Years in office, promotion incentives, and fiscal behavior: Evidence from Chinese mayors

Siwei Dai

December 6, 2021

```
library(tidyverse)
library(estimatr)
library(modelsummary)
datapath <- './ps_data/FullData.csv'
df <- as_tibble(read.csv(datapath))
biodatapath <- './ps_data/BiographicalData.csv'
biodf <- as_tibble(read.csv(biodatapath))
fiscdatapath <- './ps_data/fiscal.dta'
fiscdf <- haven::read_dta(fiscdatapath)
```

Project description: Data sources, data tidying, and data description

This project investigates how promotion incentives could impact fiscal behaviors of Chinese mayors over Cities in China are under dual-leadership. The municipal party secretary and the mayor jointly leads the city.

Data Sources

Two datasets are used for analysis. The first dataset is the "Chinese Political Elite Database" (CPED), compiled by Junyan Jiang (accessed at https://www.junyanjiang.com/data.html). This dataset includes the biographical and career information of all mayors and municipal party secretaries from 2000 to 2015, provincial governors and party secretaries (1995 - 2015). This dataset is originally constructed for Jiang (2018), a paper that investigates

Another dataset documents county-level fiscal revenue and expenditure from 1994 to 2007. The dataset is accessed from the replication data of Xu Xu (2021) who purchased and digitized the original "Fiscal Statistics of Cities and Counties" compiled by the Budget Department of the Ministry of Finance of China. The original paper concerns the effect of implementing digital surveillance programs on public security spending.

Data Tidying

Before running analyses, I tidied and merged the two datasets to produce a leader-year dataset. On the CPED dataset, I first translated and created the key variables needed for analysis; I then converted the CPED dataset to a county-year dataset using the start date and end date of politicians' terms.

Two problems arise As some politicians may only start their position closer to December, they may not have sufficient control over the fiscal expenditure and income that particular year. For politicians who came into office between July to December, I coded the first year of their tenure the next year. In the following sections, there is a figure that shows the distribution of the month in which newly-appointed mayors start their jobs.

On the county-level fiscal data, I dealt with certain coding discrepancies within the original data. In some cases, missing data is coded `NA` and in other cases, it's coded 0. I recoded the missing data points on social expenditure and fiscal revenue to `NA`s.

Finally, I merged the datasets together, and filter out only the mayors and their terms.

Data Description

The final dataset includes 1397 mayors and has 5143 leader-year observations. Each observation documents who the leaders were at a particular city in a given year, the amount of city-level fiscal expenditure and revenue.

```
df_tidy <- df %>%
    # translate the variables into English
    # exper_num refers to the career stage
    rename('identifier' = 用户编码,
           'name' = 姓名,
           'job' = 标志位,
           'position_ori' = 级别,
           'exper_num' = 经历序号,
           'start_date' = 起始时间.YYYY.MM.DD.,
           'end_date' = 终止时间..YYYY.MM.DD..,
           'prefectural_code' = 二级关键词编码,
           'provincial_code' = 一级关键词编码,
           'position_code' = 职务一级关键词编码) %>%
    select(identifier, name, job, position_ori, exper_num, start_date, end_date,
           prefectural_code, provincial_code, position_code) %>%
    # impute prefectural code from provincial code
    # if the prefectural code is missing due to being a provincial leader
    mutate(prefectural_code = ifelse(is.na(prefectural_code), provincial_code,
                                     prefectural_code)) %>%
    # recode the position from strings into numeric; larger the number, higher the position
    mutate('position_numeric' = case_when(position_ori == '无级别' ~ 0,
                                          position_ori == '小于副处' ~ 1,
                                          position_ori == '副处' ~ 2,
                                          position_ori == '正处' ~ 3,
                                          position_ori == '副厅' ~ 4,
                                          position_ori == '正厅' ~ 5,
                                          position_ori == '副部' ~ 6,
                                          position_ori == '正部' ~ 7,
                                          position_ori == '副国' ~ 8,
                                          position_ori == '正国' ~ 9)) %>%
    # Code the term length
    mutate('term_length' = (as.numeric(as.Date(end_date) - as.Date(start_date)))/365,
    # As discussed above, for those who started their job between July to December,
    # I code the next year as the start year of their term
           'start_year' = ifelse(as.numeric(format(as.Date(start_date), '%b')) >= 7,
                                 as.numeric(format(as.Date(start_date), '%Y')) + 1,
```

```
                                   as.numeric(format(as.Date(start_date), '%Y'))),
           'end_year' = as.numeric(format(as.Date(end_date), '%Y'))) %>%
    # Create variables to indicate if the official is a mayor, governor or party secretary
    mutate(governor = (job == " 省长"),
           mayor = (job == " 市长"),
           party_secretary = job %in% c(" 市委书记", " 省委书记"),
           leader = governor | mayor | party_secretary)

biodf_tidy <- biodf %>%
    rename('bio_identifier' = X,
           'birth_date' = 出生日期.YYYY.MM.DD.) %>%
    select(bio_identifier, birth_date) %>%
    mutate('birth_year' = as.numeric(format(as.Date(birth_date), '%Y')))

fiscdf_tidy <- fiscdf %>%
    filter(!is.na(admcode) & admcode %in% unique(df_tidy$prefectural_code)) %>%
    # Code the income and expenditure observations that are 0 to NA
    mutate(income_ttl = ifelse(income_ttl == 0, NA, income_ttl),
           exp_ttl = ifelse(exp_ttl == 0, NA, exp_ttl),
           exp_shbz = ifelse(exp_shbz == 0, NA, exp_shbz))
```

