Department of Mathematics
Imperial College London
London SW7 2AZ
United Kingdom

Factor Investing in the Automotive Sector

An analysis of factor investing-machine learning based investment strategies in the automotive sector

by

Henry Sorsky

CID: 01387017

September 2018

Thesis submitted as part of the requirements for the award of the
MSc in Mathematics and Finance, Imperial College London, 2017-2018

Abstract

In stage 1 of this study, we set out to find functions of companies' financial reporting items that, when used to partition our automotive universe, resulted in strategies that outperform our benchmarks. We also aim to find functions of fundamentals that, with perfect knowledge of their future values, resulted in strategies that outperformed those that used current data (i.e. no knowledge of the future). We find that there exist many functions that produce far superior returns than our strategies with no knowledge of the future, and that a yearly rebalancing frequency was generally superior to higher frequency rebalancing. We also find that those top performant strategies with no knowledge of the future generally do not benefit from clairvoyance, but that some of the less performant strategies do.

In stage 2, we set out to model the future values of those functions of fundamentals that performed better when we had knowledge of the future, from stage 1. We show that this is a difficult task, but that simple linear models exist that are able to do so.

Finally, in stage 3 we test those strategies which outperform our naïve benchmark in stage 2. We find that of those that outperform the naïve benchmark, the majority experience higher gains than their no-future-knowledge counterparts.

Declaration

The work contained in this thesis is my own work unless otherwise stated.

Signature:

Henry Sorsky

Date: 11-09-2018

Acknowledgements

I would like to thank my internal supervisor, Dr Damiano Brigo, for his help and insight throughout the writing of this Thesis; my external supervisor, Dr Johannes Bauer, for his continued help, expertise, guidance and patience whilst I worked with IHS Markit - I know I am not always the easiest person to work with; and my parents, for supporting me and enabling me to undertake the MSc Mathematics and Finance.

Contents

List of Tables

List of Figures

Key Abbreviations

- CAR - Compound Annual Returns
- COGS - Cost Of Goods Sold
- EBIT - Earnings Before Interest and Tax
- EV - Enterprise Value
- FFO - Funds From Operations
- IHSM - IHS Markit
- MSE - Mean Squared Error
- SR - Sharpe Ratio

Chapter 1

Introduction

We hear in the news every day, phrases such as "The DOW is up 2% on yesterday" or "The FTSE is experiencing 5% larger growth than this time last year", many people without a complete understanding of what these acronyms actually are. Many of us know that they indicate roughly whether a market is doing well or not, but don't understand the maths behind them or how to expose ourselves to these markets' (hopefully) growth.

To explain why the press, fund managers, CEOs and Joe Public alike are all so interested in financial market's and their indexes, we walk the reader through a brief history of asset management/portfolio theory.

1.1 Pre 1950s asset management/portfolio theory

In the early 1900s, many fund managers attitudes towards investing were very different to those that we have today. Far less thought was given to risk and much more to return. Many asset managers simply sought out stocks which they believed would have the largest return, ignoring the risks that came along with them. They were essentially placing uneducated bets on stocks, without a full understanding of their dynamics[1]. Rules on portfolio construction were also not clear[2] and there was little formalisation of each fund's aims other than to make money. How they went about this was often ill-defined and dependent on which manager within each fund you considered.

A popular book on financial theory at this point in time was [Williams, 1938]. The main focus was the discounted dividend model, where Williams hypothesises that the value of an asset should equal the present value of its future dividends (i.e. its discounted dividends, were one to hold the stock indefinitely or until removal from the market). He then suggests, as was common practice at the time, to invest in the stocks who's valuation was greater than their current market price (i.e. they traded at a discount). Whilst Williams' model is the foundation

[1] Of course, this is not true for all managers, but certainly is for a large proportion of them.

[2] When we say portfolio construction, we mean the allocation of funds between available investments.

of most valuation methods employed today, his investment strategy also epitomises the flaws in the era's thinking. Firstly, they were only considering payoffs (or expected payoffs), rather than taking risk into account. Secondly, his suggested analysis pertains to individual stocks, not to portfolios as a whole.

The final major difference between the 1930s/40s and more recent times that we mention is that of information. In particular, information was scarce, poorly reported and came in non-uniform frequencies. Many fund managers, such as Benjamin Graham, capitalised on this, finding accurate data and then analysing it properly. Graham would accumulate data on, and then analyse the "books" of a company to asses it's intrinsic value and would invest if this intrinsic value were some margin greater than the market value. Graham can, therefore, be thought of as the father of **value investing**[3]. Even with his slightly smarter approach to investing/portfolio construction, his methods were still a long way off of those which many managers employ today. So where did it all change?

1.2 1952 - The arrival of Harry Markowitz and the birth of modern portfolio theory

It was whilst reading [Williams, 1938] in the library one afternoon, that a young Harry Markowitz had what is now considered one of the most important thoughts in all of finance/economics. As previously mentioned, Williams proposed that the value of a stock should equal the present value of it's future dividends. As future dividends are a random variable, Markowitz interpreted this to be to value a stock by its **expected** future dividends. Furthermore, Markowitz considered the logic behind Williams' suggested strategy[4] somewhat flawed. This is because it leads to deciding only to invest in the stock which you believe to have the highest return, as this maximises expected portfolio returns. From experience, and some common sense, Markowitz knew that this was neither how investors **should** act nor how they **did** act. Instead they tended to diversify their portfolios, in order to reduce their risk. This lead him to, alongside expected return, consider the portfolio return variance, a measure of risk, when deciding on which portfolio was optimal. The fact that a portfolio's return variance (a.k.a. volatility) also depended on the

[3] Value investing is the methodology that Warren Buffet follows at his firm Berkshire Hathaway.

[4] Markowitz mentions [Williams, 1938, p. 55-75] in [Markowitz, 1952, p. 77], alluding to some suggested strategy of investing in portfolios with maximal discounted future returns. Upon reading [Williams, 1938], however, we found no such suggested strategy. As it seems to be common belief that this was a popular optimisation rule for this period, we continue as though such a suggestion was made, without formal proof.

covariance between its assets' returns affirmed his original thought that this was a better way to assess a portfolio. Furthermore, as there were now two variables to consider when assessing ones portfolio, he decided it natural to assume that investors should select a portfolio based on the set of Pareto optimal risk-return combinations[5] [Markowitz, 1990, paragraph 7]. Markowitz called this set the "Efficient Frontier", an example of which can be seen in Figure 1.1. For more in depth discussion of these notions and their developments, see [Markowitz, 1952].

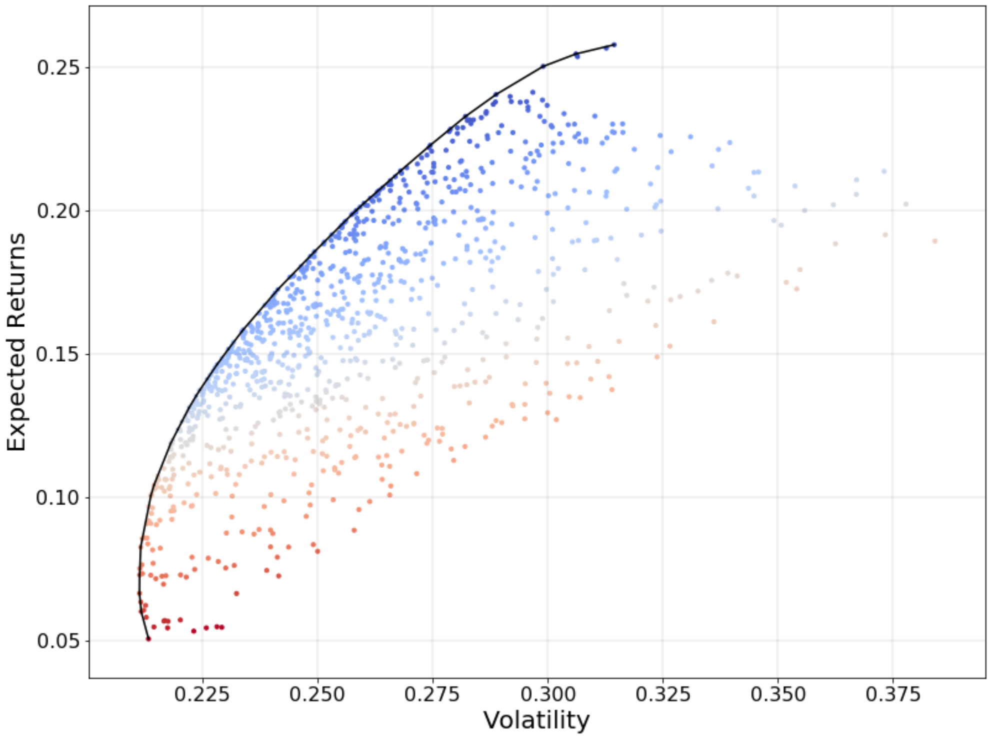


Figure 1.1: Volatility-return combinations generated by holding random weights of `AMZN`, `NFLX` and `TSLA`

In the above figure, each dot represents a portfolio. On the y-axis we have it's expected returns and on the x-axis its expected volatility. We see that the efficient frontier is represented by the curved black line that pens in all of our portfolios. We explain why in the next chapter.

As an example of why Markowitz's thinking is logical, let us consider an example:

Example 1.2.1. *Imagine you have \$1 and you can choose between two bets, each costing you \$1:*

1. *You bet \$1 and, each with probability 0.5, you can win nothing (\$0) or win \$4.*

[5] I.e. you should pick from the set of portfolios for which you cannot expect higher return without taking on more risk, or take on less risk without forfeiting some expected return. More on this in the next chapter.

2. *You bet $1 and, with probability 1, you win $2.*

Ignoring how badly priced this game is, if you were able to play this game as many times as you wish, obviously you would take the 2nd choice every time, and gain $1 every time! Why risk losing any money (which if it occurred on your first try would mean not being able to play again) when you can guarantee a profit of $1? In Markowitz's new optimisation system, given the two games have the same expected payoff, it, therefore, makes sense to consider the possible range of payoffs and their probabilities, something which is further captured by the variance rather than expected returns/payoff alone[6]. In our toy example, we see that game 2 has variance 0 (you always win $1) whereas game 1 has variance 4. As we want to minimise risk (i.e. variance) we take the game with the lowest variance and therefore choose to play game 2. Markowitz's new rules for portfolio selection, therefore, seem to comply with simple logic.

Diversification was also on the rise, with many investors realising that a diverse portfolio could give similar (if not superior returns) with less risk[7]. Let us asses the two aforementioned rules, based on how they deal with diversification. Under the old portfolio optimisation rules, if two portfolios, one of a single asset and one consisting of 30 assets had the same expected return, an investor should be indifferent between them. Under Markowitz's risk-return rule, we would usually see, for reasons explained later, that the single asset, although having the same expected return as the portfolio, will have a higher variance, and, therefore, the investor would choose to hold the 30 asset portfolio. Markowitz's rule, therefore, also complies with popular ideas on diversification.

[6]In this analysis, we have made some assumptions about investor preferences which we will go on to discuss in the next chapter.

[7]Examples of this can be seen in Figure 1.2, Figure 1.3 and Figure 1.4

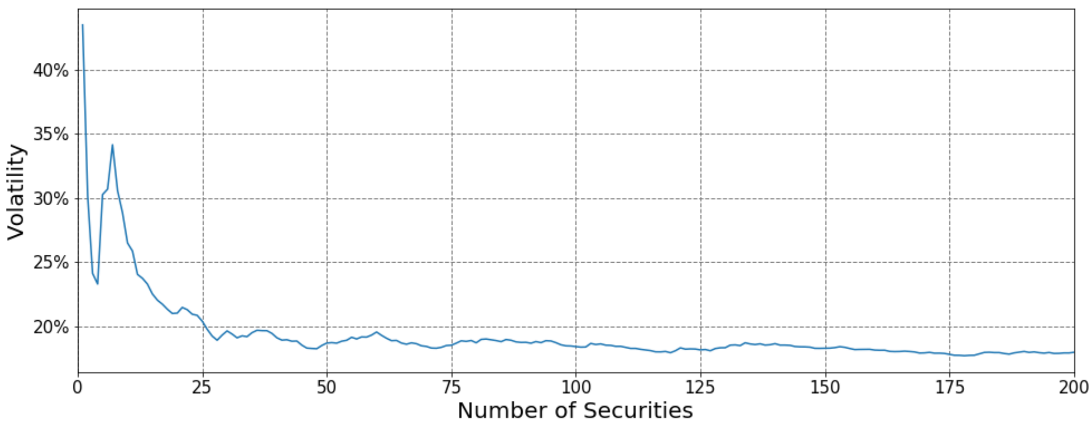


Figure 1.2: Plot of portfolio volatility vs number of securities in the portfolio, for a random selection of stocks on the NYSE, AMEX and NASDAQ, given equal weighting.

Data from https://quantquote.com/historical-stock-data

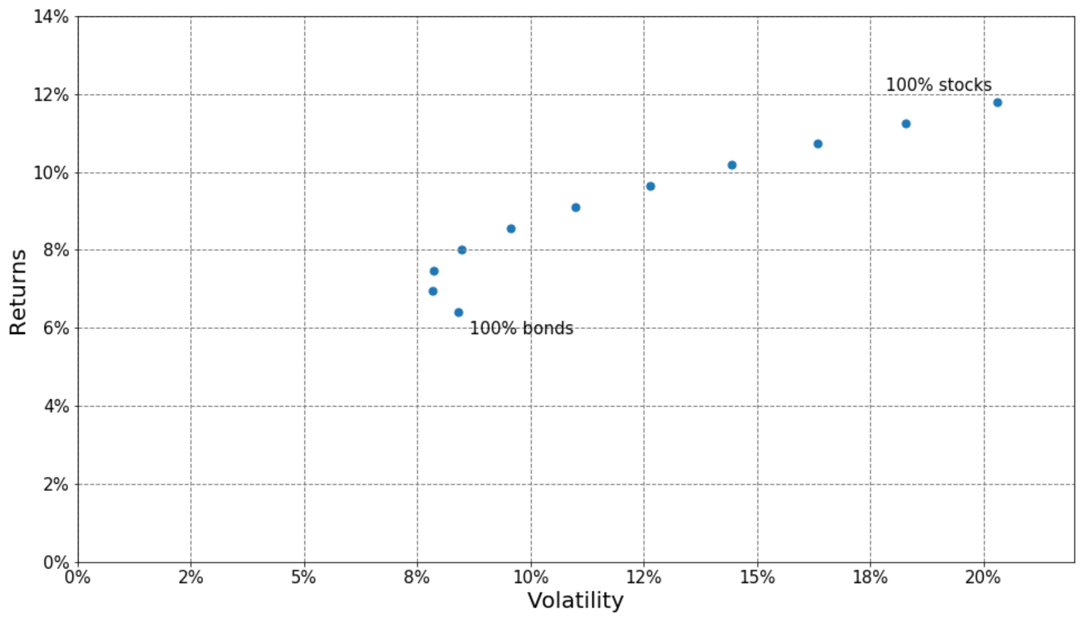


Figure 1.3: Plot of volatility-return combinations when holding portfolios of 100% bonds, 0% stocks through to 0% bonds, 100% stocks.

Data from [Elton et al., 2009, p. 60, Table 4.10]

Here we see that, as we introduce stocks to our portfolio (i.e. move away from 100% bonds), unsurprisingly, we increase our expected returns. What is surprising, however, is that, in doing so, we also decrease our expected volatility.

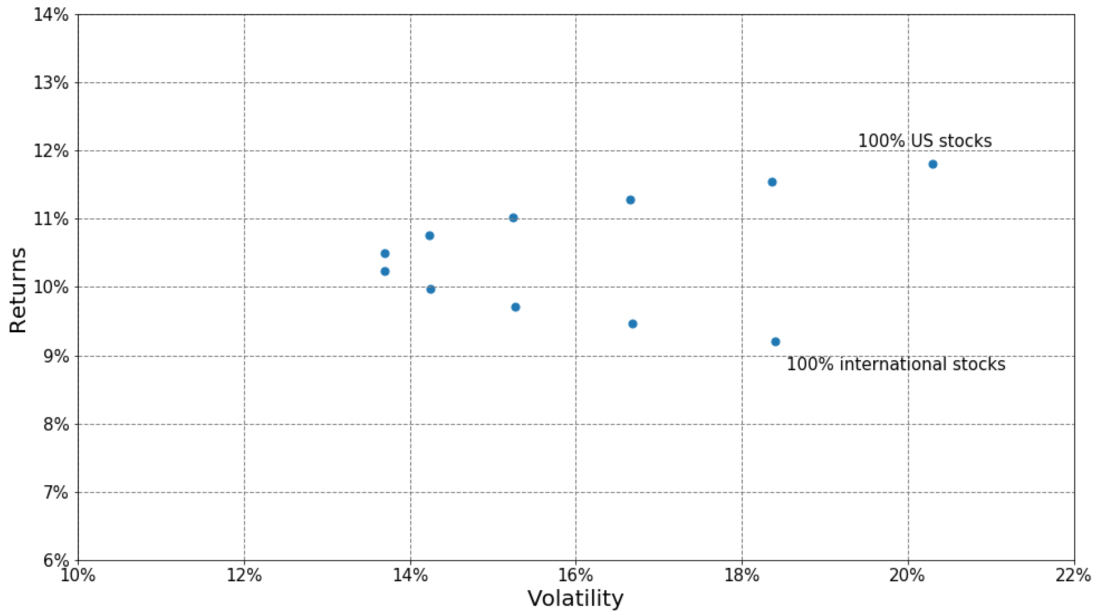


Figure 1.4: Plot of volatility-return combinations when holding portfolios of 100% international stocks, 0% US stocks through to 0% international stocks, 100% US stocks.
Data from [Elton et al., 2009, p. 61, Table 4.11]

Here we see that, moving from 100% international, 0% US, towards a 50:50 split, we are not only diversifying and reducing volatility, but we are also increasing our returns. Thus, in certain cases, increased diversification can be coupled with increased returns.

1.3 Post Markowitz and the move to passive investment

Before we continue, it will help to formally understand what an index is and how one may be constructed.

1.3.1 What is a market index?

Roughly speaking a market index tells us the level of a market. This alone isn't very useful, but the percentage change of this level (its returns) is. So how does one construct an index? Mathematically, the market index's level p_m is the weighted average of its components' prices, p_i:

$$p_m := \sum_{i=1}^{n} w_i p_i, \quad \sum_{i=1}^{n} w_i = 1.$$

The market's returns, therefore, are the returns of this level, or weighted average.

Although we are only talking about market indexes **now** in our quick walk through history, they have, in fact, been around since before the start of our tour began. The oldest continuously quoted index in the US is the Dow Jones Industrial Average Index (DJIA), having been computed and published since 1896. Although it has changed slightly throughout it's history, it largely remains the same. Simply put, it is calculated by the price weighted average[8] of 30 large stocks from USA. Although this is simple and easy to understand, there are problems with constructing an index this way:

1. As the index only contains 30 large stocks, it only tells you how those stocks are performing, and not how the the whole market is doing; and
2. This method of weighting each stock (i.e. by price) implicitly assumes that stocks trading at a higher price are more important. This seems odd; a company could split all of its, say, 1,000,000 shares into two, halving its price, and, therefore, weight, but keeping the total value of the company unchanged. This means that how much a company contributes to the value of the index depends on the somewhat arbitrary choice of how many shares it decides to split its equity into.

In reality, the average investor is more likely to invest in "large" companies (i.e. famous ones) as they know the name and feel they know more about them. It would, therefore, make more sense to give higher index weight to these companies who are larger and less to smaller, unknown, companies. An example of how to do this, would be to weight by a companies **market cap**[9]. This exposes the index more to companies who have a high market value (and are therefore more likely to be picked by the average investor/make up more of the market, by value) and vice versa. We, thus, have:

$$\text{Index Value} := \sum \text{Company Price} \times \frac{\text{Company Market Cap}}{\text{Total Market Cap}},$$

or:

$$I := \sum_{i=1}^{n} \frac{m_i}{M} \times p_i = \sum_{i=1}^{n} w_i p_i, \qquad M := \sum_{i=1}^{n} m_i,$$

where $p_i :=$ price of company i and $m_i :=$ market cap of company i.

The average investor is also likely to invest outside of a 30 stock subset of all available stocks. Therefore, to address point 1. from above, we perform this averaging over all available stocks in

[8] I.e. higher price leads to higher weight; $w_i = \frac{p_i}{\sum p_i}$.

[9] Market cap = market value of 1 share × shares outstanding.

the market that we wish to summarise. Examples of markets over which indexes are calculated include, but is by no means exhausted by:

- Stock exchanges - e.g. all stocks available on the NASDAQ;
- Countries - e.g. all stocks that are registered on USA based exchanges;
- Industry/Sector - e.g. all companies whose operations directly pertain to production of shoes; and
- Intersections, unions and subsets thereof.

As a market index somehow contains information on all of its components' prices, it is thought to summarise its market in some way; if the index experienced positive returns over some period, we would expect the constituents of the index more likely to have experienced positive returns, rather than negative, over the same period. This leads us on nicely to our next question.

1.3.2 Why is an index useful?

So now that we know what an index is and what it aims to do, how do we use them? Given we claim that:

1. Investors like diversity as it decreases return variance (volatility);
2. Market indexes give an overall summary of the stocks they contain; and
3. Having a large market-cap implies low volatility, [Fama and French, 1993, p. 10];

it makes sense to pose the question *"why not just invest in the whole market. weighted by market cap?"*. The logic behind doing so is threefold:

1. Investing in the whole market gives you maximum diversification;
2. If market indexes give a summary of stocks (i.e. stock returns), with no other knowledge, wouldn't we expect each stock to produce roughly the same returns as the market in which it lives? and
3. If we weight by market cap, we are hopefully reducing the variance of our portfolio's returns.

We admit, this is all a bit hand-wavey, but it gives us an idea of why we might decide to invest in the market as a whole rather than to pick and choose individual stocks. Furthermore, it is a very simple strategy that anyone could, in theory, implement at home.

1.3.3 The rise of the ETF

Given the simplicity and proposed superiority of the aforementioned market-investment strategy[10], this is exactly what a large number of funds decided to do. In 1989, to meet demand for index exposure, what was known as an IPS (Index Participation Share) started trading on the American Stock Exchange and Philadelphia Stock Exchange [Gastineau, 2010]. Unfortunately, due to the Chicago Mercantile Exchange (CME) and the Commodity Futures Trading Commission (CFTC), they never really took off and the IPS died an early death. In the next few years, however, many efforts were made to produce similar products, with the Toronoto Stock Exchange releasing the Toronto 35 Index Participation Units (TIPs 35) in 1990 and State Street Global Investors releasing the S&P 500 Trust ETF[11] (SPDR) in early 1993. In fact, SPDR is still one of the most actively traded ETFs to this day.

Summary 1

- We should consider both risk and reward when selecting among investment options.
- Diversification generally reduces volatility.
- Indexes summarise their markets in some way.

[10] We will show more rigorously in the next chapter why such logic holds up.
[11] Exchange Traded Fund.

Chapter 2

Technical Introduction

In this chapter we will go on to discuss some of the ideas proposed in Chapter 1 in greater detail, and to introduce some empirical facts about, and modifications to, the aforementioned passive investment strategies.

2.1 Diversification

In Chapter 1 we saw that diversifying ones portfolio by investing in a multitude of stocks, in general, decreases the portfolio variance, but offered no explanation as to why. Consider a portfolio of n assets. Following [Elton et al., 2009, p. 56], as the return of a portfolio is the weighted sum of the individual assets' returns, for an equal weighted portfolio, we have:

$$\begin{aligned}
\mathrm{Var}[r_p] &= \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} r_i\right] \\
&= \frac{1}{n^2}\sum_{i,j=1}^{n} \mathrm{Cov}[r_i, r_j] \\
&= \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}[r_i] + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} \mathrm{Cov}[r_i, r_j] \\
&= \frac{1}{n}\overline{\sigma_i^2} + \frac{(n-1)}{n}\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} \frac{\mathrm{Cov}[r_i, r_j]}{n(n-1)} \\
&= \frac{1}{n}\overline{\sigma_i^2} + \frac{(n-1)}{n}\overline{\Sigma_{ij}} \qquad (2.1)
\end{aligned}$$

where:

1. r_p denotes the portfolio returns;
2. $\overline{\sigma_i^2}$ the average variance across the assets; and
3. $\overline{\Sigma_{ij}}$ the average cross-covariance across the assets (i.e. excluding $i = j$).

