



What do we want to model?

Understanding Exogenous  
vs Endogenous Data

# Learning Objectives

---

- Distinguish between Exogenous and Endogenous factors
- Practice: What types of data to use for technical and fundamental trading models?
- Preview how you can model build with BigQuery ML

# Agenda

---

Exogenous and Endogenous  
model factors

Asking the right questions of  
your data

Case Study: Modelling CPU  
Performance by Vendor with  
BigQuery

# Trading and investing models have different goals

- Trading = predict relatively short-term changes in the price of an asset
- Investing = estimating the fair value of assets relative to their current price
- Both types of models require lots of data as input factors

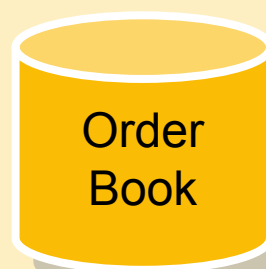
# Endogenous factors depend on trading data

## Technical Strategy Model

Use historical stock prices, order book information and trading volume for AAPL to predict future behavior



Price data



Order Book

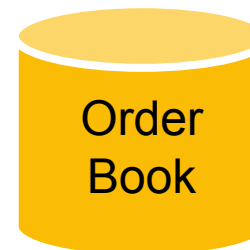


Trading Volume



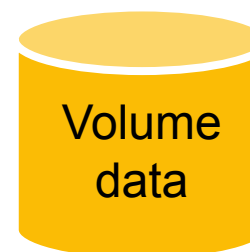
Price data

Price trends and volatility: AAPL traded higher 4 out of 5 days and daily price volatility decreased to 1%



Order Book

Size of orders at the bid and ask price, bid/ask spread: Average size increased to 3,000 shares. Bid-ask spread declined to \$0.14.



Volume data

Volume: AAPL traded to a record high on 145% of average daily volume.

Endogenous models are created using the historical price and volume data of AAPL stock. Endogenous models assume that all available fundamental data is incorporated into the current share price. Volume is used to distinguish price noise from more durable price changes.

# Exogenous factors depend on fundamental or macro data

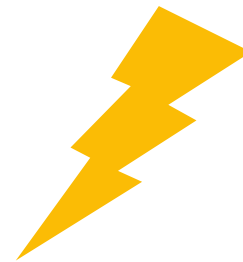
## Event-Driven Strategy Model

Affected by releases of fundamental data affecting AAPL and by unexpected changes in macro variables

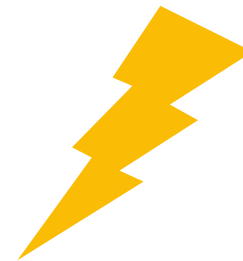
Quarterly earnings and guidance

Customer/Supplier Shocks

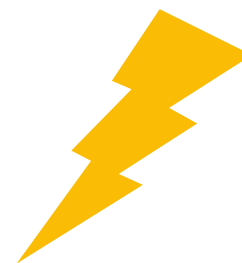
Macro Data Releases



Earnings Shock: AAPL quarterly earnings and forward guidance exceed analyst expectations



Supplier Shock: A major semiconductor supplier shuts down operations



Customer Shock: 30% tariffs are imposed on iPhones and watches

Exogenous variables affect share price but are not dependent on share price.

# Technical Model: Predicting changes in stock price for AAPL using stock price and volume data alone

Are these factors **Endogenous** or **Exogenous**?

An unanticipated change in consumer tastes for your product

The U.S. Federal Reserve changes the cost of borrowing

A competitor has major safety issues with their products

The price of critical components for making phones and computers suddenly changes

Quarterly earnings announcements

Shipping times from overseas grow due to new ocean weather patterns

Trading volume matched with price to track

Volume-Weighted-Average-Price

CEO is replaced unexpectedly

Trends in stock price volatility

# Technical Model: Predicting changes in stock price for AAPL using stock price and volume data alone

Are these factors **Endogenous** or **Exogenous**?

