**Econ 151a – Statistical Modeling with R for Economics and Finance**

Brandeis University

Summer 2022

Project 1

In project 1, you will create the Nasdaq-100 Index from scratch, using historical prices of the index constituents. The purpose of this project is to practice data management, data cleaning and merging, exploratory analysis of data, and data visualization.

For simplicity, you will only need to create the Nasdaq-100 Index for 2021 using daily data. You can ignore changes constituent list and use the list as of the end of 2021.

**Deadline: July 4, 2022**

**Expected project output: See syllabus**

This document provides detailed instructions on how to complete the project and suggested timeline.

1. **Suggested timeline**

At the end of Week 2: start the project. In Week 2, you will have learned about data resources available to you. This project will require that you use some of them. Now is a good time to find the relevant databases, download the data files, and get familiar with variables that you will work with.

Week 3: Coding.

Week 4: Finalize R Markdown and record the presentation.

1. **Instructions**
2. Read about NASDAQ-100 Index

<https://en.wikipedia.org/wiki/Nasdaq-100>

1. Obtain NASDAQ-100 constituent list

Access WRDS through Brandeis. Select “Compustat – Capital IQ” data.

On the “Compustat – Capital IQ” webpage, select “North America” -> “Index Constituents”

Input data range “2021-01” to “2021-12”

Choose “Search the entire database”

Select all query variables.

Save the output in your preferred format. The output is a very small file; it includes the constituent IDs of Nasdaq-100 index and some other indices covered by this database.

Filter the output file: keep if “Index Name” = “Nasdaq 100”, and “Effective Thru Date” is empty or in 2022. There should be 101 stocks left. How many of these stocks are constituents only part of 2021?

Company CIK is the unique identifier used by the SEC’s EDGAR database. You will need these IDs for project 2.

1. ID cross-walk

The constituent list that you obtained from the previous step contains several identifiers, including the Global Company Key (“gvkey”, unique identifier in Compustat). However, there isn’t an identifier applicable to the CRSP database, which provide stock price and return data. Hence, you need to create a mapping between gvkey and CRSP identifiers.

To do so, access WRDS through Brandeis. Select “CRSP” data.

On the “CRSP” webpage, select “CRSP/Compustat Merged” -> “Compustat CRSP Link”

Select “GVKEY” as the company code type. Upload the GVKEY of the constituents.

Choose query variables: all variables under “Link Information” and “Identifying Information”

Now you have the PERMNO, the unique ID used by CRSP.

1. Obtain the daily stock price data for the constituents

Access WRDS through Brandeis. Select “CRSP” data.

On the “CRSP” webpage, select “Stock/Security Files” -> “Daily Stock File”

Set data range to be 2021.

Upload code of the constituents. Make sure you select the appropriate format. You have obtained several applicable company code formats from the previous steps, e.g. TICKER, PERMNO, CUSIP.

Choose the following query variables: cusip, ncusip, comnam, ticker, permco, prc, vol, openprc, askhi, bidlo, ret, retx, shrout

Save the data in your preferred format.

1. Calculate the weighted average of daily return of all constituent stock. This is your replication of the daily return of the Nasdaq-100 index.

Plot your output. How does it compare to the Nasdaq-100 index daily return? You can find this through a Google search.