

Group Project Instructions

AEM 2850 / AEM 5850 - Spring 2024

Overview

In this project, you will use tools from the tidyverse to wrangle and visualize stock price and ESG ratings data for major automakers. In practice, if you were to use R for investment research in industry, you would want to use additional packages that provide specialized functions to process and visualize financial data. We do not have time to go into the details of these packages in this course, but this project should give you a taste of how R can be used in industry analysis.

Logistics and expectations

- Work in the groups posted on canvas. Here are some tips for collaborating:
 - Start by agreeing on a way to manage your shared code (e.g., a shared project on Posit Cloud, local files synced through Dropbox, etc.). We set up separate workspaces for each group with projects that allow collaborative editing, so this will probably be the best approach
 - Leave notes to yourself and others in your source code(s)
 - Set up a timeline with regular meetings/check-ins
- **Treat this assignment as a research report your manager asked you to complete**, not a problem set for your professor. Use Quarto to create a polished report in PDF format that summarizes your work, presents visualizations, and discusses takeaways. You have wide latitude in how you structure and present your work. It does not need to – probably should not! – include the verbose guidance we have provided.
- **You will be graded on both the content and its presentation.** Some tips:
 - **Please be concise** when writing text to summarize your work and discuss takeaways
 - **Make each graphic clear and comprehensible (even on its own):** follow best practices from class and readings when making visuals
 - **Do not include your source code in the memo** unless you think it is crucial for communicating your methods or findings. You can avoid printing the code in all the chunks by adding a code chunk that contains `knitr::opts_chunk$set(echo = FALSE)` at the start of the .qmd file (but below the yaml header).
 - **Follow Quarto etiquette rules and style** that we introduced throughout the semester in labs: don't output extraneous messages or warnings, use nice formatting for tables, and adjust the dimensions for your figures as needed (see lab solutions for examples)
- Submit **both** a compiled PDF and a .zip of your project directory that is self-contained, with all of: the .qmd source file, .RProj, the data files we provided, any other scripts or data you used, etc. We should be able to open your project and render your document with one click
- Do not use any packages other than those we use in class or are specifically referenced below
- Each group will submit a single deliverable and receive a single grade, but I reserve the right to penalize students who do not contribute to their group.

Context

In this assignment, you will work with data on stock prices and ESG ratings. You will look at the world's largest automobile manufacturing groups in terms of sales as of 2012 ([The Economist, 2012](#)). Among them, your focus will be on the Volkswagen group.

Back then, analysts and industry observers believed Volkswagen was poised to become the next dominant force in the auto industry, toppling giants like Toyota and General Motors ([Forbes, 2013](#)). The company had restructured its business model to break away from stagnancy in the 1980's and 1990's, streamlining its production process and branching out to new markets.

Among Volkswagen's targets was the American car market, and part of its entry strategy was the introduction of "clean" diesel light vehicles ([USA Today, 2013](#)). While diesel engines emit less carbon dioxide and other greenhouse gases than their petrol counterparts, on the other hand they produce more particulate matter and nitrogen oxides, pollutants that are harmful to human health ([The Conversation, 2017](#)). However, Volkswagen's diesel engines seemed to just be built differently: lab tests showed that their particulate matter emissions were lower than the conventional diesel, so much so that the 2009 VW Jetta was awarded the "Green Car of the Year" prize at the Los Angeles Auto Show, one of the industry's flagship event. It was the first diesel car to receive the honor ([Reuters, 2008](#)).

But, as it turned out, it was too good to be true. In September 2015, after months of investigation, the EPA publicly notified Volkswagen that it had been found cheating emission tests to make its diesel vehicles clear American environmental regulation ([EPA](#)). Volkswagen cars had software capable of detecting when a car was being tested for emissions and adjusting power output in order to reduce the amount of pollutants produced. In reality, however, emissions were 40 times higher than allowed standards when driving on the road under normal conditions ([BBC, 2015](#)). The discovery of this deliberately installed "defeat device" — the "Dieselgate" — undermined Volkswagen's reputation and yielded the company a series of sanctions worldwide, most notably a multi-billion settlement in the United States resulting from its violation of the Clean Air Act. On the other hand, it reportedly marked a turning point in the company's direction not only in terms of technology, but also corporate culture and transparency, seeking closer cooperation with environmental regulators and NGOs ([Forbes, 2022](#)).