An unanticipated change in consumer tastes for your product

The U.S. Federal Reserve changes the cost of borrowing

A competitor has major safety issues with their products

The price of critical components for making phones and computers suddenly changes

Quarterly earnings announcements

Shipping times from overseas grow due to new ocean weather patterns

Trading volume matched with price to track Volume-Weighted Average Price

CEO is replaced unexpectedly

Trends in stock prices measured by momentum



# Event-Driven Model: Predicting changes in stock price for AAPL based on release of fundamental and macro data

Are these factors **Endogenous** or **Exogenous**?

An unanticipated change in consumer tastes for your product	The U.S. Federal Reserve changes the cost of borrowing	A competitor has major safety issues with their products
The price of critical components for making phones and computers suddenly changes	Quarterly earnings announcements	Shipping times from overseas grow due to new ocean weather patterns
Trading volume matched with price to track Volume-Weighted-Average-Price	CEO is replaced unexpectedly	Trends in stock prices measured by momentum

Key takeaway: **Exogenous** and **endogenous** factors both drive changes in share prices. Technical strategies tend to focus on **endogenous** factors and event-driven strategies focus on **exogenous factors**

# Agenda

---

Exogenous and Endogenous  
model factors

Asking the right questions of  
your data

Case Study: Modelling CPU  
Performance by Vendor with  
BigQuery

# Is there a “right” set of data for financial modelling?

## Ask yourself:

- What data do I have?
- What data do others have?
- Does the freshness of my data matter a little or a lot?
- What assumptions is my model making?
- Is there a combination of things I could model?

# Is there a “right” set of data for financial modelling?

## Ask yourself:

- What data do I have?
- What data do others have?
- Does the freshness of my data matter a little or a lot?
- What assumptions is my model making?
- Is there a combination of things I could model?

# Is there a “right” set of data for financial modelling?

## Ask yourself:

- What data do I have?
- What data do others have?
- Does the freshness of my data matter a little or a lot?
- What assumptions is my model making?
- Is there a combination of things I could model?

# Is there a “right” set of data for financial modelling?

## Ask yourself:

- What data do I have?
- What data do others have?
- Does the freshness of my data matter a little or a lot?
- What assumptions is my model making?
- Is there a combination of things I could model?

# Is there a “right” set of data for financial modelling?

## Ask yourself:

- What data do I have?
- What data do others have?
- Does the freshness of my data matter a little or a lot?
- What assumptions is my model making?
- Is there a combination of things I could model?



# Agenda

---

Exogenous and Endogenous  
model factors

Asking the right questions of  
your data

Case Study: Modelling CPU  
Performance by Vendor with  
BigQuery

# Case study: Predicting CPU performance across vendors

If you knew the performance of brand new CPU chips produced by leading semiconductor manufacturers, what could you model?

- Performance vs competitor
- Performance over time
- Growth/slowing in improvements