Converting the Political Elite datase to a leader-year dataset

```
df_tidy_year <- df_tidy %>%
    # filter out NAs, otherwise there would be errors in converting to leader-year
    filter(!is.na(start_year) & !is.na(end_year) & !is.na(prefectural_code)) %>%
    mutate('year' = map2(start_year, end_year, `:`)) %>%
    unnest(cols = c(year)) %>%
    # filter out the experiences after the fiscal data is available
    filter(year >= 1994) %>%
    # Create two dummies
    mutate('term_year' = year - start_year + 1,
           'first_year' = ifelse(year == start_year, 1, 0),
           'last_year' = ifelse(year == end_year, 1, 0))
```

Merge the dataset with spending data and biographical information and filter out mayors only

```
# Merge the three data frames together into one dataset
df_merged <- fiscdf_tidy %>%
    left_join(df_tidy_year, by = c('admcode' = 'prefectural_code',
                                   'year' = 'year')) %>%
    left_join(biodf_tidy, by = c('identifier' = 'bio_identifier')) %>%
    # Create the key variables
    mutate('income_log' = log(income_ttl),
           'exp_log' = log(exp_ttl),
           'social_exp_log' = log(exp_shbz),
           'edu_exp_log' = log(exp_jy),
           'age' = year - birth_year + 1,
           'age55' = age >= 55,
```

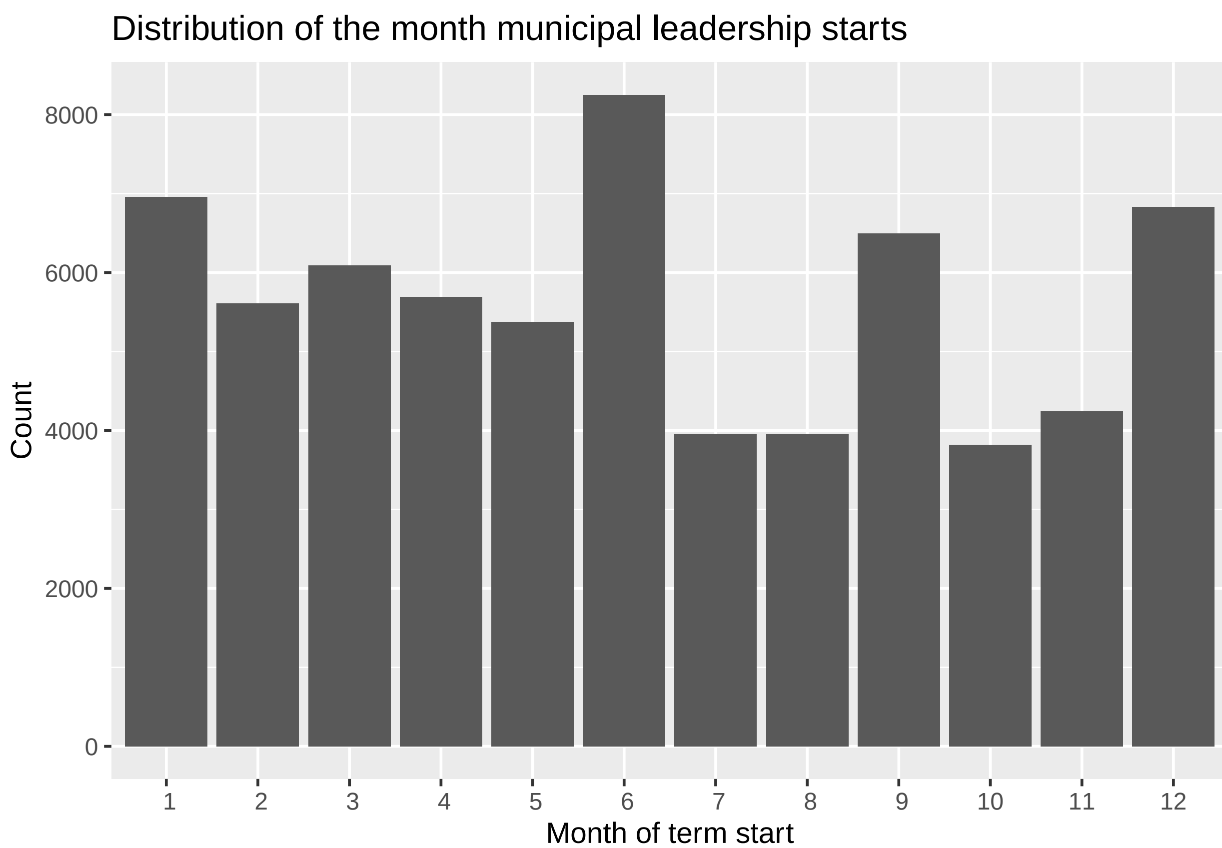
```r
           'midterm' = (term_year >= 3),
           'completeterm' = (term_year >= 5)) %>%
    # Create a term identifier
    mutate(term_identifier = paste(name, admcode)) %>%
    arrange(admcode, year)

df_merged_mayor <- df_merged %>%
    filter(mayor)
```

