**3 DESIGN AND IMPLEMENTATION**

**3.1 Financial Data – Explanation of the financial data used**

The data the dissertation will be based on data supplied by financial data company Quandl. The data that will be used is called **Core US Fundamentals Data.** This data is Accounting Metrics published by the individual companies themselves, from stock exchanges where the companies are listed or metrics calculated from published data.

**Accounting Metrics**

There are six rows and 111 columns. There are six rows because of the different dimensions, dimensions refer to how the data is reported, the distinctions are outlined below.

As Reported view (AR):

* excludes restatements
* point-in-time view with data time-indexed to the date the form 10 regulatory filing was submitted to the SEC
* presents data for the latest reporting period at that filing date
* may include multiple observations in a quarter if more than one filing is made during the quarter
* on limited occasion may not have any observations in a particular quarter. Sometimes companies are delayed in reporting for up to 18 months. On such occasions they may report multiple documents on the same date to catch up, in which case these datasets will only provide date for the most recent reporting period.
* typically suitable for back-testing

Most-Recent Reported view (MR):

* includes restatements
* time indexed to the financial/report period
* presents the most recently reported data for that reporting period
* typically suitable for assessing business performance after restatements for mergers/divestitures

The Quandl documentation states that ARY (as-reported annual) dimension is typically suitable for back-testing, which is what the dissertation is based on.

These are annual observations of one year duration. Using the as-reported dimension will more closely align with the date that information was disseminated to the market, and the corresponding market impact.

In addition there are 3 time dimensions:

Annual (Y): Annual observations of one year duration

Trailing Twelve Months (T): Quarterly observations of one year duration

Quarterly (Q): Quarterly observations of quarterly duration (available only for US domestic companies, unavailable for foreign companies)

These Accounting Metrics are only available on a monthly basis.

Below is an example of the latest Accounting Metrics for the stock **Apple (APPL)**

|  | **ticker** | **dimension** | **calendardate** | **datekey** | **reportperiod** | **lastupdated** | **accoci** | **assets** | **assetsavg** | **assetsc** | **...** | **sharesbas** | **shareswa** | **shareswadil** | **sps** | **tangibles** | **taxassets** | **taxexp** | **taxliabilities** | **tbvps** | **workingcapital** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **None** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **0** | AAPL | ARQ | 2018-09-30 | 2018-11-05 | 2018-09-29 | 2018-11-05 | -3454000000 | 365725000000 | NaN | 131339000000 | ... | 4745398000 | 4801588000 | 4847547000 | 13.100 | 365725000000 | 0 | 2296000000 | 0 | 76.168 | 14473000000 |
| **1** | AAPL | ART | 2018-09-30 | 2018-11-05 | 2018-09-29 | 2018-11-05 | -3454000000 | 365725000000 | 3.723045e+11 | 131339000000 | ... | 4745398000 | 4955377000 | 5000109000 | 53.597 | 365725000000 | 0 | 13372000000 | 0 | 76.168 | 14473000000 |
| **2** | AAPL | ARY | 2018-12-31 | 2018-11-05 | 2018-09-29 | 2018-11-05 | -3454000000 | 365725000000 | 3.723045e+11 | 131339000000 | ... | 4745398000 | 4955377000 | 5000109000 | 53.597 | 365725000000 | 0 | 13372000000 | 0 | 76.168 | 14473000000 |
| **3** | AAPL | MRQ | 2018-09-30 | 2018-09-29 | 2018-09-29 | 2018-11-05 | -3454000000 | 365725000000 | NaN | 131339000000 | ... | 4829926000 | 4801588000 | 4847547000 | 13.100 | 365725000000 | 0 | 2296000000 | 0 | 76.168 | 14473000000 |
| **4** | AAPL | MRT | 2018-09-30 | 2018-09-29 | 2018-09-29 | 2018-11-05 | -3454000000 | 365725000000 | 3.723045e+11 | 131339000000 | ... | 4829926000 | 4955377000 | 5000109000 | 53.597 | 365725000000 | 0 | 13372000000 | 0 | 76.168 | 14473000000 |

**Accounting Metrics Explanation**

Below is a table with descriptions for the Accounting Metric data.

**Table** - Quandl source table

**Indicator** - Accounting Metric code

**Isfilter** - Whether the search criteria can be applied to the Accounting Metric

**Title** - Full name of the Accounting Metric

**Description** - Full description of the Accounting Metric

**Unit Type** - Unit of measurement for the Accounting Metric

A sample of the full descriptions dataset

|  | **table** | **indicator** | **isfilter** | **isprimarykey** | **title** | **description** | **unittype** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **None** |  |  |  |  |  |  |  |
| **0** | SF1 | revenue | N | N | Revenues | [Income Statement] Amount of Revenue recognized from goods sold; services rendered; insurance premiums; or other activities that constitute an earning process. Interest income for financial institutions is reported net of interest expense and provision for credit losses. | currency |
| **1** | SF1 | cor | N | N | Cost of Revenue | [Income Statement] The aggregate cost of goods produced and sold and services rendered during the reporting period. | currency |
| **...** | ... | ... | ... | ... | ... | ... | ... |
| **109** | SF1 | reportperiod | N | Y | Report Period | [Entity] The Report Period represents the end date of the fiscal period. | date (YYYY-MM-DD) |
| **110** | SF1 | lastupdated | Y | N | Last Updated Date | [Entity] Last Updated represents the last date that this database entry was updated; which is useful to users when updating their local records. | date (YYYY-MM-DD) |