Potential inputs in a model for valuing a business

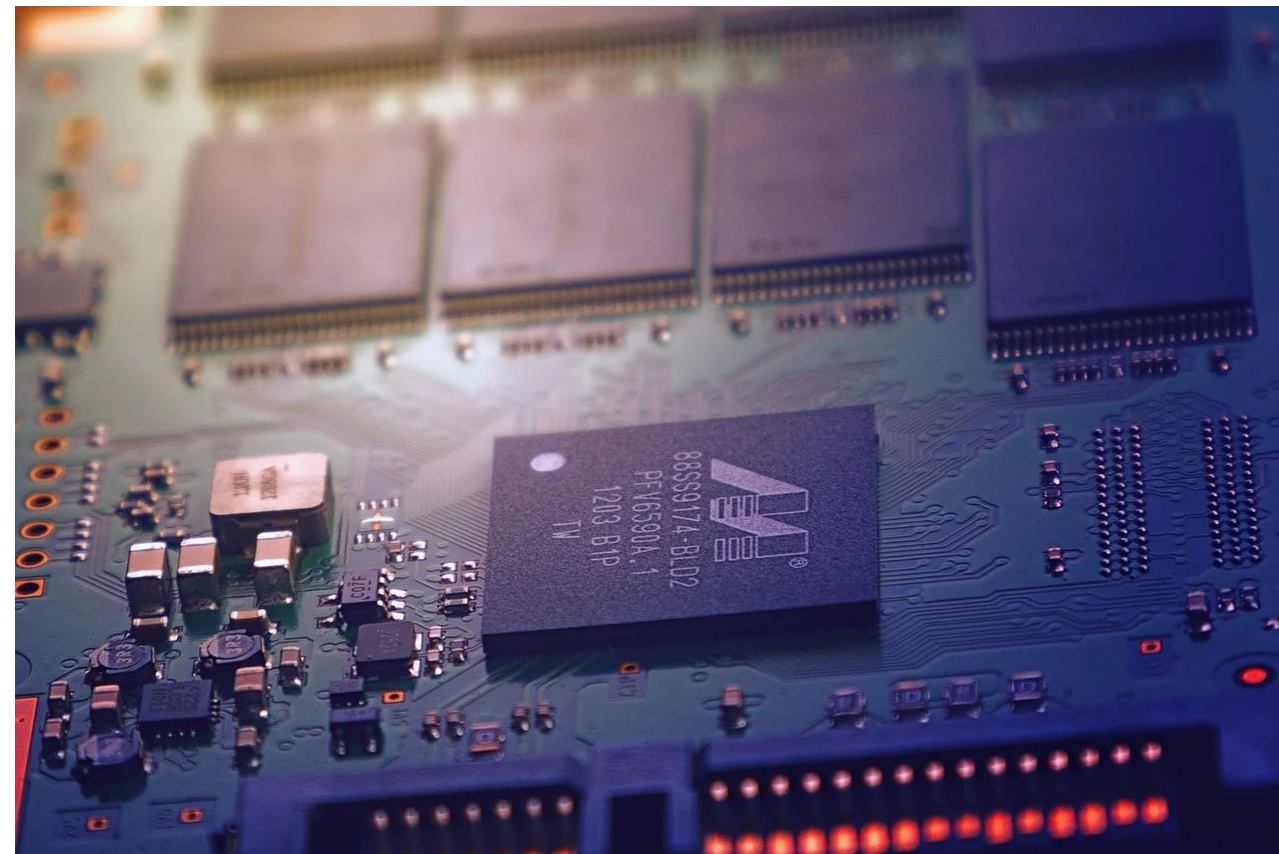
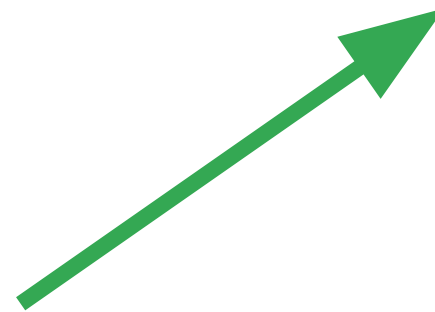
# Case study: Predicting CPU performance across vendors

You are given the following raw inputs:

- vendor,
- model\_name,
- max\_mhz,
- nominal\_mhz,
- cores,chips,
- channels,
- mem\_gb,
- mem\_speed,
- l1\_cache\_mem\_kb ,
- l2\_cache\_mem\_kb ,
- l3\_cache\_mem\_mb ,
- os,
- compiler,
- sponsor

Could you quickly predict an overall performance score?

**benchmark\_score**



# Topic Preview: Linear Regression with BigQuery Machine Learning

```
CREATE OR REPLACE MODEL
  `cpu_performance.spec_int_speed_631` Model name
OPTIONS
  (model_type='linear_reg',input_label_cols=['_631_deepsjeng_s']) AS Model type and what
column we're predicting
for
SELECT * FROM (
  SELECT
    vendor,
    model_name,
    max_mhz,
    nominal_mhz,
    cores,chips,
    channels,
    mem_gb,
    mem_speed,
    l1_cache_mem_kb ,
    l2_cache_mem_kb ,
    l3_cache_mem_mb ,
    os,
    compiler,
    sponsor,
    _631_deepsjeng_s # This is what we're predicting
  FROM
    `sara-bqml.cpu_performance.spec_int_speed`
)
```

Model inputs

- Quickly test if you can accurately model the behavior in the data in minutes
- If you're predicting a number, use a forecasting model (like linear regression)
- If you're classifying (bucket A, bucket B) use a classification model (like logistic regression)
- Spend the most time understanding the nature of your data before modeling

# The 80/20 rule of building machine learning models

