Yifan Zhang, Hailong Bao

Models on Financial Markets and Investment Strategies

Contents

I A Logistic Regression Model on the Trend of a Price Curve 1

1 Trend Identification 3
1.1 Reflection Points and Trends 3
1.2 Empirical Facts of Trends 5
    1.2.1 Stock market indices 5
        1.2.1.1 S&P 500 5
        1.2.1.2 Dow Jones Industrial 30 5

2 A Logistic Regression Model for Trend Identification 9
2.1 Generic Approach 9
2.2 Risk Driver: Forms 9
2.3 Risk Driver: List of Candidates for Stock Markets 10

Bibliography 13

Part I

A Logistic Regression Model on the Trend of a Price Curve

1

Trend Identification

CONTENTS

1.1 Reflection Points and Trends 3
1.2 Empirical Facts of Trends 4
  1.2.1 Stock market indices 5
    1.2.1.1 S&P 500 5
    1.2.1.2 Dow Jones Industrial 30 5

Statistically it has been a classic problem to distinguish *trends* from *variances* when one looks at the historical prices of any financial instrument. The price trajectory is hence usually regarded as non-stationary and treated in a stochastic way. For instance, in the classical Black-Scholes world a stock price is usually modelled as a Geometric Brownian Motion, with log-normal variations around an exponential trend. However, intuitively it seems also quite obvious that certain trends do exist, which are naturally important for investment decisions.

In this chapter we discuss a straightforward approach to automatically identify the trends in a price trajectory based on the identification of local minimums / maximums. We then investigate the empirical facts of such trends for stock indices as well as famous individual stocks. In subsequent chapters, we will build up a model based on logistic regression and other machine learning techniques to identify the potential trend using financial and economic variables as risk drivers.

It should be noted that trend identification has been investigated in the technical analysis. Conventional methods include making use of moving averages, momentum indicators, trendlines and chart patterns via visual inspection (e.g. "higher highs and higher lows"). The difference however is that those techincal analysis methods usually aim to help identify short-term market anomalies e.g. overbought or oversold signals, while in this chapter the purpose is to structurally quantify the upward / downward trends and furthermore link them with potential risk drivers.

1.1 Reflection Points and Trends

Motivated by the idea that any unilateral trend would start with a low/high point and end at a high/low point, we start with defining local minimums/maximums.

Consider a price trajectory $\{S_t\}$ is observed. We have the following definition:

Definition 1 ***(Reflection point.)*** *A price S_t is called a* ***minimum/maximum reflection point of period*** *τ (MIRP / MARP in short) if it is the minimum/maximum of the price in the period $[t-\tau, t+\tau]$.*

By definition such reflection points are local extremes within a certain time window. It's also clear that they are subject to the choice of time window: a reflection point of a longer time window is also one for a shorter time window, but not vice versa.

It's tempting to already identify a *trend* by dividing the entire trajectory into pieces according to adjacent MARPs or MIRPs. However, such definition of reflection points focuses only on the local behaviour of the price trajectory and it's far from sufficient to always be able to use them for meaningful trend identification. In particular, there is no guarantee that

- the local minimums and maximums are always distributed in a "crossed" way, i.e. it can happen that two local minimums lie adjacent with no local maximum in between (*duplicated RPs*). An example is a W-shaped price curve, with a relatively low high between two local minimums.

- a local minimum / maximum is always lower / higher than its adjacent local maximum / minimum (*irregular RPs*). Exceptional cases are rare, but can happen when a relatively short horizon is chosen.

Ideally, we would like to find *meaningful* reflection points that helps to identify trends in a reasonable way. This leads to the following definition.

Definition 2 ***(Validated reflection point.)*** *A reflection point is called a* ***validated maximum / minimum reflection point*** *(vMARP / vMIRP in short) if it has a lower minimum / higher maximum reflection point as its preceding and subsequent reflection points.*

We can then define a trend based on the validated reflection points. Note that the definition of the validated reflection point guarantees that the entire price trajectory can be divided into either upward or downward trends – except the piece before the first and after the last reflection point.

Definition 3 ***(Trend.)*** *A piece of price trajectory is called a(n)* ***upward / downward trend*** *if it starts with a vMIRP / vMARP (inclusive) and ends with a vMARP / vMIRP (exclusive).*

To get the validated RPs, the identified RPs needs to be filtered. For the first issue mentioned above (i.e. duplicate RPs), there are two ways to deal with this:

- keep the lowest MIRP / highest MARP and drop all the others in case of adjacent same local extremes;
- identify the piece of price trajectory between two MARPs/MIRPs as a separate *flat trend* next to the upward and downward ones.