Distribution of start and end date

As discussed above in the data tidying section, one concern with merging fiscal data with political cycles is how much influence a politician has over fiscal spending and revenue. If all politicians are appointed to their job in January of each year, we can assume they have substantial control over the fiscal behaviors of that year. Alternatively, if politicians are appointed in December of a particular year, they wouldn't have enough control over policies and fiscal revenue/expenditure. The following figure shows the distribution of the month that mayors start their new jobs.

```r
fig_distofstartdate <- df_merged %>%
    filter(!is.na(format(as.Date(start_date), '%b'))) %>%
    ggplot(aes(x = format(as.Date(start_date), '%b'))) +
    geom_bar() +
    labs(x = 'Month of term start',
         y = 'Count',
         title = "Distribution of the month municipal leadership starts")
fig_distofstartdate
```



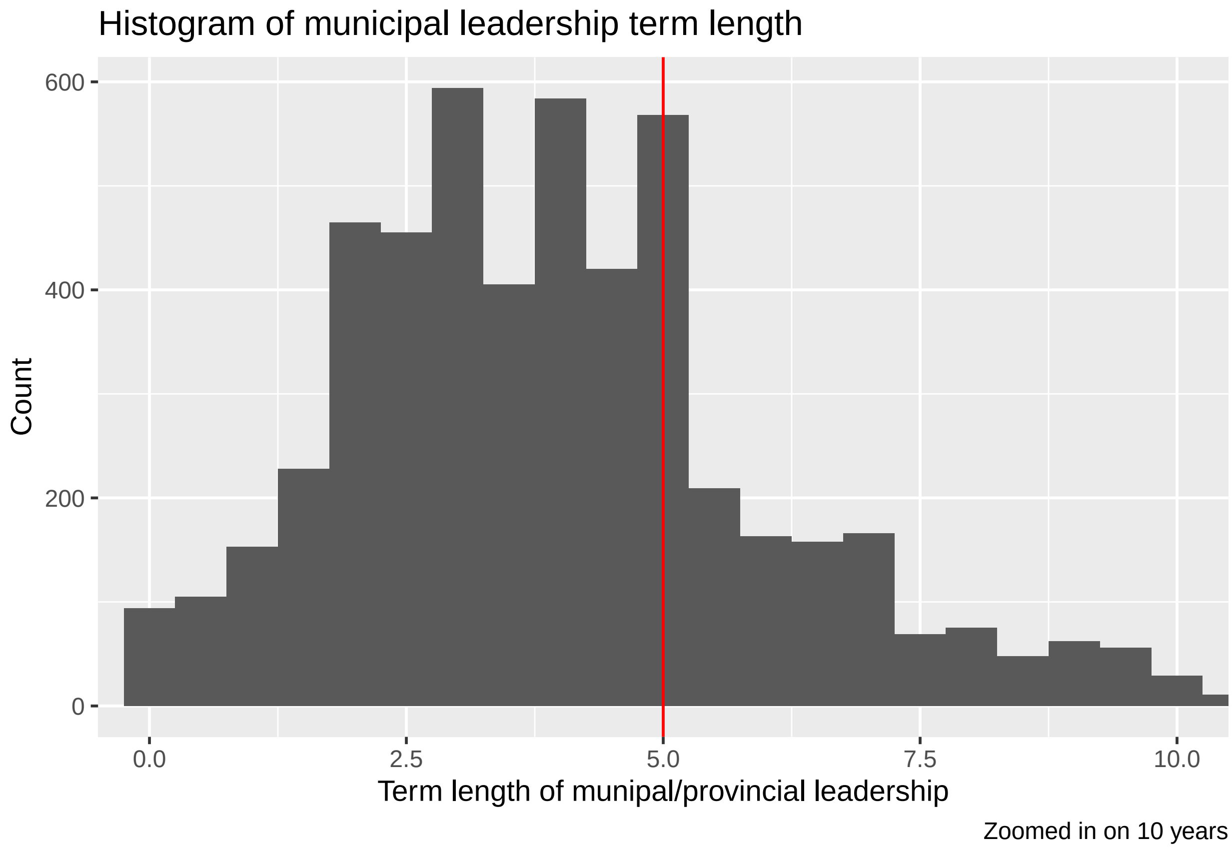

```
# ggsave("./FinalPJ/fig_distofstartdate.png")
```

