

STAT450 Real Estate Proposal

Leonie Lu, Yuetong Liu, Peter Han, Yuting Wen

Summary

Every year, all property owners in BC have to pay property taxes—the single greatest operating expense. The property tax is determined based on property assessment and property tax rate (mill rates). This project will construct a predictive model to estimate property taxes for a given property for the upcoming year. We are interested in tax prediction only in Metro Vancouver because property assessment in Vancouver is likely to be much higher than that in other areas. It might be helpful to perform separate analyses on Metro Vancouver so as to improve prediction accuracy.

Objectives

To predict the property tax for a given property in Metro Vancouver for the upcoming year.

Data

To address our problems, we will use Property Assessment Data (2016-2020) given by our client, and Municipal Budget, which can be found on the BC government website.

Variables:

- *Mill rate (2016-2019)*
- *Assessment (2016-2020)*
- *Tax class code (01, 05, 06)*
- *Area code*
- *Assert type*
- *Year*
- *Municipal Budget of Vancouver (2016-2019)*

Analysis outline (data analysis specifically for Metro Vancouver)

1. Data cleaning

Filter municipalities in greater vancouver (“Metro Vancouver” district): Burnaby, Coquitlam, Delta, Langley, Maple Ridge, North Vancouver, Pitt Meadows, Port Coquitlam, Port Moody, Richmond, Surrey, Vancouver, White Rock, West Vancouver, Bowen Island, Anmore, Belcarra and Lions Bay.

a. Aggregation of data:

Transform the raw data: for each year, compute summary statistics of assessed values (for land and improvement, i.e. current year total) within the same municipality and tax class.

Since assessed values differ a lot in scales, we decide to transform assessed values into mean growth with respect to the same property in the previous year. We will also compute percentage changes of municipal budgets.

- b. Create dummy variables for tax class and municipality code.
- c. Incorporate past mill rates into the dataset.

2. Exploratory data analysis:

- 1. **Visualize the data:** for each municipality and tax class, plot past mill rates, aggregate assessed value, and government budget against time separately.
- 2. Plot percentage change of mill rates, assessed value, government budget against time separately.
- 3. Create box plots of mill rates, grouped by region.
- 4. Test correlations between past mill rates and assessed values and government budget.
- 5. Compute autocorrelations of past mill rates to test the independence of mill rates.

3. Model considered

1. Regression Family:

- a. We plan to use Linear/Non-Linear regression with regularization, such as L1 regularization, L2 regularization, or Elastic Net (combine L1 and L2). Since L1 is robust to outliers, L2 helps to reduce variance.
- b. Define $\% \Delta$ = percentage change between year of interest and the year before that.
- c. We propose two regression models, with and without “year” as explanatory variables.
- d. $\% \Delta$ of mill rate = $\beta_0 + \beta_1 \times \text{year} + \beta_2 \times (\text{municipality class}) + \beta_3 \times (\text{tax class}) + \beta_4 \times (\% \Delta \text{ of assessed value}) + \beta_5 \times (\% \Delta \text{ of budget})$.
- e. $\% \Delta$ of mill rate = $\beta_0 + \beta_1 \times (\text{municipality class}) + \beta_2 \times (\text{tax class}) + \beta_3 \times (\% \Delta \text{ of assessed value}) + \beta_4 \times (\% \Delta \text{ of budget})$. (“percentage change” has included the time effect of previous year).

2. Black box methods:

- a. We are going to try a few black box methods, such as Neural Network, in case the regression model does not perform as expected. These black box methods have good prediction power, but low interpretability.

4. Goodness of fit

We are considering using methods such as AIC, and/or test error from cross validation to indicate the goodness of fit.