**Metric List**

Metric List A full list of all the 111 metrics is below

|  | **title** | **description** |
| --- | --- | --- |
|  |  |  |
| **0** | Revenues | [Income Statement] Amount of Revenue recognized from goods sold; services rendered; insurance premiums; or other activities that constitute an earning process. Interest income for financial institutions is reported net of interest expense and provision for credit losses. |
| **1** | Cost of Revenue | [Income Statement] The aggregate cost of goods produced and sold and services rendered during the reporting period. |
| **2** | Selling General and Administrative Expense | [Income Statement] A component of [OpEx] representing the aggregate total costs related to selling a firm's product and services; as well as all other general and administrative expenses. Direct selling expenses (for example; credit; warranty; and advertising) are expenses that can be directly linked to the sale of specific products. Indirect selling expenses are expenses that cannot be directly linked to the sale of specific products; for example telephone expenses; Internet; and postal charges. General and administrative expenses include salaries of non-sales personnel; rent; utilities; communication; etc. |
| **3** | Research and Development Expense | [Income Statement] A component of [OpEx] representing the aggregate costs incurred in a planned search or critical investigation aimed at discovery of new knowledge with the hope that such knowledge will be useful in developing a new product or service. |
| **4** | Operating Expenses | [Income Statement] Operating expenses represents the total expenditure on [SGnA]; [RnD] and other operating expense items; it excludes [CoR]. |
| **5** | Interest Expense | [Income Statement] Amount of the cost of borrowed funds accounted for as interest expense. |
| **6** | Income Tax Expense | [Income Statement] Amount of current income tax expense (benefit) and deferred income tax expense (benefit) pertaining to continuing operations. |
| **7** | Net Loss Income from Discontinued Operations | [Income Statement] Amount of loss (income) from a disposal group; net of income tax; reported as a separate component of income. |
| **8** | Consolidated Income | [Income Statement] The portion of profit or loss for the period; net of income taxes; which is attributable to the consolidated entity; before the deduction of [NetIncNCI]. |
| **9** | Net Income to Non-Controlling Interests | [Income Statement] The portion of income which is attributable to non-controlling interest shareholders; subtracted from [ConsolInc] in order to obtain [NetInc]. |
| **10** | Net Income | [Income Statement] The portion of profit or loss for the period; net of income taxes; which is attributable to the parent after the deduction of [NetIncNCI] from [ConsolInc]; and before the deduction of [PrefDivIS]. |
| **11** | Preferred Dividends Income Statement Impact | [Income Statement] Income statement item reflecting dividend payments to preferred stockholders. Subtracted from Net Income to Parent [NetInc] to obtain Net Income to Common Stockholders [NetIncCmn]. |
| **12** | Net Income Common Stock | [Income Statement] The amount of net income (loss) for the period due to common shareholders. Typically differs from [NetInc] to the parent entity due to the deduction of [PrefDivIS]. |
| **13** | Earnings per Basic Share | [Income Statement] Earnings per share as calculated and reported by the company. Approximates to the amount of [NetIncCmn] for the period per each [SharesWA]. |
| **14** | Earnings per Diluted Share | [Income Statement] Earnings per diluted share as calculated and reported by the company. Approximates to the amount of [NetIncCmn] for the period per each [SharesWADil]. |
| **15** | Weighted Average Shares | [Income Statement] The weighted average number of shares or units issued and outstanding that are used by the company to calculate [EPS]; determined based on the timing of issuance of shares or units in the period. |
| **16** | Weighted Average Shares Diluted | [Income Statement] The weighted average number of shares or units issued and outstanding that are used by the company to calculate [EPSDil]; determined based on the timing of issuance of shares or units in the period. |
| **17** | Capital Expenditure | [Cash Flow Statement] A component of [NCFI] representing the net cash inflow (outflow) associated with the acquisition & disposal of long-lived; physical & intangible assets that are used in the normal conduct of business to produce goods and services and are not intended for resale. Includes cash inflows/outflows to pay for construction of self-constructed assets & software. |
| **18** | Net Cash Flow - Business Acquisitions and Disposals | [Cash Flow Statement] A component of [NCFI] representing the net cash inflow (outflow) associated with the acquisition & disposal of businesses; joint-ventures; affiliates; and other named investments. |
| **19** | Net Cash Flow - Investment Acquisitions and Disposals | [Cash Flow Statement] A component of [NCFI] representing the net cash inflow (outflow) associated with the acquisition & disposal of investments; including marketable securities and loan originations. |
| **20** | Net Cash Flow from Financing | [Cash Flow Statement] A component of [NCF] representing the amount of cash inflow (outflow) from financing activities; from continuing and discontinued operations. Principal components of financing cash flow are: issuance (purchase) of equity shares; issuance (repayment) of debt securities; and payment of dividends & other cash distributions. |
| **21** | Issuance (Repayment) of Debt Securities | [Cash Flow Statement] A component of [NCFF] representing the net cash inflow (outflow) from issuance (repayment) of debt securities. |
| **22** | Issuance (Purchase) of Equity Shares | [Cash Flow Statement] A component of [NCFF] representing the net cash inflow (outflow) from common equity changes. Includes additional capital contributions from share issuances and exercise of stock options; and outflow from share repurchases. |