We, therefore, see that as the number of stocks we hold increases ($n \to \infty$), the variance of the portfolio returns approaches the average cross-covariation between our assets in the portfolio. Furthermore, as most stocks are usually not too correlated, $\overline{\Sigma_{ij}}$ is often lower than most of the individual $\overline{\sigma_i^2}$. We, therefore, see a decrease in our portfolio variance from σ_1^2 towards $\overline{\Sigma_{ij}}$, the average cross-covariance[1]. In Figure 2.1 we see the theoretical volatility of an equal weighted portfolio of US stocks, according to (2.1)[2].

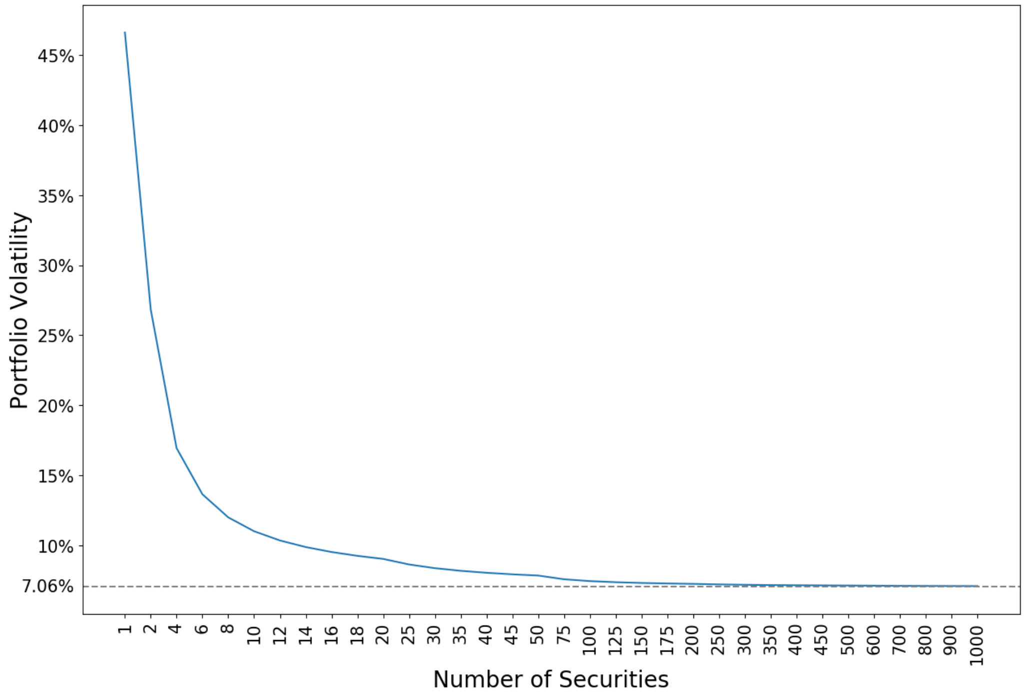


Figure 2.1: Theoretical effect of diversification on portfolio variance.
Data from [Elton et al., 2009, p. 57, Table 4.8]

2.2 Markowitz and MPT

2.2.1 Assumptions

In our toy example, Example 1.2.1, we made one (arguably two) assumption(s). The first was that you could play the game infinitely often, as long as you had the funds. The second, arguable, assumption, was that the investor/gambler is risk averse, i.e. they prefer an amount of money **with certainty** ($p = 1$) to the same amount of money **with uncertainty** (multiple outcomes, each with $p_i < 1$), i.e. in expectation.

[Markowitz, 1952] does not make the assumption of infinite trials, but does assume that the investor is risk averse. Obviously, this is not always the case. If you needed $1,000,000 to

[1] Obviously this is not always true. Consider a market of a 5-year US treasury bond, `TESLA` and `AMZN`. Obviously starting with the bond at $i = 1$ we would likely have higher volatility in holding all three than just the bond. In stock markets, however, we generally see the aformentioned behaviour.

[2] It isn't quite as nice a story for non-equally weighted portfolios, as adding a new security with low weight probably won't decrease portfolio volatility much, but the general idea is clear; adding more un/negatively-correlated stocks should decrease volatility.

stay alive, and you could choose two bets, one of $250,000 under certainty, the other with $1,000,000 gain or $500,000 loss, with equal probability. Assuming you don't want to die, you would choose the second bet, and, therefore, be risk prone, rather than risk averse. Similarly, there will exist investors who would rather take on a bit of extra risk with the possibility of achieving extra returns. From here on, we shall continue to assume risk aversion of the investor unless otherwise stated, however it is something to bear in mind.

As for the infinite trials assumption, that was more useful for our argument as to why to choose option 2. When it comes to [Markowitz, 1952], it assumes that we are considering all investments over the same, single, investment period.

2.2.2 The Efficient Frontier

Let us consider Figure 1.1 again. For convenience, we repeat it below.

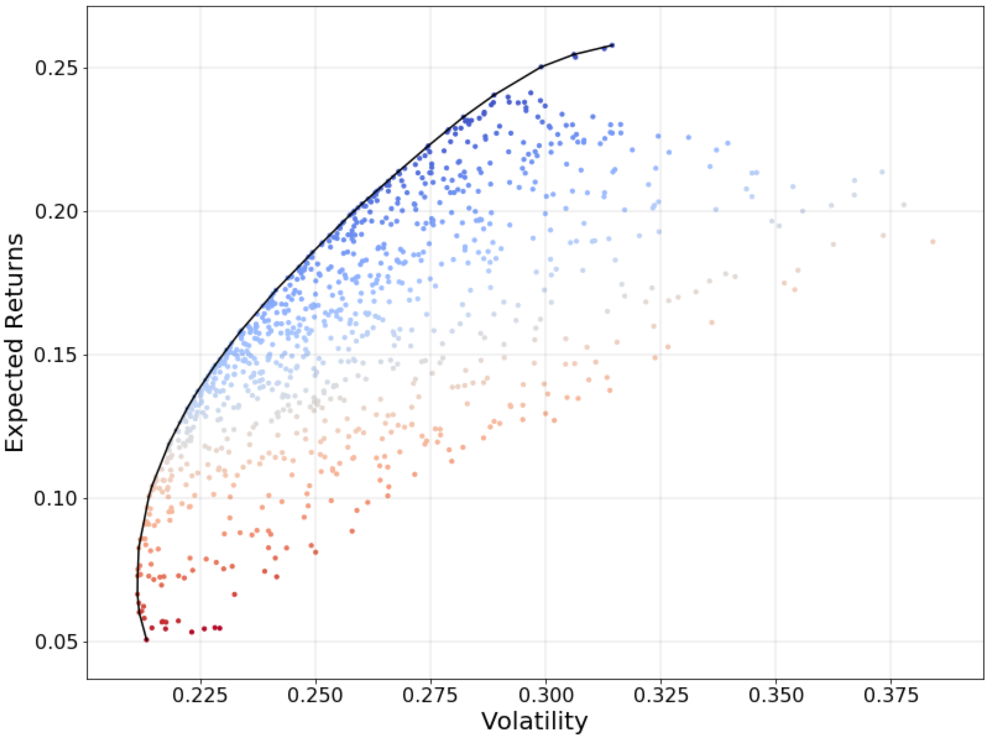


Figure 2.2

For a given volatility, if we have a portfolio that lies along the line of said volatility, we know that we **at least** want to move up said line until no observed portfolio with superior returns exists[3]. This still leaves us with two questions:

1. How do we know that there doesn't exist another combination of observed portfolios that gives the same returns as an observed portfolio, but with lower volatility?
2. What about for a volatility (or return) in between observed points? How do we know what is the optimal returns for the given volatility?

To give the reader an idea of an answer to these questions, consider Figure 2.3.

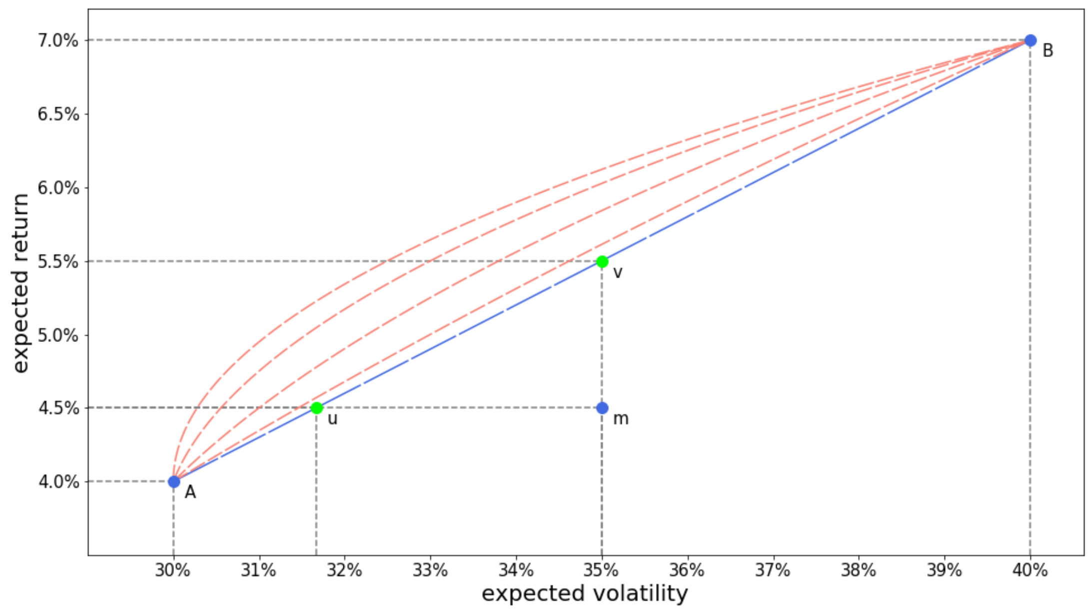


Figure 2.3: Caption

Imagine we had only observed 3 portfolios and that their return-volatility relationships were described by A, B and m. To address the first of our two questions, consider point $m = (35\%, 4.5\%)$. Given:

1. We have observed A and B, with $4\% = r_A < r_m < r_B = 7\%$; and
2. $\mathbb{E}$ is linear,

we know that we can create a new portfolio that gives the same expected returns as m, defined by holding a weighted sum of portfolios A and B.

[3]The same can be said for a given return, but moving left on our plot.

Proof.

1. $r_A < r_m < r_B \Rightarrow \exists\, t : r_m = tr_A + (1-t)r_B$
2. $\mathbb{E}[tA + (1-t)B] = t\,\mathbb{E}[A] + (1-t)\,\mathbb{E}[B] = tr_A + (1-t)r_B = r_m$

□

Now we know that we can replicate the expected returns, we need to know how our new portfolio's variance compares to that of m. Given $\text{STD}(X)$ is sub-additive[4], we have that:

$$\begin{aligned}\text{STD}[tA + (1-t)B] &\leqslant \text{STD}[tA] + \text{STD}[(1-t)B],\\ &= t\,\text{STD}[A] + (1-t)\,\text{STD}[B],\\ &= t\sigma_A + (1-t)\sigma_B,\end{aligned}$$

so in the **worst case**, our new portfolio's volatility is a linear combination of A and B's volatilities, but **could** be even lower. This tells us that, for a given return, any point to the right of the line joining A and B will be dominated in volatility by some portfolio that is a linear combination of A and B. This is shown in Figure 2.3 by the point u. Similarly, we can also see that there will exist a point between A and B that exhibits, at least, as low volatility as m, but produces greater expected returns, as shown by point v.

In answering our first question, we have also, somewhat, answered our second. We have shown that we know we can achieve return-volatility combinations at least as good as those on the upper-left side of the convex hull enclosing our points. In reality, as portfolio volatility is only a linear combination of component volatilities when we have zero covariance between the two components (A and B in our graph), we usually experience a volatility to the left of the line connecting the components, and, therefore, a frontier similar to one of the red dotted lines in Figure 2.3. In fact, Merton actually showed that the efficient frontier can always be expressed as a parabola in r and σ and derives an analytic formula in [Merton, 1972].

2.3 Capital Asset Pricing Model

Using Markowitz's work as their foundation, Treynor [Treynor, 1961, Treynor, 1962], Sharpe [Sharpe, 1964], Lintner [Lintner, 1965] and Mossin [Mossin, 1966] all independantly developed

[4]See A.1

what is now referred to as the Capital Asset Pricing Model (CAPM). As this is not the main focus of this thesis, we discuss only Sharpe's paper here.

[Sharpe, 1964] starts as an almost direct continuation of [Markowitz, 1952]. Using [Tobin, 1958] he builds on the fact that Markowitz portfolio optimisation can be split into two phases:

1. The choice of a unique optimum combination of risky assets; and
2. The choice of weight to assign to the riskless asset and the combination chosen in 1. .

On top of Markowitz's assumptions, Sharpe further assumes that:

- There exists a risk free rate (from the riskless asset), at which all investors may lend or borrow an infinite amount[5]; and
- All investors possess homogeneous views on the expected return and volatility of all assets[6].

Following [Tobin, 1958] and the same logic as in Section 2.2.2, we see that, with the introduction of a riskless asset, the choice of portfolios we wish to optimise over changes from Figure 2.4a to Figure 2.4b, where the new efficient frontier is tangent to the old one. This suggests that all investors would invest solely in a weighted combination of the risk free asset and the portfolio through which our new efficient frontier passes. Let us call the efficient frontier before the introduction of a riskless asset frontier-one and after frontier-two. Sharpe then argues that, given our assumptions, and the theory of supply-and-demand, the prices of all assets would change to reflect investors' interest or lack thereof and we would see a shift such that the new area sketched by our risk asset portfolios' return-vol combinations would have a new frontier-one that meets and is equal to a portion of the new frontier-two (i.e. frontier-one is a line segment of frontier-two), more akin to Figure 2.5.

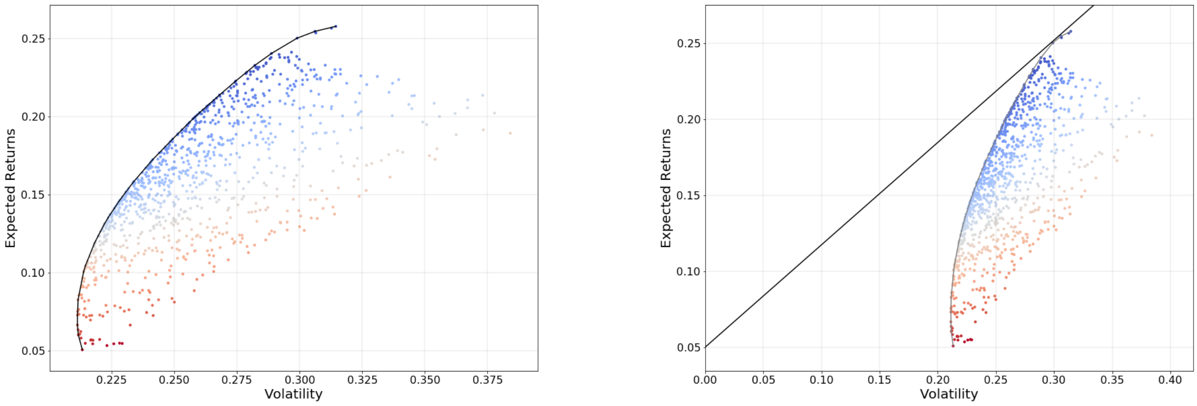
Sharpe then goes on to show that **some** portfolio consisting of all assets in the market, the market portfolio, must lie on this new efficient frontier and that investors will invest in it, along with the risk free asset. Finally, he ends up with (2.2), the Capital Asset Pricing Model

---

[5] Tobin's analysis only assumes a riskless asset exists for the particular investor in question, not a uniform rate for all investors.

[6] It is also worth noting that Sharpe somewhat implicitly assumes that:

1. All investors are price takers (i.e. they cannot influence prices);
2. There are no transaction costs;
3. All securities are infinitely divisible;
4. All information is available at the same time to all investors.



(a) No riskless asset (b) Riskless asset

Figure 2.4: Three simple graphs

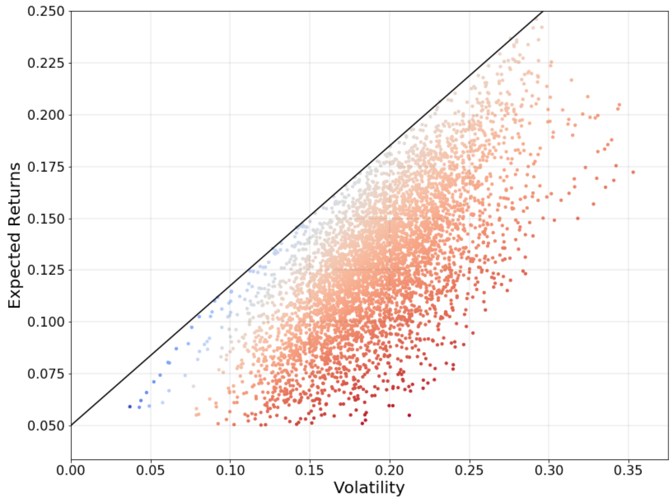


Figure 2.5: Efficient frontier at equilibrium

(CAPM) formula,

$$\mathbb{E}[r_p] = r_f + \beta_{p,m}\,\mathbb{E}[r_m - r_f], \tag{2.2}$$

where:

- r_p := portfolio returns;
- r_f := the risk free rate of return;
- r_m := the return of the market portfolio; and
- $\beta_{p,m}$:= the beta coefficient in the above regression[7].

[7]This can alternatively be written as $\rho_{p,m}\frac{\sigma_p}{\sigma_m}$, with $\rho_{p,m}$ standing for correlation between portfolio and market returns and σ_i the standard deviation of the portfolio/market returns.

2.3.1 Passive Investment Revisited

CAPM

Now that we have been better acquainted with CAPM and its formulation, let us consider one of its main implications. From the equilibrium argument, the fact that the aforementioned market portfolio lies on the efficient frontier and that it is, in fact, only this portfolio and the riskless asset that investors buy, we can argue that this must be the market-cap weighted portfolio. I think [Sun, 2003, p. 7] argues this best.

> Since every person holds the market portfolio, when aggregated across the entire market, the total holdings of each security will equal the market capitalization of that security. Thus, the percentage of each security in the market portfolio is proportional to its overall market capitalization.

Therefore, a direct result of CAPM is that, given its assumptions, everyone must hold the market-cap weighted portfolio along with the risk free asset, in a weighted amount that maximises their utility over risk and return. This is further testament as to why one would want to invest more passively than actively. We now go on to briefly discuss other rationals.

Long Term Averages

In [Sharpe, 1991], Sharpe argues that, if the market is made of two groups, passive managers and active managers[8], then **on average** active managers must experience the same return as passive managers, before costs.

Proof.

1. Asset managers can be thought of as an asset. If you give an asset manager \$x, you are effectively buying \$x of that manager, with his returns being determined by the returns he achieves for your capital.
2. We can therefore say that:

$$r_M \equiv w_P \cdot r_P + w_A \cdot r_A, \tag{2.3}$$

[8] Whether these be individual investors or institutional investors.

where:

- w_i is the % of capital invested in the market manager type i, $i \in \{A, P\}$
- A, P indicate active and passive managers respectively.

3. As passive managers invest in the market portfolio, we have that:

$$r_P \equiv r_M. \tag{2.4}$$

4. Subbing (2.4) into (2.3) we have that:

$$r_M = w_P \cdot r_M + w_A \cdot r_A.$$

5. Finally, collecting terms in r_M and given we assert that all managers are active or passive, i.e. $w_A + w_P = 1$, we have that:

$$\begin{aligned}(1 - w_P) \cdot r_M &= w_A \cdot r_A, \\ &= (1 - w_P) \cdot r_A,\end{aligned}$$

and find that $r_M = r_A$.

□

He goes on to argue that the cost of actively managing must be greater than that of passively managing as more time must be invested in research (which is paid for through employee wages) and in trading through transaction costs. He summarises that, given the before cost returns must be equal and the cost of actively managing should be greater than that of passively managing, the returns of active managers must be lower than those of passive managers, **on average**, after costs.

Efficient Market Hypothesis

The efficient market hypothesis (EMH) states that all market prices accurately reflect all available information, i.e. if you think a stock is under/overvalued, you are not pricing it correctly, either through an incorrect formula or through lack of information. A direct result of this, if true, is that one cannot consistently beat the market on a long term basis, and that additional risk-adjusted returns are only available through inside information.

Evidence

It is all well and good stating "under these assumptions, then we have that...", but as we know, many assumptions about financial markets do not hold in real life. Instead of checking the validity of the assumptions (which is very difficult), we take a look at how active managers perform vs their relevant index. [Soe and Poirier, 2016, p. 1] claims that over the whole of 2016, 84.6%, 87.9% and 88.8% of large, mid and small-cap managers underperformed the S&P 500, S&P MidCap 400 and S&P SmallCap 600 respectively. It also states that, on a 5-year time horizon, 91.9%, 87.9% and 97.6% underperformed their respective benchmarks. It is a similarly bleak story for the 10-year period (85.4%, 91.3%, 90.8%).

It is important to note, however, that this analysis is on a per-fund basis. They do not, therefore, follow the suggestions of [Sharpe, 1991] and consider a per-dollar managed approach, however his suggestion to do so was to avoid bias towards evidence **against** active managers underperforming.

In summary, although we have not shown that active managers definitely do/do-not underperform their relevant benchmarks, there is certainly enough testament toward their underperformance as to warrant further research into passive investment strategies.

2.3.2 Risk Premia

Let us, again, consider (2.2), and in particular the sign of $\mathbb{E}[r_m - r_f]$. As holding the market portfolio inherits more risk than holding the risk free asset (almost by definition), the holder must be compensated for this risk. If they were not, no one would hold any risky assets, solely the risk free asset and, by the law of supply and demand, the market would arrive at a state where you were once again compensated for holding risky assets. How large this compensation is depends on some average market view of expected market returns and some average appetite for risk. As we expect that $\mathbb{E}[r_m - r_f] > 0$, we deem $\mathbb{E}[r_m - r_f]$ to be a **risk premium**, i.e. a premium paid to the holder for taking on risk. It is worth noting that only undiversifiable risk is rewarded in such a way. As market risk (a.k.a. systematic risk) is undiversifiable (by definition all stocks are susceptable to market risk, therefore we cannot diversify it away), it is rewarded. To see a risk that is **not** compensated by some risk premium, consider the reformulation of (2.2) to:

$$r_p = r_f + \beta_{p,m}[r_m - r_f] + \epsilon_p,$$

where all variables mean the same as before, with the introduction of ϵ_p as a 0 mean random variable, explaining the difference in portfolio returns against expected portfolio returns. Under the CAPM model, ϵ_p is the idiosyncratic risk that can be diversified away/hedged[9], and is, therefore, not rewarded, due to ones ability to remove it.

2.4 Factor Models

Despite CAPM's widespread popularity, there is a large amount of empirical evidence that its results do not stand up to reality[10]. Due to this, many other researchers propose alternative models/frameworks. We discuss one of such here, but before doing so, we must introduce the notion of a factor model.

Definition 2.1 (Factor Model)**.** *Suppose we have a set of p observable random variables, $x_1, \ldots x_p$, each with mean $\mu_1, \ldots \mu_p$. We say the x_i follow a **factor model** if, for some unknown $\beta_{i,j}$ and unobservable $\mathcal{F}_j$, $i \in \{1, \ldots, p\}$, $j \in \{1, \ldots, k\}$, $k < p$, where:*

1. *$\mathcal{F}$ and ε are independent;*
2. *$\mathbb{E}[\mathcal{F}] = 0$; and*
3. *$\text{Cov}[\mathcal{F}] = I$,*

we have that:

$$x_i = \mu_i + \sum_{j=1}^{k} \beta_{i,j} \mathcal{F}_j + \varepsilon_i.$$

2.4.1 Arbitrage Pricing Theory

Due to the restrictions CAPM imposes in its formulation, Stephen Ross decided to propose his own model with much more general conditions to arrive at the same result. [Ross et al., 1972], [Ross et al., 1973] and [Ross, 1976] hypothesise and develop the idea of an arbitrage based model for predicting asset returns. Unlike CAPM, APT assumes that markets sometimes misprice assets slightly before they revert back to their true price. We don't discuss the specifics here, but, in general, Ross uses an arbitrage based argument to show that, under the assumption that a single-factor model can be used to explain asset returns, in order to avoid permanent arbitrage opportunities, we must have "in all but the most profound sort of disequilibria" a similar solution to that of CAPM.