Background: Promotion incentives, term length, and age constraint

Uncertainty in term prospects: distribution of term length for Chinese mayors

There is a formal term limit for Chinese mayors and provincial governors: 5 years. Yet this formal term limit is not strictly enforced. As shown in the following figure, most of Chinese mayors are promoted or transferred to other positions before reaching the 5-year term limit (the red vertical line on the figure marks the fifth year in office). This means that politicians face substantial uncertainty on how long they could stay at a particular position.

```
fig_tenuretime <- df_merged_mayor %>%
    group_by(name) %>%
    ggplot(aes(x = term_length)) +
    geom_histogram(binwidth = 0.5) +
    coord_cartesian(xlim = c(0, 10)) +
    geom_vline(color = "red", xintercept = 5) +
    # Change x coordinates
    labs(x = "Term length of munipal/provincial leadership",
         y = "Count",
         title = "Histogram of municipal leadership term length",
         caption = "Zoomed in on 10 years")
fig_tenuretime
```

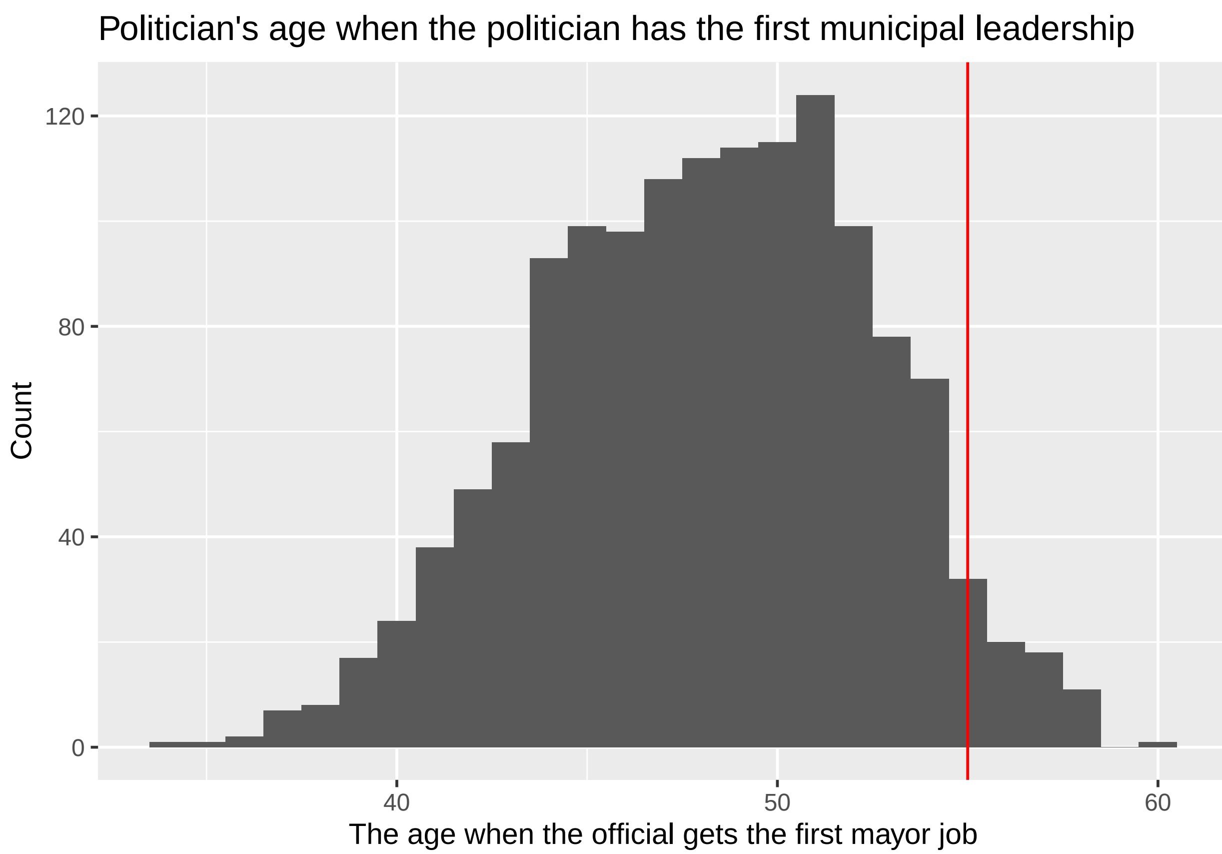



```
# ggsave("./FinalPJ/fig_tenuretime.png")
```

