies present in each region, with a particular focus on the presence of Gull. In case of interest, here are the observations in the pricewatch data collected by Sam Warburton:

| **region** | **–** | **ALTERNATE FUEL** | **GULL** | **CHALLENGE** | **CALTEX** | **MOBIL** | **Z ENERGY** | **GAS ALLEY** | **BP** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Auckland | 206 | 143 | 349 | 251 | 279 | 372 | 279 | 279 | 372 |
| Bay of Plenty | 262 | 200 | 295 | 202 | 278 | 264 | 279 | 161 | 370 |
| Coromandel | 3 | 6 | 249 | 4 | 209 | 219 | 233 | 69 | 277 |
| Waikato | 0 | 0 | 228 | 227 | 276 | 302 | 278 | 220 | 367 |
| Hawke’s Bay | 0 | 45 | 217 | 16 | 261 | 228 | 255 | 132 | 345 |
| Northland | 0 | 0 | 205 | 52 | 274 | 229 | 256 | 268 | 275 |
| Manawatū-Whanganui | 0 | 108 | 193 | 120 | 275 | 265 | 277 | 176 | 371 |
| Taranaki | 0 | 184 | 178 | 188 | 243 | 88 | 274 | 188 | 301 |
| Wellington | 0 | 91 | 78 | 252 | 278 | 277 | 279 | 213 | 372 |
| East Coast | 0 | 41 | 71 | 158 | 221 | 122 | 145 | 104 | 211 |
| Central Plateau | 0 | 0 | 44 | 51 | 159 | 244 | 275 | 0 | 295 |
| Wairarapa | 0 | 1 | 0 | 0 | 2 | 42 | 3 | 2 | 103 |
| Canterbury | 0 | 262 | 0 | 279 | 279 | 278 | 279 | 245 | 372 |
| Nelson | 0 | 50 | 0 | 34 | 216 | 237 | 263 | 163 | 237 |
| Otago | 0 | 225 | 0 | 226 | 273 | 265 | 277 | 116 | 364 |
| Southland | 0 | 114 | 0 | 197 | 249 | 195 | 226 | 98 | 245 |
| West Coast | 0 | 115 | 0 | 227 | 119 | 121 | 115 | 65 | 242 |

That table was generated by:

library(knitr)

fuel\_tidy %>%

group\_by(region, Company) %>%

summarise(freq = n()) %>%

spread(Company, freq, fill = 0) %>%

arrange(desc(GULL)) %>%

kable