Long-Term Stock Forecasting

Magnus Erik Hvass Pedersen

First published: December 17, 2020

Latest revision: February 14, 2021[1]

Abstract

It is well-known that there is a strong correlation between valuation ratios and long-term returns on certain stock-market indices such as the S&P 500. Scatter-plots of valuation ratios versus long-term stock-returns often show a characteristic downwards slope, where higher valuation ratios correspond to lower future stock-market returns, and vice versa. But a formal explanation of this phenomenon has never been given until now. In this paper we show how to properly decompose stock-returns into three components: Dividend yield, change in valuation ratio such as the P/E or P/Sales ratio, and the change in Earnings or Sales Per Share. Together with the basic formula for calculating annualized returns, this explains the characteristic curves we often see in scatter-plots of long-term stock-returns. We also derive formulas that let us forecast the mean and standard deviation for the future stock-returns from these three components. This is demonstrated on real-world data for both individual stocks as well as entire stock-market indices such as the S&P 500, 400 and 600 for U.S. stocks, and various Exchange Traded Funds (ETF) for international stock-market indices.

1 See the last page for changes.

Table of Contents

1 Introduction

The founder of fundamental investment analysis, Ben Graham, often said that the stock-market is a voting-machine in the short run and a weighing-machine in the long run. He also likened the stock-market to a manic-depressive character named Mr. Market who would often swing between being wildly ecstatic and deeply depressed. That was almost a century ago and Ben Graham was probably hardened by his own dramatic experience in the stock-market crash of 1929 and the “Great Depression” in the following decade. Ben Graham’s seminal work [1] was mostly concerned with revaluation of stocks, especially stocks that were trading at deep discounts to their tangible book-value or even liquidation value, and Graham was not so much concerned with growth, which is another major source of long-term stock-returns that is just as important to understand as revaluation.

We might think that we are a lot more rational today, with all of our modern science and technology. Yet the stock-markets seem to be more volatile and irrational than ever. It appears that more and more stocks are being traded by advanced computer programs with short-term strategies. When a person or a computer is trying to predict what the stock-market will do in the short-term, they are essentially trying to predict what other market participants – whether human or machine – are going to do next. This is like a frantic carousel where everyone is trying to get in front of everyone else. It’s a fool’s game!

When investing for the long-term you are instead trying to predict the company’s long-term future growth, and whether the stock may be revalued up or down relative to its sales or earnings. This notion of long-term investing is formalized in this paper, where we will derive a few simple formulas, that take as input a range of your guesses for the stock’s future valuation ratio, the company’s future growth in sales or earnings, and the future dividend yield. These are used to calculate two parameters for the forecasting model, so it can output the mean and standard deviation for the future stock-returns.

Considering the importance of long-term stock-returns and their forecasting, there has been very little academic research on this topic. The most well-known work is probably by Robert Shiller who received the financial Nobel prize in the year 2013, in part for his work on this topic. A paper by Shiller and Campbell from the year 2001 [2] contains several scatter-plots for the S&P 500 stock-market index, showing how different valuation ratios have historically related to the long-term future stock-market returns. Shiller’s book from the year 2015 [3] only has a single scatter-plot on this topic (Figure 11.1 in that book), but it shows more clearly how valuation ratios relate to 10-year future stock-market returns, where higher valuation ratios tend to result in lower returns for the following 10 years, and vice versa. Although this makes intuitive sense to someone following Ben Graham’s school of thought, Shiller never really gave a formal “proof” why this relation exists, he just observed it in the historical data. Shiller’s plot is made with only yearly data-points between 1881 and 2013, but it still shows a characteristic downwards-sloping shape that we will also formally explain in this paper.

A detailed study of long-term returns for both USA and international stock-markets was published by an investment professional named Keimling in the year 2016 [4], using monthly data-points between 1979 and 2015 from the MSCI database. It is an empirical study that does not give any theoretical explanation. It also studies the predictive strength of different valuation ratios, as well as considering both the future stock-returns and max draw-downs. The scatter-plots show curves that are very similar to the other research papers cited above, and whose shape will be formally explained in this paper.

An extensive review of the literature on this topic was made by Zaremba in 2019 [5]. As far as I know, none of the existing research papers have given a formal explanation for the patterns we see in long-term stock-returns, and how to derive and use proper formulas for long-term stock forecasting.

This paper derives the forecasting formulas for long-term stock-returns simply from the mathematical definition of annualized return, along with a decomposition of the stock-return into its three basic components. The paper shows that under certain conditions, this forecasting model has a very good fit to the historical data for some individual stocks and stock-indices. Examples are also given where the forecasting model does not fit the historical data, and the cause of this is explained. This paper also uses synthetic data, which makes it easy to experiment with different hypotheses and see how the forecasting model reacts to different kinds of data and random noise.

The key to using these formulas to forecast the future stock-returns, is to provide reasonably good estimates for the future valuation ratio of the stock, and the future growth of the company, and its future dividend yield. For individual stocks this can be quite difficult, but for broadly diversified stock-indices such as the S&P 500, it may be reasonable to assume that the future valuation ratios and growth-rates will be similar to their past 20 or 30 year averages.

This paper has its origin in my research book from 2015 [6], in which I provided a few plots (Figure 41 on page 83 in that book) similar to Shiller's scatter-plots, showing the relation between a valuation ratio for the S&P 500 and the return for the following 1, 3, 6, and 10 years. I used daily instead of yearly data-points as Shiller had done, so the relation between valuation ratio and future stock-returns was even more clear. A few years later I revisited this topic with the intention of merely putting the data and computer code on the internet. This turned into an entire research series as I made more discoveries, and eventually also derived the mathematical formula that explains the particular pattern we often see in the historical data for long-term stock returns. This paper assembles all of that research and also explains the puzzles that had not been solved in my previous research series.

This paper has been written so it can hopefully be understood by a large audience with diverse backgrounds. Only basic skills in investment finance, mathematics and statistics are required. All data and computer code is available on the internet (see Section 16) so everyone who can use the Python programming language can repeat all the experiments in this paper. There is also a series of video lectures on the research that eventually resulted in this paper.

2 Forecasting Model

There are two kinds of investment returns from owning shares in a company: Dividends and changes in the share-price. These can be combined into a so-called Total Return by assuming the dividends are reinvested immediately after having been paid out to shareholders, so as to buy more shares in the same company. Taxes are ignored for now and will be discussed later. Let $Total\,Return_t$ denote the number of shares that we own at time-step t multiplied by the share-price at that time-step:

$$Total\,Return_t = Shares_t \cdot Share\,Price_t \tag{1}$$

This is merely another time-series that is a scaled version of the share-price, where the scaling-factor is the number of shares that we own, which increases over time from the reinvestment of dividends. This is not the company's number of shares outstanding. It is simply a way of writing that the Total Return for shareholders grows from reinvestment of dividends into this factor of accumulating "shares".

The Annualized Return is then calculated from the Total Return between time-steps t and $t+Years$ and should perhaps be denoted more fully as $Ann\,Return_{t,t+Years}$ but for brevity we will just denote it $Ann\,Return_t$ and keep in mind that it covers the period between time-steps t and $t+Years$. This measures the shareholder's annualized investment return during this period, both from the change in share-price and from the accumulated reinvestment of dividends. It is defined as:

$$Ann\,Return_t = \left(\frac{Total\,Return_{t+Years}}{Total\,Return_t}\right)^{1/Years} - 1 \tag{2}$$

Using this with the definition of Total Return from Eq. (1) we get:

$$Ann\,Return_t = \left(\frac{Shares_{t+Years} \cdot Share\,Price_{t+Years}}{Shares_t \cdot Share\,Price_t}\right)^{1/Years} - 1 \tag{3}$$

We now want to express the annualized return as a function of some valuation ratio, so that we can use the valuation ratio to estimate and forecast the future annualized return. In this paper we will be using the valuation ratio known as P/Sales (Price-to-Sales ratio) for reasons that will be explained in Section 2.5. The ratio is defined as:

$$P/Sales_t = \frac{Share\,Price_t}{Sales\,Per\,Share_t} \tag{4}$$

We can rewrite this definition to get:

$$Share\,Price_t = P/Sales_t \cdot Sales\,Per\,Share_t \tag{5}$$

Note that this is actually just a mathematical identity because:

$$Share\ Price_t = P/Sales_t \cdot Sales\ Per\ Share_t = \frac{Share\ Price_t}{Sales\ Per\ Share_t} \cdot Sales\ Per\ Share_t = Share\ Price_t \tag{6}$$

We can then insert Eq. (5) back into Eq. (3) for the Annualized Return, to make it clear that the investment return is actually composed of the change in the number of shares from reinvestment of dividends, and the revaluation from the change in the P/Sales ratio, and the growth or change in the Sales Per Share – and all of these changes are between time-steps t and $t+Years$. The formula has been annotated with blue labels to make the three return components extra clear:

$$Ann\ Return_t = \left(\overbrace{\frac{Shares_{t+Years}}{Shares_t}}^{\text{Reinvest Dividends}} \cdot \overbrace{\frac{P/Sales_{t+Years}}{P/Sales_t}}^{\text{Revaluation}} \cdot \overbrace{\frac{Sales\ Per\ Share_{t+Years}}{Sales\ Per\ Share_t}}^{\text{Sales Growth}} \right)^{1/Years} - 1 \tag{7}$$

In other words, the investment return has three components: Dividends, change in valuation ratio such as the P/E or P/Sales ratio, and the company's growth in Earnings or Sales Per Share. That is why it is meaningless to classify stocks as either being so-called value- or growth-stocks, because both revaluation and growth may impact the investment returns either positively or negatively. It is the combination of all three components of investment returns that ultimately matters.

