Emissions Reduction from Residential Electricity Consumption

Jonathan M. Gilligan

12/4/2019

Introduction

This document explains the calculation of the estimated reduction in greenhouse gas emissions due to the reduction of per-capita residential electricity consumption, as presented by Davis (2017). Davis observed that after many decades of steady growth, per-capita residential electricity consumption in the United States leveled off around 2010 and then began dropping. Davis further attributed a large part of this departure from the long-term trend to the widespread adoption of energy-efficient light bulbs in U.S. homes.

Here, I calculate the impact of that drop on greenhouse gas emissions. First, following Davis, I obtained annual data on retail electricity sales, aggregated by state (EIA 2019) and annual state-level poopulation estimates from the U.S. census.

First, I calculate residential electric sales and population for each state for each year 1990–2018 and use these to calculate the national per-capita residential electricity consumption.

Then I fit a straight line to the trend from 1990–2006 and extrapolate this line to 2018 as a quasi-business-as-usual counterfactual without the growth of efficient lightbulb use. The pre-2007 data fits the line very nicely so the extrapolation seems reasonable.

Finally, I follow Davis (2017) in fitting a quadratic function to the full data set to represent the actual trend (Fig. 1)

Next, I repeat the regression analysis for each state, individually and calculate the difference in state-wide residential electricity sales (in MWh per year) for 2018 as estimated both by the counterfactual linear estrapolation and by the quadratic estimate of the actual trend for that state. Then I multiply this difference (reduction of use, in MWh per year) by that state's emissions factors, as reported by the EPA's eGRID 2016 analysis (EPA 2018). This givess the emissions reduction (CO_2e) due to the reduction of residential electricity demand.

The total reduction in electricity consumption for the US is 296 TWh, which corresponds to a reduction in greenhouse gas emissions of 110 million metric tons CO_2e.

Methods

Set up by loading required libraries

```r
library(pacman)
p_load(tidyverse, magrittr, lubridate, readxl, broom, janitor, censusapi)
p_load_gh("walkerke/tidycensus")

data_dir = here::here("data")
if (! dir.exists(data_dir)) dir.create(data_dir)
```

Electricity Data

Download Data Files

First, we download the electricity sales data from the Energy Information Agency (EIA 2018):
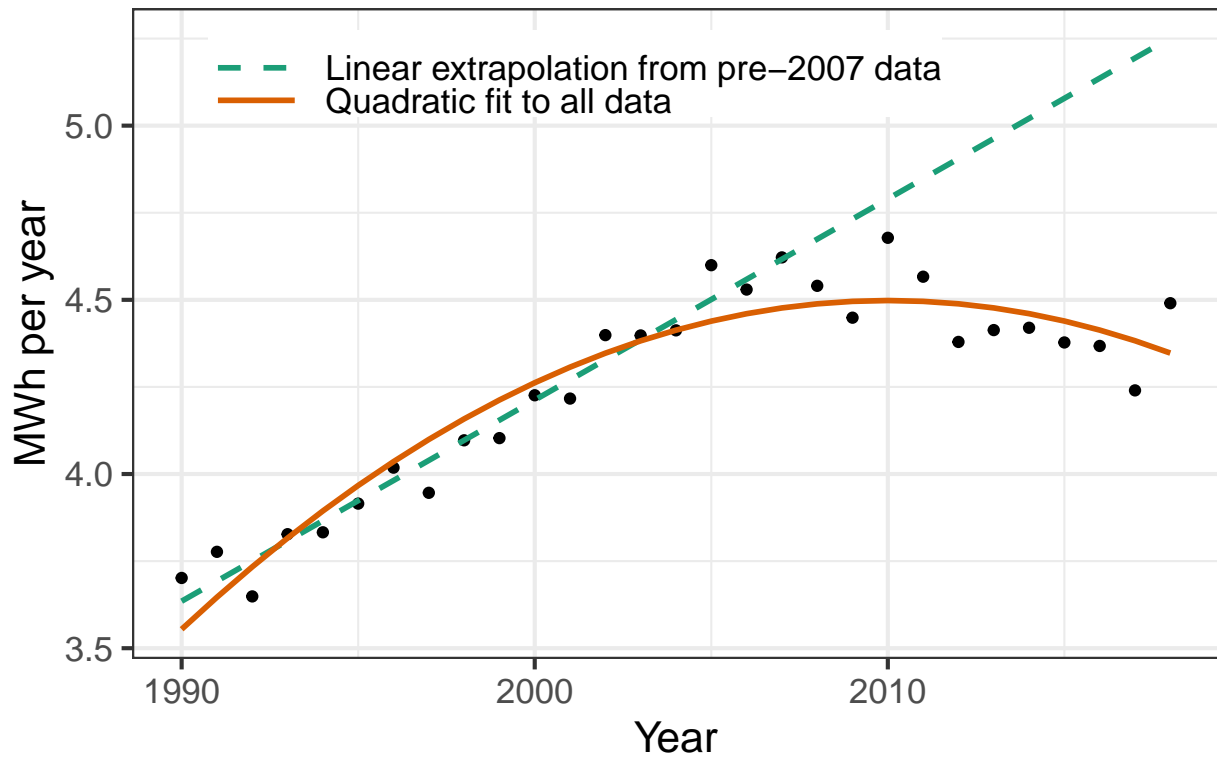
Per Capita Electricity Consumption