| **23** | Payment of Dividends & Other Cash Distributions | [Cash Flow Statement] A component of [NCFF] representing dividends and dividend equivalents paid on common stock and restricted stock units. |
| **24** | Net Cash Flow from Investing | [Cash Flow Statement] A component of [NCF] representing the amount of cash inflow (outflow) from investing activities; from continuing and discontinued operations. Principal components of investing cash flow are: capital (expenditure) disposal of equipment [CapEx]; business (acquisitions) disposition [NCFBus] and investment (acquisition) disposal [NCFInv]. |
| **25** | Net Cash Flow from Operations | [Cash Flow Statement] A component of [NCF] representing the amount of cash inflow (outflow) from operating activities; from continuing and discontinued operations. |
| **26** | Effect of Exchange Rate Changes on Cash | [Cash Flow Statement] A component of Net Cash Flow [NCF] representing the amount of increase (decrease) from the effect of exchange rate changes on cash and cash equivalent balances held in foreign currencies. |
| **27** | Net Cash Flow / Change in Cash & Cash Equivalents | [Cash Flow Statement] Principal component of the cash flow statement representing the amount of increase (decrease) in cash and cash equivalents. Includes [NCFO]; investing [NCFI] and financing [NCFF] for continuing and discontinued operations; and the effect of exchange rate changes on cash [NCFX]. |
| **28** | Share Based Compensation | [Cash Flow Statement] A component of [NCFO] representing the total amount of noncash; equity-based employee remuneration. This may include the value of stock or unit options; amortization of restricted stock or units; and adjustment for officers' compensation. As noncash; this element is an add back when calculating net cash generated by operating activities using the indirect method. |
| **29** | Depreciation Amortization & Accretion | [Cash Flow Statement] A component of operating cash flow representing the aggregate net amount of depreciation; amortization; and accretion recognized during an accounting period. As a non-cash item; the net amount is added back to net income when calculating cash provided by or used in operations using the indirect method. |
| **30** | Total Assets | [Balance Sheet] Sum of the carrying amounts as of the balance sheet date of all assets that are recognized. Major components are [CashnEq]; [Investments];[Intangibles]; [PPNENet];[TaxAssets] and [Receivables]. |
| **31** | Cash and Equivalents | [Balance Sheet] A component of [Assets] representing the amount of currency on hand as well as demand deposits with banks or financial institutions. |
| **32** | Investments | [Balance Sheet] A component of [Assets] representing the total amount of marketable and non-marketable securties; loans receivable and other invested assets. |
| **33** | Investments Current | [Balance Sheet] The current portion of [Investments]; reported if the company operates a classified balance sheet that segments current and non-current assets. |
| **34** | Investments Non-Current | [Balance Sheet] The non-current portion of [Investments]; reported if the company operates a classified balance sheet that segments current and non-current assets. |
| **35** | Deferred Revenue | [Balance Sheet] A component of [Liabilities] representing the carrying amount of consideration received or receivable on potential earnings that were not recognized as revenue; including sales; license fees; and royalties; but excluding interest income. |
| **36** | Deposit Liabilities | [Balance Sheet] A component of [Liabilities] representing the total of all deposit liabilities held; including foreign and domestic; interest and noninterest bearing. May include demand deposits; saving deposits; Negotiable Order of Withdrawal and time deposits among others. |
| **37** | Property Plant & Equipment Net | [Balance Sheet] A component of [Assets] representing the amount after accumulated depreciation; depletion and amortization of physical assets used in the normal conduct of business to produce goods and services and not intended for resale. |
| **38** | Inventory | [Balance Sheet] A component of [Assets] representing the amount after valuation and reserves of inventory expected to be sold; or consumed within one year or operating cycle; if longer. |
| **39** | Tax Assets | [Balance Sheet] A component of [Assets] representing tax assets and receivables. |
| **40** | Trade and Non-Trade Receivables | [Balance Sheet] A component of [Assets] representing trade and non-trade receivables. |
| **41** | Trade and Non-Trade Payables | [Balance Sheet] A component of [Liabilities] representing trade and non-trade payables. |
| **42** | Goodwill and Intangible Assets | [Balance Sheet] A component of [Assets] representing the carrying amounts of all intangible assets and goodwill as of the balance sheet date; net of accumulated amortization and impairment charges. |
| **43** | Total Liabilities | [Balance Sheet] Sum of the carrying amounts as of the balance sheet date of all liabilities that are recognized. Principal components are [Debt]; [DeferredRev]; [Payables];[Deposits]; and [TaxLiabilities]. |
| **44** | Shareholders Equity | [Balance Sheet] A principal component of the balance sheet; in addition to [Liabilities] and [Assets]; that represents the total of all stockholders' equity (deficit) items; net of receivables from officers; directors; owners; and affiliates of the entity which are attributable to the parent. |
| **45** | Accumulated Retained Earnings (Deficit) | [Balance Sheet] A component of [Equity] representing the cumulative amount of the entities undistributed earnings or deficit. May only be reported annually by certain companies; rather than quarterly. |