Data

We are providing three data files for you to work with:

1. **Auto makers data.xlsx** (Bloomberg): This file contains three sheets:
 - **dict_manufacturers**: a “dictionary” relating company names to tickers, exchanges, and currency their stocks are listed in.
 - **dict_variables**: a “dictionary” explaining what the different price variables mean.
 - **data**: data on daily stock prices.
2. **exchange_rates.csv** (FRED): Contains daily exchange rates of several currencies to American dollar (USD). They’re expressed in USD terms — that is, 1 USD is worth **x** in terms of the other currency.
3. **refinitiv_esg.csv** (WRDS): Six ESG measures by Refinitiv (now LSEG), ranging from 0 to 1, with 1 denoting excellent practices:
 - **EnvironmentPillarScore**: resource use, emissions, and innovation
 - **SocialPillarScore**: workforce, human rights, community, and product responsibility
 - **GovernancePillarScore**: management, shareholders and corporate social responsibility
 - **ESGScore**: a weighted average of the three pillar scores
 - **ESGControversiesScore**: reflects controversies and scandals surrounding the company
 - **ESGCombinedScore**: equal to **ESGScore** if $\text{ESGScore} < \text{ESGControversiesScore}$, a weighted average of the two otherwise

Part 1: Data wrangling

Read and tidy auto manufacturers' stock price data

- a) Read the auto stock price data from the `data` sheet in the Excel file “Auto makers data.xlsx”, accounting for non-standard NA values. Make sure all columns have the expected types (everything should be numeric except the date).

Tip: Browse the Excel file to see what these “non-standard” NA values could be. You can use the `na` argument of `read_excel` to specify that these special values should be treated as missing data. Also, check whether you need to skip rows or select a specific range of cells when importing the `data` sheet.

- b) Pivot the data frame to long format in order to separate company ticker from variable name. Put them in two separate columns: `ticker` and `varname`. Assign this new object to `prices_long`.

Print the first 10 rows of `prices_long`.

- c) Note that the `date` column is of type character. Transform it so that R understands it as a date (is it in “month/day/year” format, or something else?). Then, use it to create a `year` column and a `month` column containing just the year and month of the observation, respectively.

Tip: The `lubridate` library (loaded as part of the core tidyverse) has the convenient functions `year` and `month` you can use here.

- d) Import company names from the `dict_manufacturers` sheet in the Excel file and join with the `prices_long` data.

Present a table showing the number of non-missing observations of the highest price in a day for each company (`PX_HIGH`). This table should have two columns: one with each company's name (not stock symbol) and another with the requested calculation. Make sure your table is nicely formatted.¹ Arrange it by number of non-missing observations in ascending order.

- e) Import “exchange_rates.csv,” making sure that the date column is read correctly (is it in “month/day/year” format, or something else?). If not, convert the column to date format. Join with `prices_long` so that only dates appearing in both data frames are kept. Convert all prices to USD using the daily exchange rates.
- f) Select `date`, `year`, `month`, `company`, `varname` and prices in USD, and pivot `prices_long` so that the different types of prices are spread across different columns. This is now a tidy prices data frame for our purposes (with the level of observation: company-date).²

¹Some starter information on formatting tables in R Markdown — the language Quarto uses to produce pdfs — is available [here](#).

²For other types of analysis, leaving the data in long format rather than spreading prices across columns could be a more appropriate “tidy data” structure.

- g) Focus on the year of 2022. For each company, calculate the mean and standard deviation of the last (closing) price in USD (`PX_LAST`). Present your results in a nicely formatted table with company names on the rows and the two statistics on two different columns. Arrange it by average price in descending order.

Read and tidy auto manufacturers' ESG data

- h) Import “`refinitiv_esg.csv`,” which contains various types of ESG scores for the auto manufacturers. Create two new objects from this data: one containing letter grades and another with numeric scores. Both should have `year-company-score_type` as the level of observation. Join with `company_names` to get company names as a column (you don't need exchange or currency information).
- i) Use the data containing letter grades to make a table of ESG letter grades in 2022. Put company names on the rows and grades in three columns: one for ESG pillars, one for controversies, and one for the combined grade. Sort them by the grades in that order, then alphabetically by company name. Present the results in a nicely formatted table.

Part 2: Volkswagen's Dieselgate and stock prices

- a) Make a plot of Volkswagen's closing stock price over time. Add a vertical line at the date the US EPA issued Volkswagen a notice of violation of the Clean Air Act, marking the beginning of the Dieselgate scandal (September 18, 2015). Be sure to use a caption or figure title to explain the contents of the visualization.³

Would you say the Dieselgate had a meaningful impact on Volkswagen's stock value?

- b) Focus on 2015 only. Create a [candlestick](#) chart for Volkswagen's stock prices. For simplicity, you may just plot a vertical line between the open and close prices each day, and omit the high and low prices. Use logical color-coding for whether the price change was positive or negative each day. Also, like in the previous graph, add the Dieselgate date as a vertical line.

In 2015, prior to Dieselgate, how was Volkswagen stock performing? Do you think investors were bullish on the company, or is there any indication that investors had been losing trust in it for a while?