[9] Remember, under CAPM, all returns are explained by the market returns plus some random variable unrelated to any other common factor across all stocks.

[10] For examples, see [Fama and French, 2004].

He continues by suggesting that, rather than a single factor model like CAPM, asset returns should be explained by a multi-factor model, such as:

$$r_i = \alpha_i + \sum_{j=1}^{k} \beta_{i,j} \mathcal{F}_j + \varepsilon_i,$$

and that this leads to (2.5).

$$\mathbb{E}[r_i] = r_f + \sum_{j=1}^{k} \beta_{i,j} \left(\mathbb{E}[\mathcal{F}_j - r_f] \right) \tag{2.5}$$

It is worth noting that, unlike CAPM, APT tells us nothing about the sign of each factor's excess return $\mathbb{E}[\mathcal{F}_j - r_f]$ and, as such, $\mathbb{E}[\mathcal{F}_j - r_f]$ should not be considered a risk premium under this framework[11].

A key takeaway from APT is that, in theory, we can now explain asset returns as a combination of returns of different macro-economic factors. For a company like Amazon, these might be market returns, tech stock returns and public sentiment, for example.

2.4.2 Fama-French 3-Factor Model

As a direct result of APT, many new multi-factor models were proposed in the years after its publication. The first, widely accepted, model was that of Fama and French in [Fama and French, 1992]. Before we discuss it's details, we first introduce the reader to the concept of company reporting.

[International Accounting Standards Board, 2007] of the International Accounting Standards (IAS), as set out by the International Accounting Standards Board (IASB) states,:

> The objective of general purpose financial statements is to provide information about the financial position, financial performance, and cash flows of an entity that is useful to a wide range of users in making economic decisions. To meet that objective, financial statements provide information about an entity's: [IAS 1.9]
>
> - assets
> - liabilities
> - equity
> - income and expenses, including gains and losses

[11] We use the term excess, as is common practice when talking about returns other than of the risk free asset, but be aware that excess may mean excess in the positive or negative direction.

- contributions by and distributions to owners (in their capacity as owners)
- cash flows.

> That information, along with other information in the notes, assists users of financial statements in predicting the entity's future cash flows and, in particular, their timing and certainty.

In more general terms, financial statements are there to give their readers an idea of how a company is performing financially[12]. Within these statements, there are a number of values that are reported, e.g. net-income, revenues, total liabilities. For large companies, the list of reported items (a.k.a. fundamentals) is so vast it is not worth delving into here. Instead we introduce one of the most famous items (or values derived from these items), the book-value of a company.

Definition 2.2 (Book-value)**.** *The* ***book-value*** *of a company is defined to be the net asset value of a company, less any intangible assets (e.g. good faith agreements) and liabilities it has.*

In more general terms, it is the value of everything a company owns, minus the value of everything it owes. Now that we understand the notion of book-value, we can discuss one of the foundations of this study.

It was a well known fact that CAPM failed to adequately explain asset returns for portfolios consisting of small/large stocks, and of portfolios consisting of high/low value (high/low book-to-market ratio) stocks. It tended to underestimate returns for small or high-value stocks and overestimate them for big or low-value stocks. Using this idea, [Fama and French, 1992] shows that CAPM does not adequately explain asset returns and that, on top of market returns, one can also use the excess returns of small market-cap stocks over large market-cap stocks and of high book-to-market ratio stocks over low book-to-market ratio stocks to do so. [Fama and French, 1993] proposes the model described in (2.6) as a means to explain asset returns over the traditional CAPM. Rather than taking an equilibrium based approach as those who suggested CAPM did, Fama and French based their findings purely off of empirical

[12] This promotes transparency and helps to stop many crimes, such as fraud. As a companies stock price is related to its public perception, and its public perception often to its financial performance, a companies financial reports can substantially influence a securities price. Many investors also believe that these reports contain information that the market does not fully take into account.

statistics.

$$r_p = r_f + \beta_{mkt} \left[r_m - r_f \right] + \beta_{SMB} \cdot SMB + \beta_{HML} \cdot HML + \varepsilon. \quad (2.6)$$

$$\begin{aligned} r_p &:= \text{portfolio returns.} \\ r_f &:= \text{risk free rate.} \\ r_m &:= \text{market returns.} \\ SMB &:= \text{returns of small stocks over big stocks.} \\ HML &:= \text{returns of high value stocks over low value stocks.} \\ \varepsilon &:= 0 \text{ mean error term.} \end{aligned} \quad (2.7)$$

For a full definition of SMB and HML, see [Fama and French, 1993] or Section 4.1.3.

Throughout [Fama and French, 1992], [Fama and French, 1993] and [Fama and French, 1995] it is shown that the Fama-French 3-Factor model from (2.6) better explains cross sectional asset returns than CAPM, with their 3-factor model having 21 of 25 portfolios with adjusted-R^2 above 0.9, but only 2 of 25 using CAPM [Fama and French, 1993, p. 19-25].

2.4.3 Carhart

Another popular factor based model is the Carhart 4-Factor model, which, on top of the Fama-French 3-Factor model, includes a momentum factor, given previous high performers tend to outperform previous low performers.

$$r_i = \alpha_i + \beta_i^{mkt}[r_m - r_f] + \beta_i^{SMB} r_{SMB} + \beta_i^{HML} r_{HML} + \beta_i^{UMD} r_{UMD} + \varepsilon_i \quad (2.8)$$

2.5 Why factor investing?

At this point, it is natural for one to ask, so why do we care about risk factors? If they're just compensation for risk, we're not really gaining anything are we? Well, the facts are these; in good market conditions, indexes based on these extra risk factors **do** tend to outperform the market-cap index. In bad conditions, however, they tend to underperform the market, [Ang, 2014, p. 444]. See [Ang, 2014, p. 450] for examples. In general, however, markets seem to grow and have longer periods of strength than of weakness. In the long run then, it would make sense that the above market returns from market growth periods might cancel out the below market returns from market declines and beat the market index. In fact, this is exactly what we have seen. Since 1973, there have been multiple periods that factor indexes[13] have underperformed the market. Overall, however, $1 invested in the MSCI World index from 1973-2015 would have risen to $34, whereas $1 invested in their value index would have risen to $49, or $98 in their momentum index[14], [Authers, 2015]. In the long run then, it seems that the benefits outweigh the costs, for now at least.

Given these factors historical outperformance of the market, a new type of passive investing appeared. Rather than investing in the whole market weighted by market cap, investors decided to encorporate these findings in some way, selecting subsets of the market to invest in. based on these factors (factor investing), or using alternative weighting systems to market cap (smart-beta investing). The benefits of these methods are similar to that of passive investing:

- Large investment capacity - due to investing in indexes, the market cap of your universe is very large. It would, therefore, take an extreme amount of capital to move your market in some way (i.e. not to be a price taker). This is very attractive to large funds, e.g. pension schemes, as many smaller strategies do not scale well when dealing with portfolio values in the high millions/billions;
- Low costs - as these methods are quite simple and can be readily automated, little effort is required to execute them, thus, costs are low in terms of both research and execution; and
- Diversification - as these methods are based on index investing, we still experience, given a large enough universe to invest over, very good amounts of diversification amongst stocks.

Furthermore, despite the point that we argued earlier about actively managed dollars under-performing passively managed dollars, on average, given the strong historical performance of these methods, one would hope that they would shift you into the subset of active managers who **do** beat their passive benchmark.

Finally, coming full circle back to Markowitz, [Clarke et al., 2005] shows that, with the addition of factor investment strategies, one can expand the efficient frontier and push/rotate it north-west, thus offering higher returns for the same level of risk.

[13]Indexes constructed with weights according to some risk factor, e.g. the value factor.

[14]These are the equivalent of roughly 8.8%, 9.7% and 11.5% compound annual returns respectively.

Given the topics discussed in this section, it is clear that research into factor investing, and specifically ways to access these factors could be extremely rewarding.

Summary 2

Factor investing is an easy, cheap way to, in the long run, gain excess returns on top of market returns, whilst offering a variety of different investment options.

Chapter 3

Problem Formulation

3.1 Foundation of this study

Now that we have discussed why it is worth researching factor investing, we discuss the particulars of this study. As with most things in finance, if you know the value of something before the market, there is an opportunity to exploit this knowledge and profit from it. Furthermore, many academics and institutionalists alike have begged the question *"why book-to-market"* when determining a companies value. What if another combination of fundamentals better encapsulates the idea of a companies value? Using these ideas, [Alberg and Lipton, 2017] looked into following Fama-French-esque methods but based on future values of book-value (and other financial reporting information). Although slightly more complicated[1], their general method is, each year:

1. Rank all stocks on the NYSE, AMEX and NASDAQ exchanges based on some ratio of fundamentals (e.g. book-value to market-value)[2]; and
2. Buy those that fall in the top 50 stocks and hold them for 1 year.

They then suggest to imagine that we know the value of said ratio x-months in the future, exactly, and rank on that instead, do we experience higher returns than using current data? Their results show that, for book-to-market, EBIT-to-EV, net-income-to-EV and sales-to-EV, the further ahead we know their values, the higher returns we achieve. The greatest gain using future information is for EBIT-to-EV, increasing from 14.4% compound annual returns (CAR) at 0-clairvoyance (i.e. using today's data) to around 70% CAR at 3-years-clairvoyance. As such, they decide to try to predict the future value of EBIT-to-EV, but at a 1-year-clairvoyance, where there was a clairvoyance-based CAR of 44%. In doing so they managed to attain a consistently superior mean-squared-error in their modelling than a naïve model of assuming $x_{t+k} = x_t$. Finally, they find that, using their model, their CAR increased from 14.4% to 17.1%, with

[1] It also takes into account reinvesting dividends, funds from acquisitions etc.

[2] They exclude all financial sector companies, companies not based in the US and all with a inflation adjusted market-cap below $100 million. The final list contains 11,815 stocks.

Sharpe ratio increasing from 0.55 to 0.68. Although they don't attain the hypothetical upper limit of 44% CAR with perfect future knowledge, even a 2.7% increase is quite remarkable, using only publicly available, relatively cheap, data. Given such an improvement in CAR and Sharpe ratio is attainable using publicly available, easily attained data[3], we thought that with particularly relevant proprietary data, one should be able to access even more of these excess returns through superior models.

3.2 Our data

Our universe consists of 31 automotive companies, has a total market cap of approximately $1 trillion, covers companies based in 8 different countries and captures roughly 91% of global car sales. We choose this set due to the proprietary dataset that IHS Markit (IHSM) owns, containing, hopefully, relevant information for predicting fundamentals ratios. The exact list of companies we consider can be found in B.2.

Fundamentals For our fundamentals data, we use the Thomson Reuters Worldscope database, including its FX data to convert all values to USD.

Production data For our proprietary data, thanks to IHS Markit (IHSM), we have monthly reports of a variety of sales/production based metrics for each of 31 automotive companies. It contains information on sales volumes, production volumes, production plant utilisations, fleet ages, and market shares, among others. For more information, please contact IHS Markit.

Before continuing, please take the time to read the declaration about the IHSM data in Section B.1.

3.3 Testing pipeline

Our research pipeline consists of three main sections:

1. Stage 1 - factor identification : we need to identify if there are any ratios of fundamentals which, if we know them in advance, give us excess returns on top of those we experience without knowledge of the future.

[3]It can actually be downloaded for free if you have an affiliation with a university such that you have a .ac, .edu or similar email address.

2. Stage 2 - factor modelling/prediction : if there are any factors which, given knowledge of the future, give us excess returns, we need to come up with a model that beats the naïve predictor of assuming no change over our prediction horizon.
3. Stage 3 - model backtesting : given a suitable model, we need to run a backtest to construct our portfolios based on our model's predictions and assess its performance.

We now discuss each of these in details.

3.3.1 Stage 1 - Factor identification

Data For the factor identification portion of our research, we use only the Worldscope fundamentals data. As reports are made in the parent company's home country's currency, we use the databases daily FX rates to convert all report values to USD. Furthermore, as not all companies adhere to the same reporting frequency, varying from annual to quarterly reporting, we scale each report value as to be comparable on a quarterly basis. E.g. if a report stated revenues of $1,000,000 over the previous year, we divide it by 4 to obtain $250,000 as a rough estimate for its last quarter results[4]. Any missing data values are forward filled from the last known value. As we might be rebalancing more frequently than each company's reporting frequency, we construct a monthly timeseries by forward filling last known values, e.g. say we were rebalancing on 01/03/18 but the last BMW report was on 24/01/18, we would use the values from 24/01/18 as the 01/03/18 values.

Process The general process for stage 1 is as follows:

1. Rank our universe on some **ranking factor** (e.g. book to market).
2. Buy those with a ranking factor above some threshold. If we are shorting, short all those with a value below some threshold, e.g. go long the top 20% and short the bottom 20%, when ranked on book-to-market.
   - When longing/shorting assets, we do so according to some weighting rule, e.g. buy/short an equal cash value of each asset in the portfolio.
3. Hold these 1/2 portfolios for some period, e.g. for 1 month.
4. Repeat 1-3 over backtest period.

[4] An alternative approach could be to sum all data on a rolling, trailing 12 months to convert everything to yearly, but due to the nature of our proprietary data and the information advantage we have, we believe it more fruitful to look at the highest possible frequency.

For our backtests, we have a number of parameters to search over, which we discuss here.

- `ranking_factor` - which ratio we are using in our ranking, e.g. book-to-market. For a list of factors used, see the next section.
- `quantiles` - how many quantiles to split our ranking into before buying the top quantile and, possibly, shorting the bottom. We test over the set $\{3, 4, 5\}$, i.e. long/short the top/bottom 33%, 25%, 20% respectively.
- `shorting` - whether we are shorting the bottom quantile or not; `TRUE` or `FALSE`
- `holding_period` - how long we hold the assets for before rebalancing. We test over {monthly, quarterly, yearly} or $\{\texttt{M}, \texttt{Q}, \texttt{Y}\}$.
- `weighting` - what weighting rule we are using. We test over {market-cap weighted, equal-weighted } or $\{\texttt{M}, \texttt{E}\}$.

Example 3.3.1 (Portfolio Construction)**.** *Imagine we are ranking based on book-to-market (BM), not shorting the bottom portfolio, longing the top 20%, rebalancing monthly and weighting according to market cap (MC). Imagine it is 01/08/18 and our universe is as follows:*

Company	BM	Stock Price ($)	MC ($)
BMW	1.45	24.40	6.10E+10
VW	1.35	36.80	9.20E+10
TESLA	0.68	4.00	1.00E+10
FIAT CHRYSLER	1.49	20.00	5.00E+10
KIA	0.71	12.84	3.21E+10
HYUNDAI	0.80	8.96	2.24E+10
FORD	1.40	28.92	7.23E+10
GM	1.36	21.08	5.27E+10
HONDA	0.73	42.12	1.05E+11
MITSUBISHI	0.69	80.00	2.00E+11

1. *Rank the stocks and take top 20%.*

Company	BM	Stock Price ($)	MC ($)
FIAT CHRYSLER	1.49	20.00	5.00E+10
BMW	1.45	24.40	6.10E+10
FORD	1.40	28.92	7.23E+10
GM	1.36	21.08	5.27E+10
VW	1.35	36.80	9.20E+10
HYUNDAI	0.80	8.96	2.24E+10
HONDA	0.73	42.12	1.05E+11
KIA	0.71	12.84	3.21E+10
MITSUBISHI	0.69	80.00	2.00E+11
TESLA	0.68	4.00	1.00E+10

2. *Create long portfolio from selected stocks, weighted by proportion of market cap.*

Company	BM	Stock Price ($)	MC ($)	Weight
FIAT CHRYSLER	1.49	20.00	5.00E+10	0.45
BMW	1.45	24.40	6.10E+10	0.55

Imagine the universe on 03/09/18 (the first weekday in September) is:

Company	BM	Stock Price ($)	MC ($)
BMW	1.89	24.63	6.16E+10
VW	1.28	35.13	8.78E+10
TESLA	0.87	4.88	1.22E+10
FIAT CHRYSLER	1.31	21.64	5.41E+10
KIA	0.87	6.52	1.63E+10
HYUNDAI	0.86	4.25	1.06E+10
FORD	1.33	31.86	7.97E+10
GM	1.73	25.91	6.48E+10
HONDA	0.94	46.58	1.16E+11
MITSUBISHI	1.06	72.54	1.81E+11

Given Fiat Chrysler went from $20.00 to $21.64 and BMW from $24.40 to $24.63, and that portfolio returns over a period are the value weighted sum of the individual returns, our portfolio returns for 01/08/18 to 03/09/18 are $0.45 \times (\frac{21.64}{20.00} - 1) + 0.66 \times (\frac{24.63}{24.40} - 1) = 0.45 \times 8.2\% + 0.66 \times 0.9\% = 4.2\%$. *We then repeat the above steps again, moving forward in time.*

Clairvoyance When it comes to claivoyantly using future data in our stage 1 backtest, we first shift our fundamental data backwards by three months (e.g. 03/09/18 would map to 01/06/18) before applying identical methods as when there is no clairvoyance (those explained above). We do **not** use future values of market-cap when weighting by market-cap, as this would require us to forecast future market-cap, which is essentially forecasting future stock prices, rendering our research somewhat redundant.

Performance metrics Although the main metrics we care about are compound annual returns (CAR) and annualised vol, there are a number of other metrics we use to give evidence that the results are/aren't due to some artifact/anomaly in the data. We present a list of all metrics below:

- `CAR` - compound annualsed returns over the backtest period.

- annualised_vol - daily return volatility scaled to annual volatility, measured as STD[daily portfolio returns] $\times \sqrt{252}$. We annualise daily volatility as measuring volatility from annual returns directly is likely to under estimate.
- sharpe_ratio - CAR / annualised_vol.
- avg_monthly_spread - arithmetic average of difference in monthly returns between the top quantile portfolio and bottom quantile portfolio. The thought behind this is that if a strategy produces high returns, but this spread is low, the returns are probably not due to the ranking factor. Note that this is the monthly returns of the "top" and "bottom" portfolios we are holding, not of each months top and bottom portfolios (e.g. if we are rebalancing yearly, the stocks in these portfolios stay the same over each year).
- std_monthly_spread - STD[monthly spread]
- avg_monthly_spearman - the average monthly Spearman correlation coefficient between the rankings induced by the ranking factor and by the following month's returns. Averaging is done via transformation using Fisher z-transform, arithmetic averaging, then inverse Fisher z-transformation, following the methodology presented in [Corey et al., 1998]. Similarly to avg_monthly_spread, if the ranking factor gives us some sort of information about future returns, then we would expect this number to be higher than if not. The note mentioned with regards to monthly spread is also relevant here, the ranking due to the ranking factor is induced at the start of the rebalancing period, not on a month by month basis.
- std_monthly_spearman - STD[monthly Spearman]

Ranking factors used Given we are following many of the same methods outlined in [Alberg and Lipton, 2017], we decide to include the 4 factors they tested in our analysis, namely:

- Book-to-Market (Book-Value to Market-Cap)
- EBIT-to-EV
- Net-Income-to-EV
- Sales-to-EV

[Yan and Zheng, 2017] also performs a through data mining based search of ratios of functions of fundamentals. As such, we take some of the factors they deem to be highly performant and include them in our analysis. For the full list of ratios of (functions of) fundamentals tested in this study, please see Section B.3.2. For a full descriptions of each of the financial reporting items, please see Section B.3.1.

Time frame Due to data restrictions, our backtests will run from 2008-01-01 to 2017-03-01.

3.3.2 Stage 2 - Factor modelling/prediction

As it makes more sense to describe our methods for factor modelling once we know which ratios of fundamentals it is that we'll be modelling, we do so in the introduction to Chapter 5.

3.3.3 Stage 3 - model backtesting

Once we decide on a suitable model for the ratio we wish to predict, we generate a timeseries of ratio predictions for each company. If any missing values occur, due to lack of input features etc., we fill them with the most recent **known** value for that ratio. E.g. say we were predicting 3-month ahead book-to-market, y_{t+3}, but we didn't have a value for input x_t^1, we would use the current book-to-market ratio, y_t, to fill the missing datapoint.

Once we generate this timeseries, we follow the same process as described for Stage 1, as though it were the Worldscope data.

3.4 Tools

Our research is conducted in `Python`, and, in particular, using the `itertools`, `matplotlib`, `multiprocessing`, `numpy`, `pandas`, `scipy` and `sklearn` packages, the contributors of which we would like to give thanks to.

Chapter 4

Stage 1 - Factor Identification

4.1 No Clairvoyance

Out of our metrics, as the main two that investors care about are CAR and SR, we present a table of the top performing parameter combinations according to each, for each rebalancing frequency.

R	W	S	Q	ranking_factor	r	σ_r	$\overline{\rho}$	$\overline{\Delta}$	SR	σ_ρ	σ_Δ
Y	E	L	5	inventory_to_ev	0.408	0.2617	0.0481	0.0087	1.56	0.212	0.0764
Y	E	L	4	accpayable_to_lagged_sales	0.402	0.2701	0.0449	0.0215	1.49	0.209	0.0703
Y	E	L	5	equity_to_ev	0.401	0.2601	0.0495	0.0116	1.54	0.197	0.0809
Y	E	L	4	accpayable_to_lagged_cogs	0.399	0.2670	0.0352	0.0163	1.50	0.197	0.0707
Y	E	L	4	inventory_to_ev	0.399	0.2561	0.0487	0.0092	1.56	0.215	0.0693
Y	E	L	5	accpayable_to_lagged_cogs	0.390	0.2823	0.0273	0.0143	1.38	0.186	0.0748
Y	E	L	5	totassets_to_ev	0.379	0.2606	0.0333	0.0061	1.46	0.226	0.0789
Y	E	L	4	sales_to_ev	0.371	0.2621	0.0389	0.0108	1.41	0.205	0.0648
Y	E	L	3	accpayable_to_lagged_cogs	0.365	0.2536	0.0353	0.0149	1.44	0.199	0.0619
Y	E	L	5	accpayable_to_lagged_sales	0.362	0.2894	0.0318	0.0190	1.25	0.204	0.0786
Q	E	L	5	inventory_to_ev	0.277	0.2850	0.0476	0.0094	0.97	0.211	0.0728
Q	E	L	5	inventory_to_totassets	0.263	0.2730	0.0342	0.0095	0.96	0.208	0.0738
Q	E	L	5	accpayable_to_lagged_sales	0.249	0.2969	0.0205	0.0154	0.84	0.208	0.0800
Q	E	L	4	accpayable_to_lagged_sales	0.239	0.2800	0.0173	0.0123	0.85	0.196	0.0718
Q	E	L	5	totassets_to_ev	0.225	0.2756	0.0243	0.0058	0.82	0.204	0.0773
Q	E	L	4	totassets_to_ev	0.220	0.2613	0.0143	0.0052	0.84	0.197	0.0667
Q	E	L	4	inventory_to_ev	0.220	0.2798	0.0344	0.0074	0.79	0.209	0.0631
Q	E	L	5	inventory_to_currliab	0.216	0.2593	0.0242	0.0040	0.83	0.206	0.0654
Q	E	L	5	accpayable_to_lagged_cogs	0.208	0.2909	-0.0030	0.0076	0.71	0.191	0.0734
Q	E	L	4	accpayable_to_lagged_cogs	0.207	0.2784	0.0070	0.0098	0.74	0.191	0.0715
M	E	L	5	inventory_to_totassets	0.256	0.2711	0.0294	0.0059	0.95	0.224	0.0771
M	E	L	5	totassets_to_ev	0.228	0.2745	0.0251	0.0065	0.83	0.196	0.0742
M	E	L	5	inventory_to_ev	0.228	0.2814	0.0344	0.0051	0.81	0.212	0.0776
M	E	L	4	inventory_to_ev	0.225	0.2750	0.0312	0.0066	0.82	0.206	0.0667
M	E	L	4	totassets_to_ev	0.210	0.2633	0.0208	0.0073	0.80	0.197	0.0685
M	E	L	4	accpayable_to_lagged_sales	0.204	0.2777	0.0258	0.0106	0.73	0.202	0.0747
M	E	L	4	accpayable_to_lagged_cogs	0.202	0.2762	0.0074	0.0088	0.73	0.196	0.0729
M	E	L	5	accpayable_to_lagged_cogs	0.201	0.2894	0.0050	0.0097	0.70	0.189	0.0736
M	E	L	5	inventory_to_currliab	0.200	0.2567	0.0144	0.0009	0.78	0.212	0.0654
M	E	L	5	cogs_to_ev	0.199	0.3118	0.0213	0.0075	0.64	0.206	0.0765

Table 4.1: Each of the top 10 performing parameter combinations for different holding/rebalancing periods, according to CAR, all with 0-clairvoyance (using known data).