Figure 1: Analysis of per-capita electricity consumption in the U.S., following Davis (2017). The linear extrapolation is fit to the data from 1990–2006 and the quadratic approximation is fit to the full set of observations. The linear extrapolation reflects a counterfactual "business as usual" scenario that follows the trend through the last year before Walmart began its major effort to promote energy-efficient light bulbs.

```r
sales_file <- file.path(data_dir, "sales_annual.xlsx")
sales_url <- "https://www.eia.gov/electricity/data/state/sales_annual.xlsx"

if (! file.exists(sales_file)) {
  download.file(sales_url, sales_file, mode = "wb")
}
```

Next, we get emissions factors for states from the Environmental Protection Agency (EPA 2018):

```r
emission_file <- file.path(data_dir, "egrid2016_summarytables.xlsx")
emission_url <-
  "https://www.epa.gov/sites/production/files/2018-02/egrid2016_summarytables.xlsx"

if (! file.exists(emission_file)) {
  download.file(emission_url, emission_file, mode = "wb")
}
```

Wrangle Data

Now we read the data in and organize it:

```r
# Make a tibble of state FIPS codes. Add `00` for the US.
state_fips <- fips_codes %>% select(state, state_code, state_name) %>%
  distinct() %>%
  bind_rows(tibble(state = "US", state_code = "00",
                   state_name = "United States"))

use_data <- read_excel(sales_file, skip = 2,
                       na = c("", "NA"),
                       col_names = c("year", "state", "sector",
                                     "residential", "commercial", "industrial",
                                     "transportation", "other", "total"),
                       col_types = c("numeric", "text", "text",
                                     rep("numeric", 6))) %>%
  mutate(sector = factor(sector)) %>% left_join(state_fips, by = "state")

# Divide emission factors by 2200 to convert from pounds per MWh to
# metric tons per MWh
emission_factors <- read_excel(emission_file, "Table 3", skip = 3) %>%
  clean_names() %>% rename(state = x1) %>% mutate(e_factor = co2e / 2200) %>%
  filter(! is.na(state))

use_data <- use_data %>%
  left_join(select(emission_factors, state, e_factor), by = "state") %>%
  mutate(emissions = residential * e_factor)
```

State Population Data

Getting the census data is a bit more difficult because the census organizes its annual current population estimates differently each inter-censal decade.