Age constraint in promotions

Chinese politicians also face another constraint in promotions: age. For politicians who have reached the position of mayor and the corresponding administrative rank, the age constraint for further promotion is 55 years old (Kou & Tsai, 2014). The following figure shows the age at which politicians got their first mayor job. The red vertical line marks the 55-year-old cutoff. This figure shows for the bulk of the politicians,

```
fig_promotiontime <- df_merged_mayor %>%
    group_by(name) %>%
    filter(age == min(age)) %>%
    ggplot(aes(x = age)) +
    geom_histogram(binwidth = 1) +
    geom_vline(color = "red", xintercept = 55) +
    labs(x = "The age when the official gets the first mayor job",
         y = "Count",
         title = "Politician's age when the politician has the first municipal leadership")
fig_promotiontime
```



Promotion incentives and Fiscal Behaviors

Given the age constraint on future promotions and substantial uncertainty on how long a politician can serve as a mayor, we would predict that in the starting years of their terms, politicians are incentivised to signal their competence to the upper-level organizational department for early promotion. Promotion prospects may prompt politicians to adopt certain policies and fiscal choices in the beginning of their terms. Given the time period this project centers on (1995 to 2007), I argue Chinese mayors may focus on two objectives: boost economic performance and maintain social stability (or prevent collective action). Therefore, I choose to investigate how fiscal revenue and social expenditure changes over the years in office.

To boost economic performance, Chiense mayors may create tax exemptions or sell the state-owned land cheaply to developers for development at the beginning of their term; this would result in lower fiscal revenue in the early years of their mayorship. On preventing collective actions and maintaining social stability, Chiense mayors may opt to spend more on social welfare and distribute more resources to reduce social grievances at the beginning of their term.

Alternatively, if the politician has reached the age of 55, the impact of career incentives on policy choices and fiscal behaviors is expected to diminish as the politician is excluded from further promotion give the age constraint.

Building on these , I generate two hypotheses to test in empirical analysis: Hypothesis 1: Fiscal revenue would increase as a politicans Hypothesis 2: Social expenditure decreases over the

Dependent variables

The key dependent variables are city-level `social expenditure` and `fiscal revenue`. The dependent variables are logged in the regression models.

Independent variables

The key independent variables are `years in office` and `Age >= 55`. `Age >= 55` is a dummy variable indicating if a particular year, a polician's age have reached or exceeded 55, the cutoff age for potential promotion. I'm interested in the interaction term of `years in office` and `Age >= 55`: the hypothesis put forward above means that if

The alternative specification of using the continuous variable of `age` is not viable because `age` is perfectly collinear with `Years in office`.

In the models, I control for year fixed effects and mayoralty fixed effects.

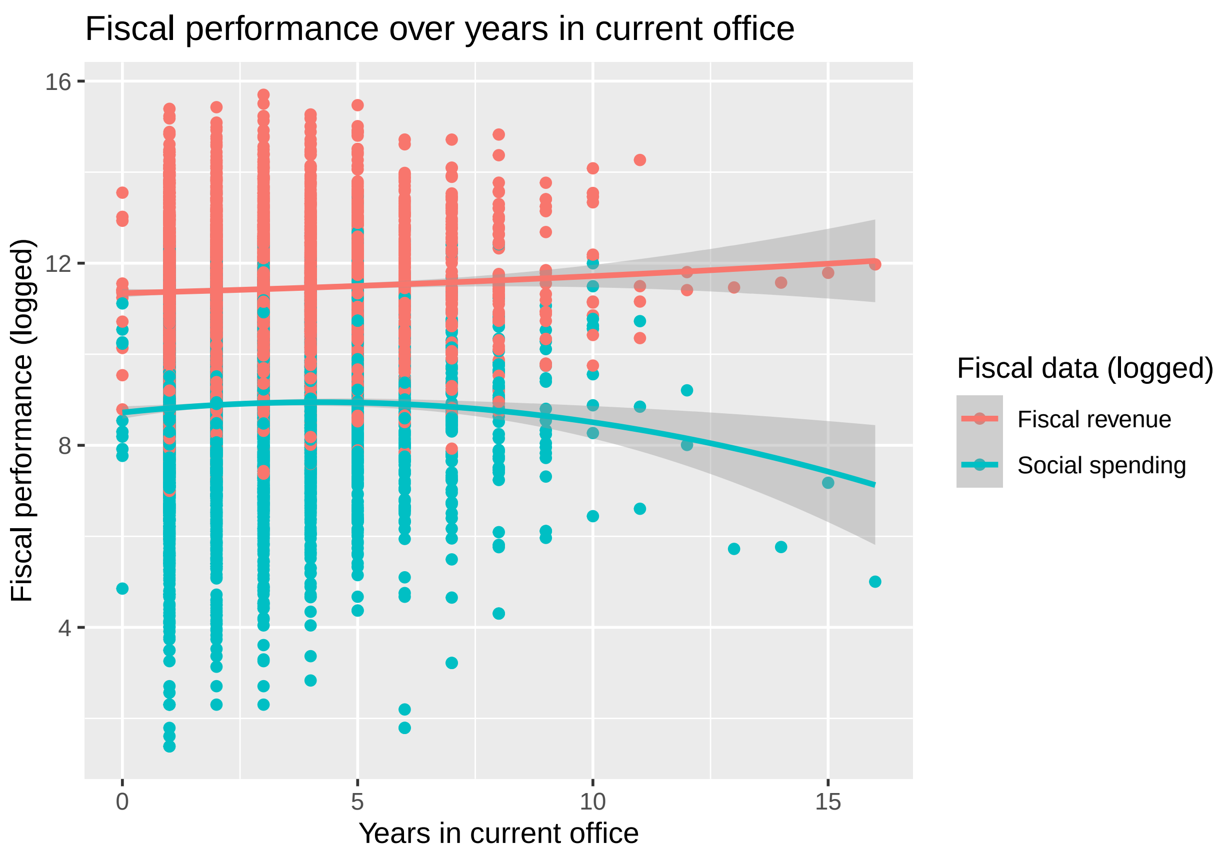
To account for the effect of age on promotion incentives, I have also