| **46** | Accumulated Other Comprehensive Income | [Balance Sheet] A component of [Equity] representing the accumulated change in equity from transactions and other events and circumstances from non-owner sources; net of tax effect; at period end. Includes foreign currency translation items; certain pension adjustments; unrealized gains and losses on certain investments in debt and equity securities. |
| **47** | Current Assets | [Balance Sheet] The current portion of [Assets]; reported if a company operates a classified balance sheet that segments current and non-current assets. |
| **48** | Assets Non-Current | [Balance Sheet] Amount of non-current assets; for companies that operate a classified balance sheet. Calculated as the different between Total Assets [Assets] and Current Assets [AssetsC]. |
| **49** | Current Liabilities | [Balance Sheet] The current portion of [Liabilities]; reported if the company operates a classified balance sheet that segments current and non-current liabilities. |
| **50** | Liabilities Non-Current | [Balance Sheet] The non-current portion of [Liabilities]; reported if the company operates a classified balance sheet that segments current and non-current liabilities. |
| **51** | Tax Liabilities | [Balance Sheet] A component of [Liabilities] representing outstanding tax liabilities. |
| **52** | Total Debt | [Balance Sheet] A component of [Liabilities] representing the total amount of current and non-current debt owed. Includes secured and unsecured bonds issued; commercial paper; notes payable; credit facilities; lines of credit; capital lease obligations; and convertible notes. |
| **53** | Debt Current | [Balance Sheet] The current portion of [Debt]; reported if the company operates a classified balance sheet that segments current and non-current liabilities. |
| **54** | Debt Non-Current | [Balance Sheet] The non-current portion of [Debt] reported if the company operates a classified balance sheet that segments current and non-current liabilities. |
| **55** | Earnings before Tax | [Metrics] Earnings Before Tax is calculated by adding [TaxExp] back to [NetInc]. |
| **56** | Earning Before Interest & Taxes (EBIT) | [Income Statement] Earnings Before Interest and Tax is calculated by adding [TaxExp] and [IntExp] back to [NetInc]. |
| **57** | Earnings Before Interest Taxes & Depreciation Amortization (EBITDA) | [Metrics] EBITDA is a non-GAAP accounting metric that is widely used when assessing the performance of companies; calculated by adding [DepAmor] back to [EBIT]. |
| **58** | Foreign Currency to USD Exchange Rate | [Metrics] The exchange rate used for the conversion of foreign currency to USD for non-US companies that do not report in USD. |
| **59** | Shareholders Equity (USD) | [Balance Sheet] [Equity] in USD; converted by [FXUSD]. |
| **60** | Earnings per Basic Share (USD) | [Income Statement] [EPS] in USD; converted by [FXUSD]. |
| **61** | Revenues (USD) | [Income Statement] [Revenue] in USD; converted by [FXUSD]. |
| **62** | Net Income Common Stock (USD) | [Income Statement] [NetIncCmn] in USD; converted by [FXUSD]. |
| **63** | Cash and Equivalents (USD) | [Balance Sheet] [CashnEq] in USD; converted by [FXUSD]. |
| **64** | Total Debt (USD) | [Balance Sheet] [Debt] in USD; converted by [FXUSD]. |
| **65** | Earning Before Interest & Taxes (USD) | [Income Statement] [EBIT] in USD; converted by [FXUSD]. |
| **66** | Earnings Before Interest Taxes & Depreciation Amortization (USD) | [Metrics] [EBITDA] in USD; converted by [FXUSD]. |
| **67** | Shares (Basic) | [Entity] The number of shares or other units outstanding of the entity's capital or common stock or other ownership interests; as stated on the cover of related periodic report (10-K/10-Q); after adjustment for stock splits. |
| **68** | Dividends per Basic Common Share | [Income Statement] Aggregate dividends declared during the period for each split-adjusted share of common stock outstanding. Includes spinoffs where identified. |
| **69** | Share Factor | [Entity] Share factor is a multiplicant in the calculation of [MarketCap] and is used to adjust for: American Depository Receipts (ADRs) that represent more or less than 1 underlying share; and; companies which have different earnings share for different share classes (eg Berkshire Hathaway - BRKB). |
| **70** | Market Capitalization | [Metrics] Represents the product of [SharesBas]; [Price] and [ShareFactor]. |
| **71** | Enterprise Value | [Metrics] Enterprise value is a measure of the value of a business as a whole; calculated as [MarketCap] plus [DebtUSD] minus [CashnEqUSD]. |
| **72** | Invested Capital | [Metrics] Invested capital is an input into the calculation of [ROIC]; and is calculated as: [Debt] plus [Assets] minus [Intangibles] minus [CashnEq] minus [LiabilitiesC]. Please note this calculation method is subject to change. |
| **73** | Average Equity | [Metrics] Average equity value for the period used in calculation of [ROE]; derived from [Equity]. |
| **74** | Average Assets | [Metrics] Average asset value for the period used in calculation of [ROE] and [ROA]; derived from [Assets]. |
| **75** | Invested Capital Average | [Metrics] Average invested capital value for the period used in the calculation of [ROIC]; and derived from [InvCap]. Invested capital is an input into the calculation of [ROIC]; and is calculated as: [Debt] plus [Assets] minus [Intangibles] minus [CashnEq] minus [LiabilitiesC]. Please note this calculation method is subject to change. |
| **76** | Tangible Asset Value | [Metrics] The value of tangibles assets calculated as the difference between [Assets] and [Intangibles]. |
| **77** | Return on Average Equity | [Metrics] Return on equity measures a corporation's profitability by calculating the amount of [NetIncCmn] returned as a percentage of [EquityAvg]. |
| **78** | Return on Average Assets | [Metrics] Return on assets measures how profitable a company is [NetIncCmn] relative to its total assets [AssetsAvg]. |
| **79** | Free Cash Flow | [Metrics] Free Cash Flow is a measure of financial performance calculated as [NCFO] minus [CapEx]. |