The drawback of the second treatment is that it will create a 3-state value space for the trends instead of a 2-state space consisting of only upwards and downwards. To optimally use the available historical data we choose to go with the first treatment, i.e. only keep the lowest MIRP / highest MARP.

It is also a possibility to keep all the identified RPs and identify a new maximum / minimum between two adjacent MIRPs / MARPs. The drawback is however that it will then lead to "false" RPs (i.e. which do not qualify to be a reflection point) and thus leads to "false" trends. Hence we choose to drop the redundant local RPs and keep only the meaningful ones.

The second issue (i.e. irregular RPs) usually happens when there are short-period turbulences in a long-period trend. For instance, a MARP is identified at an early stage of a long upward trend, and a MIRP is subsequently identified at a later stage, simply due to temporarily increasing market volatility. The strong upward trend however prevents to identify a *lower* MIRP after that MARP, and the same holds for the MIRP at the later stage. The result is that a wrong downward trend can be identified – which begins with a local maximum and ends with a local minimum – while it's actually an upward trend.

To deal with this issue we choose to drop both the reflection points in case a MARP is lower than an subsequent MIRP or a MIRP is higher than an subsequent MARP. The rationale is that, though those reflection points are local extremes, in most cases they cannot be used to identify meaningful trends in the price trajectory.

We then have an algorithm to filter the RPs to get the validated ones:

1. Divide the RPs into groups based on whether they are adjacent local MARPs / MIRPs. If there are multiple MARPs / MIRPs in any group, keep the highest / lowest one and drop all the others.
2. For the remaining RPs, drop the "irregular" ones, i.e. if a MARP is lower than its subsequent MIRP, or a MIRP is higher than its subsequent MARP.

It is not the purpose to provide a rigorous mathematical proof to show that such way to find the validated RPs is in any sense optimal. Nonetheless, it can be shown that such a way is sufficient to filter out validated RPs, based on which we can identify meaningful trends.

The trends defined in this section should however be interpreted carefully. Any trend would start with a vMIRP and end with a vMARP (or vice versa), which are all the local extremes if one looks at a period of τ before and after that time. This means that this identified trend cannot be further extended forwards or backwards in time at least within a period of τ. The duration of the trend, on the other hand, can be of any duration, even shorter than τ.

A larger τ would indicate a stronger requirement on the RP identification, and hence fewer identified trends. Potential RPs, which may be recognized in case of a smaller τ, would tend to be treated as short-term fluctuations within a longer trend.

1.2 Empirical Facts of Trends

In this section we look at what the trend identification would look like for different asset classes under different time horizons. Furthermore other empirical facts are also investigated.

The following time windows are used:

- *Short*: 21 days / 1 month;
- *Medium*: 63 days / 3 months / 1 quarter;
- *Medium-Long*: 126 days / 6 months;
- *Long*: 252 days / 1 year.

Please note that the time windows are by definition *one-sided* in this framework. For example, a 252-day window indicates that we are looking for local highs and lows in a 2-year horizon; any reflection point is a local extreme value in its preceding and subsequent 1 year.

1.2.1 Stock market indices

We first look at the stock market in the U.S.. The 3 most famous indices are examined: S&P500, Dow Jones and Nasdaq.