Do Chinese mayors change fiscal behaviors over different years in an office?

The following figure shows how

I used `pivot_longer` to reshape the dataset and

```r
# filter out who the mayors/leaders are
fig_fisc <- df_merged_mayor %>%
    pivot_longer(cols = c(social_exp_log, income_log),
                 names_to = "fiscal_type",
                 values_to = "values") %>%
    ggplot(aes(x = term_year, y = values, color = fiscal_type)) +
    geom_point() +
    geom_smooth(method = 'lm', formula = y ~ poly(x, 2)) +
    labs(x = "Years in current office",
         y = "Fiscal performance (logged)",
         title = "Fiscal performance over years in current office") +
    scale_color_discrete(name = "Fiscal data (logged)",
                         labels = c("Fiscal revenue", "Social spending"))
fig_fisc
```




```
# ggsave("./FinalPJ/fig_fisc.png")
```

Regression analysis

Income trends

```
# OLS linear model
model_inc <- lm(income_log ~ term_year + factor(year) + factor(term_identifier),
                data = df_merged_mayor)
# Estimate with robust standard errors
modelr_inc <- lm_robust(income_log ~ term_year,
                        fixed_effects = ~ year + term_identifier,
                        data = df_merged_mayor)
# Add a non-linear term into the model
modelr_inc2 <- lm_robust(income_log ~ poly(term_year, 2),
                         fixed_effects = ~ year + term_identifier,
                         data = df_merged_mayor)
# Add an interaction term of term_year with Age >= 55
modelr_inc_interage <- lm_robust(income_log ~ age55*term_year,
                                 fixed_effects = ~ year + term_identifier,
                                 data = df_merged_mayor)
mdlist_inc <- list(model_inc, modelr_inc, modelr_inc2, modelr_inc_interage)
# Create two rows indicating fixed effects have been added to the models
```

	Model 1	Model 2	Model 3	Model 4
Years in office	0.019+	0.019**	2.575**	0.005
	(0.011)	(0.007)	(0.804)	(0.008)
Years in office (squared)			0.768**	
			(0.291)	
Age >= 55				−0.046+
				(0.028)
Age >= 55 X Years in office				0.020**
				(0.006)
Num.Obs.	5089	5089	5089	4683
R2	0.984	0.984	0.984	0.984
R2 Adj.	0.978	0.978	0.978	0.978
AIC	−1579.6			
BIC	8333.8			
Log.Lik.	2306.784			
F	148.926			
Std.Errors		HC2	HC2	HC2
Mayoralty fixed effects	Y	Y	Y	Y
Year fixed effects	Y	Y	Y	Y

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

```r
tbl_fe2 <- tibble("variable" = c("Mayoralty fixed effects", "Year fixed effects"),
                 "values" = "Y") %>%
    cbind(replicate(.$values, n = length(mdlist_inc) - 1))
# Producing the regression tables
tbl_inc <- modelsummary(mdlist_inc,
             stars = TRUE,
             coef_map = c("term_year" = "Years in office",
                          "poly(term_year, 2)1" = "Years in office",
                          "poly(term_year, 2)2" = "Years in office (squared)",
                          "age55TRUE" = "Age >= 55",
                          "age55TRUE:term_year" = "Age >= 55 X Years in office"),
             add_rows = tbl_fe2)
tbl_inc
```

Interpretations

Table 1 shows the regression results when using city-level fiscal revenue (logged) as the dependent variable. In all the models, mayoralty fixed effects and year fixed effects are controlled.

`Model 1` estimates a linear model of fiscal revenue on the years a politican has been a mayor. The coefficient on the `Years in office` variable is 0.019, which means for every 1 year increase in `Years in office` is associated with a 0.019 increase in logged fiscal revenue. The estimated standard error on the OLS coefficient of `Years in office` is 0.011, which is the estimate of the standard deviation of the coefficient on `Years in office` over a sampling distribution. The p-value is 0.0714325, which means the probability that we would observe a value at least as extreme as the estimate under the null distribution is 0.0714325. This means we fail to reject the null hypothesis that the coefficient on `Years in office` is zero at a p-value $p < 0.1$.