2.1 Forecasting the Mean Return

It is important to understand that Eq. (7) above is a factual relation between the annualized return of a stock and the three return components, because it follows directly from rewriting and combining their definitions. So if you know the future dividend yield, the future P/Sales ratio, and the future Sales Per Share, then you can predict the future stock-return with complete accuracy.

But we cannot know these three return components in advance, so we are instead interested in forecasting a range of possible outcomes for the annualized return, when we only know a range of possible values for these three components. We can then summarize a range of outcomes using the mean as the centre-value, and the standard deviation as the spread of outcomes. The mean is also known as the average or expected value and is typically denoted $E[X]$ for some random variable X and the standard deviation is denoted $Std[X]$.

The mean annualized return between time-steps t and $t+Years$ is denoted $E[Ann\ Return_t]$ and is calculated directly from Eq. (7):

$$E[Ann\ Return_t] = E\left[\left(\frac{Shares_{t+Years}}{Shares_t} \cdot \frac{P/Sales_{t+Years}}{P/Sales_t} \cdot \frac{Sales\ Per\ Share_{t+Years}}{Sales\ Per\ Share_t} \right)^{1/Years} - 1 \right] \tag{8}$$

Because $P/Sales_t$ is a value that is known at time-step t it follows from the properties of the mean operator that we can rewrite the formula as:

$$E[Ann\,Return_t] = \frac{E\left[\left(\frac{Shares_{t+Years}}{Shares_t} \cdot P/Sales_{t+Years} \cdot \frac{Sales\ Per\ Share_{t+Years}}{Sales\ Per\ Share_t}\right)^{1/Years}\right]}{P/Sales_t^{\ 1/Years}} - 1 \tag{9}$$

We can use this formula directly if we supply a range of possible outcomes for A) the future growth in share-count from reinvestment of dividends, B) the future P/Sales ratio, and C) the future growth in Sales Per Share. This formula also takes into account any dependency between the three components.

We can simplify the formula if we assume that the three return components are independent. This is probably not a completely valid assumption, as we might imagine that companies with high sales-growth also tend to trade at higher valuation ratios, so these three components are probably somewhat dependent upon each other. But we will estimate both the mean and standard deviation for the range of outcomes, so small estimation errors are acceptable, and it makes the formula easier to calculate.

Assuming independence between the three return components, the formula becomes:

$$E[Ann\,Return_t] = \frac{E\left[\left(\frac{Shares_{t+Years}}{Shares_t}\right)^{1/Years}\right] \cdot E\left[P/Sales_{t+Years}^{1/Years}\right] \cdot E\left[\left(\frac{Sales\ Per\ Share_{t+Years}}{Sales\ Per\ Share_t}\right)^{1/Years}\right]}{P/Sales_t^{\ 1/Years}} - 1 \tag{10}$$

At first glance this might look even more intimidating but let us discuss the three components in turn:

- $Shares_{t+Years}$ is the number of future shares we own that has grown from reinvestment of dividends, so $E\left[\left(\frac{Shares_{t+Years}}{Shares_t}\right)^{1/Years}\right]$ is the mean annualized growth-rate from reinvestment of dividends, for all possible outcomes. If we assume the dividend is paid out regularly and immediately invested back into buying more shares of the same company, then the mean annual growth-rate from reinvestment of dividends is roughly equal to the mean Dividend Yield, which is defined as the dividend for the trailing year divided by the share-price. They are only “roughly” equal because the payout of dividends typically only occurs between 1 and 4 times a year, and it is the share-price on those particular days that determines how many new shares we can buy for the dividend, whereas we calculate the Dividend Yield for all trading days. But these numbers should be roughly equal on average. So we have:

$$E\left[\left(\frac{Shares_{t+Years}}{Shares_t}\right)^{1/Years}\right] \simeq E[Dividend\,Yield + 1] \tag{11}$$

- $P/Sales_{t+Years}$ is the future valuation ratio. We will simply write $P/Sales$ without the time-step to indicate a distribution of possible values. Furthermore, if we only know the mean P/Sales ratio but not its distribution, then we can estimate the mean of the exponent by moving the exponent outside of the parentheses, however, this increases the estimation error slightly due to a mathematical phenomenon known as "Jensen's Inequality", so it is preferred to calculate the exponent inside the parentheses when possible:

$$E\left[P/Sales_{t+Years}^{1/Years}\right] = E\left[P/Sales^{1/Years}\right] \simeq E\left[P/Sales\right]^{1/Years} \tag{12}$$

- $Sales\,Per\,Share_{t+Years}$ is for the future time-step $t+Years$ so $E\left[\left(\frac{Sales\,Per\,Share_{t+Years}}{Sales\,Per\,Share_t}\right)^{1/Years}\right]$ is the mean annualized growth-rate in Sales Per Share for all possible outcomes. If we assume that the sales-growth is independent from one year to the next, then we can simply use the one-year sales-growth as an estimate:

$$E\left[\left(\frac{Sales\,Per\,Share_{t+Years}}{Sales\,Per\,Share_t}\right)^{1/Years}\right] \simeq E\left[Sales\,Per\,Share\,Growth\,Rate+1\right] \tag{13}$$

Putting all this together we get the first of the two main formulas for the forecasting model:

$$E\left[Ann\,Return_t\right] = \frac{a}{P/Sales_t^{1/Years}} - 1 \tag{14}$$

where the parameter a can be estimated from the three components described above:

$$a \simeq E\left[Dividend\,Yield+1\right] \cdot E\left[P/Sales^{1/Years}\right] \cdot E\left[Sales\,Per\,Share\,Growth\,Rate+1\right] \tag{15}$$

We can use the historical means for the Dividend Yield, P/Sales ratios, and annual growth in Sales Per Share, or we can use other numbers if we think the future will be different from the past.

Note that when using the forecasting formula in practice, we often drop the time-step t in Eq. (14). It is probably a bit confusing that *P/Sales* then denotes an actual value in Eq. (14), while it is a stochastic or random variable in Eq. (15). Hopefully the notation will become more intuitive once you start using it.[2]

2 You may also notice that I often – but not always – capitalize certain words such as Dividend Yield, Sales Per Share, Share-Price, Annualized Return, etc. This is meant to emphasize that it refers to particular data variables or formula definitions. But it would be awkward to repeatedly capitalize words such as "share-price" or "annualized return" when they are used in a more general context. Another inconsistency is that I sometimes switch between British and American spelling. My reason for using this writing style is explained in more detail here.

2.2 Forecasting the Standard Deviation of Return

Let us now consider the standard deviation for the forecasted annualized returns. This measures the spread of the distribution around its mean. It is denoted $Std[Ann\,Return_t]$ and is calculated directly from Eq. (7):

$$Std[Ann\,Return_t] = Std\left[\left(\frac{Shares_{t+Years}}{Shares_t} \cdot \frac{P/Sales_{t+Years}}{P/Sales_t} \cdot \frac{Sales\,Per\,Share_{t+Years}}{Sales\,Per\,Share_t}\right)^{1/Years} - 1\right] \tag{16}$$

Because $P/Sales_t$ is a value that is known at time-step t it follows from the properties of the standard deviation that we can rewrite it as follows, where we have also removed the -1 at the end:

$$Std[Ann\,Return_t] = \frac{Std\left[\left(\frac{Shares_{t+Years}}{Shares_t} \cdot P/Sales_{t+Years} \cdot \frac{Sales\,Per\,Share_{t+Years}}{Sales\,Per\,Share_t}\right)^{1/Years}\right]}{P/Sales_t^{\;1/Years}} \tag{17}$$

Letting the numerator be the parameter b we then get the second of the two main forecasting formulas:

$$Std[Ann\,Return_t] = \frac{b}{P/Sales_t^{\;1/Years}} \tag{18}$$

where the parameter b can be estimated from the Dividend Yield, the P/Sales distribution, and the annual growth-rate in Sales Per Share, using the same arguments as above for the parameter a in Eq. (15):

$$\begin{aligned} b &= Std\left[\left(\frac{Shares_{t+Years}}{Shares_t} \cdot P/Sales_{t+Years} \cdot \frac{Sales\,Per\,Share_{t+Years}}{Sales\,Per\,Share_t}\right)^{1/Years}\right] \\ &\simeq Std\left[(Dividend\,Yield+1) \cdot P/Sales^{1/Years} \cdot (Sales\,Per\,Share\,Growth\,Rate+1)\right] \end{aligned} \tag{19}$$

But even if we assume that these random variables are independent, we cannot easily split this formula into three factors for the individual standard deviations, because the closed-form formula is very complicated. It is much easier to either calculate this standard deviation directly from the historical data, or assume independence of the random variables and perform a simple Monte Carlo simulation.