As we had to abbreviate the column headings to allow the table to fit, here are their full names and contents:

1. R - rebalancing/holding period;
2. W - weighting;
3. S - shorting;
4. Q - number of quantiles used to split ranking
5. r - `CAR`
6. σ_r - `annualised_vol`
7. $\overline{\rho}$ - `avg_monthly_spearman`
8. $\overline{\Delta}$ - `avg_monthly_spread`
9. SR - `sharpe_ratio`
10. σ_ρ - `std_monthly_spearman`
11. σ_Δ - `std_monthly_spread`

R	W	S	Q	ranking_factor	r	σ_r	$\overline{\rho}$	$\overline{\Delta}$	SR	σ_ρ	σ_Δ
Y	E	L	5	inventory_to_ev	0.408	0.262	0.0481	0.0087	1.56	0.212	0.0764
Y	E	L	4	inventory_to_ev	0.399	0.256	0.0487	0.0092	1.56	0.215	0.0693
Y	E	L	5	equity_to_ev	0.401	0.260	0.0495	0.0116	1.54	0.197	0.0809
Y	E	L	4	inventory_to_currliab	0.359	0.234	0.0421	0.0106	1.53	0.217	0.0589
Y	E	L	4	accpayable_to_lagged_cogs	0.399	0.267	0.0352	0.0163	1.50	0.197	0.0707
Y	E	L	4	accpayable_to_lagged_sales	0.402	0.270	0.0449	0.0215	1.49	0.209	0.0703
Y	E	L	5	totassets_to_ev	0.379	0.261	0.0333	0.0061	1.46	0.226	0.0789
Y	E	L	5	inventory_to_currliab	0.356	0.247	0.0381	0.0099	1.44	0.209	0.0651
Y	E	L	3	accpayable_to_lagged_cogs	0.365	0.254	0.0353	0.0149	1.44	0.199	0.0619
Y	E	L	3	equity_to_ev	0.338	0.236	0.0408	0.0073	1.43	0.216	0.0666
Q	E	L	5	inventory_to_ev	0.277	0.285	0.0476	0.0094	0.97	0.211	0.0728
Q	E	L	5	inventory_to_totassets	0.263	0.273	0.0342	0.0095	0.96	0.208	0.0738
Q	E	L	5	accpayable_to_lagged_sales	0.249	0.297	0.0205	0.0154	0.84	0.208	0.0800
Q	E	L	4	accpayable_to_lagged_sales	0.239	0.280	0.0173	0.0123	0.85	0.196	0.0718
Q	E	L	5	totassets_to_ev	0.225	0.276	0.0243	0.0058	0.82	0.204	0.0773
Q	E	L	4	totassets_to_ev	0.220	0.261	0.0143	0.0052	0.84	0.197	0.0667
Q	E	L	4	inventory_to_ev	0.220	0.280	0.0344	0.0074	0.79	0.209	0.0631
Q	E	L	5	inventory_to_currliab	0.216	0.259	0.0242	0.0040	0.83	0.206	0.0654
Q	E	L	5	accpayable_to_lagged_cogs	0.208	0.291	-0.0030	0.0076	0.71	0.191	0.0734
Q	E	L	4	accpayable_to_lagged_cogs	0.207	0.278	0.0070	0.0098	0.74	0.191	0.0715
M	E	L	5	inventory_to_totassets	0.256	0.271	0.0294	0.0059	0.95	0.224	0.0771
M	E	L	5	totassets_to_ev	0.228	0.274	0.0251	0.0065	0.83	0.196	0.0742
M	E	L	5	inventory_to_ev	0.228	0.281	0.0344	0.0051	0.81	0.212	0.0776
M	E	L	4	inventory_to_ev	0.225	0.275	0.0312	0.0066	0.82	0.206	0.0667
M	E	L	4	totassets_to_ev	0.210	0.263	0.0208	0.0073	0.80	0.197	0.0685
M	E	L	4	accpayable_to_lagged_sales	0.204	0.278	0.0258	0.0106	0.73	0.202	0.0747
M	E	L	4	accpayable_to_lagged_cogs	0.202	0.276	0.0074	0.0088	0.73	0.196	0.0729
M	E	L	5	accpayable_to_lagged_cogs	0.201	0.289	0.0050	0.0097	0.70	0.189	0.0736
M	E	L	5	inventory_to_currliab	0.200	0.257	0.0144	0.0009	0.78	0.212	0.0654
M	E	L	5	cogs_to_ev	0.199	0.312	0.0213	0.0075	0.64	0.206	0.0765

Table 4.2: Each of the top 10 performing parameter combinations for different holding/rebalancing periods, according to SR, all with 0-clairvoyance (using known data).

From Table 4.1 and Table 4.2, we can clearly see that the optimal holding period for our factor based strategies corresponds to yearly rebalancing. Unfortunately, investors are rarely willing to be locked into holding stocks for such long periods of time without continual assurance that it is correct to be doing so. As such, we only devote further investigative efforts to parameter combinations with a quarterly or monthly rebalancing period.

4.1.1 Benchmarks

Whilst it is all nice and well being able to say "we can achieve m% returns doing x, y, z", a strategy is only as good as its performance against a benchmark. The first benchmark we will be using to assess our strategies is the market-cap weighted portfolio across all our stocks, similar to the `CARZ` ETF. The second benchmark we will be using is an equal weighted portfolio across all of our stocks. Their performance is summarised in Table 4.3.

R	W	r	σ_r	SR
D	M	0.009	0.265	0.033
M	M	0.008	0.264	0.032
Q	M	0.016	0.260	0.062
Y	M	0.004	0.252	0.017
D	E	0.120	0.225	0.534
M	E	0.100	0.224	0.444
Q	E	0.115	0.219	0.525
Y	E	0.126	0.231	0.545

Table 4.3: Summary of benchmark performance metrics.

From our benchmarks, if we decide to be harsh on ourselves and compare our strategies to the best benchmark across all weighting-rebalancing combinations, the numbers to beat are 12.6% CAR and 0.545 SR, both from the equal weighted, yearly rebalancing benchmark. We, therefore, see that we have a plethora of 0-clairvoyance strategies that outperform our benchmarks.

[0]The D in the rebalancing column of Table 4.3, unsurprisingly, stands for daily.

(a) Y.E.L.q5.c0.inventory_to_ev

(b) Q.E.L.q5.c0.totassets_to_ev

(c) M.E.L.q5.c0.cogs_to_ev

Figure 4.1: Daily returns plots of some of the top performing strategies, according to CAR, for each rebalancing period. The naming format follows "rebalancing.weighting.shorting.quantiles.clairvoyance.ranking_factor".

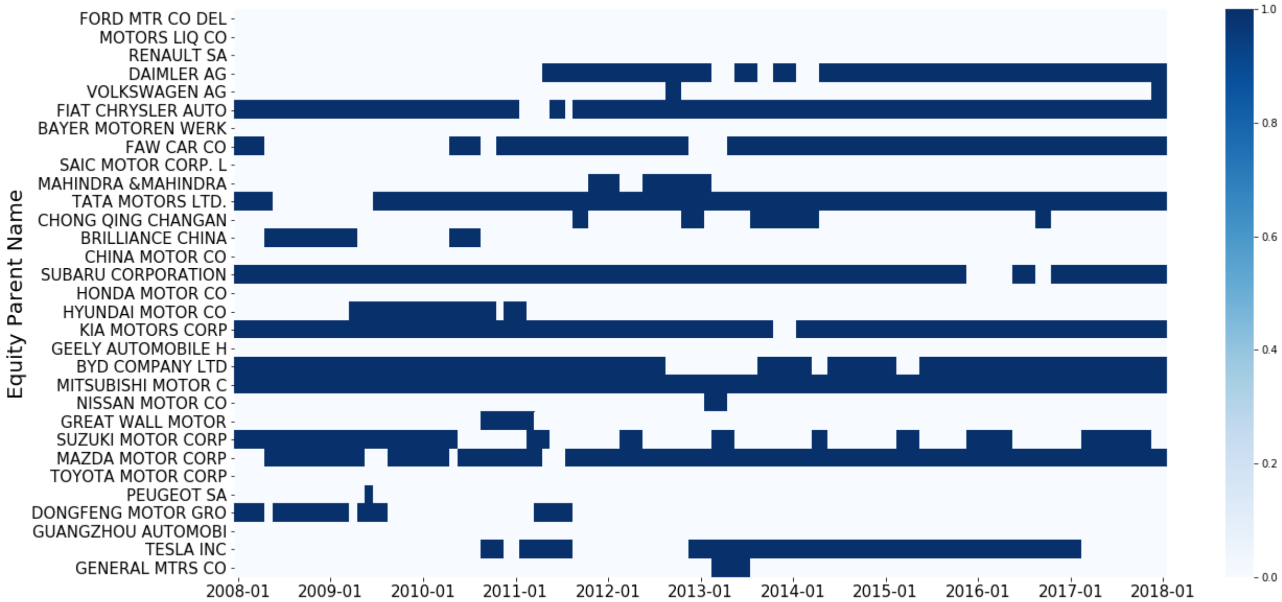


Figure 4.2: Heatmap of which companies' stocks would be in our long portfolio when ranking on inventory_to_totassets using 3 quantiles.

4.1.2 Other information

Finally, it will later be useful to know how many stocks within our top portfolio stay there each month. As such:

Definition 4.1 (Retention rate)**.** *We define the x period retention rate of a portfolio at time t to be the proportion of stocks in the portfolio that were also in the portfolio x periods ago at $t-x$.*

$$r_x^t := \frac{\sum\limits_{s_i \in \mathcal{P}^t} \mathbb{1}_{\{s_i \in \mathcal{P}^{t-x}\}}}{\sum\limits_{s_i \in \mathcal{P}^{t-x}} 1}$$

where s_i is represents the stock of company i and $\mathcal{P}^t$ represents our portfolio at time t.

We provide a table of the above strategies' retention rates in Table B.3. We also provide the above strategies' turnover rates in Table B.1.

4.1.3 Factor Regression

Given the the investment strategies we propose are factor based, the next natural form of analysis to perform is regression vs the Fama-French factors. What is not so natural, however, is the calculation of said factors. On the one hand, we could regress against the "official" factors published on French's Dartmouth webpage, [Fama and French, 2018], however these are calculated over a universe of US based stocks and indicate the relevant factor values over that universe. On the other hand, we could regress against our own calculations of the same factors, but over our universe of 31 automotive stocks. Each method has its own merits, e.g. the official factors for having a large amount of data to estimate the factor values and our own calculations for being directly relevant to the universe we consider, but equally, each has its own downfalls, e.g. the similarity in market-caps amongst our universe probably making it difficult to identify a strong SMB factor, or the lack of inclusion of any non-US stocks for the Fama-French universe. As such, we perform both regressions here and comment on the results of each. Again, as the slope coefficients we obtain in our regressions don't mean much on their own, we compare them to our yearly-rebalanced equal-weighted benchmark, but also to the monthly-rebalanced market-cap weighted benchmark, as this is closest to a "standard" benchmark that you might find in academia or industry. The regressions performed assume the same model as (2.6).

Official Fama-French factors

Table B.6 summarises the results of regressing the strategies from Table 4.1 against the official Fama-French factors published online. As we can see, most of our strategies have a more positive β_m, more negative β_{SMB} and more negative β_{HML} than our benchmarks. The relative statistics can be seen in Table 4.5, but the sign of the benchmarks' βs must be taken into account when analysing them.

R	W	β_m	β_{SMB}	β_{HML}
M	C	1.019	-0.013	-0.026
Y	E	1.415	-0.199	-0.169

Table 4.4: Benchmark official factor regression coefficients.

β_i	% > Y.E.bench	% > M.C.bench
β_m	90	100
β_{SMB}	16	13
β_{HML}	13	3

Table 4.5: Top performing 0-clairvoyance strategy official regression summary stats. The "% > ..." columns tell us how many of our strategies have a higher β_i than the relevant benchmark.

Automotive Fama-French factors

To attempt to imitate the official Fama-French factors, we follow the same procedure as [Fama and French, 1993, p. 9-10], partitioning our automotive universe by book-to-market ratio into 3 groups and by market-cap into 2 groups and then taking the partions' cartesian product, i.e. $\{\texttt{high}, \texttt{medium}, \texttt{low}\} \times \{\texttt{big}, \texttt{small}\} = \{\texttt{high-big}, \ldots, \texttt{low-small}\}$. From this cross-sectional partitioning, we rebalance yearly from the first business day in June and define:

$$\begin{aligned} SMB(t) &:= \frac{1}{3}\left(r_{\texttt{high-small}}(t) + r_{\texttt{medium-small}}(t) + r_{\texttt{low-small}}(t)\right) \\ &- \frac{1}{3}\left(r_{\texttt{high-big}}(t) + r_{\texttt{medium-big}}(t) + r_{\texttt{low-big}}(t)\right); \text{and} \\ HML(t) &:= \frac{1}{2}\left(r_{\texttt{high-small}}(t) + r_{\texttt{high-big}}(t)\right) - \frac{1}{2}\left(r_{\texttt{low-small}}(t) + r_{\texttt{low-big}}(t)\right); \end{aligned}$$

where $SMB(t)$ and $HML(t)$ are calculated on a monthly basis. For $r_{mkt} - r_f$, we calculate r_{mkt} as usual from a market-cap weighted, yearly-rebalanced portfolio of all stocks in our universe and use the r_f provided in the Fama-French data on French's website mentioned above. Our calculated factors can be found in Table B.5. It is worth noting the average and standard deviations of these factors from 2008-01 to 2017-06, as shown in Table 4.6, and that 0 is within 1 standard deviation of each of their means over the period.

	$r_{mkt} - r_f$	r_{SMB}	r_{HML}
$\overline{r_i}$	0.0031	0.0059	0.0019
σ_i	0.0701	0.0475	0.0615
1-σ CI	[-0.0670,0.0732]	[-0.0416,0.0534]	[-0.0005,0.0011]

Table 4.6: Summary statistics for the automotive Fama-French factors we calculated.

Table B.7 summarises the results of regressing the strategies from Table 4.1 against our automotive Fama-French factors. Unlike with the official factors, the statistics of our strategies when compared to the benchmarks very much depends on which benchmark you compare them to. As such, we don't attempt to summarise these relationships here, but simply present the data in Table 4.8. Similarly, care with the sign of the benchmarks' βs must be taken into account when analysing these summary statistics.

R	W	β_m	β_{SMB}	β_{HML}
M	C	0.967	-0.024	-0.091
Y	E	1.096	0.436	-0.126

Table 4.7: Benchmark automotive factor regression coefficients.

β_i	% > Y.E.bench	% > M.C.bench
β_m	57	97
β_{SMB}	93	100
β_{HML}	70	53

Table 4.8: Top performing 0-clairvoyance strategy automotive factor regression summary stats.

4.2 Clairvoyance

As we have seen, many of our 0-clairvoyance strategies perform extremely well across our our performance metrics. This does not seem to be any indication, however, of performance using future values of the ranking factors. For example, lets look at the top 3 performing monthly strategies. From Figure 4.3 we can clearly see that introducing clairvoyance into the fundamentals used actually decreases our CAR.

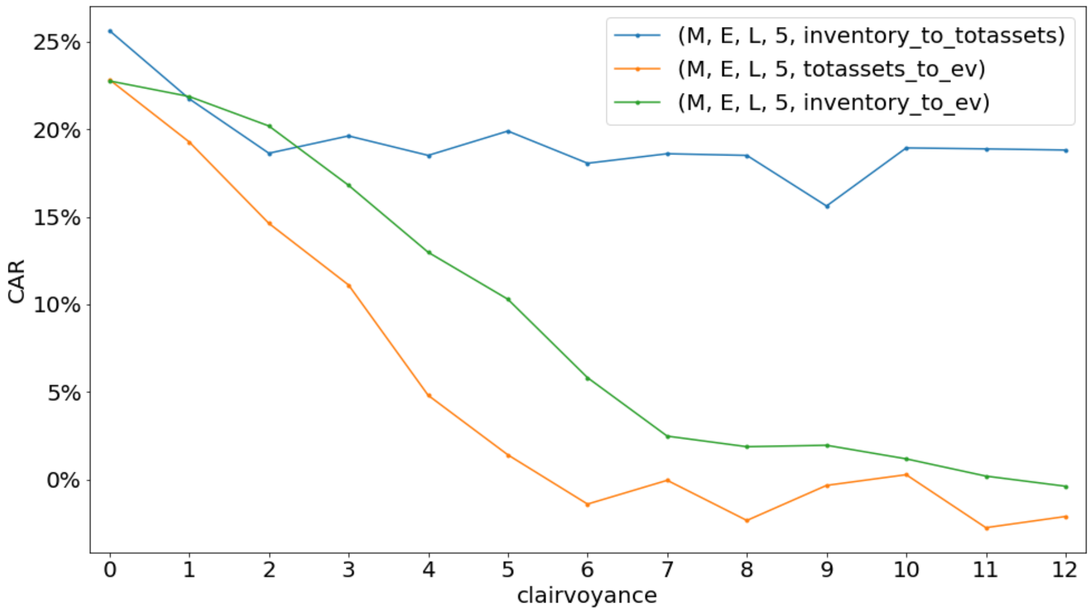


Figure 4.3: Comparing performance of top 3 performing strategies by CAR, as we increase clairvoyance from 0 to 12 months.

Instead of looking at the best performers, and then investigating the effect of clairvoyance on them, let us first pivot on clairvoyance and then assess the best performers. Before we continue, however, there is a point we must stress.

Consider Figure 4.4. We see that, although some of the strategies are highly performant at 6-months clairvoyance, their performance as a function of clairvoyance doesn't seem to be particularly smooth, even increasing, decreasing, increasing and decreasing again for `M, C, L, 5, ffo_to_totassets`. If the knowledge of future information really did give us an advantage here, one would expect CAR to increase with clairvoyance up to a point, and to decrease thereafter. Thinking practically, given markets are reacting to information, it is unlikely that knowledge of a companies financials 12-months in advance, with no knowledge of what they are x-months

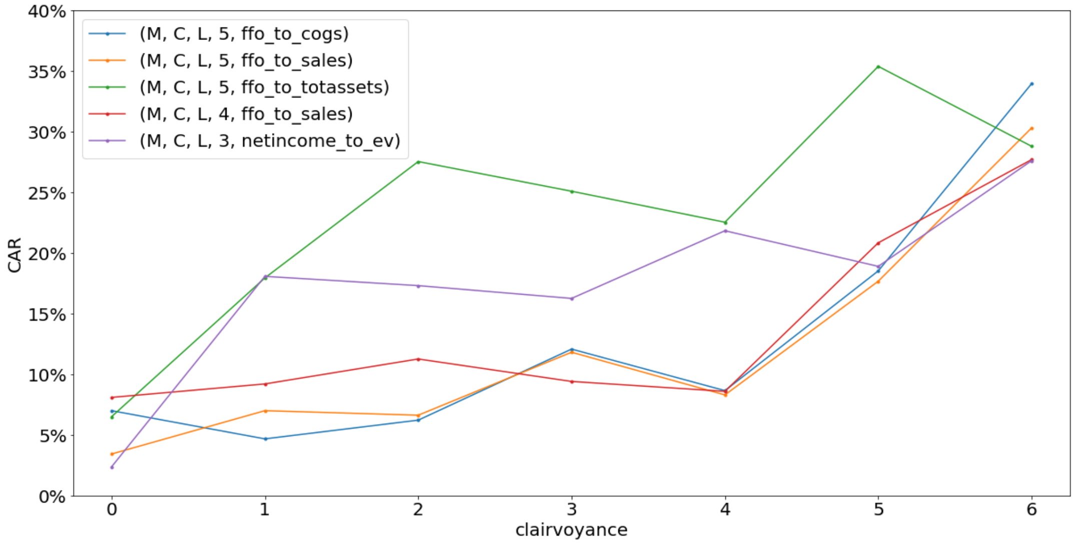


Figure 4.4: Plot of CAR vs clairvoyance for the top performing strategies, based on CAR, at a clairvoyance of 6-months.

in advance for $x \in \{1, \ldots, 11\}$, would tell us much about the returns over the next couple of months[1]. If this is the case, then it is likely that those strategies with a somewhat oscillating clairvoyance-CAR plot after some point are merely picking up on artifacts/anomalies after said point. As such, we decide not to consider strategies that, up to the point of clairvoyance considered, decrease by more than 1% between levels of clairvoyance, or decrease twice in a row. Although this thinking might be somewhat overly conservative, it is better to be conservative and have a robust model/strategy than to base investments based purely on historical data and to invest in artifacts/anomalies[2].

After applying our filters, we then took the 5 best performing strategies for each level of clairvoyance up to 6-months and summarised them in Table 4.9. As, for modelling, the rebalancing period, weighting scheme and whether we short don't actually matter, we only really need consider the ranking factor, the clairvoyance (i.e. prediction) distance and, to a lesser extent, the number of quantiles[3]. As such, the combinations we consider modelling in the next phase are `ffo_to_ev`, `inventory_to_ev`, `accpayable_to_lagged_sales`, `inventory_to_totassets`, `ffo_to_totassets`, `ffo_to_inventory` and `netincome_to_ev` at clairvoyances $\{1, 2, 3\}$.

[1]What we are trying to get at here is that knowledge of the distant future is not always useful in the short-term. In the long-term, yes, it is likely extremely useful, but not necessarily in the short-term.

[2]See `http://www.tylervigen.com/spurious-correlations` for some rather humerous examples of historical anomalies.

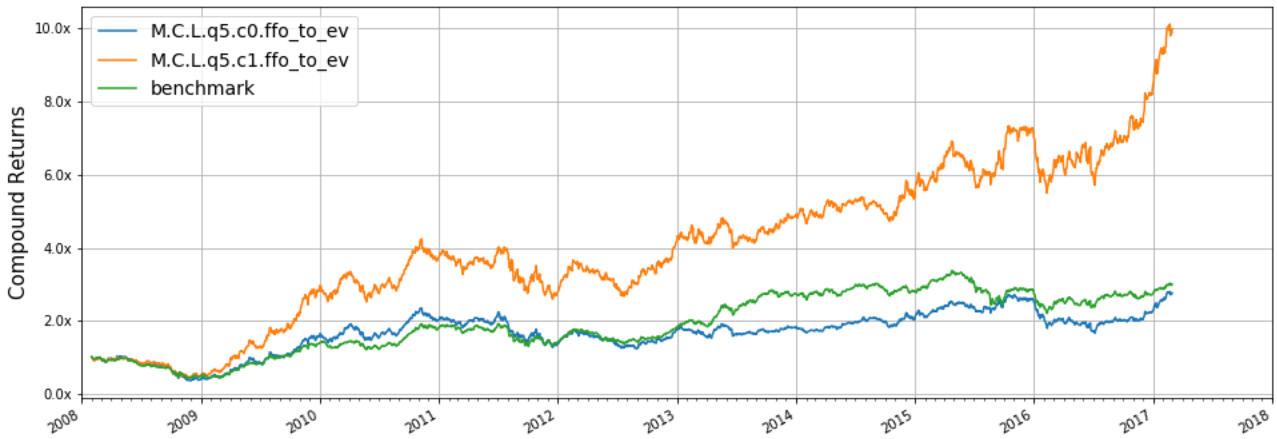
[3]We don't consider the number of quantiles as important as, ideally, our model would be robust across each

R	W	S	Q	ranking_factor	X	$c0$	cX
M	C	L	5	ffo_to_ev	1	0.114	0.279
Q	E	L	4	inventory_to_ev	1	0.220	0.264
Q	E	L	5	accpayable_to_lagged_sales	1	0.249	0.259
M	C	L	5	inventory_to_totassets	1	0.163	0.258
Q	E	L	5	inventory_to_totassets	1	0.263	0.255
M	C	L	5	ffo_to_ev	2	0.114	0.277
Q	E	L	4	inventory_to_ev	2	0.220	0.277
M	C	L	5	ffo_to_totassets	2	0.065	0.276
M	E	L	5	ffo_to_ev	2	0.178	0.250
M	C	L	4	ffo_to_totassets	2	0.098	0.244
M	E	L	5	ffo_to_inventory	3	0.182	0.262
M	C	L	4	netincome_to_ev	3	0.045	0.256
Q	C	L	4	netincome_to_ev	3	0.021	0.256
Q	C	L	5	ffo_to_ev	3	0.085	0.250
Q	E	L	4	netincome_to_ev	3	0.127	0.244
Q	C	L	4	netincome_to_ev	4	0.021	0.263
M	C	L	4	ffo_to_totassets	4	0.098	0.248
Q	C	L	4	ffo_to_totassets	4	0.091	0.244
Q	C	L	4	inventory_to_totassets	4	0.099	0.235
M	C	L	4	ffo_to_inventory	4	0.055	0.232
M	C	L	4	ffo_to_totassets	5	0.098	0.255
Q	C	L	3	accpayable_to_lagged_cogs	5	0.065	0.254
Q	C	L	4	inventory_to_totassets	5	0.099	0.244
Q	C	L	5	ffo_to_inventory	5	0.031	0.225
Q	C	L	3	inventory_to_totassets	5	0.080	0.223
M	C	L	3	ebit_to_ev	6	0.061	0.264
M	C	L	4	ffo_to_totassets	6	0.098	0.261
Q	C	L	4	inventory_to_totassets	6	0.099	0.235
M	E	L	3	ffo_to_inventory	6	0.142	0.234
Q	C	L	4	inventory_to_currliab	6	0.065	0.221

Table 4.9: Best performers for each level of clairvoyance up to 6-months. X denotes the level of clairvoyance, c0 the 0-clairvoyance CAR and cX the X-clairvoyance CAR.

possible top portfolio size.