1.2.1.1 S&P 500

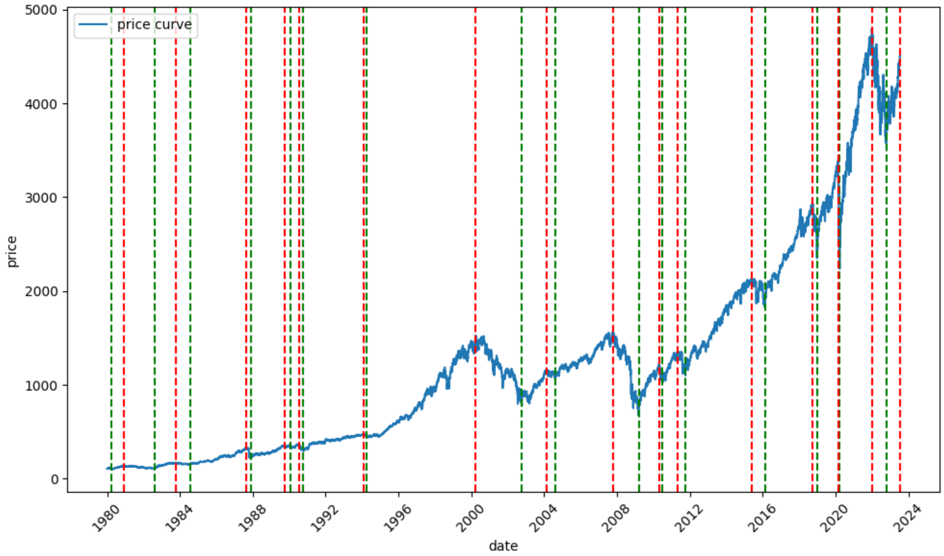

FIGURE 1.1
S&P 500 Index: curve and trend identification based on 126-day horizon.

1.2.1.2 Dow Jones Industrial 30

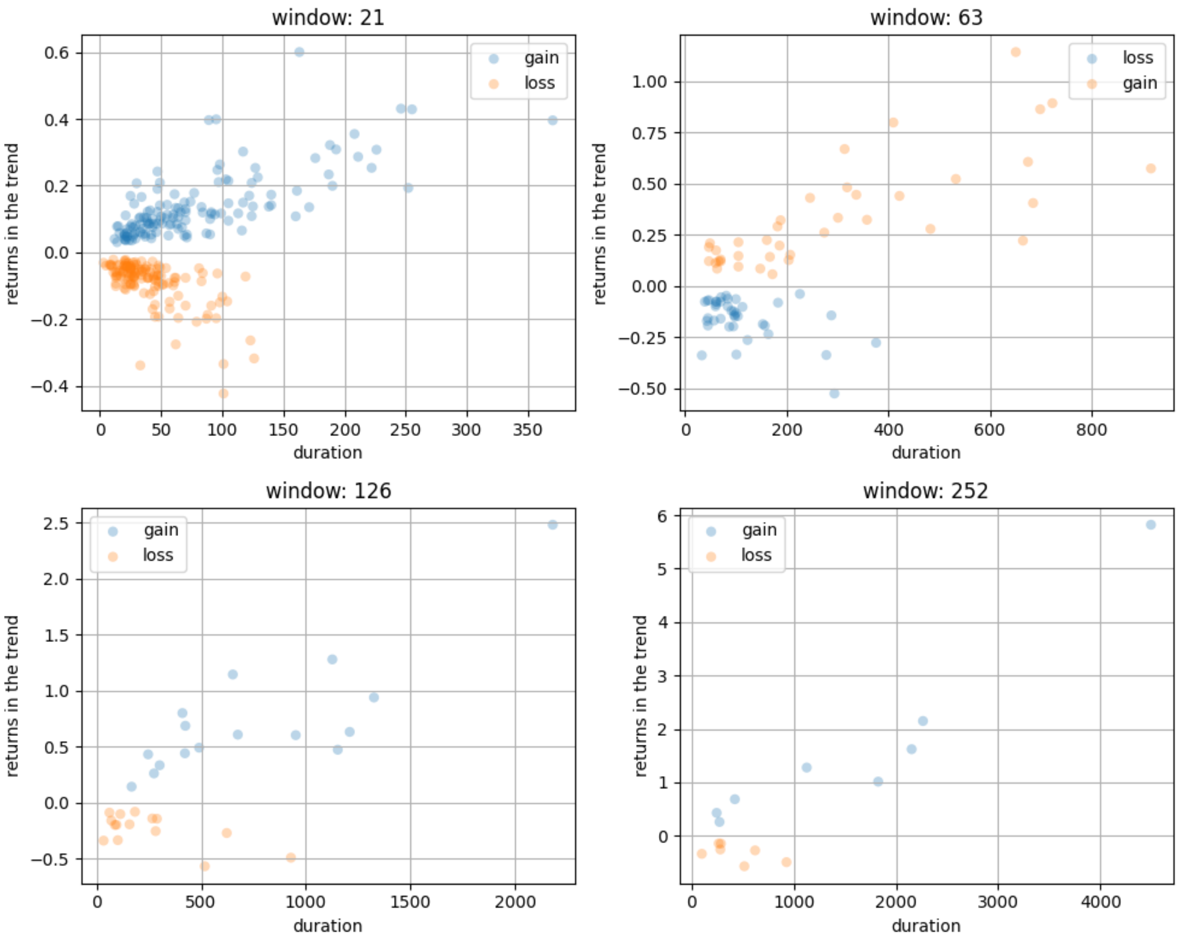

FIGURE 1.2
Scatter plots of the returns and durations of the trend for S&P 500 index, with different horizons.