| **80** | Return on Invested Capital | [Metrics] Return on Invested Capital is ratio estimated by dividing [EBIT] by [InvCapAvg]. [InvCap] is calculated as: [Debt] plus [Assets] minus [Intangibles] minus [CashnEq] minus [LiabilitiesC]. Please note this calculation method is subject to change. |
| **81** | Gross Profit | [Income Statement] Aggregate revenue [Revenue] less cost of revenue [CoR] directly attributable to the revenue generation activity. |
| **82** | Operating Income | [Income Statement] Operating income is a measure of financial performance before the deduction of [IntExp]; [TaxExp] and other Non-Operating items. It is calculated as [GP] minus [OpEx]. |
| **83** | Gross Margin | [Metrics] Gross Margin measures the ratio between a company's [GP] and [Revenue]. |
| **84** | Profit Margin | [Metrics] Measures the ratio between a company's [NetIncCmn] and [Revenue]. |
| **85** | EBITDA Margin | [Metrics] Measures the ratio between a company's [EBITDA] and [Revenue]. |
| **86** | Return on Sales | [Metrics] Return on Sales is a ratio to evaluate a company's operational efficiency; calculated by dividing [EBIT] by [Revenue]. ROS is often a component of DuPont ROE analysis. |
| **87** | Asset Turnover | [Metrics] Asset turnover is a measure of a firms operating efficiency; calculated by dividing [Revenue] by [AssetsAVG]. Often a component of DuPont ROE analysis. |
| **88** | Payout Ratio | [Metrics] The percentage of earnings paid as dividends to common stockholders. Calculated by dividing [DPS] by [EPSUSD]. |
| **89** | Enterprise Value over EBITDA | [Metrics] Measures the ratio between [EV] and [EBITDAUSD]. |
| **90** | Enterprise Value over EBIT | [Metrics] Measures the ratio between [EV] and [EBITUSD]. |
| **91** | Price Earnings (Damodaran Method) | [Metrics] Measures the ratio between [MarketCap] and [NetIncCmnUSD] |
| **92** | Price to Earnings Ratio | [Metrics] An alternative to [PE] representing the ratio between [Price] and [EPSUSD]. |
| **93** | Sales per Share | [Metrics] Sales per Share measures the ratio between [RevenueUSD] and [SharesWA]. |
| **94** | Price to Sales Ratio | [Metrics] An alternative calculation method to [PS]; that measures the ratio between a company's [Price] and it's [SPS]. |
| **95** | Price Sales (Damodaran Method) | [Metrics] Measures the ratio between [MarketCap] and [RevenueUSD]. |
| **96** | Price to Book Value | [Metrics] Measures the ratio between [MarketCap] and [EquityUSD]. |
| **97** | Debt to Equity Ratio | [Metrics] Measures the ratio between [Liabilities] and [Equity]. |
| **98** | Dividend Yield | [Metrics] Dividend Yield measures the ratio between a company's [DPS] and its [Price]. |
| **99** | Current Ratio | [Metrics] The ratio between [AssetsC] and [LiabilitiesC]; for companies that operate a classified balance sheet. |
| **100** | Working Capital | [Metrics] Working capital measures the difference between [AssetsC] and [LiabilitiesC]. |
| **101** | Free Cash Flow per Share | [Metrics] Free Cash Flow per Share is a valuation metric calculated by dividing [FCF] by [SharesWA]. |
| **102** | Book Value per Share | [Metrics] Measures the ratio between [Equity] and [SharesWA]. |
| **103** | Tangible Assets Book Value per Share | [Metrics] Measures the ratio between [Tangibles] and [SharesWA]. |
| **104** | Share Price (Adjusted Close) | [Entity] The price per common share adjusted for stock splits but not adjusted for dividends; used in the computation of [PE1]; [PS1]; [DivYield] and [SPS]. |
| **105** | Ticker Symbol | [Entity] The ticker is a unique identifer for an issuer in the database. Where a ticker contains a . or a - this is removed from the ticker we use. For example BRK.B is BRKB in this dataset. We include the BRK.B ticker in the Related Tickers field of the ticker listing. Where a company is delisted and the ticker is recycled; we utilise that ticker for the currently active company and append a number to the ticker of the delisted company. For example GM represents the current actively traded General Motors entity; and GM1 represents the entity that filed for bankruptcy in 2009. Where we have identified that multiple classes of shares exist for a company; we include the alternative share classes in the Related Tickers field of the ticker listing. For example we provide data for GOOGL; an... |
| **106** | Dimension | [Entity] The dimension field allows you to take different dimensional views of data over time. ARQ: Quarterly; excluding restatements; MRQ: Quarterly; including restatements; ARY: annual; excluding restatements; MRY: annual; including restatements; ART: trailing-twelve-months; excluding restatements; MRT: trailing-twelve-months; including restatements. |
| **107** | Calendar Date | [Entity] The Calendar Date represents the normalized [ReportPeriod]. This provides a common date to query for which is necessary due to irregularity in report periods across companies. For example; if the report period is "2015-09-26"; the calendar date will be "2015-09-30" for quarterly and trailing-twelve-month dimensions (ARQ;MRQ;ART;MRT); and "2015-12-31" for annual dimensions (ARY;MRY). We also employ offsets in order to maximise comparability of the period across companies. For example consider two companies: one with a quarter ending on 2018-07-24; and the other with a quarter ending on 2018-06-28. A naive normalization process would assign these to differing calendar quarters of 2018-09-30 and 2018-06-30 respectively. However, we assign these both to the 2018-06-30 calendar qua... |
| **108** | Date Key | [Entity] The Date Key represents the SEC filing date for AR dimensions (ARQ;ART;ARY); and the [REPORTPERIOD] for MR dimensions (MRQ;MRT;MRY). In addition; this is the observation date used for [Price] based data such as [MarketCap]; [Price] and [PE]. |
| **109** | Report Period | [Entity] The Report Period represents the end date of the fiscal period. |
| **110** | Last Updated Date | [Entity] Last Updated represents the last date that this database entry was updated; which is useful to users when updating their local records. |