- c) For each company, not just Volkswagen, calculate the average closing stock price in two periods: within 30 days before and 30 days after the Dieselgate date. Filter your data to this time range, and treat the Dieselgate announcement date as being in the "after" period. Organize your results in a data frame with the two averages side by side. From these averages, calculate the return for each company as the relative change in price in the 30 days after the Dieselgate date compared to 30 days before: $(\bar{p}_{\text{after}} - \bar{p}_{\text{before}})/\bar{p}_{\text{before}}$. Finally, create a bar chart of the return (y) for each company (x), with companies ordered by return in ascending order (make sure their names are legible). Have your y-axis show return as a percentage and use logical color-coding for whether it was positive or negative.

In the immediate aftermath of the Dieselgate scandal news, were there losses across the entire auto industry?

³One way to label figures in a more formal report/memo format is by specifying figure captions within your code chunks using the `fig-cap`, as in [this example](#). Then you could reference the figures by number in your writeup.

Part 3: ESG scores and stock prices

- a) Using the numeric ESG scores data frame, create a line plot of year (x) and score (y). Each company should have its own line, but they should all be light gray except for Volkswagen's, which should be highlighted in black or some other obvious color. Add a vertical line to denote the Dieselgate year (2015). Make one graph for each of the six types of scores and arrange them in two rows.

Given what you read in the Context section of this assignment and the definition of each ESG score in the Data section, do Volkswagen's scores behave as you'd expect?

*Tip: The key here is to use the **group** aesthetic in combination with **color**. For **group**, you use **company**, but for **color** you use a new variable that is equal to "Volkswagen" for Volkswagen and "Others" for the rest. This way, you can set the color of the lines based on this new variable, while still having a line for each company. Then you can customize the line colors using **scale_color_manual**.*

- b) Focus on the "ESGCombinedScore" and switch to the ESG letter grades data frame. Join with the tidy stock prices data frame, keeping only companies and years that appear in both data frames. Use the closing price (PX_LAST) to compute average monthly prices and, with that, compute monthly returns by calculating the percent change in average price from the previous month (do this just across months within a same year, not across years, since the letter grades change discontinuously between years). Create a ridgeline plot of monthly returns, with ESG grade on the y-axis (descending order) and monthly return on the x-axis. Use logical color-coding for whether the company is Volkswagen or not.

Compared to other auto makers, what (if anything) would you say happened to Volkswagen's monthly stock returns as its ESG score deteriorated over time?

*Tip: To calculate monthly returns, take a look at the **dplyr::lag** function. Also, in the graph, you will have the Volkswagen density overlapping with that of other companies for some grades. Use the **alpha** aesthetic to make the densities semi-transparent.*

- c) Using ESG numeric scores, once again focus only on "ESGCombinedScore" and join with the tidy stock prices data frame, keeping only companies and years that appear in both data frames. Calculate the average annual price for each company and year using closing prices (PX_LAST). Normalize these annual average prices by subtracting the mean and dividing by the standard deviation across years (i.e., calculate these statistics for each company using all years of data). Create a scatter plot of normalized prices (y) against ESG scores (x), and fit a linear regression line.

What does this plot suggest about the relationship between ESG scores and stock prices? Is it consistent with the ridgeline plot you created above?

Part 4: AEM 5850 Groups Only

Quarto supports many different output formats beyond the PDF documents we have used for assignments in this class. For example, one can [produce presentations](#) in HTML, PPTX, or PDF formats. After completing the basic report above, please produce a second, separate slide deck deliverable to summarize and extend your analysis:

- a) Extend 3.c by estimating two linear regression models. In the first one, estimate the relationship between closing prices and ESG (combined) scores, including company fixed effects. In the second one, estimate the same relationship, including both company and year fixed effects.

Comment on the results: How do the estimate coefficients on ESG scores compare to one another, and how do they compare to the plot in 3.c? How do you interpret these results? Do you think one, both, or neither could be interpreted as the causal effect of ESG scores on automaker stock prices? How do the estimated standard errors affect your interpretation of the results?

Tip: The functions `fixest::feols`, `broom::tidy`, and/or `modelsummary::modelsummary` may be helpful for estimating and presenting your results.

- b) Create at least one extra visualization based on the data used in this project.
- c) Use Quarto to produce a PowerPoint that summarizes your key findings from this project: include the 2 visualizations from Parts 1-3 that you think are most effective, at least 1 slide of bullet points describing your findings from those visualizations, your regression results from 4.a, and the extra visualization(s) from 4.b. Please submit your rendered .pptx file as well as the .qmd script used to produce it. If you would like to do any manual post-processing of the .pptx file, that is permissible, but please submit the final version as a second file in addition to rendered version.

Grading breakdown

Component	Points
1.a	1
1.b	1
1.c	1
1.d	1
1.e	1
1.f	1
1.g	2
1.h	1
1.i	2
2.a	3
2.b	5
2.c	6
3.a	6
3.b	6
3.c	6
Code submitted	1
Code renders document	3
Overall presentation	5
Total - AEM 2850	52
4.a	5
4.b	5
4.c	10
Total - AEM 5850	72