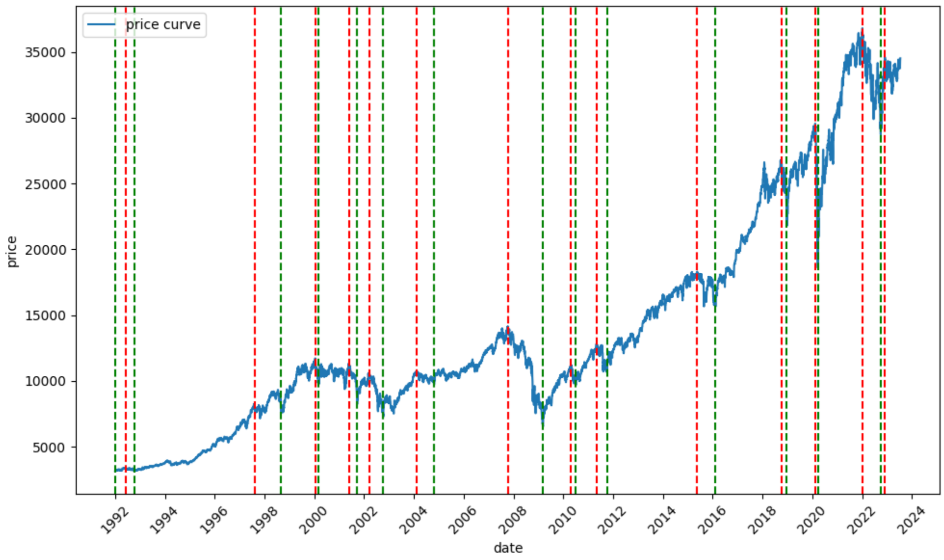

FIGURE 1.3
DJI Index: curve and trend identification based on 126-day horizon.