**Daily Accounting Metrics**

Some Accounting Metrics are available daily as well as monthly and are provided on a separate table.

Below are the metrics provided daily for Apple.

|  | **ticker** | **date** | **lastupdated** | **ev** | **evebit** | **evebitda** | **marketcap** | **pb** | **pe** | **ps** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **None** |  |  |  |  |  |  |  |  |  |  |
| **0** | AAPL | 2019-01-17 | 2019-01-17 | 828187.7 | 11.4 | 9.9 | 739617.7 | 6.9 | 12.4 | 2.8 |
| **1** | AAPL | 2019-01-16 | 2019-01-16 | 823822.0 | 11.3 | 9.8 | 735252.0 | 6.9 | 12.4 | 2.8 |

**Pricing**

Pricing data is available from another Quandl table - Sharadar Equity Prices,SEP . Updated daily, this database provides End-Of-Day (EOD) price data with coverage corresponding to Sharadar's Fundamental dataset, SF1.

Below is an example of the latest pricing information for Apple.

|  | **ticker** | **date** | **open** | **high** | **low** | **close** | **volume** | **dividends** | **closeunadj** | **lastupdated** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **None** |  |  |  |  |  |  |  |  |  |  |
| **0** | AAPL | 2019-01-17 | 154.2 | 157.66 | 153.26 | 155.86 | 29336202.0 | 0.0 | 155.86 | 2019-01-17 |

This data can be a date range, below are Apple prices from the begging of the year

|  | **ticker** | **date** | **open** | **high** | **low** | **close** | **volume** | **dividends** | **closeunadj** | **lastupdated** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **None** |  |  |  |  |  |  |  |  |  |  |
| **0** | AAPL | 2019-01-17 | 154.20 | 157.66 | 153.26 | 155.86 | 29336202.0 | 0.0 | 155.86 | 2019-01-17 |
| **1** | AAPL | 2019-01-16 | 153.08 | 155.88 | 153.00 | 154.94 | 29626386.0 | 0.0 | 154.94 | 2019-01-16 |
| **2** | AAPL | 2019-01-15 | 150.27 | 153.39 | 150.05 | 153.07 | 28227737.0 | 0.0 | 153.07 | 2019-01-15 |
| **3** | AAPL | 2019-01-14 | 150.85 | 151.27 | 149.22 | 150.00 | 31805772.0 | 0.0 | 150.00 | 2019-01-14 |
| **4** | AAPL | 2019-01-11 | 152.88 | 153.70 | 151.51 | 152.29 | 26711076.0 | 0.0 | 152.29 | 2019-01-11 |
| **5** | AAPL | 2019-01-10 | 152.50 | 153.97 | 150.86 | 153.80 | 35065010.0 | 0.0 | 153.80 | 2019-01-10 |
| **6** | AAPL | 2019-01-09 | 151.29 | 154.53 | 149.63 | 153.31 | 44641758.0 | 0.0 | 153.31 | 2019-01-09 |
| **7** | AAPL | 2019-01-08 | 149.56 | 151.82 | 148.52 | 150.75 | 39977341.0 | 0.0 | 150.75 | 2019-01-08 |
| **8** | AAPL | 2019-01-07 | 148.70 | 148.83 | 145.90 | 147.93 | 54308097.0 | 0.0 | 147.93 | 2019-01-07 |
| **9** | AAPL | 2019-01-04 | 144.53 | 148.55 | 143.80 | 148.26 | 56884249.0 | 0.0 | 148.26 | 2019-01-04 |
| **10** | AAPL | 2019-01-03 | 143.98 | 145.72 | 142.00 | 142.19 | 90466553.0 | 0.0 | 142.19 | 2019-01-03 |
| **11** | AAPL | 2019-01-02 | 154.89 | 158.85 | 154.23 | 157.92 | 31738194.0 | 0.0 | 157.92 | 2019-01-02 |

