STOCK INTRINSIC VALUE ESTIMATION: PROJECT SUMMARY (VER.2024-04-16)

This document provides a non-technical summary of the project. The project code is developed in Python and kept in the GitHub repository. Intrinsic value model is defined in docs/intrinsic_value_model.pdf. Some notes about how to run the scripts are in the repository readme file.

OUTLINE OF THE PROJECT

Purpose of this project is to provide a basic tool for stock intrinsic value estimation. The tool is based on the discounted cash flow (DCF) model and uses historical financial data to estimate the intrinsic value of a stock. A trading strategy that is based on the intrinsic value is being developed.

VISION

In long term, the project aims to fulfill the following four goals:

1. To get a deeper understanding of economic processes through the analysis of financial data. Success of the trading strategy then provides a feedback on the quality of such analysis and understanding.
2. To create an investment tool that would help to make better investment decisions. Ideally, the tool would be able to beat some passive investing benchmark like investing in stock index.
3. To get a deeper understanding of reinforcement learning by employing these methods in trading decision-making. See also my other project “Digger” for some of my research in this area.
4. To be an addition to my portfolio of projects to demonstrate my skills to potential employees.

HOW FAR WE'VE COME: A SUMMARY OF DEVELOPMENT PROGRESS

I started with gathering the data. I knew I wanted to use SEC filings data as the primary data source for the project. The reason for this decision was that the data comes with some history, there is a plenty of tutorials on how to work with these data and it is for free. Also, these data are standardized across companies and they come with metadata such as links to definitions of variables in the financial statements. If a variable is not available from SEC filings, other data sources are used - see part Data for more details.

The next step was to develop the intrinsic value discounted cash flow model. I used mainly the various resources provided by Aswath Damodaran (videos, lectures, books) as well as some other resources. My aim was to come up fast with a rough prototype model working for as much companies from the SP500 index as possible. Although the concept of the intrinsic value is not very complicated, the challenging part was (and still is) to feed the proper variables from financial statements into the model. The SEC data are standardized to some extent as most of the statements follow the GAAP accounting standards. It is however not always clear which variable from the GAAP standard to use, some companies miss some of the variables in their statements and sometimes variables in the statements are company-specific (i.e. do not follow the GAAP standard). There are dozens of these company-specific issues in the data. I decided to tackle some of them to get some minimum amount of data that would allow me to start building the model. Stock-year combinations that failed to be fed into the model from any reason are discarded at the moment and will be treated gradually.

The next step is development of an algorithm, that would use the intrinsic value estimates to make trading decisions. This part is not developed yet. I only have some very preliminary results, see part “trading algorithm”.

DATA

Data for this project is obtained from various sources, the main one being the SEC filings data. Form 10-K is used as the primary source of the data since 2009. The form is collected once a year and hence the data is currently on the yearly basis for each company. Concrete date of data publishing is kept to compare the share price movements with respect to this date.

At first, I developed some API calls to get the data from the SEC web. Then I realized I can use some of the preprocessed databases available for offline download. Although there is some time lag between publish date of companies' statements and the database update, for the backtesting purposes that would serve me well. I can still use the API call procedures to get the more up-to-date data, although it remains to be analyzed whether the publish date on the SEC website corresponds to the publish date elsewhere.

I'm working with a dataset that encompasses around 7,300 potential stock-year combinations. However, there's still a significant amount of work required for data cleansing and aligning it with the model. Currently, many stock-year combinations are being

discarded. Despite this, I've managed to gather approximately 2,900 observations containing both intrinsic and market values. Additionally, I'm mindful that even these estimated values are built upon various assumptions which should be verified during further development.

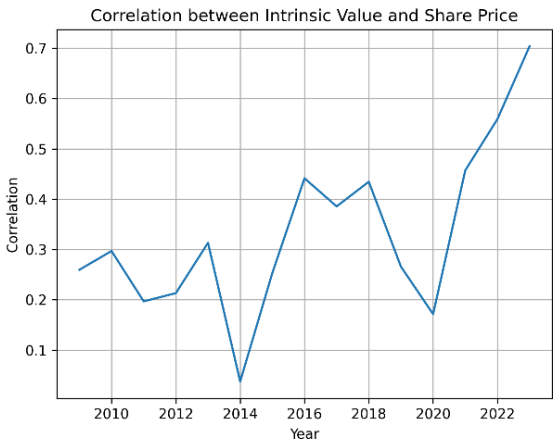
DATA SOURCES

1. **SEC filings Data:** The data is obtained from the SEC website (https://www.sec.gov/dera/data/financial-statement-data-sets) in form of quarterly collections of various financial reports in .txt files containing multiple tables. These data are then post-processed to extract only the information relevant for the model (like the 10-K reports) and only the required tickers. The postprocessed data are materialized in the form of company-level .csv files.
2. **Yahoo Finance Data:** Data on companies' betas (CAPM model) and share prices are obtained from Yahoo Finance. Currently the beta is taken as constant throughout the years and is not updated.
3. **Data from Aswath Damodaran's website:** Data on estimated equity risk premium is obtained from Aswath Damodaran's website (https://pages.stern.nyu.edu/~adamodar/New_Home_Page/datafile/implpr.html).
4. **FED FRED Data:** Data on government bond yields is obtained from the FED FRED database. These data are used as the risk-free rate in the model.