(a) M.C.L.q5.c1.ffo_to_ev



(b) Q.E.L.q4.c2.inventory_to_ev

(c) M.C.L.q4.c3.netincome_to_ev

Figure 4.5: Daily returns plots of some of the top performing clairvoyant strategies, according to CAR, for clairvoyances up to 3-months. The naming format follows "rebalancing.weighting.shorting.quantiles.clairvoyance.ranking_factor".

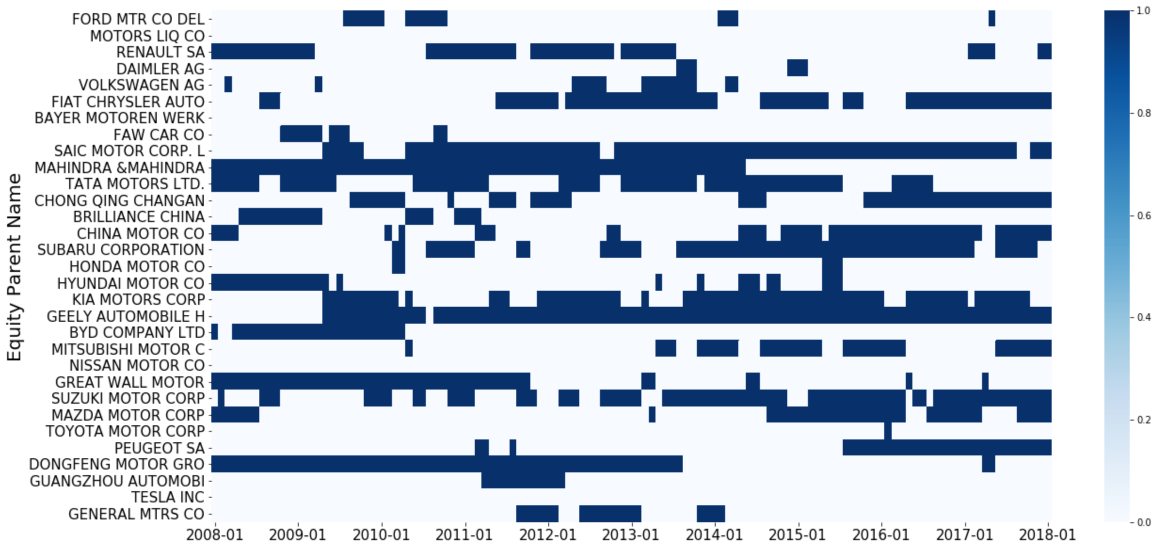


Figure 4.6: Heatmap of which companies' stocks would be in our long portfolio when ranking on ebit_to_ev using 3 quantiles.

4.2.1 Other information

Once again, we provide a table of the above strategies' retention rates in Table B.4, and their turnover rates in Table B.2.

4.2.2 Factor Regression

As before, it is worth looking at the results of regressing our top performing strategies vs the official and automotive Fama-French market, size, and value factors.

Official Fama-French factors

Table B.8 summarises the results of regressing the strategies from Table 4.9 against the official Fama-French factors. The results are largely similar to for the best performing 0-clairvoyance strategies, with most having more positive, more negative and more negative β_m, β_{SMB} and β_{HML} than our benchmarks, respectively. Once again, we must take care with the sign of the benchmarks' βs when analysing the summary statistics.

R	W	β_m	β_{SMB}	β_{HML}
M	C	1.02	-0.01	-0.03
Y	E	1.44	-0.27	-0.18

Table 4.10: Benchmark official factor regression coefficients.

β_i	% > Y.E.bench	% > M.C.bench
β_m	83	100
β_{SMB}	13	10
β_{HML}	27	7

Table 4.11: Top performing x-clairvoyance strategy official factor regression summary stats.

When it comes to the strategies that are performant at some clairvoyance other than 0, however, it makes more sense to compare each x-clairvoyance strategy to the corresponding 0-clairvoyance strategy rather than vs the benchmark. When comparing $c0$ strategies against cX strategies, our findings are as follows:

- 87% of cX strategies have more positive β_m than their corresponding $c0$ strategies;
- 60% of cX strategies have more positive β_{SMB} than their corresponding $c0$ strategies; and
- 70% of cX strategies have more positive β_{HML} than their corresponding $c0$ strategies.

Testing these results under the hypothesese that:

$$H_0 : \mathbb{P}[\beta_i^X > \beta_i^0] = 0.5 = \Pr[\beta_i^X < \beta_i^0];\ \text{vs}$$
$$H_a : \mathbb{P}[\beta_i^X > \beta_i^0] > 0.5;$$

we find that these numbers have p-values of 3e-5, 0.18 and 0.02 respectively. This tells us that, up to (around) the 2% significance level, we would reject, fail to reject, and reject the hypothesis that the clairvoyant strategies do not increase the β_i of their 0-clairvoyance counterparts, for $i \in \{m, SMB, HML\}$. It is worth noting, at the 5% level, the range in which $\mathbb{P}[\beta_i^X > \beta_i^0]$ could fall and that we would reject the null hypothesis. For β_m it is $[0, 0.72)$, and for β_{HML} it is $[0, 53]$%. We, therefore, conclude that these clairvoyant strategies, on average, have higher β_m and higher β_{HML} than their 0-clairvoyance counterparts, at the 5% level, but fail to conclude a difference in their β_{SMB}.

We also test the similar, but slightly different, hypothesis that adding clairvoyance increases/decreases the resulting strategies β_i, in the direction of the 0-clairvoyance β_i. Our results show that:

1. 87% of cX strategies have more extreme β_m than their corresponding $c0$ strategies[4];
2. 43% of cX strategies have more extreme β_m than their corresponding $c0$ strategies; and
3. 20% of cX strategies have more extreme β_m than their corresponding $c0$ strategies.

[4] I.e. more positive if the $c0$ β_m is positive, and vice versa.

As we have more strategies with LESS extreme β_i for the SMB and HML factors, we test the alternative that adding clairvoyance DECREASES the resulting β_i, and vice versa for the market factor, i.e:

$$H_0 : \mathbb{P}[\frac{\beta_i^X}{\beta_i^0} > 1] = 0.5 = \Pr[\frac{\beta_i^X}{\beta_i^0} < 1]; \text{ vs}$$
$$H_{a_1} : \mathbb{P}[\frac{\beta_i^X}{\beta_i^0} > 1] > 0.5.$$
$$H_{a_2} : \mathbb{P}[\frac{\beta_i^X}{\beta_i^0} < 1] > 0.5.$$

Here we find that our results lead to p-values of 3e-5, 0.29 and 0.001 respectively. This tells us that, up to the 1% level, we would reject, fail to reject, and reject the hypothesis that that the clairvoyant strategies do not increase, decrease and decrease the extremeness of β_i of their 0-clairvoyance counterparts. Again, it is owrth noting, at the 5% level, the ranges within which $\mathbb{P}[\beta_i^X > \beta_i^0]$ could fall and that we would reject the null hypothesis. For β_m it is $[0, 0.72)$ again, and for β_{HML} it is $(0.36, 1]$. We, therefore, conclude that these clairvoyant strategies, on average, have more extreme β_m and less extreme β_{HML} than their 0-clairvoyance counterparts, at the 5% level, but fail to conclude a difference in their β_{SMB}.

Automotive Fama-French factors

Table B.9 summarises the results of regressing the strategies from Table 4.9 against the official Fama-French factors. Once again, we must take care with the sign of the benchmarks' βs when analysing the summary statistics.

R	W	β_m	β_{SMB}	β_{HML}
M	C	0.97	-0.02	-0.09
Y	E	1.10	0.43	-0.13

Table 4.12: Benchmark automotive factor regression coefficients.

β_i	% > Y.E.bench	% > M.C.bench
β_m	73	97
β_{SMB}	86	100
β_{HML}	43	33

Table 4.13: Top performing x-clairvoyance strategy automotive factor regression summary stats.

We perform the same tests as before with respect to positivity, and find that:

- 63% of cX strategies have more positive β_m than their corresponding $c0$ strategies;
- 90% of cX strategies have more positive β_{SMB} than their corresponding $c0$ strategies; and
- 73% of cX strategies have more positive β_{HML} than their corresponding $c0$ strategies.

Under the same hypothesis tests, we find p-values of 0.1, 4e-6 and 0.008 respectively, and rejection ranges of $[0, 0.76]$ and $[0, 0.57)$ for β_{SMB} and β_{HML} respectively. We, therefore, conclude that, when regressed against our automotive Fama-French factors, these clairvoyant strategies have more positive β_{SMB} and β_{HML}, at the 5% level, than their 0-clairvoyance counterparts, but fail to conclude a difference in their β_m.

Similarly, we consider the extremeness of each clairvoyant strategies β_i when compared to its 0-clairvoyance counterpart and find that:

1. 63% of cX strategies have more extreme β_m than their corresponding $c0$ strategies;
2. 90% of cX strategies have more extreme β_m than their corresponding $c0$ strategies; and
3. 27% of cX strategies have more extreme β_m than their corresponding $c0$ strategies.

Here we now test that the clairvoyant strategies produce more, more and less extreme β_m, β_{SMB} and β_{HML} respectively, than their 0-clairvoyance counterparts. The p-values are 0.1, 4e-6 and 0.008 respectively, giving us rejection ranges of $[0, 0.76]$ and $(0.43, 1]$ for β_{SMB} and β_{HML}. We, therefore, conclude that, when regressed against our automotive Fama-French factors, these clairvoyant strategies have more extreme β_{SMB} and less extreme β_{HML}, at the 5% level, than their 0-clairvoyance counterparts, but fail to conclude a difference in their β_m.

A link to CSVs of all results can be found in Appendix B.2

Chapter 5

Stage 2 - Factor Modelling/Prediction

Now that we know which ratios we wish to predict, we explain our methodology in building a model.

5.1 Methodology

For our modelling, we attempt to predict each of the ratios mentioned in Section 4.2 c-months ahead, for c in $\{1, 2, 3\}$. Let y_t be the ratio we are trying to predict, at time t, and x_t^i the value of input i at time t. We are, therefore, trying to find a relationship such that

$$y_{t+c} = f(x_t^1, \ldots, x_t^n) + \varepsilon_t,$$

with ε_t being zero-mean. There are many ways in which we could attempt this, e.g.:

1. Predict ratios directly - have the ratio we wish to predict as our target in our model training:
$$\hat{y}_{t+c} = f(x_t^1, \ldots, x_t^n).$$
2. Predict the constituents of the ratio and combine - have two separate models (or perhaps a multi-target regression), make predictions for the numerator and denominator of our ratio and then divide to get the actual value of interest:
$$\hat{y}_{t+c}^1 = f^1(x_t^1, \ldots, x_t^n),$$
$$\hat{y}_{t+c}^2 = f^2(x_t^1, \ldots, x_t^n),$$
$$\hat{y}_{t+c} = \frac{\hat{y}_{t+c}^1}{\hat{y}_{t+c}^2}.$$
3. Predict changes in the ratio - rather than predict y_{t+c}, we could attempt to predict:
$$\Delta y_{t+c} := y_{t+c} - y_t, \tag{5.1}$$

and use this to predict y_{t+c}:

$$\begin{aligned}\hat{y}_{t+c} &= y_t + \widehat{\Delta y}_{t+c},\\ &= y_t + f(x_t^1, \ldots, x_t^n).\end{aligned}$$

4. Predict changes in the constituents and combine - combine 2. and 3. .
5. Predict percentage changes - the same as 3. and 4., but predict:

$$\Xi y_{t+c} := \frac{y_{t+c}}{y_t} - 1. \tag{5.2}$$

As all of these methods are ultimately trying to do the same thing, and it isn't particularly clear whether any is a better choice than the others, we attempt all of them and see which yields the strongest results.

5.2 Timeframe

As the proprietary dataset from ISHM only has clean data on all companies back to 2010-01, the timeframe for our model testing/calibration will be 2010-01 to 2017-03. We believe it is not a problem that the two timeframes in stage 1 and stage 2 are not identical given stage 1 pertained to factor identification and stage 2 to factor modelling. For factor identification, we want to identify factors that have a long standing outperformance of the benchmark. For factor modelling, the results of stage 1 are completely irrelevant, in that we could choose to model non-performant factors if we so choose. Factor modelling can, therefore, be thought of as a completely separate task, where it just so happens that we are modelling the factors that performed well in stage 1. As such, it shouldn't matter that the two testing timeframes are not the same.

5.3 Data preperation

As mentioned in Section 3.2, as well as the financial reporting data from the Worldscope database, we also have a proprietary dataset, collected and curated by IHSM, detailing information on various aspects of each of the 31 automotive companies production and sales numbers, broken down on a month-by-month basis. We first merge these two datasets and then perform some feature engineering to aid us in our modelling. The feature engineering process is as follows[1]:

1. Aggregate each feature in the IHSM dataset into a trailing x-month sum, for x in $\{1,\dots,6\}$. We do this because each of the features in the IHSM dataset refer to a 1-month timeframe each but the financial reporting to a 3-month timeframe.
2. For each feature calculate the x-month Δ-difference[2] and x-month Ξ-difference[3], on itself, for x in $\{1,\dots,3\}$. We do this because it is more likely that some sort of past change in one of the features would be indicative of some sort of future change in the target, rather than the past/current level of a feature being indicative of the future change of the target (i.e. for the Δy_{t+c} and Ξy_{t+c} predictions).
3. For each feature, lag it by x-months for x in $\{1,\dots,3\}$.

As for the naming, the structure of each features name is:
`<root feature>-<aggregation>-<difference>-<lag>`.
E.g. sales_volume_prev_3m_sum_prev_1m_pct_change_bshift_2m would correspond to, taking the sales volume and:

1. Applying a trailing three month sum to it;
2. Calculating the %-difference (Ξ) from 1-month previously on the result of 1. ; and
3. Lagging the resulting timeseries of 2. by 2-months.

5.4 Feature selection

Now that we have a large number of features to aid us in our predictions, $8,624$ in fact, it is time to select those that we believe will be useful in attempting to predict y_{t+c}. For linear regression, it stands to reason that the higher a feature's correlation with the target, the more useful it will be in predicting said target. We don't, however, want any of our features to be easily predicted by any combination of the other features (i.e. we want to avoid multicollinearity). As such, for any of our linear regression based models, our feature selection process is as follows:

Denote our target variable y_{t+c} and our possible input features x_t^i, for i in $\{1,\dots,n\}$. Let the set of feature variables we use as our inputs to the linear regression be $\mathcal{X} := \varnothing$. Without loss of generality, assume the x^i are sorted such that x^1 is most correlated with y and x^n least

[1] For each step in the feature generation, when we say "for each feature", we mean for each feature in the newest set of features, i.e. including those generated in previous steps, unless otherwise stated.
[2] Akin to the Δ operator from (5.1)
[3] Akin to the Ξ operator from (5.2)

correlated with y. For i in $\{1, \ldots, n\}$, if $|\rho(x^i, y)| > 0.05$ and $|\rho(x^i, x^j)| < 0.5$ for all x^j in $\mathcal{X}$, add x^i to $\mathcal{X}$.

This process has now done two things:

1. Filtered our features such that only those with a high enough correlation with the target remain in $\mathcal{X}$.
2. Ensured that no two variables within our set of input features, $\mathcal{X}$, have high correlation with each other, reducing the chance of multicollinearity.

The correlations calculated in the above process were originally done in two different ways, each one to be tested separately:

1. Find the sample correlations over the whole timeframe 2010-01 to 2017-03; and
2. Find the rolling correlations over the last 3-years, moving the window from [2010-01, 2012-12] to [2014-04,2017-03] and average using the Fisher z-transform method described in Section 3.3.1;

however, given the Fisher z-transform method is attempting to estimate this whole timeframe correlation, we get very similar results for the correlations and, therefore, for the features we choose to use in our model. As such, we decide just to use method 1.

Another suitable method would be to individually regress each x^i against y, and add the highest performing feature to $\mathcal{X}$, as long as the maximal performance of regressing x^i on all possible combinations of the x^j in $\mathcal{X}$ did not exceed some threshold.

5.5 Models

When it comes to regression of any kind, one must always try to implement a simple linear model before expanding. As such, we start with OLS regression before attempting anything more complicated. Also, as our dataset is quite small[4], we don't attempt to use neural nets or any other complex models.

[4] 30 companies reporting 12 times a year for roughly 8 years, so around 2,800 datapoints in total for training and testing.

5.5.1 The naïve benchmark

Similar to trading strategies, any machine learning model is only as good as its performance when compared to some naïve benchmark. The naïve model we compare to here is the forward-fill model, i.e. we assume that $y_{t+c} = y_t$.

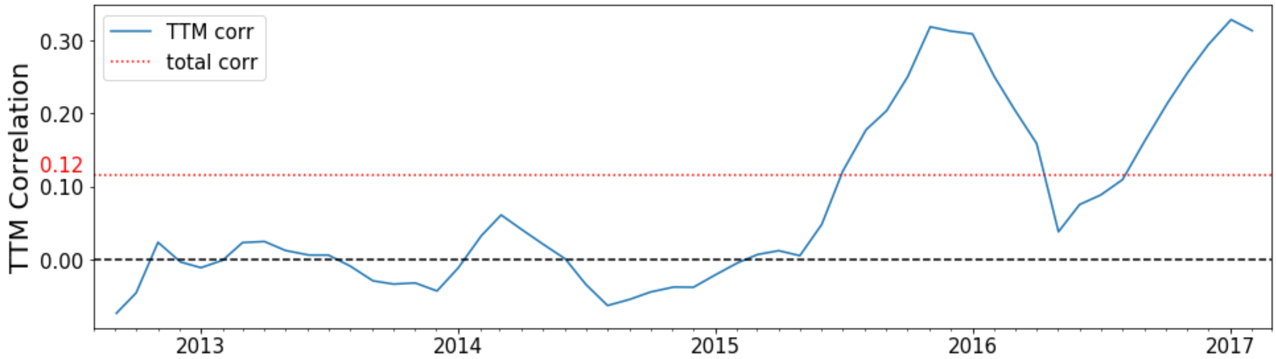
5.5.2 Training and testing

At each point in time that we test our model, we train it on the past 3-years most recent data and try to predict the data that is c-months into the future. We then move forward one month and do the same again.

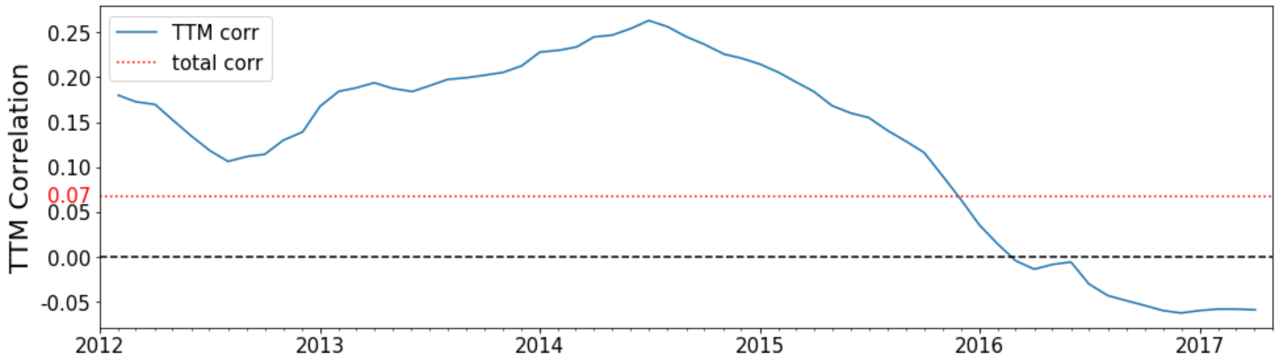
5.6 Results

Following the feature selection method outlined in Section 5.4, the suggested linear models varied between 3 and 10 input features. Unfortunately, we found that none of our models meaningfully outperformed the benchmark. Some were found to be worse on average, but better in times of extreme uncertainty, i.e. when the target ratio changed a lot, but none were found to be better on average, all with respect to MSE. Furthermore, we found that even of the best performing models with respect to MSE, all underperformed the naïve model with respect to their sensitivity[5]in classifying which stocks end up in our portfolio. Upon further inspection, this turned out to be because the naïve model was quite good at this indirect classification most of the time, often classifying with between 80-90% sensitivity across each of the ratios we attempted to predict. This is not entirely surprising, given the heatmaps we saw in Chapter 4 and the retention rates in Table B.4, averaging 76% for the strategies with clairvoyances of up to 3-months. We also found that, despite many of the features having high correlation with the target, either on average or over the whole testing period, these correlations were not stable, and often changed signs when looking at a rolling/expanding basis. Section 5.6 shows 2 examples of this, each having positive overall correlation but trailing twelve month correlations that fluctuate, and even dip below zero at points.

[5]When we say sensitivity here, we mean the number of true positives divided by the number of total positives when classifying stocks as being in our top portfolio. E.g. if the real portfolio should be (a,b,c) and we predict (a,b,p), our sensitivity is $\frac{2}{3}$, as we correctly predicted a and b out of a,b and c.



(a) Trailing 12-month correlation between netincome_to_ev_prev_3m_diff_fshift_3m and inventory_to_ev_prev_3m_diff_bshift_3m.



(b) Trailing 12-month correlation between netincome_to_ev_fshift_1m and ave_age_months.

Figure 5.1

5.7 Model/feature selection 2

Given this naïve predictor is so highly performant, we decide instead to take a different approach. We choose not to follow the previous feature selection method outlined in Section 5.4, nor to try to predict changes of any sort. Instead, given y_{t+c} seems to often be very similar to y_t we decide to follow the model outlined in (5.3) for different x_t, and build it up from there.

$$y_{t+c} = \beta_0 y_t + \beta_1 x_t \tag{5.3}$$

Our method of selecting x_t to test is similar to before. For each y_{t+c} that we wish to predict, we test all x_t such that $|\rho(x, y)| > 0.05$.

5.8 Results 2

As our training method minimises the overall MSE for each training set, whereas we only really care about whether it is predicting which stocks would be in our portfolio correctly, we also measure how many of said portfolio the model correctly predicts for each of our 3 portfolio sizes (i.e. the classification sensitivity).

Given that a lower average MSE than the naïve model does not necessarily mean better performance, as discussed earlier, we instead decide to consider a model a success only if it had a lower classification sensitivity than the naïve model, and had at least one timeperiod where it had a better sensitivity. Requiring it is never worse across any of our choice of quantiles is too restrictive[6], however, so we only require it to never be worse, for the choice quantiles we are considering in that backtest.