Note that when using the forecasting formula in practice, we often drop the time-step t in Eq. (18) like we did for the mean forecast in Eq. (14). It is probably a bit confusing that *P/Sales* then denotes an actual value in Eq. (18), while it is a stochastic or random variable in Eq. (19). Hopefully the notation will become more intuitive once you see some of the examples below and also start using it yourself.

2.3 Simpler Forecasting Formula

The forecasting formula for the mean annualized return from Eq. (14) and its parameter a from Eq. (15) can be simplified even further for use in daily investment analysis that can be calculated more quickly in your head. First note that for values X, Y and Z that are close to zero, their sum is approximately equal to the following:

$$X+Y+Z \simeq (X+1)\cdot(Y+1)\cdot(Z+1)-1 \tag{20}$$

Because the Dividend Yield, and the annual growth-rate in Sales Per Share, and the annualized return from the revaluation, are all values that are typically below, say, 0.2 or 20%, we can use the above formula to estimate the mean annualized return from Eq. (14) and and Eq. (15) as follows:

$$E[Ann\,Return_t] \simeq E[Dividend\,Yield] + E[Sales\,Per\,Share\,Growth\,Rate] + E\left[\left(\frac{P/Sales}{P/Sales_t}\right)^{1/Years}\right] - 1 \tag{21}$$

So the mean annualized return can be estimated from the sum of the mean Dividend Yield, the mean annual growth-rate in Sales Per Share, and the mean annualized change in valuation ratio. We will see an example of how to apply this simple formula in Section 3. This may often be good enough for quick daily investment analysis, but keep in mind that it introduces more estimation errors than Eq. (14).

2.4 Tax

When using the forecasting formulas in practice, it may be interesting to know the returns after taxes. But the forecasting formulas can become very complicated when adjusting for taxes, not only because of the usual difference between tax-rates for dividends and capital gains, but also because the taxation may occur at different points in time, e.g. capital gains may either be taxed when the shares are sold, or the tax may have to be paid at the end of the year, or in some countries the capital gains tax has to be paid annually even if the shares have not yet been sold. The tax-rate may also depend on the number of years you have owned the shares. The tax-laws may also change over time.

Even if we were to consider just one simple tax-scenario, the proper forecasting formula from Eq. (14) and its parameter a from Eq. (15) still become quite complicated when adjusting for taxes. Because we will ignore taxes in this paper, we will just modify the simpler forecasting formula from Eq. (21), to demonstrate how you can quickly estimate the effect of taxes on the mean annualized return.

Let the tax-rate for dividends be denoted *Tax Dividend*, and let the tax-rate for capital gains be denoted *Tax Capital Gains*. We can then modify the simpler forecasting formula from Eq. (21) as follows:

$$\begin{aligned} E[Ann\,Return_t] \;\; \simeq \;\; & E[Dividend\,Yield]\cdot(1-Tax\,Dividend)+ \\ & E[Capital\,Gains]\cdot(1-Tax\,Capital\,Gains) \end{aligned} \tag{22}$$

The mean Capital Gains is also a growth-rate here, that is defined from the mean annual growth in Sales Per Share plus the mean annualized return from revaluation:

$$E[Capital Gains] = E[Sales Per Share Growth Rate] + E\left[\left(\frac{P/Sales}{P/Sales_t}\right)^{1/Years}\right] - 1 \tag{23}$$

The tax-scenario modelled in these two formulas, is that dividends are taxed immediately, and capital gains are taxed annually regardless of whether the shares have been sold or not – otherwise the capital gains tax should have been calculated inside the exponent for annualization. In case the capital gains are negative so they are actually losses, we would typically not adjust the capital gains for taxes in Eq. (22) but instead get a tax-credit that could be deducted from the capital gains on other investments. Note how complicated the formula quickly becomes when adjusting for taxes.

Let us consider an example of how taxes affect the forecasted stock-return when using Eq. (22). Assume the future Dividend Yield has been forecasted to 5% on average, and the future growth in Sales Per Share has been forecasted to 10% per year on average, and the future loss from revaluation has been forecasted to -4% per year because the current P/Sales ratio is assumed to be 3 but the future P/Sales ratio in 10 years is only assumed to be 2. If we add these three return components using Eq. (21) we get a mean annualized return of 11%.

Now let the tax-rate for dividends be $Tax Dividend = 20\% = 0.2$ and let the tax-rate for capital gains be $Tax Capital Gains = 30\% = 0.3$. Using the tax-adjusted formula in Eq. (22) we then get a 4% Dividend Yield plus 4.2% for the Capital Gains, for a total of 8.2% compared to 11% if there were no taxes.

2.5 Choice of Valuation Ratio

The reason that we are using the P/Sales ratio in these formulas, is that it is well-defined when the Sales Per Share is greater than zero, which is the case for most publicly traded stocks. The Sales Per Share is usually also far more stable than the Earnings Per Share. The problem with the P/Sales ratio is that it can be difficult to interpret and compare across different stocks.

Although it is much easier to interpret and compare the P/E (Price-to-Earnings) ratio across different stocks, the problem is that the Earnings Per Share typically fluctuates much more than the Sales Per Share, and the Earnings Per Share may also be zero or even negative, which causes the P/E ratio to be ill-defined. That is why the raw P/E ratio is a poor choice of predictor variable in forecasting formulas.

The solution proposed by Robert Shiller [3] is to use the Cyclically Adjusted P/E ratio (CAPE) defined as the share-price divided by the average of the trailing 10 years of Earnings Per Share, so as to smoothen the year-to-year volatility in the reported earnings. The problem with the CAPE ratio is that it also implicitly uses the average Sales Per Share for the trailing 10 years, and for a company that has grown its sales significantly during that period, it can greatly underestimate its future sales.

A better solution would be to use the average Net Profit Margin for the trailing 10 years and multiply that with the Sales Per Share for the trailing 12 months, so as to get a more accurate estimate of the expected Earnings Per Share for the company's current level of sales. We will not do that in this paper, but it would be a fine research project for you to do. We will merely continue to use the P/Sales ratio in this paper.

Because of how the forecasting formulas are derived, we could actually use any valuation ratio you might think of. For example, we could use the infamous P/T ratio where T is the temperature in Timbuktu. This is a joke of course, but why would this not make sense? Because in order for the valuation ratio to be useful in forecasting, we must be able to estimate what the valuation ratio might reasonably be in the future. The temperature in Timbuktu has nothing to do with the sales and earnings of a company in USA, so we could essentially drop the temperature T from the valuation ratio P/T, and therefore we would be left with only the share-price P, which cannot be used to predict itself, so the P/T valuation ratio is useless in forecasting. There must be a somewhat predictable relation between the share-price and the denominator in the ratio.

2.6 Return Curves

In the case-studies further below, we will see plenty of examples where historical long-term stock-returns follow a characteristic downwards-sloping curve as a function of the valuation ratio. This actually arises from the very definition of Annualized Return in Eq. (2) so both the forecasted mean in Eq. (14) and the standard deviation in Eq. (18) exhibit this characteristic pattern. We call these for Return Curves.

Let us now show how these Return Curves look for different choices of parameters. We base this on Eq. (14) for the forecasted mean, and we have set the parameter $a=1$ in these plots, so it is easy to see the effect of varying the P/Sales ratio at the time of buying a stock. The parameter a is actually just a linear scale that moves the curves up or down. So the formula is defined as follows and the plot has the P/Sales ratio on the x-axis and the Annualized Return on the y-axis:

$$Annualized\ Return = \frac{1}{P/Sales^{1/Years}} - 1 \tag{24}$$

The plot is shown in Figure 1. Note the characteristic downwards slope of all the curves, which is because the formula divides with the annualized P/Sales ratio, so the annualized return decreases in a reciprocal manner as the P/Sales increases. The steepness of the curve is determined by the number of investment years, where the curve becomes more flat as the number of investment years increases. This is because the revaluation of the P/Sales ratio is spread over more years, so its effect on the annualized return is lower. The formula for the standard deviation in Eq. (18) has the same type of curve as in Figure 1, except it does not subtract 1 as done in this plot.