To run the code that follows, you will need to get a free API key from the U.S. Census at http://api.census.gov/data/key_signup.html and install it in R, as described by the tidycensus package. documentation for the census_api_key() function

Now that we have a census API key, we can get the data for 1990–1999. One complication is that for the

1990s intercensal, the Census bureau reports annual population estimates by county, but does not, as far as I can tell, give an easy way to get estimates aggregated to the state level. No matter, because we can add up the county data to obtain state totals.

```r
state_pop_1990s_file <- file.path(data_dir, "state_pop_1990s.Rds")

if (file.exists(state_pop_1990s_file)) {
  state_pop_1990s <- read_rds(state_pop_1990s_file)
} else {
  county_pop_1990s <- map_df(fips,
                             ~getCensus("pep/int_charagegroups", 1990,
                                        vars=c("POP", "YEAR", "STATE",
                                               "COUNTY", "AGEGRP",
                                               "RACE_SEX", "HISP"),
                                        region="county:*",
                                        regionin=str_c("state:", .x))
                             ) %>%
    as_tibble() %>% clean_names() %>% mutate(pop = as.numeric(pop))

  # YEAR is two digit, so add 1900
  state_pop_1990s <- county_pop_2 %>%
    group_by(year, state) %>% summarize(pop = sum(pop)) %>% ungroup() %>%
    mutate(year = 1900 + as.integer(year)) %>%
    filter(year >= 1990) %>% rename(state_code = state)

  write_rds(state_pop_1990s, state_pop_1990s_file)
}

sp_1990 <- state_pop_1990s
```

Data for the 2000s intercensal decade is much easier because it"s published at the state level. One complication is that this data includes both the actual 2000 decennial census (population on April 1) and the estimated July 1 population. For simplicity, and to keep the times consistent from year to year, we keep the July 2000 data and discard the April 2000 data, but there isn't an appreciable difference at the level of precision and accuracy of this calculation.

```r
state_pop_2000s_file <- file.path(data_dir, "state_pop_2000s.Rds")

if (file.exists(state_pop_2000s_file)) {
  state_pop_2000s <- read_rds(state_pop_2000s_file)
} else {
  state_pop_2000s <- getCensus("pep/int_charagegroups", 2000,
                               vars=c("POP", "DATE_", "DATE_DESC", "GEONAME"),
                               region="state:*") %>%
    as_tibble() %>% clean_names()
  write_rds(state_pop_2000s, state_pop_2000s_file)
}

sp_2000 <- state_pop_2000s %>%
  mutate(date = date_desc %>% str_replace_all("([0-9]) *[A-Za-z].*$", "\\1") %>% mdy) %>%
  filter(date != "2000-04-01") %>%
  mutate(year = year(date), pop = as.numeric(pop)) %>%
  select(year, state_code = state, pop)
```

Finally, we get the data for the 2010 decennial census and the 2011–2019 annual estimates. See documentation

for time-series format at https://www.census.gov/data/developers/data-sets/popest-popproj/popest/popest-vars/2018.html. The date code is an integer where 3–11 represent estimates of the July 1 population for 2010–2018, respectively.

```
state_pop_2010s_file <- file.path(data_dir, "state_pop_2010.Rds")

if (file.exists(state_pop_2010s_file)) {
  state_pop_2010s <- read_rds(state_pop_2010s_file)
} else {
  state_pop_2010s <- get_estimates("state", "population", time_series = TRUE) %>%
    clean_names() %>% rename(state = geoid)
  write_rds(state_pop_2010s, state_pop_2010s_file)
}

sp_2010 <- state_pop_2010s %>% filter(date >= 3) %>%
  mutate(year = 2010 + date - 3, date = make_date(year, 7, 1)) %>%
  spread(key = variable, value = value) %>% clean_names() %>%
  select(year, state_code = state, pop)
```

```
sp <- bind_rows(sp_1990, sp_2000, sp_2010) %>%
  filter(state_code != "00")

sp_us <- sp %>% group_by(year) %>%
  summarize(state_code = "00", pop = sum(pop)) %>%
  ungroup()

sp <- bind_rows(sp, sp_us)
```