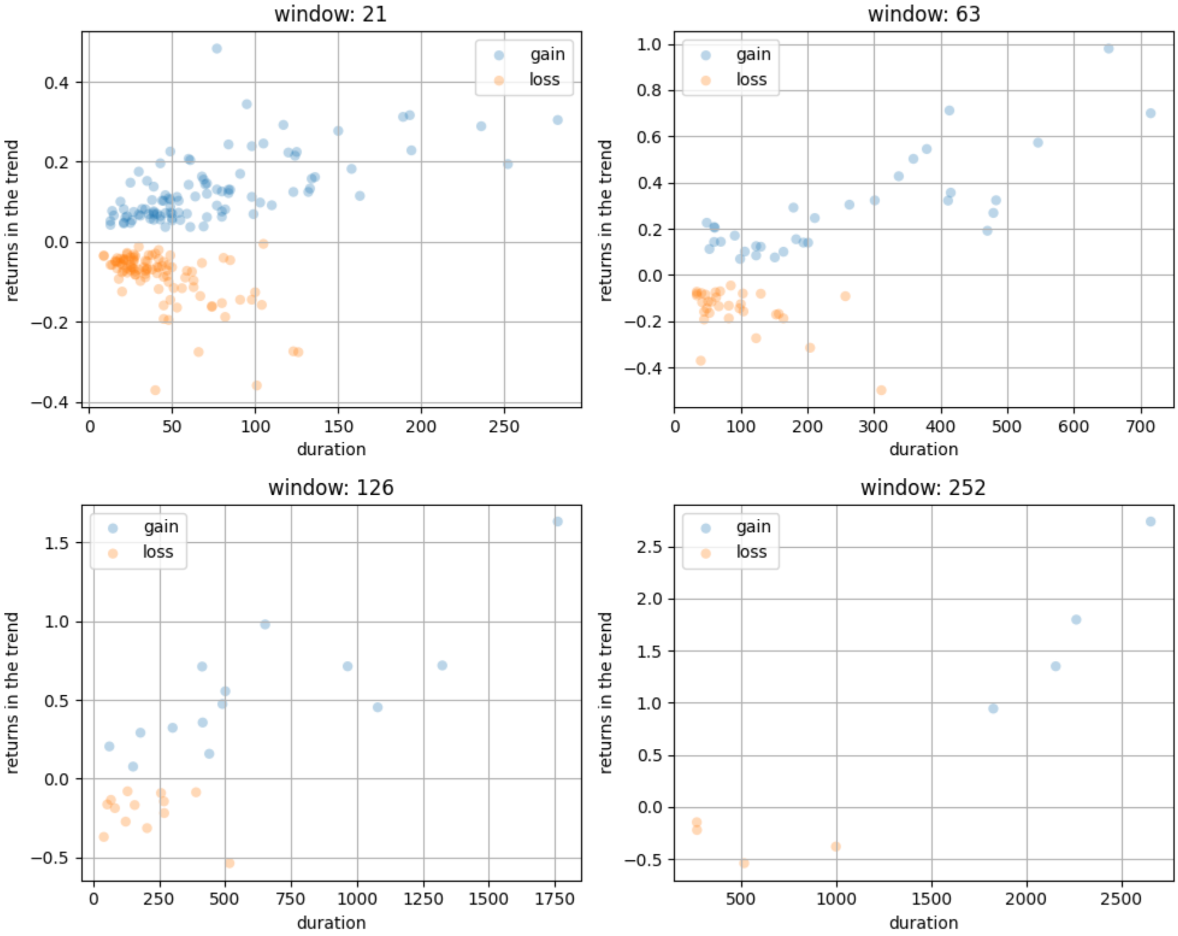

FIGURE 1.4
Scatter plots of the returns and durations of the trend for DJI index, with different horizons.

2

A Logistic Regression Model for Trend Identification

CONTENTS

2.1 Generic Approach 9
2.2 Risk Driver: Forms 9
2.3 Risk Driver: List of Candidates for Stock Markets 10

One of the most important purposes for the trend identification is to forecast the type of trend given a set of selected risk drivers. That is, by making use of the information from historical data, we need a model that tells the most probable trend given the current market conditions. Given the two possible values for a trend (either upwards or downwards), the *logistic regression model* can be a suitable candidate.

2.1 Generic Approach

The binomial logistic regression (i.e. dependent variable takes only 2 states) usually takes the following form:

$$p(x) = \frac{1}{1 + e^{-\beta \cdot \mathbf{X}}},$$

where

- $p(x)$ is the probability that the dependent variable Y takes the value 1,
- β is a vector of coefficients to be estimated, and
- $\mathbf{X}$ is the vector of risk drivers.

We will start with the logistic regression in its simplest form. Building the model would consist of the following steps:

1. *Data collection and processing*: relevant data should be schematically collected, including a.o. the risk driver data. The data are then processed to arrive at the final formats, e.g. data frequency, MA horizon, etc.
2. *Risk driver selection*: the logit model is fit and the optimal set of risk drivers are chosen based on some criteria on e.g. goodness of fit or a performance measure.
3. *Final model fit and prediction*: The model is then calibrated based on the final risk driver selection and can be used for prediction.

2.2 Risk Driver: Forms

Intuitively, economic and financial variables are the most suitable candidates since the prices of financial assets are usually strongly influenced by them. Other variables, like a binary geopolitical indicator, should also be considered. A key question is however how these variables, which are usually numerical, should be included in the model. For instance, should we include the CPI by itself, or the relative changes w.r.t.

last month / year? Should it be a warning signal if the market interest rates are quite high and the 10Y-3M spread has been deep negative? How do you further judge whether a rate level is "high" or not? Does a burnout effect exist in the market regarding a crisis or a market anomaly? These questions should be considered in order to transform the risk drivers in the right form.

In general, the following can be considered: [1]

1. **Unchanged format**: the variable is used as it is, i.e. by the value itself and without any further processing.
2. **Relative change**: the relative change of the variable w.r.t. the last month / year is used.
3. **Quantiles**: the relative position of the variable in the past certain period (e.g. 10 years) is used.
4. **Scenario-dependent**: this can be in many forms. One possibility is to count the duration from the start of a certain event, e.g. the outburst of a crisis or the start of a(n) upward / downward trend.
5. **Comparison to expectation**: It is used whether the actual variable is higher or lower than the market expectation for it. However, it should be noted that such market expectations data are usually difficult to access and may not be

2.3 Risk Driver: List of Candidates for Stock Markets

In this section we include plausible candidates as the risk driver, which can be further classified into 2 groups: economic and financial. The purpose is to outline a generic list, independent of certain countries / regions, and hence it should be applicable to most developed markets. The focus is primarily on the stock markets. We first focus on the general market indices, after which we then delve into the cases of specific companies on an individual level.