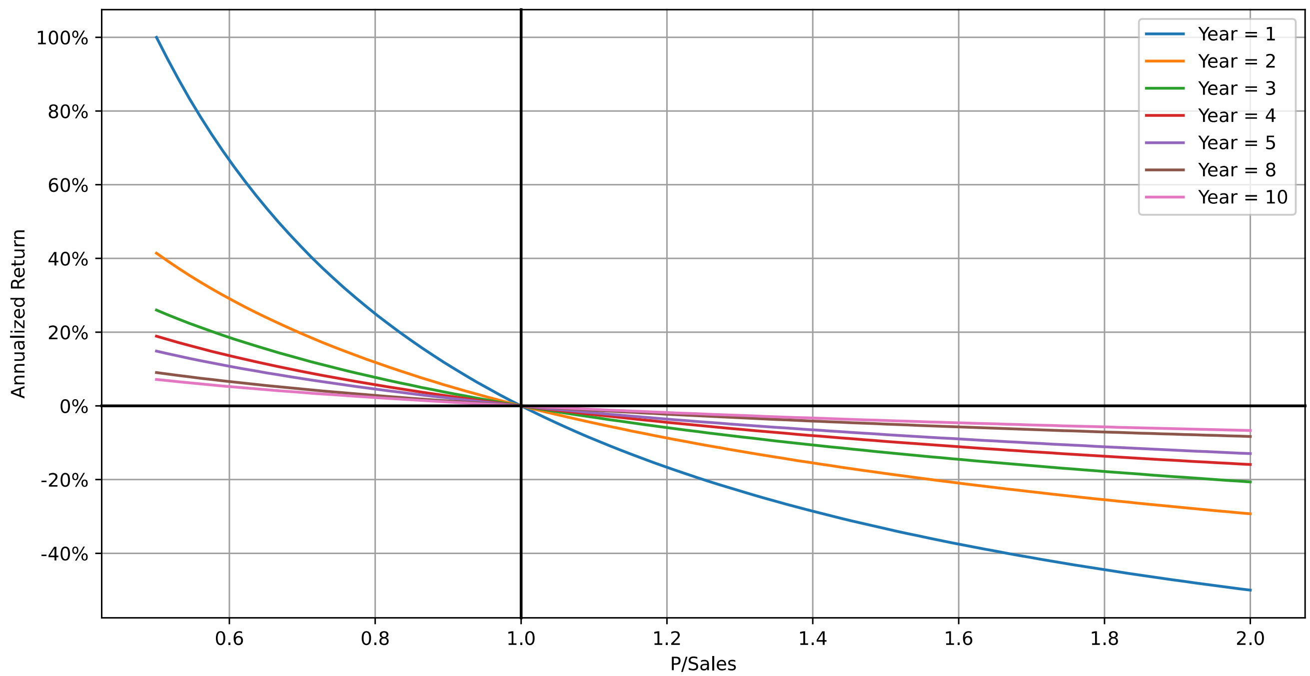


Figure 1: Return Curves for predicting the Annualized Return from the P/Sales ratio using Eq. (24)

2.7 Inflation

In his empirical research on the relation between valuation ratios and long-term stock-returns, Robert Shiller typically uses the inflation-adjusted stock-returns, see e.g. [3], which are also called the *real* stock-returns, while the non-adjusted stock-returns are called *nominal*. On the surface this seems reasonable because an investor should be concerned about the inflation-adjusted returns on their investment. If the nominal stock-return is 5% but the inflation is 7%, then you would actually lose purchasing power from making the investment, as you would not be able to purchase the same amounts of goods and services afterwards, because the prices had generally increased more than your stock-return. So it makes sense to consider the *real* or inflation-adjusted stock-returns.

However, when we think about this a bit more deeply, it seems strange to consider how different valuation ratios affect the *real* stock-returns. For example, if the inflation rate is expected to be 3% next year, and then it unexpectedly jumps to 15% in 5 years, and then it falls back down to 3% in 10 years, then how is the 15% inflation in 5 years supposed to be relevant for the valuation ratio today and the future valuation ratio in 10 years when we sell the stock? The inflation does not seem to be relevant for the valuation ratio. It would make more sense if the inflation rate was a component of the growth-rate in Earnings or Sales Per Share. But this discussion gets rather philosophical and it can easily become a bit hazy to think about.

It is much easier to use a mathematical framework, where we can precisely define the inflation-adjustment and derive correct formulas. So we start by defining a new time-series called CPI_t for the Consumer Price Index, which is a standardized measure of the prices of goods and services at time-step t in a given country such as USA. This time-series could be normalized to begin with the value 1 at time-step 1 so that $CPI_1 = 1$, but we are ultimately interested in the inflation over time, that is, the change in CPI-values over time, so it is not necessary to normalize the time-series like this.

We then define the *Real* Total Return of a stock or stock-index similar to Eq. (1) but now divided by the CPI so as to adjust for inflation:

$$Real\,Total\,Return_t = \frac{Shares_t \cdot Share\,Price_t}{CPI_t} \tag{25}$$

By itself, the Real Total Return is a strange number that is hard to interpret directly, because we have divided the Total Return with the CPI values. But we are actually only interested in how the Real Total Return changes over time, so we define the Real Annualized Return similar to Eq. (2):

$$Real\,Ann\,Return_t = \left(\frac{Real\,Total\,Return_{t+Years}}{Real\,Total\,Return_t} \right)^{1/Years} - 1 \tag{26}$$

Using this formula with the definition in Eq. (25) and the decomposition of Share-Price into the product of the P/Sales ratio and the Sales Per Share, we get the following formula for the Real Annualized Return between time-steps t and $t+Years$:

$$Real\,Ann\,Return_t = \left(\frac{Shares_{t+Years}}{Shares_t} \cdot \frac{P/Sales_{t+Years}}{P/Sales_t} \cdot \frac{Sales\,Per\,Share_{t+Years}}{Sales\,Per\,Share_t} \cdot \frac{CPI_t}{CPI_{t+Years}} \right)^{1/Years} - 1 \tag{27}$$

This formula is almost the exact same as Eq. (7) which is the basis of the *nominal* forecasting model that we will be using throughout this paper – except that the formula in Eq. (27) also has a factor for the change in CPI so as to adjust for inflation. Also note that the time-steps in the CPI fraction are inverted compared to the other fractions in this formula, because we divided instead of multiplied with the CPI in Eq. (25).

We then take the mean of Eq. (27) and get the same forecasting formula for the mean Real Annualized Return as we had in Eq. (14):

$$E[Real\,Ann\,Return_t] = \frac{a}{P/Sales_t^{\,1/Years}} - 1 \tag{28}$$

But now the parameter a is defined slightly differently in order to adjust for the inflation. Using similar arguments and derivations as in Section 2.1, we get the following formula for the new parameter a:

$$\begin{aligned} a \quad \simeq \quad & E[Dividend\,Yield + 1] \cdot E[P/Sales^{1/Years}] \cdot \\ & E[Sales\,Per\,Share\,Growth\,Rate + 1] \cdot E\left[(Inflation\,Rate + 1)^{-1} \right] \end{aligned} \tag{29}$$

Note that we use the reciprocal of the annual Inflation Rate because the time-steps are inverted in the CPI fraction, as mentioned above. We can make this derivation a bit more explicit by defining the annualized Inflation Rate between time-steps *t* and *t*+*Years* as follows:

$$Inflation\ Rate_t = \left(\frac{CPI_{t+Years}}{CPI_t} \right)^{1/Years} - 1 \tag{30}$$

And then using a little algebra we get:

$$\left(Inflation\ Rate_t + 1 \right)^{-1} = \left(\frac{CPI_t}{CPI_{t+Years}} \right)^{1/Years} \tag{31}$$

We then take the expectation operator and get the inflation-adjustment we had in Eq. (29), where we also removed the time-step *t* because it is understood that the Inflation Rate is a stochastic variable:

$$E\left[\left(Inflation\ Rate_t + 1 \right)^{-1} \right] = E\left[\left(\frac{CPI_t}{CPI_{t+Years}} \right)^{1/Years} \right] \tag{32}$$

Similarly, we can take the standard deviation of Eq. (27) and get the same forecasting formula for the standard deviation of the Real Annualized Return as in Eq. (18):

$$Std\left[Real\ Ann\ Return_t \right] = \frac{b}{P/Sales_t^{\,1/Years}} \tag{33}$$

But once again the parameter *b* is defined slightly differently from its *nominal* definition in Eq. (19), because we also have to adjust for inflation:

$$b \simeq Std\left[\begin{aligned} &\left(Dividend\ Yield + 1 \right) \cdot P/Sales^{1/Years} \cdot \\ &\quad \left(Sales\ Per\ Share\ Growth\ Rate + 1 \right) \cdot \left(Inflation\ Rate + 1 \right)^{-1} \end{aligned} \right] \tag{34}$$

We now have a precise definition of how to adjust for inflation in the forecasting model. So if we want to plot the historical *real* annualized returns and see how well the forecasting model fits, then we should also take the historical inflation into account when calculating the model parameters *a* and *b*.