INTRINSIC VALUE MODEL

The intrinsic value model, utilizing the discounted cash flow approach, assesses the true worth of an asset by forecasting its future cash flows and discounting them back to their present value. This model is used to assess value of each stock in each period in which data are available. Please refer to docs/intrinsic_value_model.pdf for full intrinsic value model definition.

The plot on the right shows correlation between the estimated intrinsic value and the share price over each year in the data. If we assume that the market price perfectly reflects value of each company, we would be looking for a model where the correlation is 1 all the time.

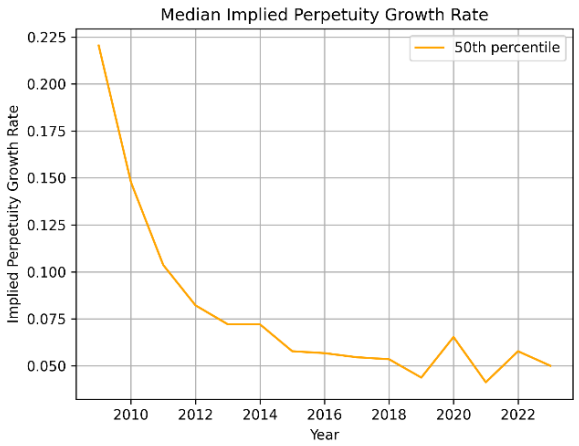


Deviation between intrinsic value and share price, on the other hand, can signal potential opportunities for trading. When the share price deviates significantly from its intrinsic value, investors may buy undervalued stocks or sell overvalued ones, aiming to capitalize on the eventual convergence of price towards its true worth.

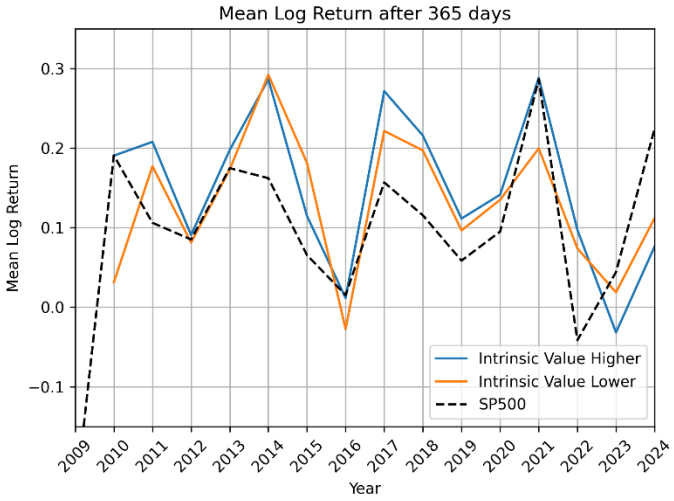
PERPETUITY GROWTH RATE ESTIMATION

The perpetuity growth rate turned out to be an important factor in the DCF model, having a significant impact on the intrinsic value estimate but being tricky to estimate at the same time. I approached this issue by using implied perpetuity growth rate estimated from the previous period's data. The estimate is obtained by finding perpetuity growth rate value that minimizes the difference between the intrinsic value estimate and the market price of the stock.

For majority of cases, the estimated growth rates have meaningful values, with median around 5 %. However, in about one fifth of cases the estimates are higher than 100 %, sometimes reaching as much as 1E6. This issued requires some further attention.



TRADING ALGORITHM



So far, I only have some very preliminary results in terms of comparisons of average annual returns for both groups of stock-year combinations with higher and with lower intrinsic value compared to the market value. I also compare this with return of SP500. On average the return (mean log yearly return) is higher for the model, though there are some years where SP500 performs better.

However, I approach these results with a high caution. Tickers that enter to the intrinsic value calculation should ideally be those in the SP500 index. In that case the SP500 line should lie between the two lines representing mean log return for both groups of stock-year combinations (higher and lower intrinsic value). Due to issues in the data this does not hold in most years. The reason is that the intrinsic value was not estimated for some of the stock-year combinations, usually due to some missing input to the model. This underlines the necessity to clean the data thoroughly.

My vision here is to develop an algorithm that would dynamically adjust the portfolio allocation given total available resources, stock values, dividends, risk appetite, expected fund withdrawal etc. An idea is to employ reinforcement learning techniques to do (a part of) this trading-related decision-making. This idea however needs further research, especially given the limited amount of data available for training. This contrasts with the reliance of reinforcement learning on large number of training iterations. Small amount of input data, which are moreover static, pose a risk of overfitting the model.

ISSUES, LIMITATIONS AND TODO'S

1. **Extraction of relevant data from financial statements:** The data extraction from the SEC filings is not straightforward. The data is not always in the same format, some columns are missing, some are named differently, etc. This makes the data extraction process quite complex. Ideally, we would need to go through the data for each company manually to make sure we extract the correct information.
2. **More frequent data:** Financial statements data are currently acquired from the 10-K forms. It would be beneficial to use quarterly 10-Q to get higher frequency of data.
3. **Other data sources:** Incorporate any other relevant data sources
4. **Model overfitting:** Need for a careful analysis of possible overfitting of any trading strategy. This is given by short-lived market imperfections, short history of data etc.
5. **Short history of data:** data since 2009 contains no US recessions except the one caused by Covid pandemic.
6. **Trading strategy**