A link to CSVs of all results can be found in Appendix B.2

[6]If we require that the model is never worse than the naïve model for all three of our choices of quantiles, we end up with only the models that produce an identical ranking to the naïve model, thus we relax our constraints slightly.

Chapter 6

Stage 3 - Model Backtesting

Now that we have a filtered set of models that we believe to be more predictive than our naïve model, we test each to assess its real world performance when used in a trading strategy. For each model that was superior to our naïve benchmark in the modelling phase, we generate a timeseries of predictions and use these as our "fundamental data" in each of the relevant strategies.

Example 6.0.1. *Imagine we have a model to predict ffo_to_ev, 2-months in advance, using the current ffo_to_ev ratio and the production_growth_trend feature. Imagine this model is superior to the naïve model when using 3 quantiles. We first move month-by-month, predicting the 2-month ahead ffo_to_ev, given "current" information. After generating a complete timeseries, for any strategy that:*

- *Didn't fail our filtering criteria from Section 4.2;*
- *Uses ffo_to_ev with 2-month clairvoyance; and*
- *Uses the top 33% of stocks ranked on 2-month ahead ffo_to_ev as it's long portfolio;*

we use this generated timeseries as the timeseries we produce our rankings from, following the same steps as in Section 3.3.1.

Once we have tested the strategy using our predicted timeseries, we compare it to the results of using the same strategy with 0-clairvoyance and with x-clairvoyance, e.g. 2-months clairvoyance for Example 6.0.1.

6.1 Backtest results

In our backtest, there were 2347 strategies where we had both that the model was outperformant of the naïve model, and where the original backtest with perfect knowledge of the future met our filters imposed in Section 3.3.1. Of these 2347 stage 3 backtests, there are 355 where the CAR using our predictions is between the 0-clairvoyance CAR and the x-clairvoyance CAR, roughly 15%. When we take into account how many of the models were equivalent to the naïve model[1], however, this rate increases to about 44%. Furthermore, again ignoring the equivalent models, 74% of models outperform their 0-clairvoyance equivalent. Table 6.1 provides some summary statistics about the relative positions of the 0-clairvoyance, x-clairvoyance and model CARs[2].

Position	n
$c0 < \widehat{cX} < cX$	355
$cX < \widehat{cX} < c0$	31
$c0 < cX < \widehat{cX}$	68
$cX < c0 < \widehat{cX}$	169
$\widehat{cX} < c0 < cX$	182
$\widehat{cX} < cX < c0$	30
$\widehat{cX} = c0$	1543

Table 6.1: Number of backtest CARs in each relative position, with 0-clairvoyance ($c0$), x-clairvoyance (cX) and the relevant model ($\widehat{cX}$).

To assess how much of the potential excess gains we can achieve through prediction, for each strategy where $\widehat{cX}$ was between $c0$ and cX, we calculate t, such that $\widehat{cX} = (1-t)\cdot c0 + t \cdot cX$. We find that $\bar{t} = 22.5\%$, with $t_{max} = 95.1\%$ and $t_{min} = 0.5\%$.

[1] Originally it was $\frac{355}{2347}$, changing to $\frac{355}{804}$ when we remove 1543 of them.

[2] It is worth noting that, although in our stage 1 backtest, all of the strategies tested here had x-clairvoyance CAR greater than their 0-clairvoyance CAR, as the backtesting window for stage 3 is different, there is no guarantee that the same holds here. Should this be the case, this does not mean that it is not a strategy worth considering, just that this was a period of poor performance for the strategy, which we believe to have generally strong performance, given the longer backtesting window in stage 1. As such, even if we have 0-clairvoyance CAR, x-clairvoyance CAR and model CAR of 15.1, 10.3 and 11.5, that should still be considered a success, as we managed to achieve 75% of the excess returns that were theoretically attainable.

R	W	S	Q	y	x_1	x_2	X	$c0$	cX	$\widehat{cX}$
M	C	L	5	accpayable_to_lagged_sales_fshift_1m	accpayable_to_lagged_sales	revenues_sales_growth_trend_prev_5m_sum_prev_1m_diff_bshift_3m	1	0.157	0.171	0.247
Q	C	L	5	ffo_to_totassets_fshift_2m	ffo_to_totassets	ave_age_months_prev_3m_pct_change	2	0.188	0.269	0.197
Q	C	L	5	ffo_to_totassets_fshift_2m	ffo_to_totassets	ave_age_months_prev_5m_sum_prev_3m_pct_change	2	0.188	0.269	0.197
M	C	L	5	accpayable_to_lagged_sales_fshift_1m	accpayable_to_lagged_sales	production_growth_trend_bshift_1m	1	0.157	0.171	0.170
M	C	L	5	accpayable_to_lagged_sales_fshift_1m	accpayable_to_lagged_sales	production_growth_trend_prev_4m_sum	1	0.157	0.171	0.166
Q	E	L	3	ffo_to_ev_fshift_2m	ffo_to_ev	production_growth_trend	2	0.151	0.096	0.155
Q	E	L	3	ffo_to_ev_fshift_2m	ffo_to_ev	production_growth_trend_bshift_1m	2	0.151	0.096	0.155
Q	C	L	4	ffo_to_totassets_fshift_3m	ffo_to_totassets	sales_prev_2m_pct_change_bshift_3m	3	0.151	0.138	0.152
Q	C	L	4	ffo_to_totassets_fshift_2m	ffo_to_totassets	sales_prev_2m_pct_change_bshift_1m	2	0.151	0.215	0.152
Q	C	L	4	ffo_to_totassets_fshift_2m	ffo_to_totassets	lagged_sales_prev_2m_pct_change	2	0.151	0.215	0.152
Q	C	L	4	ffo_to_totassets_fshift_2m	ffo_to_totassets	sales_prev_2m_pct_change	2	0.151	0.215	0.152
Q	C	L	4	ffo_to_totassets_fshift_3m	ffo_to_totassets	sales_prev_2m_pct_change_bshift_2m	3	0.151	0.138	0.152
Q	C	L	4	ffo_to_totassets_fshift_3m	ffo_to_totassets	lagged_sales_prev_2m_pct_change_bshift_2m	3	0.151	0.138	0.152
Q	C	L	4	ffo_to_totassets_fshift_3m	ffo_to_totassets	lagged_sales_prev_2m_pct_change_bshift_1m	3	0.151	0.138	0.152
Q	E	L	4	ffo_to_totassets_fshift_3m	ffo_to_totassets	entvalue_prev_2m_pct_change_bshift_2m	3	0.129	0.155	0.146
Q	E	L	4	ffo_to_totassets_fshift_3m	ffo_to_totassets	ev_prev_2m_pct_change_bshift_2m	3	0.129	0.155	0.146
Q	E	L	4	ffo_to_totassets_fshift_3m	ffo_to_totassets	ev_prev_2m_pct_change_bshift_1m	3	0.129	0.155	0.143
Q	E	L	4	ffo_to_totassets_fshift_3m	ffo_to_totassets	entvalue_prev_2m_pct_change_bshift_1m	3	0.129	0.155	0.143
Q	E	L	4	ffo_to_totassets_fshift_2m	ffo_to_totassets	ev_prev_2m_pct_change	2	0.129	0.177	0.141
Q	E	L	4	ffo_to_totassets_fshift_2m	ffo_to_totassets	entvalue_prev_2m_pct_change	2	0.129	0.177	0.141

Table 6.2: Top 10 performing models, according to backtest CAR. y, x_1, x_2 indicate the target we are trying to model, input 1, and input 2, respectively.

(a) M.E.L.q3.ffo_to_ev_fshift_2m-ffo_to_ev-production_growth_trend

(b) Q.E.L.q4.ffo_to_totassets_fshift_2m-ffo_to_totassets-ev_prev_2m_pct_change

(c) Q.C.L.q5.ffo_to_totassets_fshift_2m-ffo_to_totassets-ave_age_months_prev_3m_pct_change

Figure 6.1: Plots of $c0$, cX and $\widehat{cX}$ CAR for 3 of the models where $c0 < \widehat{cX} < cX$. In blue we have the no clairvoyance ($c0$) CAR, in orange we have the x-month clairvoyance CAR (cX) and in green we have the model CAR ($\widehat{cX}$).

6.2 Factor regression

Given we regressed our 0-clairvoyance and x-clairvoyance strategies against the Fama-French factors, it would, therefore, also make sense to do so here for the returns produced using our models.

6.2.1 Official Fama-French factors

Of the non-equivelant models, 51% had a β_m between their $c0$ and cX β_m's, 33% a β_{SMB} between their β_{SMB}'s and 47% a β_{HML} between their β_{HML}'s.

For a more detailed summary of breakdown of relative positions, see Table 6.3

relative position	mkt	SMB	HML
$\beta_i^0 < \widehat{\beta_i^X} < \beta_i^X$	397	63	191
$\beta_i^X < \widehat{\beta_i^X} < \beta_i^0$	15	202	189
$\beta_i^0 < \beta_i^X < \widehat{\beta_i^X}$	100	13	2
$\beta_i^X < \beta_i^0 < \widehat{\beta_i^X}$	30	346	256
$\widehat{\beta_i^X} < \beta_i^0 < \beta_i^X$	256	96	103
$\widehat{\beta_i^X} < \beta_i^X < \beta_i^0$	2	80	59

relative position	mkt	SMB	HML
$\beta_i^0 < \widehat{\beta_i^X} < \beta_i^X$	0.50%	0.08%	0.24%
$\beta_i^X < \widehat{\beta_i^X} < \beta_i^0$	0.02%	0.25%	0.24%
$\beta_i^0 < \beta_i^X < \widehat{\beta_i^X}$	0.13%	0.02%	0.00%
$\beta_i^X < \beta_i^0 < \widehat{\beta_i^X}$	0.04%	0.43%	0.32%
$\widehat{\beta_i^X} < \beta_i^0 < \beta_i^X$	0.32%	0.12%	0.13%
$\widehat{\beta_i^X} < \beta_i^X < \beta_i^0$	0.00%	0.10%	0.07%

Table 6.3: The number of β_is from the non-equivalent models that fall into each position relative to their $0/x$-clairvoyance counterparts. The table on the left shows absolute numbers and the right shows the proportion of each column falling into each relative position.

From Table 6.3 we can see that the majority of models have a β_m between their $c0$ and cX β_m's. As for their β_{SMB} and β_{HML}'s, the split is far less clear.

6.2.2 Automotive Fama-French factors

Of the non-equivelant models, 43% had a β_m between their $c0$ and cX β_m's, 58% a β_{SMB} between their β_{SMB}'s and 50% a β_{HML} between their β_{HML}'s.

relative position	mkt	SMB	HML
$\beta_i^0 < \widehat{\beta_i^X} < \beta_i^X$	179	387	326
$\beta_i^X < \widehat{\beta_i^X} < \beta_i^0$	163	81	76
$\beta_i^0 < \beta_i^X < \widehat{\beta_i^X}$	4	25	70
$\beta_i^X < \beta_i^0 < \widehat{\beta_i^X}$	231	12	43
$\widehat{\beta_i^X} < \beta_i^0 < \beta_i^X$	210	293	267
$\widehat{\beta_i^X} < \beta_i^X < \beta_i^0$	13	2	18

relative position	mkt	SMB	HML
$\beta_i^0 < \widehat{\beta_i^X} < \beta_i^X$	22%	48%	41%
$\beta_i^X < \widehat{\beta_i^X} < \beta_i^0$	20%	10%	10%
$\beta_i^0 < \beta_i^X < \widehat{\beta_i^X}$	1%	3%	9%
$\beta_i^X < \beta_i^0 < \widehat{\beta_i^X}$	29%	2%	5%
$\widehat{\beta_i^X} < \beta_i^0 < \beta_i^X$	26%	37%	33%
$\widehat{\beta_i^X} < \beta_i^X < \beta_i^0$	2%	0%	2%

Table 6.4: The number of β_is from the non-equivalent models that fall into each position relative to their $0/x$-clairvoyance counterparts. The table on the left shows absolute numbers and the right shows the proportion of each column falling into each relative position.

From Table 6.4 we can see that the majority of models have a β_{SMB} between their $c0$ and cX β_{SMB}'s and a β_{HML} between their $c0$ and cX β_{HML}'s. As for their β_m's, the split is far less clear, however there is almost a majority that have their β_m between, but not quite, at 42%.

A link to CSVs of all results can be found in Appendix B.2

Chapter 7

Conclusion & Future Work

7.1 Summary of thesis achievements

7.1.1 Stage 1 - factor identification

In chapter 4 we attempted to find a number of different ranking factors which outperform our benchmarks in backtests. We succeeded in showing that a large number of ratios of company fundamentals produced excess returns over the market when used in our factor based strategy, across a variety of different rebalancing periods, number of quantiles used and weighting schemes. We saw that yearly rebalancing schemes tend to outperform strategies using a shorter rebalancing period, in terms of both compound annual returns and Sharpe ratio. We also saw that, when regressed against the official Fama-French factors, our top performing portfolios general had higher β_m, but lower β_{SMB} and β_{HML} than M.C.benchmark and Y.E.benchmark. When regressed against our own calculations of the factors for our automotive universe, we saw that our top performing strategies generally had; higher β_m than M.C.benchmark, but similar (i.e. equal above and below) β_m to Y.E.benchmark; higher β_{SMB} than both benchmarks; and higher β_{HML} than Y.E.benchmark.

We then went on to show that there exist a number of different strategies that achieve greater compound annual returns when using exactly known future data. We saw that these improvements often did not exist for the highest performing strategies based on known data, but for others that we could theoretically achieve higher returns using future data than the best strategies using known data did. Of these strategies, we saw that using future data generally increased the extremity of β_m vs their 0-clairvoyance counterparts, and reduced the extremity of β_{HML} when regressed vs the official Fama-French factors. When regressed against our automotive factors, we saw that the addition of clairvoyance usually increased the extremity of β_{SMB} and decreased the extremity of β_{HML}.

7.1.2 Stages 2 & 3 - factor modelling and model backtesting

Once we identified some future ratios worth predicting, we attempted to do so using our proprietary IHSM dataset. We found that, due to the high retention rates of our strategies' top portfolios, it was difficult to beat the naïve benchmark by much, but that using some features from the IHSM dataset, it was possible to do so. Of those models that outperformed the benchmark in terms of classification sensitivity, we found that the resulting strategy returns outperformed the 0-clairvoyance strategy in 74% of cases. As for the resulting returns timeseries regressions vs the Fama-French factors, more analysis would need to be done.

7.1.3 Overall

Overall we have show that there exist multiple ratios of fundamentals which, if known in advance, result in strategies that outperform even the best of our benchmarks, and that knowledge of said ratios in advance does increase our strategy performance, all within the automotive sector. We have also shown that, using theoretically possible to know information[1], we can build models that outperform the naïve model of assuming no change, and that, when backtested inside a trading strategy, these more often than not result in outperformance of strategies where no predictive model is used.

7.2 Applications

As discussed in Section 2.5, the benefits of factor investing over regular passive investing, and even most active investing strategies, are numerous. As for the particular applications of this study, we would like to see a bit more robustness in the results before implementation in a live trading strategy. The results do, however, show that there is value in investing time and effort into modelling these future ratios and that more effort should be done to do so.

7.3 Future Work

Whilst undertaking this study, there were many things we came up with which we thought would be interesting to research, but didn't have the time to implement.

[1] See declaration in Appendix B.

7.3.1 Factor identification

Rather than using market-cap or equal weighting schemes, we thought an interesting idea might be to weight by the ranking factor used to create the portfolio. We had this idea because the current weighting scheme, once the stocks that enter our portfolio have been decided, doesn't take into account the magnitude of the ranking factor for each of our stocks. Would it not make sense that, given the ranking factor is meant to give us information on future returns, we would want to give higher weight to those which it ranks higher?

Alternatively to the previous suggestion, we also thought it might be interesting to follow the market-cap/equal weighting scheme but then apply some sort of tilts based on the ranking factor you are using.

Given each of these factors used in the ranking is meant to give information on the future returns of the companies' stocks, we thought it would make sense to somehow combine two factors, e.g. ffo_to_ev and inventory_to_sales, and construct our portfolios from this. One suggestion might be to take cross sections of the rankings and only to buy stocks that are ranked in the top $x\%$ by both rankings, similar to Fama and French's methodology in constructing their 6 portfolios for their HML and SMB calculation. The thinking behind this is that, if both factors give superior returns when used individually, when used in unison, if a stock is ranked in the top $x\%$ for both, then it really is likely to have strong future performance. Another idea might be to combine the metrics used to induce the rankings in some way, e.g. scale each to [0,1] and take the weighted average, based on the historical `avg_monthly_spearman`. We could write a whole other thesis on how to combine ranking systems, so we'll stop with this idea here.

As we saw, yearly rebalanced portfolios were able to achieve far superior returns than any of the quarterly or monthly rebalanced portfolios. The problem was that investors usually don't want to be locked into holding stocks for a year at a time. One, sort of, solution that we had for this, would be to split the funds into 12 equal sized portfolios, rebalance them yearly, but stagger them so that there is a different portfolio rebalanced each month. This has the added bonus that it makes us less susceptible to our choice of when we rebalance within the year. Were we only to rebalance once within the year, we might be doing so on "stale" data; if we rebalance a few days before a company, which reports annually, released their financial statement, for example. How to redistribute funds such that the portfolios were all of roughly equal cash size is slightly more difficult.

Similar to the above suggestion, one could stagger the initial investments by 1-month each, and, after a year, apply weighting tilts, up to a maximum of some amount, so that our weights are constantly updated (more like a monthly strategy), but we are still exposed to companies for a year at a time. This is roughly the same as the previous suggestion, but deals with the reallocation of funds. E.g. set the weight such that each month, the new weight is the arithmetic average of the trailing 12-months' weights, had we been rebalancing monthly[2][3]. For a numerical example of this, see Example B.9.1.

It would also be interesting to perform some analysis to see how much each strategy's performance is due-to/affected-by the performance of USD. Given we are converting both prices and financial statements to USD in our analysis, the strength of USD vs other currencies might play a big part in explaining returns. Toyota stock might not change in price, but if USD weakens vs JPY, this would show up to us as positive return on Toyota, for example. Perhaps, for purely research purposes, it might be more useful to perform the analysis based purely on each stock's returns, within its own currency.

Finally, although it was not used here, we are strong supporters of the `backtrader` package, availble from `https://www.backtrader.com/`, with its plethora of features. Although it can be difficult to use at times with a steep learning curve, the pros far outweigh the cons in its use in analysing backtested trading strategies. Had we had more time to work on this study, porting our weights timeseries into the backtrader system would have been our next course of action in improving our analysis/code.

7.3.2 Factor modelling

Unfortunately, due to a late discovery of some errors in the data we received, we were not able to devote as much time to tuning of our predictive models as we would have liked and had many more approaches/techniques we would have like to use. We discuss some of thes here.

For any input/target that is scale dependant, e.g. Δ-differences or a level, it would make sense to scale these by some notion of the relevant companies size. This makes your data more comparable and should help in training. An example would be to follow the methodology in

[2]I.e. apply `.rolling(12, min_periods=1).mean()` to the monthly rebalancing timeseries of weights, for those of you familiar with `pandas`.

[3]Although we didn't perform the analysis on all of our strategies, as a quick example, if we apply this TTM transformation to the top performing strategy at 0-clairvoyance (Y.E.L.q5.inventory_to_ev), we get a CAR of 38% and a sharpe of 1.48, only slightly worse than the yearly rebalanced version, but now with monthly weight adjustments.

[Alberg and Lipton, 2017] and divide all balance sheet items by the most recently available market cap for each training period. similar adjustments would need to be made for the IHSM data.

Another approach from [Alberg and Lipton, 2017] that we were not able to attempt was multi-target regression. Multi-target regression is a way of reducing noise and hopefully preventing overfitting. It is particularly useful in neural networks when complex relationships between variables are developed as hidden layer features, not so much in multi-target linear regression, where the optimisation boils down to n-seperate single-target linear regressions.

Although we did not find any models outperformant of the naïve benchmark in our first attempt at modelling, one method which seemed to improve test accuracy was, due to the infrequency of our data, splining it in some way to artificially create training data. Due to time restrictions, this was not implemented in the second round of modelling, but we would be interested to see its effect, perhaps in a more general framework.

As mentioned earlier, the modelling we performed essentially boiled down to a classification of being in or out of the top quantile. Although we attempted to do so through regression, with hindsight, doing so through classification models probably would have yielded better results, as our models were not optimising for high classification rates. Doing so would require a large amount of re-engineering of features, making them all relative to the other 30 companies at that point in time. Failing to do so could create one-to-many mappings in the training set, which is generally to be avoided if possible[4].

Finally, on the topic of alternative methods of optimisation, this is also something we considered implementing in our tests. If we wished to stick with a regression model, we could still have swapped the loss function to be based on the final classification rate of the regression model and optimised this instead. An example might be to optimise the same loss function as OLS, but where the loss is only counted if it is "close" to the cut off of the top quantile in some way. This way we are not overfitting on the data from companies that are nowhere near being classified as inside the top quantile.

[4] Imagine company A's fundamentals don't change from one month to the next, but all the others do. It could be that A was previously classified as being in the top quantile and now as not. Structuring your data such that this can/is likely to occur in your training set is generally a bad idea.

7.3.3 Other suggestions

Finally, although it was not done here, we would highly recommend the use of the `Julia` programming language, for its flexibility, like `Python`, and its high performance, like `C`. Many of the tests in this study were over an extremely large grid of combinations and took many hours to run[5]. With the use of `Julia`, this likely could have been greatly reduced. Many will complain of its lack of libraries/packages at this point in time, but it is only through the widespread adoption of a language that these libraries are built, and so we implore the reader to give `Julia` a try.

[5]The longest of which took roughly 50 hours whilst being run on 4 threads simultaneously.

Appendix A

Proofs

A.1 STD is sub-additive

Proof.

1.

$$\begin{aligned}\mathrm{Var}[X+Y] &= \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\rho(X,Y)\,\mathrm{STD}[X]\,\mathrm{STD}[Y] \\ &\leqslant \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\,\mathrm{STD}[X]\,\mathrm{STD}[Y]\end{aligned}$$

2. As $\mathrm{Var}[X] = \mathrm{STD}[X]^2$:

$$\begin{aligned}\mathrm{STD}[X+Y]^2 &\leqslant \mathrm{STD}[X]^2 + \mathrm{STD}[Y]^2 + 2\,\mathrm{STD}[X]\,\mathrm{STD}[Y] \\ &= (\mathrm{STD}[X] + \mathrm{STD}[Y])^2\end{aligned}$$

3. Taking square roots of both sides:

$$\mathrm{STD}[X+Y] \leqslant \mathrm{STD}[X] + \mathrm{STD}[Y]$$

□

Appendix B

Data

B.1 IHSM data declaration

During the collection and curation of the IHSM data, the information is not made available at the same time for each country. For example, some of the information for BMW in its German production plants is made available after a 1-month delay, where as it is only made available after a 2-month delay for its China plants. This would mean that, come 2010-03-01, we could, in theory, know all of the data on production from 2010-02-01 to 2010-02-28. Due to the reporting delays, however, for the plants in Germany, we do not receive this information until 2010-04-01. For the plants in China, we do not receive this information until 2010-05-01. When we attempted to model future ratios using what would be known at the time of training, e.g. including Germany 2010-02 production data when training on 2010-04-01 but not including China 2010-02 data, we failed to find any models more performant than the naïve benchmark. As such, we instead formatted the data such that it is indexed by the date it could **theoretically** be known, i.e. the first business day of the month following the relevant period, 2010-03-01 for 2010-02 data, for example. Although this means that any of our predictive strategies could not be implemented in real time as a trading strategy right now, this is something that may change in the future. With the ever increasing demand for accurate and fast data streams and the fact that these delays are government imposed, rather than by feasibility, we are confident that, at some point in the future, this information will be available the day (or perhaps days) following the period in question.