When we are using the forecasting model to make predictions about the future *real* stock-returns, we can either use the above formulas to calculate the parameters *a* and *b*, or we can simply decrease the expected annual growth-rate for the Sales Per Share by the expected inflation rate, as a quick estimate.

However, we will not be adjusting for inflation in the real-world case-studies in this paper, because the inflation-rate in USA has been fairly stable during the 30-year period we consider here. But it would be an interesting research study for you to make, if you consider periods with more unstable inflation.

3 Case Study: Tesla

> *"It kind of sucks running a public company. The [Tesla] stock goes through these huge gyrations for seemingly arbitrary reasons and then I'm asked to explain why it changed ... and I have no idea!"* – Elon Musk[3]

Let us now consider the electric car manufacturer Tesla founded by Elon Musk, who is probably one of the most versatile and industrious geniuses in human history, as he is also the founder of Space-X that reinvented space exploration, he is also involved in several other important companies, and he started his impressive career as the co-founder of PayPal which is probably still the most widely used payment platform on the internet. So I can understand the hype around his person and companies.

Tesla's stock with ticker TSLA has increased more than 10-fold during the year leading up to August 2020. At the time of this writing in early September 2020, TSLA had reached an all-time high market-capitalization over USD 450 billion. Because the TSLA stock has recently had a 5:1 stock-split which may be confusing, we will use the total sales and market-cap instead of per-share numbers. The question is whether a market-cap of USD 450b is justified for the Tesla company and what it might imply for future long-term stock-returns?

Instead of using the main forecasting formula from Eq. (14), let us try and use the simpler and more intuitive formula from Eq. (21) which is reprinted here for easy reference:

$$E\left[Ann\,Return_t\right] \simeq E\left[Dividend\,Yield\right] + E\left[Sales\,Per\,Share\,Growth\,Rate\right] + E\left[\left(\frac{P/Sales}{P/Sales_t}\right)^{1/Years}\right] - 1$$

Remember that this formula introduces more estimation errors than the proper forecasting formula in Eq. (14). Since we are trying to forecast future numbers to insert into this formula, it is all very uncertain anyway, so we might think that a bit more estimation errors are not very important. But as we will see in this example, the estimation errors can actually become quite large when some of the return components are large.

Let us start with the formula's first part, the Dividend Yield. Tesla currently does not pay any dividend and probably will not pay any dividend for several years to come. So we will assume the Dividend Yield is zero.

3 https://youtu.be/ODpuYSG4eQc?t=140

Now let us consider the mean annual growth-rate in Sales Per Share. Table 1 shows the company's annual Sales and Net Income between 2015 and 2019. As mentioned above, these are total numbers and not per-share numbers, but they will work just as well in the forecasting formula. The company has had explosive sales-growth from USD 4b in 2015 to USD 25b in 2019, which gives an annualized growth-rate of 58% calculated as $(25/4)^{1/4}-1$. Probably due to the "Corona-Virus Panic" during 2020, the company's sales decreased to USD 22.9b for the 12 months ending June 30, 2020, but its Net Income just barely turned positive for the last 6 months after several years of large losses.[4]

USD/Millions	2019	2018	2017	2016	2015
Sales	24,578	21,461	11,759	7,000	4,046
Net Income	-862	-976	-1,962	-675	-889

Table 1: Financial data for the company Tesla.[5]

So what should we forecast for the future sales-growth? Between 2015 and 2019 the sales-growth was 58% per year on average, but between 2018 and 2019 it was only 14%, and in the most recent year there was a small sales-decline. So which number should we use in the forecast? And can we even use these historical growth-rates to forecast the future? Why not turn the question around and instead ask how much sales-growth is necessary for the share-price to be justified?

So let us turn to the last part of the forecasting formula instead, which is the change in valuation ratio and how that affects the overall shareholder return. The formula uses the P/Sales ratio, but that is difficult to reason about intuitively, so we will use the P/E ratio instead. But rather than calculating the current P/E ratio from the raw financial data where the Net Income has only barely turned positive and would therefore give an extremely high P/E ratio, let us consider what the P/E ratio would be if the company had a more normal profit margin.

Historically car manufacturers have actually had quite low profit margins, probably less than 5% on average, and many car manufacturers have gone bankrupt at least once. If we think that Tesla is a truly special car-manufacturer, let us assume that it will have a 10% average Net Profit Margin in the future. This would mean that Tesla had net profits of around USD 2.5b in 2019 with sales around USD 25b. At a market-cap of USD 450b this would give a P/E ratio of 180. This is probably more than 10 times higher than the historical average P/E ratio for all stocks, so the investors who are willing to pay such a high valuation ratio for the Tesla stock, must clearly anticipate significant growth in Tesla's future sales and earnings. Once Tesla has become a mature company and the sales-growth has declined, it will probably trade at a more normal P/E ratio.

4 Form 10-Q quarterly report ending June 30, 2020 filed with US SEC: www.sec.gov/ix?doc=/Archives/edgar/data/1318605/000156459020033670/tsla-10q_20200630.htm

5 Form 10-K annual report for 2019 filed with US SEC: www.sec.gov/ix?doc=/Archives/edgar/data/1318605/000156459020004475/tsla-10k_20191231.htm#Item_6

Let us again be generous in our assumptions and say that Tesla will trade at a P/E ratio of 20 in 10 years. We then calculate the contribution to the annualized return from this change in valuation ratio:

$$Annualized\ Return\ of\ Revaluation = \left(\frac{P/E_{t+10\,Years}}{P/E_t} \right)^{1/10\,Years} - 1 = \left(\frac{20}{180} \right)^{1/10} - 1 \simeq -20\%$$

So the average sales-growth should be greater than 20% per year for the next 10 years, just to make up for the large decrease in the valuation ratio. If instead the revaluation would occur during only 5 years, then the annualized change in valuation ratio would be about -36%, so the sales-growth would have to be greater than 36% per year for 5 years, just to make up for the large decrease in valuation ratio.

This is a good example of how the simple forecasting formula in Eq. (21) creates estimation errors. The original forecasting formula in Eq. (14) uses multiplication instead of addition between the components of the stock-return. So when the revaluation gives a loss of -36% it actually requires a sales-growth of $1/(1-0.36)-1 \simeq +56\%$ in order to make up for the revaluation loss, and not just a +36% sales-growth. But for quick calculations in your head, the simple formula is often sufficient.

We had also assumed a generous 10% Net Profit Margin in these calculations. What if it is only 5%? Then the current P/E ratio would be 360 instead of 180, and for that to go down to a more normal P/E ratio of 20, would mean a decrease of nearly 95%. If this decrease in valuation ratio is spread over 5 years, it would give an annualized loss of about -44% which would have to be made up from an equivalent sales-growth in order for the investment to merely break even. If we are using the proper valuation formula which is multiplicative instead of additive, we would need a sales-growth of 79% per year for 5 years, calculated as $1/(1-0.44)-1 \simeq +79\%$. So the company's sales would have to grow from around USD 23b in 2020 to nearly USD 425b in 2025, and that is just for the investment to break-even if the shares are bought at a market-cap of USD 450b today, and assuming they will trade at a P/E ratio of 20 in 2025 and the company has a 5% profit margin then.

So the current share-price for TSLA seems to be extremely optimistic with little room for error and disappointment in the implied growth-assumptions. This might seem like an excellent stock for "short-selling", but a few months after I wrote this case-study, Tesla's share-price was nearly 90% higher!

Although a company has explosive growth for several years into the future, if the valuation ratio is much too high when the shares are bought today, we will most likely end up with a poor long-term investment result, or maybe even a big loss. But it is impossible to predict the exact timing of this.

We could do the above calculations for many different assumptions and consider which scenarios might be most likely. But this would be a big job to do manually and that is why we forecast both the mean and standard deviation using Eq. (14) and Eq. (18), so as to estimate the distribution of possible investment returns when buying a stock at different valuation ratios. This will be demonstrated in the following case-studies.

4 Case Study: Procter & Gamble

Procter & Gamble is a large American company founded in 1837, which sells a wide range of consumer products especially for personal care and hygiene. The company has almost 100,000 employees. The company's stock-ticker is PG.

It is important that you read this case study, even if you are not interested in this particular company, as it explains how to calculate the forecasting formulas and interpret the data and plots. Note that all plots for this stock are gathered towards the end of this section for easy comparison with each other.