`Model 2` is estimated with robust standard errors and the same specification as model 1. The coefficient on the `Years in office` variable remains the same as the one in model 1. The estimated robust standard error on the coefficient of `Years in office` is 0.007, which is smaller than the standard error estimated in model 1. The p-value is 0.0036624; in other words, we can reject the null hypothesis that the coefficient on `Years in office` is zero at a p-value $p < 0.01$.

`Model 3` uses a second-degree polynomial of `Years in office` in the specification. It is also estimated with robust standard errors. The coefficient on `Years in office` is 2.575, and it is larger than the coefficient on `Years in office` in model 1 and model 2. The corresponding p-value is 0.001379; this means we can reject the null hypothesis that the coefficient on `Years in office` is zero at a p-value $p < 0.01$. On the second-degree term of `Years in office`, the coefficient is 0.768 The corresponding p-value is 0.0084328. This p-value also means that we can reject the null hypothesis that the coefficient on second-degree term of `Years in office` is zero at a p-value $p < 0.01$. These results suggest that fiscal revenue may have a quadratic relationship

`Model 4` adds an interaction term of years in office and `Age >= 55`, and is estimated with robust standard errors. The coefficient on the `Years in office` variable is 0.005, with a corresponding p-value of 0.5162814. The coefficient on `Age >= 55` is -0.046 with a p-value of 0.0991564.

The coefficient on the interaction term (`Age >= 55` * `term_year`) is 0.02 with a p-value of 0.0018224. The p-values suggest that we fail to reject the null hypotheses that the coefficient on each term is 0.

Regression using social expenditure (logged) as dependent variable

```
# This set of regressions use social expenditure (logged) as DV
model_socexp <- lm(social_exp_log ~ term_year  + factor(year) + factor(term_identifier),
                   data = df_merged_mayor)
# Produce robust standard errors
modelr_socexp <- lm_robust(social_exp_log ~ term_year,
                           fixed_effects = ~ year + term_identifier,
                           data = df_merged_mayor)
# Add a non-linear term
modelr_socexp2 <- lm_robust(social_exp_log ~ poly(term_year, 2),
                            fixed_effects = ~ year + term_identifier,
                            data = df_merged_mayor)
# add an interaction with age
modelr_socexp_interage <- lm_robust(social_exp_log ~ age55*term_year,
                                    fixed_effects = ~ year + term_identifier,
                                    data = df_merged_mayor)
mdlist_socexp <- list(model_socexp, modelr_socexp, modelr_socexp2, modelr_socexp_interage)
# Create a tibble to be added into the regression table,
# reflecting if fixed effects have been added into the model
tbl_fe <- tibble("variable" = c("Mayoralty fixed effects", "Year fixed effects"),
                 "values" = "Y") %>%
    cbind(replicate(.$values, n = length(mdlist_socexp) - 1))
# Produce the regression table
tbl_socexp <- modelsummary(mdlist_socexp,
             stars = TRUE,
             # output = "./FinalPJ/tbl_socexp.jpg",
             coef_map = c("term_year" = "Years in office",
                          "poly(term_year, 2)1" = "Years in office",
                          "poly(term_year, 2)2" = "Years in office (squared)",
                          "age55TRUE" = "Age",
                          "age55TRUE:term_year" = "Age X Years in office"),
             add_rows = tbl_fe)
tbl_socexp
```

	Model 1	Model 2	Model 3	Model 4
Years in office	0.034	0.034	4.471	0.056
	(0.033)	(0.022)	(2.812)	(0.035)
Years in office (squared)			0.696	
			(1.556)	
Age				0.077
				(0.079)
Age X Years in office				0.008
				(0.018)
Num.Obs.	4425	4425	4425	4123
R2	0.951	0.951	0.951	0.952
R2 Adj.	0.928	0.928	0.928	0.929
AIC	6786.9			
BIC	15 759.1			
Log.Lik.	−1990.452			
F	41.782			
Std.Errors		HC2	HC2	HC2
Mayoralty fixed effects	Y	Y	Y	Y
Year fixed effects	Y	Y	Y	Y

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Interpretation part: coef, standard errors, p-values

Table 2 reports the regression results using logged social spending as the dependent variable.

Model 1 is a linear model of regressing social expenditure on years in current office, while controlling for learder and year fixed effects. The coefficient on the `Years in office` variable is 0.034, which means for every 1 year increase in `Years in office` is associated with a 0.034 increase in logged social expenditure. The estimated standard error on the OLS coefficient of `Years in office` is 0.033, which is the estimate of the standard deviation of the coefficient on `Years in office` over a sampling distribution. The p-value is 0.289401, which means the probability that we would observe a value at least as extreme as the estimate under the null distribution is 0.289401. This means we fail to reject the null hypothesis that the coefficient on `Years in office` is zero.