B.2 Companies

1. Bayer Motoren Werke AG
2. Brilliance China Automotive Holdings Ltd
3. BYD Company Ltd
4. China Motor Corp
5. Chongqing Changan Automobile
6. Daimler AG
7. Dongfeng Motor Group Co Ltd
8. Faw Car Co Ltd
9. Fiat Chrysler Automobiles NV
10. Ford Motor Co
11. Geely Automobile Holdings Ltd
12. Motors Liquidation Company
13. General Motors Co
14. Great Wall Motor Co Ltd
15. Guangzhou Automobile Group Co Ltd
16. Honda Motor Co Ltd
17. Hyundai Motor Co
18. Kia Motors Corp
19. Mahindra & Mahindra Ltd
20. Mazda Motor Corp
21. Mitsubishi Motors Corp
22. Nissan Motor Co Ltd
23. Peugeot SA
24. Renault SA
25. SAIC Motor Corp Ltd
26. Subaru Corp
27. Suzuki Motor Corp
28. Tata Motors Ltd
29. Tesla Inc
30. Toyota Motor Corp
31. Volkswagen AG

B.3 Fundamentals

B.3.1 Description of financial statement items

- accpayable: Accounts payable - When a company purchases goods on credit which needs to be paid back in a short period of time, it is known as Accounts Payable. It is treated as a liability and comes under the head current liabilities. Accounts Payable is a short-term debt payment which needs to be paid to avoid default.
- book: Book-value - Book value is the net asset value of a company calculated as total assets minus intangible assets (patents, goodwill) and liabilities.
- cogs: Cost of goods sold (COGS) - COGS is the direct costs attributable to the production of the goods sold in a company. This amount includes the cost of the materials used in creating the good along with the direct labor costs used to produce the good. It excludes indirect expenses such as distribution costs and sales force costs.
- currliab: Current liabilities - Liabilities are funds owed by the business, and are broken down into current and long-term categories. Current liabilities are those due within one year and includes items such as:
  - Accounts payable
  - Wages
  - Income tax deductions
  - Pension plan contributions
  - Medical plan payments
  - Building and equipment rents
  - Customer deposits
  - Utilities
  - Temporary loans, lines of credit, or overdrafts
  - Interest
  - Maturing debt
  - Sakes tax and/or goods and services tax charged on purchases
- ebit: Earnings before interest and tax (EBIT) - EBIT = Net Income + Interest + Taxes or EBIT = Revenue - Operating Expenses
- equity: Shareholders equity - Shareholders equity is broken down into two parts:
  - Share capital - This is the value of funds that shareholders have invested in the

company. When a company is first formed, shareholders will typically put in cash. For example, an investor starts a company and seeds it with $10M. Cash (an asset) rises by $10M, and Share Capital (an equity account) rises by $10M, balancing out the balance sheet.

  – Retained earnings - This is the total amount of net income the company decides to keep. Every period, a company may pay out dividends from its net income. Any amount remaining (or exceeding) is added to (deducted from) retained earnings.

- ev: Enterprise value (EV) - EV is a measure of a company's total value, often used as a more comprehensive alternative to equity market capitalization. Enterprise value is calculated as the market capitalization plus debt, minority interest and preferred shares, minus total cash and cash equivalents. EV = market value of common stock + market value of preferred equity + market value of debt + minority interest - cash and investments.
- ffo: Funds from operations (FFO) - FFO represents the sum of net income and all non-cash charges or credits. It is the cash flow of the company.
- inventory: Inventory - Inventory includes amounts for raw materials, work-in-progress goods and finished goods. The company uses this account when it reports sales of goods, generally under cost of goods sold in the income statement.
- netincome: Net income - Net income is equal to net earnings (profit) calculated as sales less cost of goods sold, selling, general and administrative expenses, operating expenses, depreciation, interest, taxes and other expenses.
- sales: Sales/revenues - Revenue is the amount of money that a company actually receives during a specific period, including discounts and deductions for returned merchandise. It is the top line or gross income figure from which costs are subtracted to determine net income. Revenue is calculated by multiplying the price at which goods or services are sold by the number of units or amount sold. Revenue is also known as sales on the income statement.
- totassets: Total assets - Total assets represent the sum of total current assets, long term receivables, investment in unconsolidated subsidiaries, other investments, net property plant and equipment and other assets. An asset is a resource with economic value that an individual, corporation or country owns.
- totliab: Total liabilities - Total liabilities refer to the aggregate of all debts for which an individual or company is liable.

B.3.2 Ratios used

- book_to_market - book-value divided by market-cap;
- ffo_to_sales - funds from operations (FFO) divided by sales;
- ffo_to_totassets - FFO divided by total assets;
- ffo_to_currliab - FFO divided by current liabilities;
- ffo_to_cogs - FFO divided by cost of goods sold (COGS);
- ffo_to_inventory - FFO divided by inventory;
- ffo_to_ev - FFO divided by enterprise value (EV);
- inventory_to_sales - inventory divided by sales;
- inventory_to_totassets - inventory divided by total assets;
- inventory_to_currliab - inventory divided by current liabilities;
- inventory_to_cogs - inventory divided by COGS;
- inventory_to_ebit - inventory divided by earnings before interest and tax (EBIT);
- inventory_to_ev - inventory divided by EV;
- cogs_to_ev - COGS divided by EV;
- ebit_to_ev - EBIT divided by EV;
- equity_to_ev - common equity divided by EV;
- netincome_to_ev - net income divided by EV;
- sales_to_ev - sales divided by EV;
- totassets_to_ev - total assets divided by EV;
- pd_accpayable_to_pd_totassets - percentage change in accounts payable since previous report divided by percentage change in total assets since previous report;
- d_totliab_to_lagged_assets - difference in total liabilities since previous report divided by total assets at time of previous report;
- d_totliab_to_lagged_equity - difference in total liabilities since previous report divided by common equity at time of previous report;
- accpayable_to_lagged_sales - accounts payable divided by sales at time of previous report; and
- accpayable_to_lagged_cogs - accounts payable divided by COGS at time of previous report.

B.4 Results

All results can be found HERE.

https://drive.google.com/file/d/1IB1PAEveqQcCHSFgvIh7ouXjaGbd5zSg/view?usp=sharing

- `benchmarks` - timeseries of returns of each of the benchmarks.
- `drop_rate` - retention rates (and drop rates = 1 - retention rate) of the top ranked strategies.
- `FF_factors` - published Fama-French factors and our calculated factors timeseries.
- `model_factor_regressions` - regressions of each our models against the Fama-French factors.
- `s1_results` - results of all parameter combinations tested in stage 1
- `turnover` - similar to `drop_rate` but with strategy turnovers rather than retention rates.
- `two_input_regression_results` - all results from regressing $y_{t+c} = \beta_1 y_t + \beta_2 x_t$.

B.5 Turnovers

R	W	S	Q	ranking_factor	turnover
Y	E	L	5	inventory_to_ev	0.42
Y	E	L	4	accpayable_to_lagged_sales	0.43
Y	E	L	5	equity_to_ev	0.39
Y	E	L	4	accpayable_to_lagged_cogs	0.44
Y	E	L	4	inventory_to_ev	0.43
Y	E	L	5	accpayable_to_lagged_cogs	0.51
Y	E	L	5	totassets_to_ev	0.49
Y	E	L	4	sales_to_ev	0.34
Y	E	L	3	accpayable_to_lagged_cogs	0.36
Y	E	L	5	accpayable_to_lagged_sales	0.49
Q	E	L	5	inventory_to_ev	0.23
Q	E	L	5	inventory_to_totassets	0.21
Q	E	L	5	accpayable_to_lagged_sales	0.29
Q	E	L	4	accpayable_to_lagged_sales	0.23
Q	E	L	5	totassets_to_ev	0.27
Q	E	L	4	totassets_to_ev	0.25
Q	E	L	4	inventory_to_ev	0.22
Q	E	L	5	inventory_to_currliab	0.24
Q	E	L	5	accpayable_to_lagged_cogs	0.31
Q	E	L	4	accpayable_to_lagged_cogs	0.26
M	E	L	5	inventory_to_totassets	0.10
M	E	L	5	totassets_to_ev	0.12
M	E	L	5	inventory_to_ev	0.10
M	E	L	4	inventory_to_ev	0.10
M	E	L	4	totassets_to_ev	0.11
M	E	L	4	accpayable_to_lagged_sales	0.11
M	E	L	4	accpayable_to_lagged_cogs	0.12
M	E	L	5	accpayable_to_lagged_cogs	0.14
M	E	L	5	inventory_to_currliab	0.11
M	E	L	5	cogs_to_ev	0.12

Table B.1: Top performing 0-clairvoyance strategies turnover rates. Note: these are turnovers between rebalancing periods.

R	W	S	Q	ranking_factor	turnover
M	C	L	5	ffo_to_ev	0.25
Q	E	L	4	inventory_to_ev	0.22
Q	E	L	5	accpayable_to_lagged_sales	0.29
M	C	L	5	inventory_to_totassets	0.10
Q	E	L	5	inventory_to_totassets	0.21
M	C	L	5	ffo_to_ev	0.25
Q	E	L	4	inventory_to_ev	0.22
M	C	L	5	ffo_to_totassets	0.21
M	E	L	5	ffo_to_ev	0.18
M	C	L	4	ffo_to_totassets	0.24
M	E	L	5	ffo_to_inventory	0.17
M	C	L	4	netincome_to_ev	0.23
Q	C	L	4	netincome_to_ev	0.49
Q	C	L	5	ffo_to_ev	0.52
Q	E	L	4	netincome_to_ev	0.32
Q	C	L	4	netincome_to_ev	0.49
M	C	L	4	ffo_to_totassets	0.24
Q	C	L	4	ffo_to_totassets	0.59
Q	C	L	4	inventory_to_totassets	0.16
M	C	L	4	ffo_to_inventory	0.29
M	C	L	4	ffo_to_totassets	0.24
Q	C	L	3	accpayable_to_lagged_cogs	0.26
Q	C	L	4	inventory_to_totassets	0.16
Q	C	L	5	ffo_to_inventory	0.59
Q	C	L	3	inventory_to_totassets	0.17
M	C	L	3	ebit_to_ev	0.18
M	C	L	4	ffo_to_totassets	0.24
Q	C	L	4	inventory_to_totassets	0.16
M	E	L	3	ffo_to_inventory	0.15
Q	C	L	4	inventory_to_currliab	0.28

Table B.2: Top performing X-clairvoyance strategies turnover rates. Note: these are turnovers between rebalancing periods.

B.6 Retention rates

R	W	S	Q	ranking_factor	r_1	r_2	r_3	r_4	r_5	r_6
Y	E	L	5	inventory_to_ev	0.70	0.67	0.67	0.71	0.92	
Y	E	L	4	accpayable_to_lagged_sales	0.73	0.69	0.70	0.79	0.92	
Y	E	L	5	equity_to_ev	0.75	0.70	0.69	0.76	0.89	
Y	E	L	4	accpayable_to_lagged_cogs	0.73	0.69	0.72	0.79	0.92	
Y	E	L	4	inventory_to_ev	0.70	0.72	0.73	0.80	0.94	
Y	E	L	5	accpayable_to_lagged_cogs	0.63	0.61	0.71	0.76	0.92	
Y	E	L	5	totassets_to_ev	0.62	0.63	0.67	0.67	0.92	
Y	E	L	4	sales_to_ev	0.78	0.68	0.69	0.75	0.94	
Y	E	L	3	accpayable_to_lagged_cogs	0.79	0.77	0.75	0.84	0.95	
Y	E	L	5	accpayable_to_lagged_sales	0.63	0.59	0.67	0.74	0.89	
Q	E	L	5	inventory_to_ev	0.83	0.78	0.73	0.71	0.69	0.66
Q	E	L	5	inventory_to_totassets	0.85	0.81	0.79	0.79	0.75	0.74
Q	E	L	5	accpayable_to_lagged_sales	0.77	0.70	0.66	0.65	0.62	0.62
Q	E	L	4	accpayable_to_lagged_sales	0.84	0.80	0.76	0.73	0.72	0.72
Q	E	L	5	totassets_to_ev	0.78	0.71	0.66	0.62	0.58	0.56
Q	E	L	4	totassets_to_ev	0.81	0.76	0.73	0.69	0.66	0.67
Q	E	L	4	inventory_to_ev	0.84	0.80	0.78	0.75	0.75	0.75
Q	E	L	5	inventory_to_currliab	0.82	0.79	0.75	0.80	0.75	0.76
Q	E	L	5	accpayable_to_lagged_cogs	0.75	0.67	0.65	0.64	0.60	0.60
Q	E	L	4	accpayable_to_lagged_cogs	0.81	0.75	0.72	0.72	0.70	0.70
M	E	L	5	inventory_to_totassets	0.93	0.89	0.85	0.83	0.82	0.80
M	E	L	5	totassets_to_ev	0.92	0.85	0.79	0.75	0.72	0.70
M	E	L	5	inventory_to_ev	0.93	0.89	0.85	0.83	0.81	0.79
M	E	L	4	inventory_to_ev	0.94	0.89	0.85	0.83	0.81	0.80
M	E	L	4	totassets_to_ev	0.92	0.87	0.81	0.79	0.77	0.75
M	E	L	4	accpayable_to_lagged_sales	0.93	0.89	0.85	0.83	0.81	0.80
M	E	L	4	accpayable_to_lagged_cogs	0.92	0.86	0.82	0.79	0.77	0.76
M	E	L	5	accpayable_to_lagged_cogs	0.90	0.82	0.76	0.73	0.70	0.69
M	E	L	5	inventory_to_currliab	0.92	0.87	0.83	0.81	0.80	0.79
M	E	L	5	cogs_to_ev	0.91	0.86	0.82	0.78	0.76	0.75
				mean	0.81	0.77	0.75	0.76	0.79	0.72

Table B.3: Top performing 0-clairvoyance strategies x-month retention rates for $x \in \{1, \ldots, 6\}$

R	W	S	Q	X	ranking_factor	r_X
M	C	L	5	1	ffo_to_ev	0.85
Q	E	L	4	1	inventory_to_ev	0.84
Q	E	L	5	1	accpayable_to_lagged_sales	0.77
M	C	L	5	1	inventory_to_totassets	0.93
Q	E	L	5	1	inventory_to_totassets	0.85
M	C	L	5	2	ffo_to_ev	0.73
Q	E	L	4	2	inventory_to_ev	0.80
M	C	L	5	2	ffo_to_totassets	0.78
M	E	L	5	2	ffo_to_ev	0.73
M	C	L	4	2	ffo_to_totassets	0.77
M	E	L	5	3	ffo_to_inventory	0.69
M	C	L	4	3	netincome_to_ev	0.74
Q	C	L	4	3	netincome_to_ev	0.67
Q	C	L	5	3	ffo_to_ev	0.56
Q	E	L	4	3	netincome_to_ev	0.67
Q	C	L	4	4	netincome_to_ev	0.68
M	C	L	4	4	ffo_to_totassets	0.65
Q	C	L	4	4	ffo_to_totassets	0.66
Q	C	L	4	4	inventory_to_totassets	0.83
M	C	L	4	4	ffo_to_inventory	0.66
M	C	L	4	5	ffo_to_totassets	0.63
Q	C	L	3	5	accpayable_to_lagged_cogs	0.79
Q	C	L	4	5	inventory_to_totassets	0.81
Q	C	L	5	5	ffo_to_inventory	0.63
Q	C	L	3	5	inventory_to_totassets	0.83
M	C	L	3	6	ebit_to_ev	0.74
M	C	L	4	6	ffo_to_totassets	0.63
Q	C	L	4	6	inventory_to_totassets	0.82
M	E	L	3	6	ffo_to_inventory	0.67
Q	C	L	4	6	inventory_to_currliab	0.79

Table B.4: Top performing X-clairvoyance strategies retention rates. r_X stands for the strategies average x-month retention rate.

B.7 Fama-French historical factors

Date	mkt-rf	SMB	HML	rf
2008-01	-0.0674	0.0078	-0.0158	0.0042
2008-02	-0.0211	0.0017	-0.0157	0.0032
2008-03	-0.0287	-0.0493	0.0774	0.0032
2008-04	0.0273	-0.0007	0.0289	0.0034
2008-05	-0.0113	-0.0023	0.1160	0.0027
2008-06	-0.1157	-0.0419	0.1192	0.0021
2008-07	-0.0398	0.0901	-0.0379	0.0013
2008-08	-0.0022	-0.0513	0.0450	0.0017
2008-09	-0.0637	-0.0033	-0.1105	0.0018
2008-10	-0.1639	-0.1899	-0.1768	0.0018
2008-11	-0.1862	0.0620	0.1004	0.0017
2008-12	0.0718	-0.0503	0.0017	0.0015
2009-01	-0.0797	0.1319	-0.1107	0.0013
2009-02	-0.0728	0.0480	0.0616	0.0015
2009-03	0.2037	-0.0046	0.1099	0.0008
2009-04	0.2386	0.0722	0.1535	0.0003
2009-05	0.0727	0.0220	0.0198	0.0000
2009-06	-0.0135	-0.0189	-0.0438	0.0000
2009-07	0.1782	0.0115	-0.0555	0.0001
2009-08	-0.1226	0.0549	0.0851	0.0002
2009-09	-0.0306	-0.0107	-0.1036	0.0001
2009-10	0.0352	0.0372	-0.0757	0.0000
2009-11	-0.0069	0.1237	-0.1628	0.0001
2009-12	0.0094	0.0279	-0.0609	0.0001
2010-01	-0.0905	0.0457	0.0389	0.0001
2010-02	-0.0248	0.0455	-0.1291	0.0001
2010-03	0.0898	0.0616	0.0807	0.0000
2010-04	-0.0077	-0.0400	0.0518	0.0000
2010-05	-0.0959	-0.0105	0.0049	0.0001
2010-06	-0.0260	0.0213	0.0301	0.0000
2010-07	0.1164	-0.0626	-0.0220	0.0000
2010-08	-0.0388	-0.0230	0.0010	0.0001
2010-09	0.1316	0.1347	-0.0017	0.0001
2010-10	0.0525	-0.0575	-0.0153	0.0001
2010-11	0.0666	-0.0327	0.0080	0.0001
2010-12	0.0209	-0.1070	0.0583	0.0001
2011-01	0.0371	0.0205	0.0915	0.0001
2011-02	0.0250	-0.0546	-0.0013	0.0001
2011-03	-0.0548	-0.0934	-0.0880	0.0001
2011-04	0.0793	-0.0642	0.1102	0.0001
2011-05	-0.0371	0.0162	0.0695	0.0001
2011-06	0.0301	0.0274	0.0303	0.0001
2011-07	-0.0168	0.0341	-0.0618	0.0001
2011-08	-0.1642	-0.0004	-0.0748	0.0001
2011-09	-0.1097	-0.0783	-0.0106	0.0000
2011-10	0.0986	-0.0042	0.0414	0.0000
2011-11	-0.0334	0.0033	-0.0043	0.0000
2011-12	-0.0341	-0.0190	0.0127	0.0000

Date	mkt-rf	SMB	HML	rf
2012-01	0.1735	-0.0314	0.0943	0.0001
2012-02	0.0654	0.0006	-0.0219	0.0000
2012-03	0.0162	-0.0456	0.0288	0.0000
2012-04	-0.0488	-0.0377	-0.0412	0.0000
2012-05	-0.1087	0.0120	0.0009	0.0000
2012-06	0.0044	-0.0088	0.0492	0.0000
2012-07	-0.0118	-0.0872	0.0474	0.0000
2012-08	0.0122	0.0063	0.0046	0.0000
2012-09	0.0176	-0.0156	-0.0138	0.0000
2012-10	0.0203	0.0058	0.0504	0.0001
2012-11	0.0734	0.0261	-0.0066	0.0000
2012-12	0.0785	0.0320	0.0560	0.0000
2013-01	0.0345	0.1001	0.1431	0.0001
2013-02	-0.0106	0.0091	-0.0754	0.0001
2013-03	-0.0259	-0.0112	-0.0153	0.0001
2013-04	0.0655	0.0749	-0.0333	0.0001
2013-05	0.0668	0.0732	-0.0149	0.0001
2013-06	-0.0169	-0.0322	-0.0812	0.0000
2013-07	0.0633	-0.0096	0.0760	0.0000
2013-08	-0.0216	0.0146	0.0750	0.0000
2013-09	0.0731	0.0337	-0.0086	0.0000
2013-10	0.0162	0.0114	0.0104	0.0000
2013-11	0.0068	0.0318	0.0478	0.0000
2013-12	-0.0091	0.0062	-0.0128	0.0000
2014-01	-0.0587	0.0276	0.0145	0.0000
2014-02	0.0339	0.0007	-0.0340	0.0000
2014-03	0.0217	0.0271	0.0420	0.0000
2014-04	-0.0145	-0.0079	-0.0155	0.0000
2014-05	0.0198	0.0043	-0.0130	0.0000
2014-06	0.0377	0.0002	-0.0505	0.0000
2014-07	-0.0384	0.0488	-0.0241	0.0000
2014-08	-0.0039	-0.0022	-0.0554	0.0000
2014-09	-0.0439	0.0428	-0.0240	0.0000
2014-10	-0.0226	0.0278	0.0708	0.0000
2014-11	0.0706	-0.0378	0.0290	0.0000
2014-12	-0.0224	-0.0127	0.0589	0.0000
2015-01	0.0330	0.0137	0.0339	0.0000
2015-02	0.0634	0.0231	-0.0026	0.0000
2015-03	-0.0059	0.0086	-0.0007	0.0000
2015-04	0.0189	0.0515	0.0303	0.0000
2015-05	-0.0192	0.0010	-0.0956	0.0000
2015-06	-0.0344	-0.0397	-0.0593	0.0000
2015-07	-0.0397	-0.0397	-0.0840	0.0000
2015-08	-0.1052	-0.0011	0.0564	0.0000
2015-09	-0.0139	0.0819	-0.0124	0.0000
2015-10	0.1054	0.0605	0.0311	0.0000
2015-11	0.0270	-0.0131	-0.0082	0.0000
2015-12	-0.0153	-0.0149	-0.0083	0.0000

Table B.5: Our calculated automotive universe Fama-French factors.

Date	mkt-rf	SMB	HML	rf
2016-01	-0.1206	-0.0250	-0.0429	0.0000
2016-02	-0.0382	0.0102	0.1000	0.0000
2016-03	0.0413	0.0333	-0.0425	0.0000
2016-04	0.0129	-0.0204	0.0246	0.0000
2016-05	-0.0063	0.0162	-0.0420	0.0001
2016-06	-0.0624	0.0061	-0.0102	0.0001
2016-07	0.1002	0.0458	-0.0080	0.0002
2016-08	0.0219	0.0167	-0.0078	0.0002
2016-09	-0.0217	-0.0031	-0.0128	0.0001
2016-10	0.0134	0.0368	0.0213	0.0001
2016-11	-0.0105	-0.0093	0.0004	0.0002
2016-12	0.0426	-0.0040	0.0157	0.0002
2017-01	0.0202	0.0244	-0.0345	0.0002
2017-02	0.0056	0.0780	0.0219	0.0002
2017-03	-0.0196	-0.0278	-0.0903	0.0002
2017-04	0.0122	-0.0166	-0.0196	0.0001
2017-05	0.0118	0.0239	-0.0231	0.0003
2017-06	-0.0009	0.0490	-0.0452	0.0004

Table B.5: Our calculated automotive universe Fama-French factors.