4.1 Basic Data

Figure 2 shows the basic financial data we will be using. First we have the Share-Price and Sales Per Share plotted on a logarithmic scale, which shows both of them trending upwards, with the Share-Price being far more volatile than the Sales Per Share.

The next plot shows the P/Sales ratio which is just the Share-Price divided by the Sales Per Share. The P/Sales ratio was at its lowest value around 1.4 in year 1994 and at its highest value around 4.5 in the early year 2020 (the data in the plot ends on June 30, 2020). On average the P/Sales ratio was around 2.7 for this 26-year period, and it looks to have been somewhat mean-reverting during this period.

The third plot shows the annual Sales Growth which ranged between +21% in year 2004 and -13% in year 2016 with an overall average of 3.7% for this 26-year period. There is no overall trend in the annual Sales Growth, which also looks to have been somewhat mean-reverting during this period.

The final plot shows the Dividend Yield which was nearly 10% in 1994 and its lowest value was around 1.8% in 2006. For most of this 26-year period the Dividend Yield was fairly stable around its mean value of 3.5%.

4.2 Forecasting Model

We will now show how to use the forecasting model to see if it fits the historical data. We are actually cheating here, because we will use the average historical data for the entire 26-year period to fit the forecasting model, to test if it would have worked properly, if we had known these values in advance.

We will forecast the mean and standard deviation of the stock's annualized return, whose mean is forecast using Eq. (14), which is reprinted here for easy reference:

$$E[Ann\,Return_t] = \frac{a}{P/Sales_t^{\,1/Years}} - 1$$

The parameter *a* is estimated using Eq. (15), where we can use the average numbers from the historical data that we found in the previous section:

$$\begin{aligned} a &\simeq E[\mathit{Dividend\,Yield}+1]\cdot E[P/\mathit{Sales}^{1/\mathit{Years}}]\cdot E[\mathit{Sales\,Per\,Share\,Growth\,Rate}+1] \\ &\simeq 1.035\cdot E[P/\mathit{Sales}^{1/\mathit{Years}}]\cdot 1.037 \end{aligned} \tag{35}$$

For 1-year investment periods we have that $E[P/\mathit{Sales}^{1/1\,\mathit{Year}}]=E[P/\mathit{Sales}]\simeq 2.7$ using the historical data from the previous section, so the parameter becomes $a\simeq 1.035\cdot 2.7\cdot 1.037\simeq 2.90$ and the forecasting formula for 1-year annualized returns is therefore:

$$E[\mathit{Ann\,Return}_t]=\frac{2.90}{P/\mathit{Sales}_t}-1 \tag{36}$$

For 5-year investment periods we really need to calculate the annualized P/Sales ratio for every single historical data-point before taking the average, otherwise we get small errors in the calculation due to “Jensen’s Inequality”, but for simplicity when using a pocket calculator, we may calculate an approximation using just the mean P/Sales as $E[P/\mathit{Sales}^{1/5\,\mathit{Years}}]\simeq E[P/\mathit{Sales}]^{1/5}\simeq 2.7^{1/5}\simeq 1.22$ and inserting this back into Eq. (35) gives the parameter $a\simeq 1.035\cdot 1.22\cdot 1.037\simeq 1.31$ so the forecasting formula for 5-year annualized returns is:

$$E[\mathit{Ann\,Return}_t]=\frac{1.31}{P/\mathit{Sales}_t^{1/5}}-1 \tag{37}$$

Similarly for 10-year investment periods we get the parameter $a\simeq 1.035\cdot 1.10\cdot 1.037\simeq 1.18$ so the forecasting formula for 10-year annualized returns is:

$$E[\mathit{Ann\,Return}_t]=\frac{1.18}{P/\mathit{Sales}_t^{1/10}}-1 \tag{38}$$

The forecasting formula is easy to calculate for the mean annualized return. Unfortunately it is not so easy to calculate for the standard deviation, because it cannot be separated into calculations of the standard deviation for the individual components, but must be computed from all the historical data at once. This is easy enough with a computer and example code is made available, see Section 16.

4.3 Annualized Returns

Figure 3 shows the historical annualized returns for the PG stock overlaid with the Return Curves from the forecasting models that we just calculated in the previous section. The plots have a lot of information so let us explain their content.

The multi-coloured dots show the historical annualized returns for the PG stock. For example, on January 1, 1995 the PG share-price was USD 15.55.[6] A year later on January 1, 1996 the PG share-

6 The share-prices are adjusted for stock-splits, and all missing days including weekends and holidays have been interpolated to make the data easier to work with, as discussed in Section 16.3. January 1 is normally not a trading day.

price was USD 20.77 which gives a return of $20.77/15.55-1\simeq 33.6\%$ but there were also dividends paid to shareholders during the year, so the Total Return when assuming dividends were reinvested without taxes was about 45.8%. On January 1, 1995 the P/Sales ratio was 1.34. So we now have a data-point to show in the top plot in Figure 3. On the x-axis we put the P/Sales value of 1.34 and on the y-axis we put the 1-year annualized return of 45.8%. Another example is on January 1, 2000 where the P/Sales ratio was 3.6 and the Total Return for the following one year was a loss of -25.2%, so we put another dot in the top plot in Figure 3 with a P/Sales ratio of 3.6 on the x-axis and the annualized return of -25.2% on the y-axis. We do this for all days in our data-set between 1994 and 2020 for a total of nearly 9,500 data-points.

Also shown in Figure 3 are the so-called Return Curves for the forecasting model. The black curve is the forecast mean, the green areas are the forecast mean ±1 standard deviation, and the red areas are the forecast mean ±2 standard deviation.

In the top plot for 1-year annualized returns, the forecast mean is the same as we calculated in Eq. (36) above, which is reprinted here for easy reference:

$$E[Ann\,Return]=\frac{2.90}{P/Sales}-1$$

and the standard deviation from Eq. (18) has been calculated by the computer as:

$$Std[Ann\,Return]=\frac{0.690}{P/Sales}$$

Note that the forecasting formulas shown above and in all the plots omit the time-step *t*, because it is implicitly understood that *P/Sales* is a variable that you can set to some number, and *Ann Return* is a stochastic variable that can take on many different values, because the future stock-return is uncertain.

Also shown in the legend of the plot is the so-called coefficient of determination R^2 which was -0.58 for the forecast mean of the stock's 1-year annualized return. This essentially compares the forecast mean to the historical annualized returns and gives us a measure of how well the forecast mean fit the historical data. Normally we say that the worst possible fit has an R^2 value of 0, but that is for linear models. For non-linear models such as this forecasting model, the R^2 can indeed be negative, so the ordinary mean of the historical data on the y-axis was a much better fit than the forecast mean. In other words, the downwards-sloping curve of the forecast mean fit the historical 1-year annualized returns much worse than their mean value of 14.1%, which is shown as the dotted black line in the plot.

5-Year Annualized Returns

Now consider the middle plot in Figure 3 which is for 5-year annualized returns. It has all the same information that we discussed above for the plot with 1-year annualized returns, except that it uses the forecasting model from Eq. (37) and the multi-coloured dots are now the historical 5-year annualized returns.

For example, on January 1, 1995 the PG share-price was USD 15.55 and 5 years later on January 1, 2000 the share-price was USD 54.39 which is an annualized return of $(54.39/15.55)^{1/5}-1\simeq 28.5\%$ but once again there was reinvestment of dividends during those 5 years which brings the annualized return up to 35.5%. On January 1, 1995 the P/Sales ratio was 1.34 so we put a dot for this data-point in the middle plot in Figure 3 with the P/Sales ratio 1.34 on the x-axis and the annualized return 35.5% on the y-axis.

The forecasting model for 5-year annualized returns is much better than for 1-year annualized returns. Whereas the 1-year historical annualized returns did not fit the forecasting model at all, the 5-year annualized returns have an R^2 value of 0.61 which is quite good, as the maximum possible value is 1 which would mean a perfect fit of the forecasting model to the historical data. We can also see from the plot that most of the historical data-points are within 1 standard deviation of the forecast mean.

Note that the forecasting model in Eq. (37) has a slightly different parameter $a=1.31$ compared to $a=1.30$ for the formula used in the middle plot in Figure 3. This is due to small rounding errors when manually calculating the parameter a from rounded mean values, and there may also be a small error due to “Jensen’s Inequality” which states that $E[P/Sales^{1/Years}]\neq E[P/Sales]^{1/Years}$. In this case the difference between the two a parameters is very small, but it could be a greater difference for some data, so it is best to calculate the parameter a using the proper formula $E[P/Sales^{1/Years}]$ instead of the approximation $E[P/Sales]^{1/Years}$.

Also note that the multi-colouring of the dots show the progression of time for the data-points. This is not of any importance in this particular plot, but it will give crucial insight in the following case study on the company Walmart. However, we can still note from the multi-colouring here, that the data-points are scattered nicely over time, so there are different periods in time that have had similar P/Sales ratios, which shows us that that the forecasting model fits the data for different periods in time. Further note that the colours cannot be compared between plots, so e.g. the yellow and blue colours may refer to different time-periods in two different plots, even though the plots are for the same stock.