**Pricing Data Explanation**

Below is a table with descriptions and unit types for all the price data types. Generally 'close' price is used for projects such as this

|  | **table** | **indicator** | **isfilter** | **isprimarykey** | **title** | **description** | **unittype** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **None** |  |  |  |  |  |  |  |
| **0** | SEP | ticker | Y | Y | Ticker Symbol | The ticker is a unique identifer for an issuer in the database. Where a ticker contains a "." or a "-" this is removed from the ticker. For example BRK.B is BRKB. We include the BRK.B ticker in the Related Tickers field. Where a company is delisted and the ticker is recycled; we use that ticker for the currently active company and append a number to the ticker of the delisted company. eg GM is the current actively traded entity; & GM1 is the entity that filed for bankruptcy in 2009. | text |
| **1** | SEP | date | Y | Y | Price Date | The trade date of the price observations. | date (YYYY-MM-DD) |
| **2** | SEP | open | N | N | Open Price - Split Adjusted | The opening share price, adjusted for stock splits and stock dividends. | USD/share |
| **3** | SEP | high | N | N | High Price - Split Adjusted | The high share price, adjusted for stock splits and stock dividends. | USD/share |
| **4** | SEP | low | N | N | Low Price - Split Adjusted | The low share price, adjusted for stock splits and stock dividends. | USD/share |
| **5** | SEP | close | N | N | Close Price - Split Adjusted | The open share closing, adjusted for stock splits and stock dividends. | USD/share |
| **6** | SEP | volume | N | N | Volume - Split Adjusted | The traded volume, adjusted for stock splits and stock dividends. | numeric |
| **7** | SEP | dividends | N | N | Dividends per Share - Split Adjusted | The dividends per share unit, adjusted for stock splits and stock dividends. Includes spinoffs where identified. | USD/share |
| **8** | SEP | closeunadj | N | N | Close Price - Unadjusted | The closing share price, not adjusted for stock splits and stock dividends. | USD/share |
| **9** | SEP | lastupdated | Y | N | Last Updated Date | The last date at which this line item was updated, typically used to filter date to be retrieved for syncing to local records. | date (YYYY-MM-DD) |

**3.2 Python packages – Explanation of the Python packages that will be used to implement Neural Nets**

**3.2.1 Implementation – Code and issues surrounding implementation**

The implementation of the RNNs was carried out using the Python library Keras. Keras runs on top of numerical computation and large-scale machine learning libraries TensorFlow, CNTK or Theano.

Gathering the data was done by accessing Quandl APIs, for price data and accounting metric data. Price data needed to be manipulated using the Pandas library into firstly monthly price from a daily pricing dataset then changed into monthly returns. Returns being the change (positive or negative) in the monthly share price.

Factor data (accounting metrics) was also returned from a Quandl API. As not all factor data is available on a monthly basis quarterly data was maintained until new data became available. Returns data is then added to the factor data. After running a correlation matrix and factors with a correlation of > 0.95 are dropped to reduce model complexity.

As Keras only accepts data in the form of arrays the Pandas dataframe is amended to a Numpy array. As the factor data is composed of different scales these values are normalised to that they take small values on a similar scale. The SKlearn module MinMax scalar is used to pass a feature range between 0 and 1 for the data.

3.2.2 Training the RNN models

A neural network is based around four objects. Layers (combined into a network or model), input data and corresponding targets, loss function that defines the feedback signal used for learning and the optimiser which determines how learning proceeds. A layer is a data-processing module that takes as input one or more tensors and outputs the same. Some layers are stateless but more often that have a state: the layer’s weights, one or several tensors learned with stochastic gradient descent, which together contain the network’s knowledge.

Different layers are appropriate for different tensor formats and different types of data processing. For instance, simple vector data, stored in 2D tensors of shape (samples, features), is often processed by densely connected layers, also called fully connected or dense layers (the Dense class in Keras). Sequence data, stored in 3D tensors of shape (samples, timesteps, features), is typically processed by recurrent layers such as an LSTM layer. Image data, stored in 4D tensors, is usually processed by 2D convolution layers (Conv2D).

The dataset is a timeseries dataset therefore a 3D tensor shape is required for the project.

After the architecture is defined there are still choices to be made with regard to

the loss function and the optimiser. As this is fundamentally a regression the chosen loss function used can be mean absolute error (mae) or mean squared error (mse). Mae being the absolute value of the difference between the predictions and the targets and mse being the square of the difference. This is a widely used loss function for regression problems. The default in the keras documentation is mse and this was used at initiation.

The dataset will be data for one stock ‘A’ for the time period 2009-13-31 to 2019-01-01 a matrix (36, 72) in shape. Train data will be the first 32 instances of the data set. Initially a very small network was used consisting of one hidden layer with 32 units. The design of neural networks is often described as more of an art than a science, a process of trial and error but also adopting know best practices was the approach taken in the design stage. Small networks are often seen as a way of mitigating overfitting. The network ends with a single unit and no activation, a linear layer. This will single layer will be maintained for all architectures used, as this is a regression problem and we are trying to predict a single continuous value. Applying an activation function would constrain the range the output could take. The optimiser is ‘rmsprop’, is the default optimiser in the keras documentation. The implementation also uses the mean absolute error as the metric to be monitored by the model. To validate the network K-fold cross validation is used splitting the data into K partitions, instantiating the K identical models and training each one on K-1 partitions while evaluation is based on the remaining partition. Taking the average of the scores returns a single measurement. The epochs, the length of time the model is trained for, is set to 100 in and attempt to restrain processor requirements. Parameters that will be adjusted are the number and size of the hidden layers. After the parameters have been amended a model is trained on all the training data and its performance evaluated on the test data. A simple model will be used as a baseline for comparison with more complicated models.