B.8 Regressions vs Fama-French factors

R	W	S	Q	ranking_factor	R^2	β_{mkt}	β_{SMB}	β_{HML}
M	C			benchmark	0.563	1.019	-0.013	-0.026
Y	E			benchmark	0.512	1.415	-0.199	-0.169
Y	E	L	5	inventory_to_ev	0.306	1.801	-0.402	-0.874
Y	E	L	4	accpayable_to_lagged_sales	0.320	1.903	-0.331	-0.952
Y	E	L	5	equity_to_ev	0.306	1.836	-0.684	-0.868
Y	E	L	4	accpayable_to_lagged_cogs	0.329	1.898	-0.302	-0.982
Y	E	L	4	inventory_to_ev	0.365	1.855	-0.524	-0.875
Y	E	L	5	accpayable_to_lagged_cogs	0.334	1.888	-0.068	-0.760
Y	E	L	5	totassets_to_ev	0.301	1.800	-0.330	-0.955
Y	E	L	4	sales_to_ev	0.376	1.865	-0.360	-0.949
Y	E	L	3	accpayable_to_lagged_cogs	0.346	1.791	-0.202	-1.009
Y	E	L	5	accpayable_to_lagged_sales	0.336	2.074	-0.493	-0.984
Q	E	L	5	inventory_to_ev	0.461	1.649	-0.359	-0.368
Q	E	L	5	inventory_to_totassets	0.403	1.324	0.159	-0.203
Q	E	L	5	accpayable_to_lagged_sales	0.368	1.845	-0.648	-0.274
Q	E	L	4	accpayable_to_lagged_sales	0.397	1.767	-0.669	-0.404
Q	E	L	5	totassets_to_ev	0.360	1.558	-0.546	-0.451
Q	E	L	4	totassets_to_ev	0.442	1.675	-0.571	-0.462
Q	E	L	4	inventory_to_ev	0.507	1.688	-0.440	-0.336
Q	E	L	5	inventory_to_currliab	0.560	1.568	0.190	-0.209
Q	E	L	5	accpayable_to_lagged_cogs	0.421	1.765	-0.491	-0.058
Q	E	L	4	accpayable_to_lagged_cogs	0.385	1.706	-0.619	-0.212
M	E	L	5	inventory_to_totassets	0.366	1.252	0.069	-0.069
M	E	L	5	totassets_to_ev	0.406	1.647	-0.483	-0.498
M	E	L	5	inventory_to_ev	0.497	1.727	-0.428	-0.348
M	E	L	4	inventory_to_ev	0.525	1.718	-0.402	-0.392
M	E	L	4	totassets_to_ev	0.455	1.713	-0.546	-0.492
M	E	L	4	accpayable_to_lagged_sales	0.401	1.714	-0.769	-0.284
M	E	L	4	accpayable_to_lagged_cogs	0.417	1.786	-0.662	-0.200
M	E	L	5	accpayable_to_lagged_cogs	0.477	1.924	-0.628	-0.046
M	E	L	5	inventory_to_currliab	0.500	1.356	0.221	0.021
M	E	L	5	cogs_to_ev	0.409	1.624	-0.357	-0.556

Table B.6: Top performing 0-clairvoyance strategies regressed vs official Fama-French factors results.

R	W	S	Q	ranking_factor	R^2	β_{mkt}	β_{SMB}	β_{HML}
M	C			benchmark	0.987	0.967	-0.024	-0.091
Y	E			benchmark	0.905	1.096	0.436	-0.126
Y	E	L	5	inventory_to_ev	0.655	1.120	0.685	-0.134
Y	E	L	4	accpayable_to_lagged_sales	0.645	1.142	0.792	-0.187
Y	E	L	5	equity_to_ev	0.668	1.100	0.844	-0.210
Y	E	L	4	accpayable_to_lagged_cogs	0.648	1.134	0.771	-0.198
Y	E	L	4	inventory_to_ev	0.694	1.113	0.671	-0.093
Y	E	L	5	accpayable_to_lagged_cogs	0.638	1.096	0.862	-0.052
Y	E	L	5	totassets_to_ev	0.644	1.108	0.741	-0.082
Y	E	L	4	sales_to_ev	0.782	1.263	0.701	-0.124
Y	E	L	3	accpayable_to_lagged_cogs	0.699	1.100	0.760	-0.219
Y	E	L	5	accpayable_to_lagged_sales	0.647	1.222	0.821	-0.238
Q	E	L	5	inventory_to_ev	0.704	0.996	0.585	0.079
Q	E	L	5	inventory_to_totassets	0.690	0.997	0.456	-0.068
Q	E	L	5	accpayable_to_lagged_sales	0.685	1.207	0.869	-0.192
Q	E	L	4	accpayable_to_lagged_sales	0.704	1.154	0.785	-0.187
Q	E	L	5	totassets_to_ev	0.626	0.974	0.575	0.034
Q	E	L	4	totassets_to_ev	0.710	1.021	0.600	0.080
Q	E	L	4	inventory_to_ev	0.779	1.021	0.638	0.071
Q	E	L	5	inventory_to_currliab	0.735	1.022	0.437	0.012
Q	E	L	5	accpayable_to_lagged_cogs	0.710	1.164	0.749	-0.052
Q	E	L	4	accpayable_to_lagged_cogs	0.719	1.164	0.753	-0.122
M	E	L	5	inventory_to_totassets	0.679	0.985	0.425	-0.037
M	E	L	5	totassets_to_ev	0.633	1.032	0.516	0.027
M	E	L	5	inventory_to_ev	0.715	1.062	0.530	0.115
M	E	L	4	inventory_to_ev	0.766	1.030	0.613	0.143
M	E	L	4	totassets_to_ev	0.697	1.085	0.517	0.070
M	E	L	4	accpayable_to_lagged_sales	0.692	1.136	0.763	-0.135
M	E	L	4	accpayable_to_lagged_cogs	0.725	1.186	0.801	-0.092
M	E	L	5	accpayable_to_lagged_cogs	0.712	1.234	0.781	-0.038
M	E	L	5	inventory_to_currliab	0.749	0.963	0.400	0.018
M	E	L	5	cogs_to_ev	0.729	1.203	0.607	-0.121

Table B.7: Top performing 0-clairvoyance strategies regressed vs automotive Fama-French factors results.

R	W	S	Q	C	ranking_factor	R^2	β_{mkt}	β_{SMB}	β_{HML}
Y	E				benchmark	0.521	1.437	-0.274	-0.183
M	C				benchmark	0.563	1.019	-0.013	-0.026
M	C	L	5	0	ffo_to_ev	0.352	1.474	-0.551	-0.399
M	C	L	5	1	ffo_to_ev	0.335	1.528	-0.454	-0.315
Q	E	L	4	0	inventory_to_ev	0.507	1.688	-0.440	-0.336
Q	E	L	4	1	inventory_to_ev	0.495	1.671	-0.476	-0.390
Q	E	L	5	0	accpayable_to_lagged_sales	0.368	1.845	-0.648	-0.274
Q	E	L	5	1	accpayable_to_lagged_sales	0.392	1.992	-0.857	-0.292
M	C	L	5	0	inventory_to_totassets	0.518	1.375	-0.009	-0.297
M	C	L	5	1	inventory_to_totassets	0.450	1.434	-0.039	-0.258
Q	E	L	5	0	inventory_to_totassets	0.403	1.324	0.159	-0.203
Q	E	L	5	1	inventory_to_totassets	0.381	1.357	0.073	-0.039
M	C	L	5	0	ffo_to_ev	0.352	1.474	-0.551	-0.399
M	C	L	5	2	ffo_to_ev	0.388	1.592	-0.610	-0.324
Q	E	L	4	0	inventory_to_ev	0.507	1.688	-0.440	-0.336
Q	E	L	4	2	inventory_to_ev	0.490	1.646	-0.468	-0.356
M	C	L	5	0	ffo_to_totassets	0.315	1.585	-0.638	-0.327
M	C	L	5	2	ffo_to_totassets	0.303	1.621	-0.472	-0.461
M	E	L	5	0	ffo_to_ev	0.317	1.636	-0.883	-0.537
M	E	L	5	2	ffo_to_ev	0.330	1.690	-0.951	-0.359
M	C	L	4	0	ffo_to_totassets	0.295	1.430	-0.643	-0.167
M	C	L	4	2	ffo_to_totassets	0.331	1.514	-0.582	-0.144
M	E	L	5	0	ffo_to_inventory	0.368	1.818	-0.616	-0.460
M	E	L	5	3	ffo_to_inventory	0.375	1.842	-0.737	-0.312
M	C	L	4	0	netincome_to_ev	0.500	1.776	-0.536	0.005
M	C	L	4	3	netincome_to_ev	0.481	1.919	-0.336	-0.451
Q	C	L	4	0	netincome_to_ev	0.504	1.672	-0.592	0.312
Q	C	L	4	3	netincome_to_ev	0.443	1.833	-0.344	-0.488
Q	C	L	5	0	ffo_to_ev	0.407	1.613	-0.491	-0.474
Q	C	L	5	3	ffo_to_ev	0.393	1.658	-0.642	-0.180
Q	E	L	4	0	netincome_to_ev	0.392	1.737	-0.891	-0.143
Q	E	L	4	3	netincome_to_ev	0.432	1.994	-0.595	-0.584
Q	C	L	4	0	netincome_to_ev	0.504	1.672	-0.592	0.312
Q	C	L	4	4	netincome_to_ev	0.431	2.022	0.107	-0.261
M	C	L	4	0	ffo_to_totassets	0.295	1.430	-0.643	-0.167
M	C	L	4	4	ffo_to_totassets	0.368	1.879	-0.584	0.223
Q	C	L	4	0	ffo_to_totassets	0.344	1.484	-0.342	-0.136
Q	C	L	4	4	ffo_to_totassets	0.350	1.775	-0.391	0.304
Q	C	L	4	0	inventory_to_totassets	0.606	1.481	0.031	-0.373
Q	C	L	4	4	inventory_to_totassets	0.555	1.603	0.032	-0.251

Table B.8: Top performing X-clairvoyance strategies regressed vs official Fama-French factors results.

R	W	S	Q	C	ranking_factor	R^2	β_{mkt}	β_{SMB}	β_{HML}
M	C	L	4	0	ffo_to_inventory	0.433	1.556	-0.132	-0.073
M	C	L	4	4	ffo_to_inventory	0.384	1.674	-0.164	-0.094
M	C	L	4	0	ffo_to_totassets	0.295	1.430	-0.643	-0.167
M	C	L	4	5	ffo_to_totassets	0.352	1.872	-0.359	0.149
Q	C	L	3	0	accpayable_to_lagged_cogs	0.491	1.514	-0.249	-0.253
Q	C	L	3	5	accpayable_to_lagged_cogs	0.473	1.565	-0.126	-0.204
Q	C	L	4	0	inventory_to_totassets	0.606	1.481	0.031	-0.373
Q	C	L	4	5	inventory_to_totassets	0.552	1.549	0.105	-0.257
Q	C	L	5	0	ffo_to_inventory	0.486	1.780	-0.078	-0.254
Q	C	L	5	5	ffo_to_inventory	0.362	1.864	0.057	0.138
Q	C	L	3	0	inventory_to_totassets	0.643	1.538	0.049	-0.370
Q	C	L	3	5	inventory_to_totassets	0.564	1.545	0.135	-0.345
M	C	L	3	0	ebit_to_ev	0.508	1.554	-0.537	-0.260
M	C	L	3	6	ebit_to_ev	0.143	0.959	0.201	-0.030
M	C	L	4	0	ffo_to_totassets	0.295	1.430	-0.643	-0.167
M	C	L	4	6	ffo_to_totassets	0.377	1.962	-0.458	0.249
Q	C	L	4	0	inventory_to_totassets	0.606	1.481	0.031	-0.373
Q	C	L	4	6	inventory_to_totassets	0.554	1.576	0.000	-0.207
M	E	L	3	0	ffo_to_inventory	0.366	1.616	-0.549	-0.349
M	E	L	3	6	ffo_to_inventory	0.383	1.704	-0.420	-0.292
Q	C	L	4	0	inventory_to_currliab	0.681	1.757	-0.026	-0.428
Q	C	L	4	6	inventory_to_currliab	0.549	1.645	-0.012	-0.240

Table B.8: Top performing X-clairvoyance strategies regressed vs official Fama-French factors results.

R	W	S	Q	C	ranking_factor	R^2	β_{mkt}	β_{SMB}	β_{HML}
Y	E				benchmark	0.905	1.097	0.434	-0.126
M	C				benchmark	0.987	0.967	-0.024	-0.091
M	C	L	5	0	ffo_to_ev	0.718	1.182	0.534	-0.240
M	C	L	5	1	ffo_to_ev	0.696	1.130	0.723	-0.124
Q	E	L	4	0	inventory_to_ev	0.779	1.021	0.638	0.071
Q	E	L	4	1	inventory_to_ev	0.772	1.016	0.587	0.091
Q	E	L	5	0	accpayable_to_lagged_sales	0.685	1.207	0.869	-0.192
Q	E	L	5	1	accpayable_to_lagged_sales	0.685	1.244	0.905	-0.144
M	C	L	5	0	inventory_to_totassets	0.620	0.953	0.226	-0.054
M	C	L	5	1	inventory_to_totassets	0.612	1.019	0.298	-0.082
Q	E	L	5	0	inventory_to_totassets	0.690	0.997	0.456	-0.068
Q	E	L	5	1	inventory_to_totassets	0.695	1.018	0.517	-0.124
M	C	L	5	0	ffo_to_ev	0.718	1.182	0.534	-0.240
M	C	L	5	2	ffo_to_ev	0.679	1.185	0.574	-0.033
Q	E	L	4	0	inventory_to_ev	0.779	1.021	0.638	0.071
Q	E	L	4	2	inventory_to_ev	0.762	1.026	0.538	0.070
M	C	L	5	0	ffo_to_totassets	0.705	1.211	0.681	-0.390
M	C	L	5	2	ffo_to_totassets	0.668	1.186	0.829	-0.312
M	E	L	5	0	ffo_to_ev	0.694	1.206	0.760	-0.260
M	E	L	5	2	ffo_to_ev	0.676	1.259	0.714	-0.155
M	C	L	4	0	ffo_to_totassets	0.708	1.215	0.548	-0.287
M	C	L	4	2	ffo_to_totassets	0.723	1.195	0.664	-0.265
M	E	L	5	0	ffo_to_inventory	0.734	1.338	0.736	-0.222
M	E	L	5	3	ffo_to_inventory	0.713	1.298	0.754	-0.186
M	C	L	4	0	netincome_to_ev	0.741	1.281	0.381	-0.088
M	C	L	4	3	netincome_to_ev	0.709	1.183	0.710	-0.156
Q	C	L	4	0	netincome_to_ev	0.769	1.341	0.349	-0.221
Q	C	L	4	3	netincome_to_ev	0.701	1.143	0.695	-0.096
Q	C	L	5	0	ffo_to_ev	0.751	1.251	0.533	-0.205
Q	C	L	5	3	ffo_to_ev	0.716	1.124	0.677	0.014
Q	E	L	4	0	netincome_to_ev	0.722	1.231	0.659	-0.264
Q	E	L	4	3	netincome_to_ev	0.703	1.210	0.788	-0.137
Q	C	L	4	0	netincome_to_ev	0.769	1.341	0.349	-0.221
Q	C	L	4	4	netincome_to_ev	0.670	1.342	0.669	-0.057
M	C	L	4	0	ffo_to_totassets	0.708	1.215	0.548	-0.287
M	C	L	4	4	ffo_to_totassets	0.692	1.348	0.704	-0.049
Q	C	L	4	0	ffo_to_totassets	0.755	1.262	0.543	-0.341
Q	C	L	4	4	ffo_to_totassets	0.714	1.332	0.757	-0.096
Q	C	L	4	0	inventory_to_totassets	0.732	0.999	0.234	-0.038
Q	C	L	4	4	inventory_to_totassets	0.707	1.072	0.298	-0.019

Table B.9: Top performing X-clairvoyance strategies regressed vs automotive Fama-French factors results.

R	W	S	Q	C	ranking_factor	R^2	β_{mkt}	β_{SMB}	β_{HML}
M	C	L	4	0	ffo_to_inventory	0.735	1.255	0.298	-0.180
M	C	L	4	4	ffo_to_inventory	0.680	1.282	0.414	-0.115
M	C	L	4	0	ffo_to_totassets	0.708	1.215	0.548	-0.287
M	C	L	4	5	ffo_to_totassets	0.641	1.314	0.683	-0.095
Q	C	L	3	0	accpayable_to_lagged_cogs	0.736	1.069	0.407	-0.073
Q	C	L	3	5	accpayable_to_lagged_cogs	0.678	1.075	0.450	-0.084
Q	C	L	4	0	inventory_to_totassets	0.732	0.999	0.234	-0.038
Q	C	L	4	5	inventory_to_totassets	0.698	1.060	0.286	-0.048
Q	C	L	5	0	ffo_to_inventory	0.737	1.316	0.321	-0.164
Q	C	L	5	5	ffo_to_inventory	0.629	1.391	0.487	0.031
Q	C	L	3	0	inventory_to_totassets	0.808	1.048	0.215	0.022
Q	C	L	3	5	inventory_to_totassets	0.753	1.067	0.276	0.010
M	C	L	3	0	ebit_to_ev	0.687	1.163	0.160	-0.207
M	C	L	3	6	ebit_to_ev	0.553	1.122	0.226	-0.226
M	C	L	4	0	ffo_to_totassets	0.708	1.215	0.548	-0.287
M	C	L	4	6	ffo_to_totassets	0.657	1.377	0.678	-0.120
Q	C	L	4	0	inventory_to_totassets	0.732	0.999	0.234	-0.038
Q	C	L	4	6	inventory_to_totassets	0.698	1.084	0.279	-0.024
M	E	L	3	0	ffo_to_inventory	0.762	1.244	0.685	-0.254
M	E	L	3	6	ffo_to_inventory	0.714	1.166	0.743	-0.181
Q	C	L	4	0	inventory_to_currliab	0.737	1.138	0.175	0.014
Q	C	L	4	6	inventory_to_currliab	0.756	1.167	0.341	0.020

Table B.9: Top performing X-clairvoyance strategies regressed vs automotive Fama-French factors results.

B.9 Examples

Example B.9.1. *Imagine we have 5 companies, A,B,C,D,E, and that the weight we hold of each, were we to be rebalancing monthly, is as described in Table B.10a.*

Table B.10

Date	A	B	C	D	E
2008-01	0.000	0.000	0.333	0.333	0.333
2008-02	0.333	0.000	0.333	0.000	0.333
2008-03	0.000	0.333	0.333	0.000	0.333
2008-04	0.333	0.000	0.333	0.000	0.333
2008-05	0.333	0.333	0.000	0.000	0.333
2008-06	0.000	0.000	0.333	0.333	0.333
2008-07	0.333	0.000	0.333	0.333	0.000
2008-08	0.000	0.000	0.333	0.333	0.333
2008-09	0.333	0.000	0.333	0.333	0.000
2008-10	0.333	0.333	0.333	0.000	0.000
2008-11	0.333	0.333	0.000	0.333	0.000
2008-12	0.333	0.333	0.000	0.333	0.000
2009-01	0.333	0.333	0.000	0.000	0.333
2009-02	0.333	0.000	0.333	0.000	0.333
2009-03	0.333	0.000	0.333	0.333	0.000
2009-04	0.333	0.333	0.000	0.000	0.333
2009-05	0.333	0.333	0.000	0.000	0.333

(a) Monthly rebalancing weights.

Date	A	B	C	D	E
2008-01	0.000	0.000	0.333	0.333	0.333
2008-02	0.167	0.000	0.333	0.167	0.333
2008-03	0.111	0.111	0.333	0.111	0.333
2008-04	0.167	0.083	0.333	0.083	0.333
2008-05	0.200	0.133	0.267	0.067	0.333
2008-06	0.167	0.111	0.278	0.111	0.333
2008-07	0.190	0.095	0.286	0.143	0.286
2008-08	0.167	0.083	0.292	0.167	0.292
2008-09	0.185	0.074	0.296	0.185	0.259
2008-10	0.200	0.100	0.300	0.167	0.233
2008-11	0.212	0.121	0.273	0.182	0.212
2008-12	0.222	0.139	0.250	0.194	0.194
2009-01	0.250	0.167	0.222	0.167	0.194
2009-02	0.250	0.167	0.222	0.167	0.194
2009-03	0.278	0.139	0.222	0.194	0.167
2009-04	0.278	0.167	0.194	0.194	0.167
2009-05	0.278	0.167	0.194	0.194	0.167

(b) Trailing 12-month (TTM) transformation of Table B.10a.

To expose us to the fact that yearly rebalanced portfolios seem to achieve higher returns than monthly rebalanced ones, without rebalancing yearly, we propose altering the monthly weights to a trailing 12-month average of the weights, had we been rebalancing monthly. If there isn't weight data available as far back as 12-months, we take the average as far back as we can go. Doing so transforms Table B.10a into Table B.10b. We see that, after the point of trailing 12-month transformation, the adjustments to the weights are now ± 0.028*, rather than* ± 0.333*.*

Bibliography

[Alberg and Lipton, 2017] Alberg, J. and Lipton, Z. C. (2017). Improving factor-based quantitative investing by forecasting company fundamentals. *arXiv preprint arXiv:1711.04837*.

[Ang, 2014] Ang, A. (2014). *Asset management: A systematic approach to factor investing.* Oxford University Press.

[Authers, 2015] Authers, J. (2015). Why factor investing is flavour of the month. *Financial Times*.

[Clarke et al., 2005] Clarke, R. G., de Silva, H., and Murdock, R. (2005). A factor approach to asset allocation. *Journal of Portfolio Management*, 32(1):10.

[Corey et al., 1998] Corey, D. M., Dunlap, W. P., and Burke, M. J. (1998). Averaging correlations: Expected values and bias in combined pearson rs and fisher's z transformations. *The Journal of general psychology*, 125(3):245–261.

[Elton et al., 2009] Elton, E. J., Gruber, M. J., Brown, S. J., and Goetzmann, W. N. (2009). *Modern portfolio theory and investment analysis*. John Wiley & Sons.

[Fama and French, 2018] Fama, E. and French, K. (2018). Fama-French historical factors data. `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`. Accessed: 2018-09-02.

[Fama and French, 1992] Fama, E. F. and French, K. R. (1992). The cross-section of expected stock returns. *the Journal of Finance*, 47(2):427–465.

[Fama and French, 1993] Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1):3–56.

[Fama and French, 1995] Fama, E. F. and French, K. R. (1995). Size and book-to-market factors in earnings and returns. *The journal of finance*, 50(1):131–155.

[Fama and French, 2004] Fama, E. F. and French, K. R. (2004). The capital asset pricing model: Theory and evidence. *Journal of economic perspectives*, 18(3):25–46.

[Gastineau, 2010] Gastineau, G. L. (2010). *The Exchange-Traded Funds Manual*, volume 186. John Wiley & Sons.

[International Accounting Standards Board, 2007] International Accounting Standards Board (2007). IAS1. `https://www.iasplus.com/en/standards/ias/ias1`. Accessed: 2018-09-09.

[Lintner, 1965] Lintner, J. (1965). Security prices, risk, and maximal gains from diversification. *The journal of finance*, 20(4):587–615.

[Markowitz, 1952] Markowitz, H. (1952). Portfolio selection. *The journal of finance*, 7(1):77–91.

[Markowitz, 1990] Markowitz, H. (1990). About Harry Markowitz. `http://hmarkowitz.com/about-harry-markowitz/`. Accessed: 2018-07-04.

[Merton, 1972] Merton, R. C. (1972). An analytic derivation of the efficient portfolio frontier. *Journal of financial and quantitative analysis*, 7(4):1851–1872.

[Mossin, 1966] Mossin, J. (1966). Equilibrium in a capital asset market. *Econometrica: Journal of the econometric society*, pages 768–783.

[Ross, 1976] Ross, S. A. (1976). The arbitrage theory of capital asset pricing.

[Ross et al., 1972] Ross, S. A. et al. (1972). Portfolio and capital market theory with arbitrary preferences and distributions: The general validity of the mean-variance approach in large markets. Technical report, Wharton School Rodney L. White Center for Financial Research.

[Ross et al., 1973] Ross, S. A. et al. (1973). *Return, risk and arbitrage*. Rodney L. White Center for Financial Research, The Wharton School, University of Pennyslvania.

[Sharpe, 1964] Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, 19(3):425–442.

[Sharpe, 1991] Sharpe, W. F. (1991). The arithmetic of active management. *Financial Analysts Journal*, 47(1):7–9.

[Soe and Poirier, 2016] Soe, A. M. and Poirier, R. (2016). SPIVA US Year-End 2016 Scorecard. `https://us.spindices.com/documents/spiva/spiva-us-year-end-2016.pdf`.

[Sun, 2003] Sun, W. (2003). Relationship between trading volume and security prices and returns.

[Tobin, 1958] Tobin, J. (1958). Liquidity preference as behavior towards risk. *The review of economic studies*, 25(2):65–86.

[Treynor, 1961] Treynor, J. L. (1961). Market value, time, and risk.

[Treynor, 1962] Treynor, J. L. (1962). Toward a theory of market value of risky assets.

[Williams, 1938] Williams, J. B. (1938). *The theory of investment value*, volume 36. Harvard university press Cambridge, MA.

[Yan and Zheng, 2017] Yan, X. and Zheng, L. (2017). Fundamental analysis and the cross-section of stock returns: A data-mining approach. *The Review of Financial Studies*, 30(4):1382–1423.