`Model 2` uses the same specification as model 1, but it is estimated with robust standard errors. The estimated robust standard error on the coefficient of `Years in office` is 0.022, which is smaller than the standard error estimated in model 1. The p-value is 0.116041.

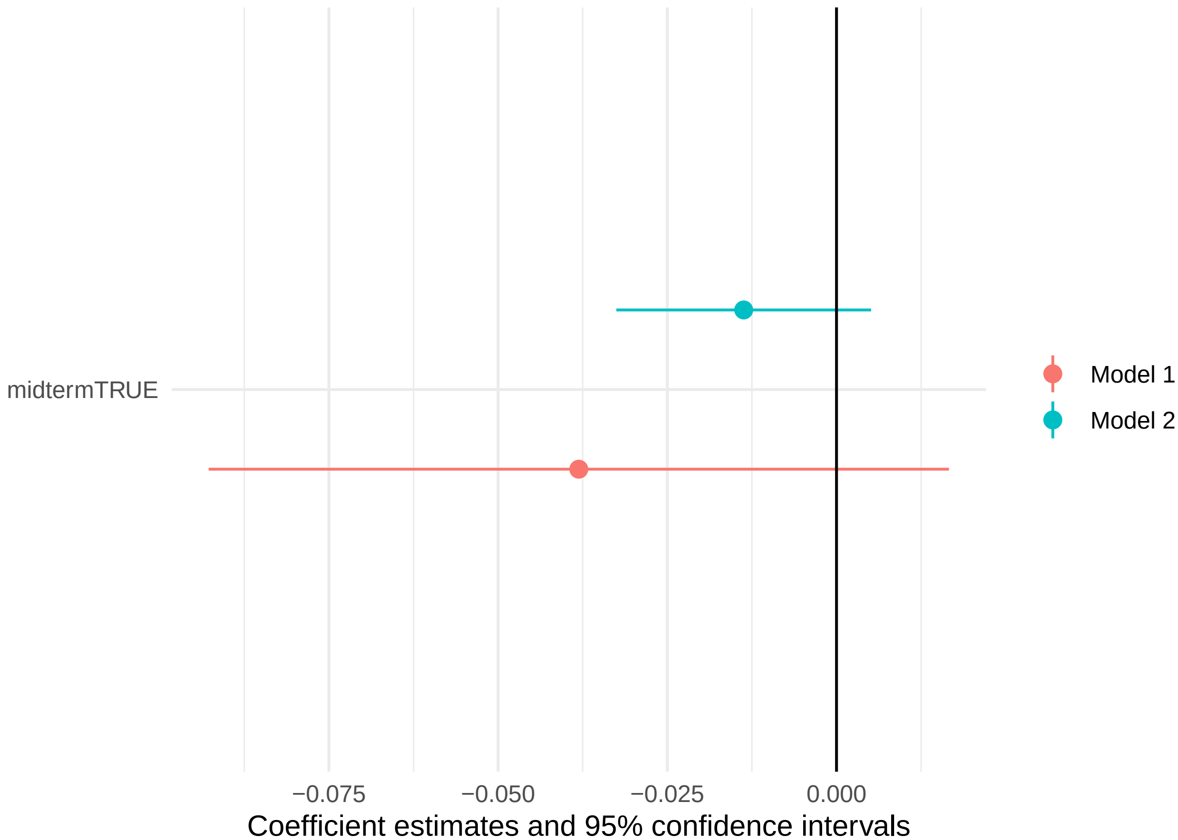
`Model 3` changes the specification and uses a second-degree polynomial of `Years in office`. It is also estimated with robust standard errors. The coefficient on `Years in office` is 4.471, and it is larger than the coefficient on `Years in office` in model 1 and model 2. The corresponding p-value is 0.1119167. On the second-degree term of `Years in office`, the coefficient is 0.696. The corresponding p-value is 0.6544329.

`Model 4` adds an interaction term of years in office and age, and is estimated with robust standard errors. The coefficient on the `Years in office` variable is −0.140, with a corresponding p-value of 0.1081126. The coefficient on `Age >= 55` is 0.077 with a p-value of 0.3312408. The coefficient on the interaction term (`Age >= 55` * `term_year`) is 0.008 with a p-value of 0.6516265. The p-values suggest that we fail to reject the null hypotheses that the coefficient on each term is 0.

Alternative specifications

In the following regressions, I used ### Using indicator variables

```r
# Using indicator variables
model_fac_socexp <- lm_robust(social_exp_log ~ midterm,
                              fixed_effects = ~ year + term_identifier,
                              data = df_merged_mayor)
model_fac_income <- lm_robust(income_log ~ midterm,
                              fixed_effects = ~ year + term_identifier,
                              data = df_merged_mayor)
mdlist_fac <- list(model_fac_socexp, model_fac_income)
modelplot(mdlist_fac) +
    geom_vline(xintercept = 0)
```



As the 95% confidence intervals don't intersect with the red vertical line