Merge Data

Now we merge the electricity sales data with the state populations:

```
per_capita_use <- use_data %>% filter(sector == "Total Electric Industry") %>%
  left_join(sp, by = c("state_code", "year")) %>%
  mutate(per_capita_use = residential / pop,
         per_capita_emissions = emissions / pop)

us_total <- per_capita_use %>% filter(state != "US") %>%
  group_by(year, sector) %>%
  summarize(residential = sum(residential),
            emissions = sum(emissions),
            pop = sum(pop),
            per_capita_use = residential / pop,
            per_capita_emissions = emissions / pop) %>%
  ungroup()
```

Analyze Data

Analyze the data using two methods:

1. Linear regression for the data from 1990–2006. This represents the trend before 2007. I pick 2007 because that is the year when WalMart made its big push to sell 100 million compact fluorescent light bulbs in a single year and thus, this marks the beginning of the concerted private-sector initiative to put efficient light bulbs into America's homes.
2. Fit a quadratic function to the full data set. This follows Davis's

(2017) analysis.

```r
fits <- us_total %>%
  mutate(early = lm(per_capita_use ~ year, data = .,
                    subset = year < 2007) %>%
           augment(newdata = tibble(year = year)) %$% .fitted,
         full = lm(per_capita_use ~ poly(year, 2), data = .) %>%
           augment(newdata = tibble(year = year)) %$% .fitted) %>%
  select(year, early, full) %>%
  pivot_longer(cols=-year, names_to = "key",
               values_to = "per_capita_use") %>%
  mutate(Fit = factor(key, levels = c("early", "full"),
                      labels = c("Linear extrapolation from pre-2007 data",
                                 "Quadratic fit to all data")))
```

Now, calculate the reduction in emissions for the US:

```r
us_pop_latest <- us_total %>% top_n(1, year) %$% pop

delta_consumption_us <- fits %>%
  pivot_wider(id_cols = year, names_from = key, values_from=per_capita_use) %>%
  mutate(delta = early - full) %>% arrange(year) %>%
  top_n(1, year) %$% delta * us_pop_latest
```

To calculate the emissions change, we can't just multiply the total reduction in residential electricity consumption by the average emissions factor for the US because emissions factors and emissions reduction both vary considerably from state to state, so we repeat the regression analysis state by state, calculate the change in electricity consumption for each state and then multiply by that state's emissions factor and finally, add up each state's change in emissions:

```r
emissions_change <- per_capita_use %>% filter(state != "US") %>%
  select(year, state, sector, per_capita_use, e_factor, pop) %>%
  group_by(state, sector) %>%
  mutate(early = lm(per_capita_use ~ year, data = ., subset = year < 2007) %>%
           augment(newdata = tibble(year = year)) %$% .fitted,
         full = lm(per_capita_use ~ poly(year, 2), data = .) %>%
           augment(newdata = tibble(year = year)) %$% .fitted,
         delta = early - full,
         delta_total = pop * delta) %>%
  top_n(1, year) %>%
  mutate(delta = delta * e_factor, delta_total = delta_total * e_factor) %>%
  ungroup() %>%
  group_by(sector) %>%
  summarize(delta_total = sum(delta_total), pop = sum(pop),
            delta_per_capita = delta_total / pop) %>%
  ungroup()

delta_emissions_us <- emissions_change$delta_total
```

References

Davis, Lucas W. 2017. "Evidence of a decline in electricity use by U.S. households," Economics Bulletin **37**, 1098–1105. http://www.accessecon.com/Pubs/EB/2017/Volume37/EB-17-V37-I2-P96.pdf Retrieved Nov. 26, 2019.

EIA 2019. *Retail Sales of Electricity by State by Sector by Provider*, U.S. Energy Information Agency.

https://www.eia.gov/electricity/data/state/ and . Retrieved Nov. 26, 2019.

EPA 2018. *Emissions & Generation Resource Integrated Database (eGRID 2016)*, U.S. Environmental Protection Agency. https://www.epa.gov/energy/emissions-generation-resource-integrated-database-egrid, Summary tables: . Retrieved Nov. 26, 2019.