The next step in the implementation process was to move from just a densely connected neural networks to recurring neural networks, adding memory to the network. RNNs process sequences by iterating through the sequence elements and maintaining a state containing information relative to what it has seen so far. Keras has a SimpleRNN layer which processes batches of sequences and takes the inputs of shape (batch\_size, timesteps, input\_features), rather than just (timesteps, input\_features.)

A SimpleRNN layer was used in the next testing with changes to the numbers of layers and the units the parameters that could be amended. As is captured in the literature simple RNNs are generally too simplistic to be of real use. SimpleRNN has a major issue: although it should theoretically be able to retain at time*t* information about inputs seen many timesteps before, in practice, such long-term dependencies are impossible to learn. This is due to the vanishing gradient problem

 LSTM and GRU layers are designed to solve this problem, and these are applied next. A single layer GRU is added with a single dense output layer, keras has an implementation for this, again the parameters are the number of layers and the units. Next a single layer LSTM is applied in the same way.

**4 TESTING AND RESULTS**

Evaluation of all models will be 4-fold cross validation with 100 epochs with mean absolute error (mae) as the validation metric, the mean for these folds is calculated. The mae is then calculated on the entire test data set providing a final test mae score.

To determine a baseline for testing, a simple small densely connected network was created.

|  |  |
| --- | --- |
| **Model Layer** | Dense |
| **Number of Layers** | 1 |
| **Number of Units** | 32 |
| **Optimiser** | rmsprop |
| **Loss** | mse |
| **Metrics** | mae |
| **Validation** |  |
| **Fold** |  |
| **1** | 0.343737632 |
| **2** | 0.134157419 |
| **3** | 0.175569564 |
| **4** | 0.177201182 |
| **Validation Mean** | 0.207666449 |
| **Test Score** | **0.108901992** |

The results for the Simple RNN

|  |  |
| --- | --- |
| **Model Layer** | Simple RNN |
| **Number of Layers** | 1 |
| **Number of Units** | 32 |
| **Optimiser** | rmsprop |
| **Loss** | mse |
| **Metrics** | mae |
| **Validation** |  |
| **Fold** |  |
| **1** | 0.287335515 |
| **2** | 0.236223578 |
| **3** | 0.156718925 |
| **4** | 0.206536397 |
| **Validation Mean** | 0.221703604 |
| **Test Score** | **0.302181363** |

The Simple RNN results are a large degradation on the baseline.

GRU results improve on the baseline slightly.

|  |  |
| --- | --- |
| **Model Layer** | GRU |
| **Number of Layers** | 1 |
| **Number of Units** | 32 |
| **Optimiser** | rmsprop |
| **Loss** | mse |
| **Metrics** | mae |
| **Validation** |  |
| **Fold** |  |
| **1** | 0.284213245 |
| **2** | 0.147503138 |
| **3** | 0.100005403 |
| **4** | 0.151451439 |
| **Validation Mean** | 0.170793306 |
| **Test Score** | **0.102192424** |

LSTM score is a further improvement

|  |  |
| --- | --- |
| **Model Layer** | LSTM |
| **Number of Layers** | 1 |
| **Number of Units** | 32 |
| **Optimiser** | rmsprop |
| **Loss** | mse |
| **Metrics** | mae |
| **Validation** |  |
| **Fold** |  |
| **1** | 0.282652259 |
| **2** | 0.131771699 |
| **3** | 0.098661974 |
| **4** | 0.125670373 |
| **Validation Mean** | 0.159689076 |
| **Test Score** | **0.096809559** |

Plotting this we can see that the baseline simple model is difficult to beat although some gains can be made with models that have the concept of ‘memory.’

Parameters that are configurable are number of layers and number of units. Changing the layers to two (excluding the output layer) had the following effects on the test scores, retaining all other model metrics.

|  |  |
| --- | --- |
| Model | MAE Score |
| Baseline | 0.10890199 |
| Simple RNN (1) | 0.30218136 |
| Simple RNN (2) | 0.1022668 |
| GRU (1) | 0.10219242 |
| GRU (2) | 0.13110155 |
| LSTM (1) | 0.09680956 |
| LSTM (2) | 0.10668621 |

Adding another layer only improved the Simple RNN model and although this was an improvement on the baseline score it did not better the one-layer GRU or LSTM models. Adding layers did not improve accuracy so the single layer was retained when adjusting the number of units.

Maintaining the single layer on the GRU and LSTM model the number of units can be adjusted. The results of amending the layers are shown below.

Units of 32 do appear to be the optimal number for both GRU and LSTM layers with LSTM(32) fairing the best of all with a MAE

With the major parameters reviewed testing now looked to amending the loss function. The other loss functions available for regression problems are mean squared log error (msle) and mean absolute error (mae). Msle has the effect of relaxing the punishing effect of large differences in large predicted values. When predicting a large value, you may not want to punish a model as heavily as mean squared error. Mae is a loss function that is generally more robust to outliers, large or small values far from the mean value.  When using these for GRU and LSTM single layer models with 32 units the results for the MAE were as below.

A change to the loss functions did not improve the score from the mse

Another parameter is the optimiser. Below are all the optimisers available to keras.

Again, the default RMSProp has the lowest MAE for both GRU and LSTM

**5 ANALYSIS AND DISCUSSION**