10-Year Annualized Returns

Now consider the bottom plot in Figure 3 which is for 10-year annualized returns of the PG stock. The forecast mean fits the historical data very well with $R^2 = 0.76$ and most of the data-points are within 1 standard deviation of the forecast mean.

Remember that this forecasting model was fitted to the historical average data in this entire 26-year period between 1994 and 2020. Whether the model's forecasts will also be accurate in the future, depends on whether the future averages for the Dividend Yield, annual growth in Sales Per Share, and P/Sales ratio will be roughly the same as they were in the past 26 years. If we feel reasonably confident that is the case, then we can use the forecasting formulas shown above the plots in Figure 3, or we can adjust the parameters *a* and *b* according to how we expect the future might be different.

Let us try and use the forecasting formulas that have been fitted to the historical average data, and see what they suggest for the future returns on the PG stock. The formulas are taken from the bottom plot in Figure 3 and the stock's P/Sales ratio has been calculated to be about 4.84 at the time of this writing in early September 2020, which is significantly higher than it has ever been before, as can be seen in Figure 2, so the forecasting model is extrapolating the historical data according to these formulas:

$$E[Ann\,Return] = \frac{1.18}{P/Sales^{1/10}} - 1 = \frac{1.18}{4.84^{1/10}} - 1 \simeq 0.8\,\%$$

$$Std[Ann\,Return] = \frac{0.036}{P/Sales^{1/10}} = \frac{0.036}{4.84^{1/10}} \simeq 3.1\,\%$$

This means that the PG stock is forecast to have an average annualized return of 0.8% with standard deviation of 3.1% for the 10 years from September 2020 to September 2030. The low average return is because the forecasting model expects there will be a loss from revaluation back to a lower P/Sales ratio, but the model also expects the loss will be made up for by the future dividends and sales growth. Once again, this assumes the future average Dividend Yield, annual growth in Sales Per Share, and P/Sales ratio will be roughly the same in the next 10 years as they have been in the past 26 years. If you believe the future will be different, then you simply change the model's parameter *a* according to your beliefs about the future.

The forecasting model is not completely sure what will happen, and that is expressed by the standard deviation, which measures the degree of uncertainty around the forecast mean. The standard deviation is calculated from the uncertainty in the historical data for the past 26 years, so that is also just an estimate that might be very different from the actual future.

4.4 Annualized Returns (Mean)

In the middle plot in Figure 3 for 5-year annualized returns on the PG stock, there are some clear outliers for the historical data-points with low P/Sales ratios, which had much higher 5-year returns than the forecasting model would suggest. One situation where this typically occurs is if the P/Sales ratio was very low at the beginning of a 5-year period and very high at the end of that 5-year period – or vice versa. This creates abnormally high or low stock-returns for that particular 5-year period, which become outliers in the scatter-plot. The PG stock mostly had mean-reverting P/Sales and annual growth in Sales Per Share, so its historical data fits the forecasting model very well for longer investment periods, but for other stocks and stock-indices such outliers can be much more pronounced.

The problem arises because we are considering investment periods of an *exact* duration such as 5 years. But when we are doing long-term investing, we are typically not interested in the return for an *exact* period such as 5 years, instead we want an estimate of the *average* return we might expect over, say, 4 to 6 years. So let us instead see some examples of forecasting such mean annualized returns.

2 to 4 Years Annualized Returns

The top plot in Figure 4 shows the historical 2-4 year *average* annualized returns for the PG stock as multi-coloured dots. Note how the multi-coloured dots in these plots are somewhat smoother than for the plots in Figure 3 which were for investment periods of *exactly* 1, 5 or 10 years. This smoothing is the result of averaging over a range of investment periods.

For example, in the 2 years between January 1, 2000 and January 1, 2002 the PG stock had annualized return of about -11.2%, while in the 3 years between January 1, 2000 and January 1, 2003 the PG stock had annualized return of about -3.6%, and in the 4 years between January 1, 2000 and January 1, 2004 the PG stock had annualized return of about 1.9%. The average of these 3 numbers is $(-11.2\% - 3.6\% + 1.9\%)/3 \simeq -4.3\%$. A computer can easily calculate the average for all possible investment periods between 2 and 4 years, not just the 3 points we calculated manually here. This gives an average annualized return of about -2.9% for all possible 2-4 year investment periods starting on January 1, 2000 and ending somewhere between January 1, 2002 and January 1, 2004.

On January 1, 2000 the P/Sales ratio for the PG stock was about 3.66, so now we have a single data-point with the P/Sales value 3.66 on the x-axis and the mean annualized return -2.9% on the y-axis. We let the computer calculate such data-points for all possible start-dates and 2-4 year investment periods between 1996 and 2020, which results in the top plot in Figure 4.

Also shown in that plot is the mean forecast as a solid black curve. This is calculated exactly the same way as we would do it for investment periods of a fixed duration. We merely set the number of years to be the mid-point for the duration-range, so in this case it would be the mid-point between 2-4 years which is 3 years. We then calculate the parameter *a* using Eq. (15) exactly as we did in the previous section for fixed investment periods. The resulting formula is shown in the top plot in Figure 4 and is:

$$E[Ann\,Return] = \frac{1.48}{P/Sales^{1/3}} - 1$$

It is the same formula we use whether we want to forecast for 3-year investment periods or for the average of 2-4 year investment periods. The difference lies in how well the forecast mean fits the historical data. Although we don't show that plot here, the forecast mean has $R^2 = 0.64$ when compared to the historical annualized returns for 3-year investment periods, while the forecast mean has a much higher $R^2 = 0.78$ when compared to the historical *average* annualized returns for 2-4 year investment periods. This is an exceptionally good fit, especially considering that it is for fairly short investment periods and covers 26 years of data. That is indeed why we use the PG stock as an example here, to show a stock where it works really well. You should not expect such a good fit in general.

The standard deviation is also calculated from Eq. (18) by setting the number of investment years to 3, just as we would do for a fixed investment period of that duration. The computer has calculated the parameter *b* using Eq. (19) and the resulting formula is shown above the top plot in Figure 4 and is:

$$Std[Ann\,Return] = \frac{0.130}{P/Sales^{1/3}}$$

At the time of this writing in early September 2020 the P/Sales ratio is around 4.84 and inserting this into the two forecasting formulas above, we get a forecast mean of about -12.5% with standard deviation of 7.7%. These are the same numbers that we use to forecast the annualized return between September 2020 and exactly 3 years later in September 2023, as well as forecasting the *average* annualized return between September 2020 and all possible 2-4 year investment periods between September 2022 and September 2024.

However, there is an important difference. When forecasting for a range of investment periods such as 2-4 years, the standard deviation should probably be significantly lower. This is because the parameter *b* in Eq. (19) really should be calculated for 2-year averages of the P/Sales ratio, growth in Sales Per Share, and Dividend Yields, which would have a lower standard deviation than their daily numbers. For the rest of this paper, however, we will continue calculating the parameter *b* using the standard formula in Eq. (19), but it would be a fine topic for you to research if this calculation can be improved.

Also note that the dotted black line in Figure 4 may be a bit confusing, because it shows the *mean of the mean* annualized returns. It is basically the mean of the values on the y-axis, which are the mean annualized returns for 2-4 year periods, hence the dotted black line is the *mean of the mean* annualized returns. Because the R^2 is so high, the downwards-sloping solid black curve is a much better fit for the historical data than this mean of the y-axis values shown as the black dotted line.

4 to 6 Years Annualized Returns

The middle plot in Figure 4 shows the historical 4-6 year *average* annualized returns for the PG stock, with the Return Curves of the forecasting model overlaid on the plot. The solid black curve for the forecast mean has a very good fit to the historical data-points with $R^2 = 0.75$.

As explained in the previous section, these forecasting formulas are actually calculated for 5 years which is the mid-point between 4-6 year periods. This plot can be compared directly to the plot for 5-year annualized returns in Figure 3, where it is obvious how the averaging has smoothened the historical annualized returns shown as the multi-coloured dots.

6 to 10 Years Annualized Returns

The bottom plot in Figure 4 shows the historical 6-10 year *average* annualized returns for the PG stock, which averages over 4 years instead of just 2 years as done in the previous two plots in Figure 4. The multi-coloured data-points are now very smooth and clearly follow the curvature of the Return Curves from the forecasting model. The historical data-points are clearly within 1 standard deviation of the forecast mean, which has an exceptionally good fit to the data-points with $R^2 = 0.80$.