1. **Economic variables**:
    - *GDP growth*: Gross Domestic Product (GDP) growth reflects the overall health and expansion of the economy. Strong economic growth typically translates to higher corporate earnings, which can drive stock prices higher.
    - *Inflation*: Inflation measures the general increase in prices over time. Moderate inflation is generally considered healthy for the economy and the stock market, but high or rapidly rising inflation can erode purchasing power and lead to uncertainties. Most importantly, it's usually partly the target of the central banks to keep the inflation on a certain level. A too high or too low inflation may have important implications on the monitory policy.
    - *Unemployment rate*: The level of unemployment is an indicator of the overall labor market conditions. Lower unemployment rates suggest a healthier economy, as more people are employed and have disposable income to invest, which can positively influence the stock market.
    - *Money supply*: An increase in the money supply can lead to increased liquidity in the financial system. When there is more money available, investors may have a greater capacity to purchase stocks, which can drive up stock prices. Moreover, an abundance of liquidity can contribute to positive investor sentiment and a willingness to invest in riskier assets like stocks.
    - *Geopolitical factors*: International trade policies, tariffs, and political events can have a significant impact on global economic conditions and market sentiment. Trade disputes or geopolitical tensions can introduce uncertainty and volatility into the stock market.
    - *Consumer confidence index*: The CCI serves as a leading indicator for consumer spending, which has a significant impact on economic growth and corporate earnings. The CCI is a measure of consumers' perceptions and sentiments about the current and future state

---

[1] This section is motivated and facilitated by ChatGPT (july 2023).

of the economy and their personal financial situation. It is usually based on surveys that ask consumers about their outlook on economic conditions, job prospects, income expectations, and buying intentions.

- *Purchasing Managers' Index*: The Purchasing Managers' Index (PMI) is considered important for the stock market because it provides valuable insights into the health and performance of the manufacturing and services sectors of the economy. The PMI is an economic indicator that measures business activity, production levels, new orders, employment, and supplier deliveries in these sectors. It is based on surveys of purchasing managers in various industries.

2. **Financial variables**:
    - *Interest rate levels*: Changes in interest rates can affect borrowing costs, corporate profitability, and investment decisions. Lower interest rates tend to stimulate economic activity and can be positive for stocks, while higher rates can increase borrowing costs and potentially dampen stock market performance.
    - *Interest rate term structure*: the term structure of interest rates usually implies the market opinion on the economy expectations. It also typically indicates the costs of funds for different time horizons.
    - *Exchange rate*: FX rates affect the competitiveness of exports and imports for an open economy. Fluctuations in exchange rates can impact the trade balances and hence have an impact on the economy.
    - *Commodity prices*: Commodities constitute the raw materials for the industrial production as well as for the economy. Shocks or irregular changes in the commodity prices may hence introduce volatilities in the stock market.
    - *Real estate prices*: The relationship between real estate prices and the stock market can be complicated. Still it is included in the list, as it has been an important factor in the current economy, in particular, real estates serve as an important source of collateral for borrowings, and deterioration in the real estate markets has been the key trigger for the 2008 financial crisis.

It should be noted that all these variables can be inter-correlated or, at least, they may arguably contain the same information. Multicolinearity should be checked in case of econometric analysis.

The aforementioned variables can be further classified into 3 categories:

1. *Leading indicators*: such variables tend to change before the overall economy changes. They can provide early signals about the direction of the economy and are often used to predict future economic trends. Examples of leading indicators include:
    - New housing permits
    - Consumer confidence index
    - Purchasing Managers' Index (PMI)
    - Most financial market variables, including a.o. stock market indices and interest rate term structure.
2. *Lagging indicators*: such variables change after the overall economy has already started to follow a particular trend. They confirm or validate the direction of the economy and are useful for assessing the current state of the economy. Examples of lagging indicators include:
    - Unemployment rate
    - Consumer Price Index (CPI)
    - Gross Domestic Product (GDP)

    Hence, when using such variables, the lagging effect must be taken into account.

Specifically for the U.S. stock market, the following table lists the variables that are included as candidates, with relevant properties.

Variable	**Category**	**Code**	**Source**	**Frequency**	**Lag**	**Start date**
CPI	Econ / Inflation	CPIAUCSL	FRED	Monthly	1M	
GDP	Econ / GDP					
Unemployment rate	Econ / Unemployment rate	UNRATE	FRED	Monthly		
M2	Econ / Money supply	M2SL	FRED	Monthly		
Fed Policy rate	Fin / IR					

TABLE 2.1

List of candidate variables for the U.S. stock market

Bibliography