Project #1

Commercial Loan Rate Estimation

# Purpose

This project will allow us to practice various R features using live data to support a decision regarding the provision of captive financing to customers at the beginning of this chapter. We will focus on translating regression statistics into R, plotting results, and interpreting ordinary least squares regression outcomes.

# Problem

As we researched how to provide captive financing and insurance for our customers, we found that we needed to understand the relationships among lending rates and various terms and conditions of typical equipment financing contracts.

We will focus on one question:

*What is the influence of terms and conditions on the lending rate of fully committed commercial loans with maturities greater than one year?*

# Data

The data set commloan.csv contains data from the St. Louis Federal Reserve Bank’s [FRED](https://fred.stlouisfed.org/categories/32407) website, which we will use to get some high level insights. The quarterly data extends from the first quarter of 2003 to the second quarter of 2016 and aggregates a survey administered by the St. Louis Fed. There are several time series included. Each loan record is collected by the time that pricing terms were set and by commitment, with maturities more than 365 days from a survey of all commercial banks. Here are the definitions.

|  |  |  |
| --- | --- | --- |
| Variable | Description | Units of Measure |
| rate | Weighted-Average Effec-  tive Loan Rate | percent |
| prepay | Percent of Value of Loans  Subject to Prepayment Penalty | percent |
| maturity | Weighted-Average Matu-  rity/Repricing Interval in Days | days |
| size | Average Loan Size | thousands USD |
| volume | Total Value of Loans | millions USD |

# Work Flow

1. Prepare the data.

* Visit the [FRED](https://fred.stlouisfed.org/categories/32407) website. Include any information on the site to enhance the interpretation of results.
* Use read.csv to read the data into R. Be sure to set the working directory where the data resides. Use

na.omit() to clean the data.

*# setwd('C:/Users/Bill*

*# Foote/bookdown/bookdown-demo-master')*

*# the project directory*

1

x.data <- **read.csv**("data/commloans.csv") x.data <- **na.omit**(x.data)

* Assign the data to a variable called x.data. Examine the first and last five entries (lookup head()). Run a summary of the data set.
* What anomalies appear based on these procedures?

1. Explore the data.

* Let’s plot the time series data using this code:

**require**(ggplot2) **require**(reshape2)

*# Use melt() from reshape2 to build*

*# data frame with data as id and*

*# values of variables*

x.melted <- **melt**(x.data[, **c**(1:4)], id = "date")

**ggplot**(data = x.melted, **aes**(x = date,

y = value)) + **geom\_point**() + **facet\_wrap**(~variable, scales = "free\_x")

* Describe the data frame that melt() produces.
* Let’s load the psych library and produce a scatterplot matrix. Interpret this exploration.

1. Analyze the data.

* Let’s regress rate on the rest of the variables in x.data. To do this we form a matrix of independent variables (predictor or explanatory variables) in the matrix X and a separate vector y for the dependent (response) variable rate. We recall that the 1 vector will produce a constant intercept in the regression model.

y <- **as.vector**(x.data[, "rate"])

X <- **as.matrix**(**cbind**(1, x.data[, **c**("prepaypenalty", "maturity", "size", "volume")]))

**head**(y) **head**(X)

* Explain the code used to form y and X.
* Calculate the *β*ˆ coefficients and interpret their meaning.
* Calculate actual and predicted rates and plot using this code.

*# Insert comment here* **require**(reshape2) **require**(ggplot2) actual <- y

predicted <- X %\*% beta.hat residual <- actual - predicted

results <- **data.frame**(actual = actual, predicted = predicted, residual = residual)

*# Insert comment here*

min\_xy <- **min**(**min**(results$actual), **min**(results$predicted)) max\_xy <- **max**(**max**(results$actual), **max**(results$predicted))

*# Insert comment here*

plot.melt <- **melt**(results, id.vars = "predicted")

*# Insert comment here*

plot.data <- **rbind**(plot.melt, **data.frame**(predicted = **c**(min\_xy,

max\_xy), variable = **c**("actual", "actual"), value = **c**(max\_xy, min\_xy)))

*# Insert comment here*

p <- **ggplot**(plot, **aes**(x = predicted,

y = value)) + **geom\_point**(size = 2.5) +

**theme\_bw**()

p <- p + **facet\_wrap**(~variable, scales = "free") p

* Insert explanatory comments into the code chunk to document the work flow for this plot.
* Interpret the graphs of actual and residual versus predicted values of rate.
* Calculate the standard error of the residuals. Interpret its meaning.

1. Interpret and present results.

* We will produce an R Markdown document with code chunks to document and interpret our results.
* The format will introduce the problem to be analyzed, with sections that discuss the data to be used, and which follow the work flow we have defined.
* We will use the following rubric to assess our performance in producing this summary document.

# Rubric

## General

The assignment solution is due *24 hours prior to the next Live Session*, after the Live Session, for which this assignment is posted.

You will only receive credit for this assignment if the assignment is on time, barring any crises.

You will only receive credit for this assignment if you attempt to answer all questions and address all sections with substantive answers, relevant code, and graphics as needed.

## Specific

Grades for assignments will follow this rubric:

* **Words:** The text is laid out cleanly, with clear divisions and transitions between sections and sub- sections. The writing itself is well-organized, free of grammatical and other mechanical errors, divided into complete sentences, logically grouped into paragraphs and sections, and easy to follow from the presumed level of knowledge.
* **Numbers:** All numerical results or summaries are reported to suitable precision, and with appropriate measures of uncertainty attached when applicable.
* **Pictures:** All figures and tables shown are relevant to the argument for ultimate conclusions. Figures and tables are easy to read, with informative captions, titles, axis labels, and legends, and are placed near the relevant pieces of text.
* **Code:** The code is formatted and organized so that it is easy for others to read and understand. It is indented, commented, and uses meaningful names. It only includes computations which are actually needed to answer the analytical questions, and avoids redundancy. Code borrowed from the notes, from books, or from resources found online is explicitly acknowledged and sourced in the comments. Functions or procedures not directly taken from the notes have accompanying tests which check whether

the code does what it is supposed to. All code runs, and the R Markdown file knits to pdf\_document

output, or other output agreed with the instructor.

* **Modeling:** Model specifications are described clearly and in appropriate detail. There are clear explanations of how estimating the model helps to answer the analytical questions, and rationales for all modeling choices. If multiple models are compared, they are all clearly described, along with the rationale for considering multiple models, and the reasons for selecting one model over another, or for using multiple models simultaneously.
* **Inference:** The actual estimation and simulation of model parameters or estimated functions is technically correct. All calculations based on estimates are clearly explained, and also technically correct. All estimates or derived quantities are accompanied with appropriate measures of uncertainty.
* **Conclusions:** The substantive, analytical questions are all answered as precisely as the data and the model allow. The chain of reasoning from estimation results about the model, or derived quantities, to substantive conclusions is both clear and convincing. Contingent answers (for example, “if X, then Y, but if A, then B, else C”) are likewise described as warranted by the model and data. If uncertainties in the data and model mean the answers to some questions must be imprecise, this too is reflected in the conclusions.
* **Sources:** All sources used, whether in conversation, print, online, or otherwise, are listed and acknowl- edged where they used in code, words, pictures, and any other components of the analysis.