4.5 Summary

This section explained in some detail how to use the forecasting model with data for the PG stock. We fitted the forecasting model using the average Dividend Yield, P/Sales ratio and annual growth in Sales Per Share for the 26 years between 1994 and 2020. We then compared the Return Curves produced by the forecasting model to the actual annualized returns during the same 26-year period. This is a form of cheating, because the data used to fit the forecasting model was taken from “the future”. But it allows us to test whether the forecasting model works if we know the correct input parameters for the future.

We saw that the forecasting model clearly did not work for 1-year annualized returns, because the P/Sales ratio had no predictive power for such short investment periods. But for investment periods of just a few years, the forecasting model fit the historical data exceptionally well with high R^2 values. The reason it worked so well for this particular stock, is because its P/Sales ratio and annual growth in Sales Per Share had been somewhat mean-reverting during the past 26 years. You should generally not expect the forecasting model to be so accurate on individual stocks.

Whether these forecasting models will also be so accurate in predicting the future returns on the PG stock, depends entirely on whether the future P/Sales ratios, the future growth in Sales Per Share, and the future Dividend Yield will revert to their historical averages for the past 26 years, which were used in these forecasting models.

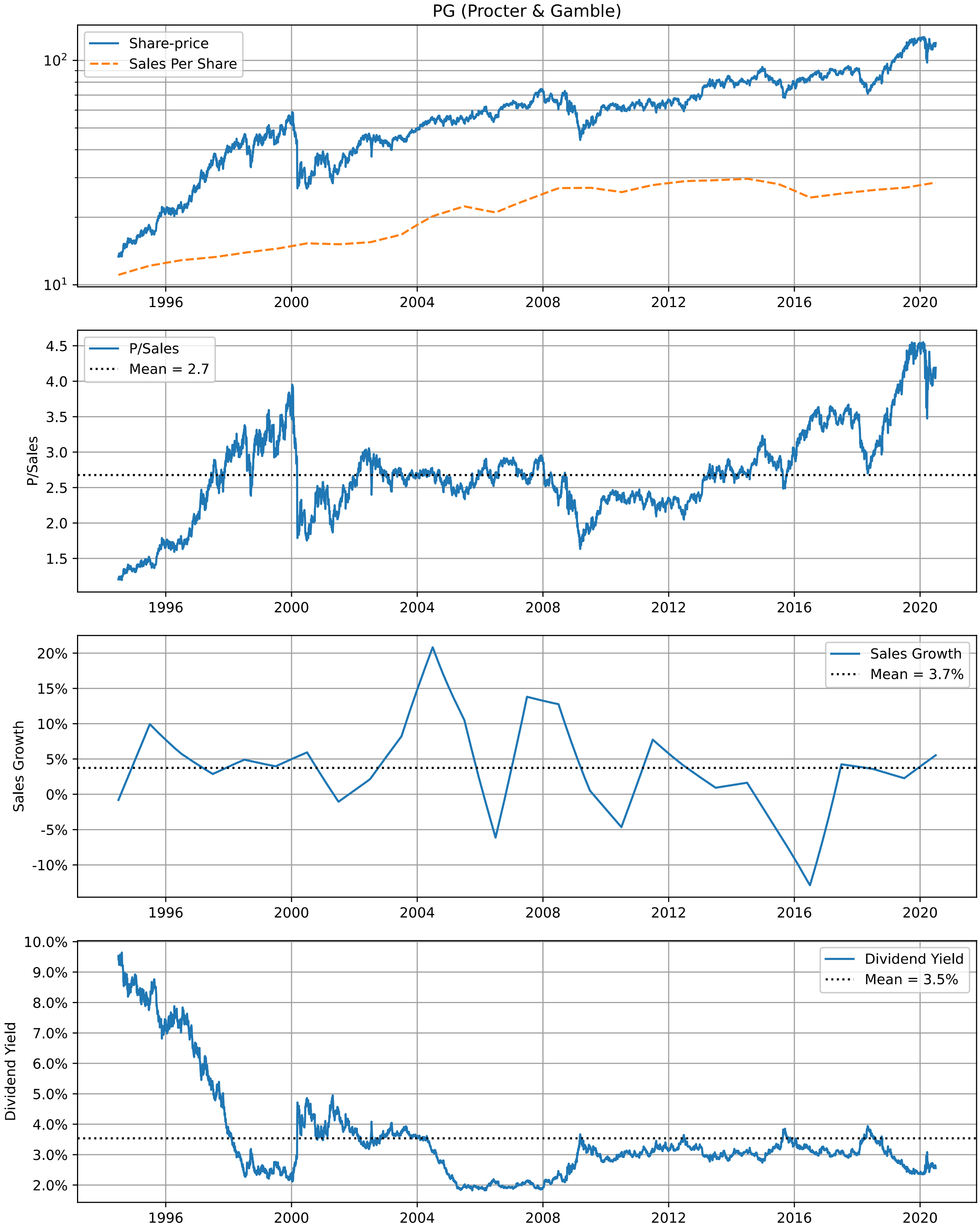


Figure 2: Basic financial data for the company Procter & Gamble (PG).

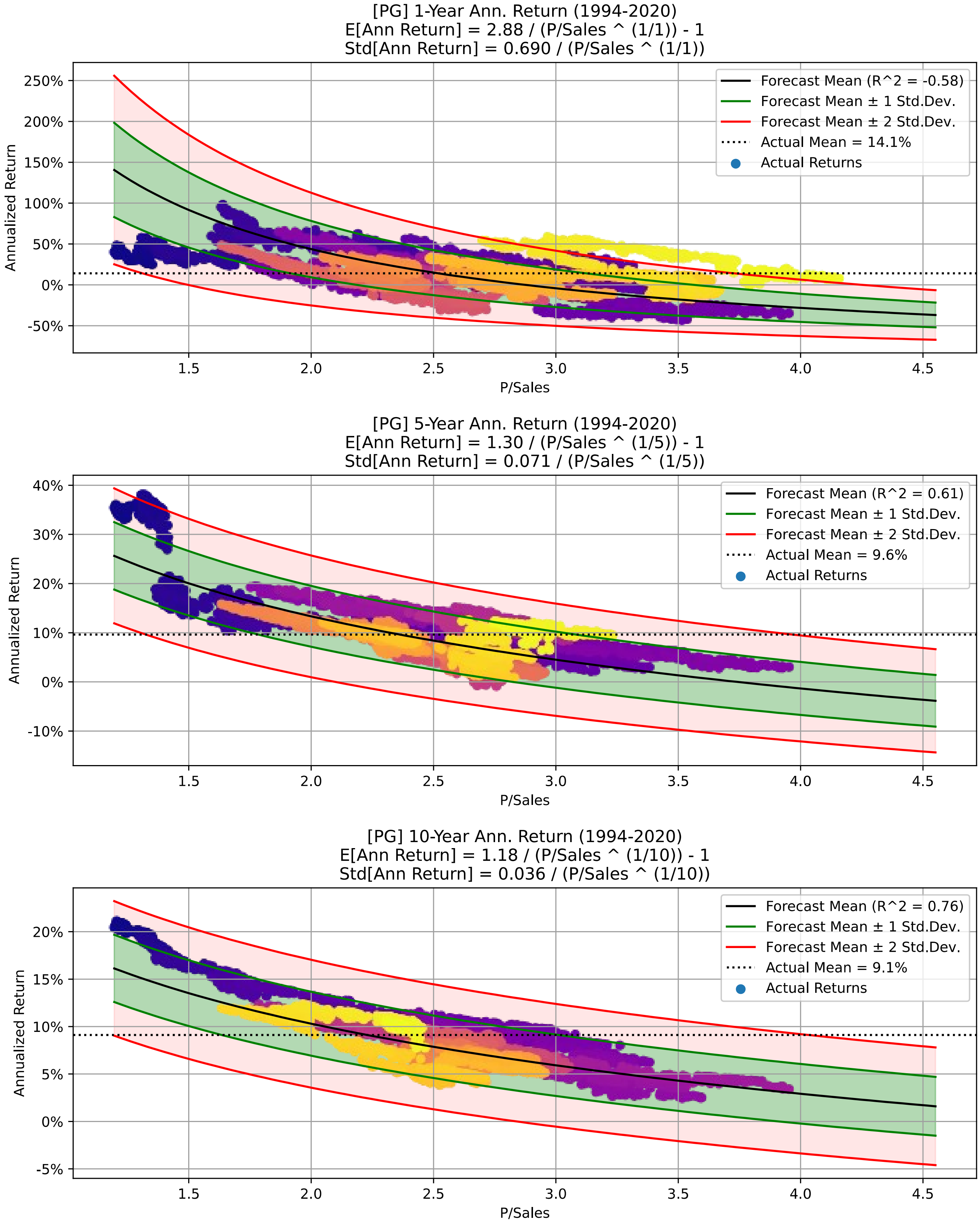


Figure 3: Historical annualized returns for the company Procter & Gamble (PG).

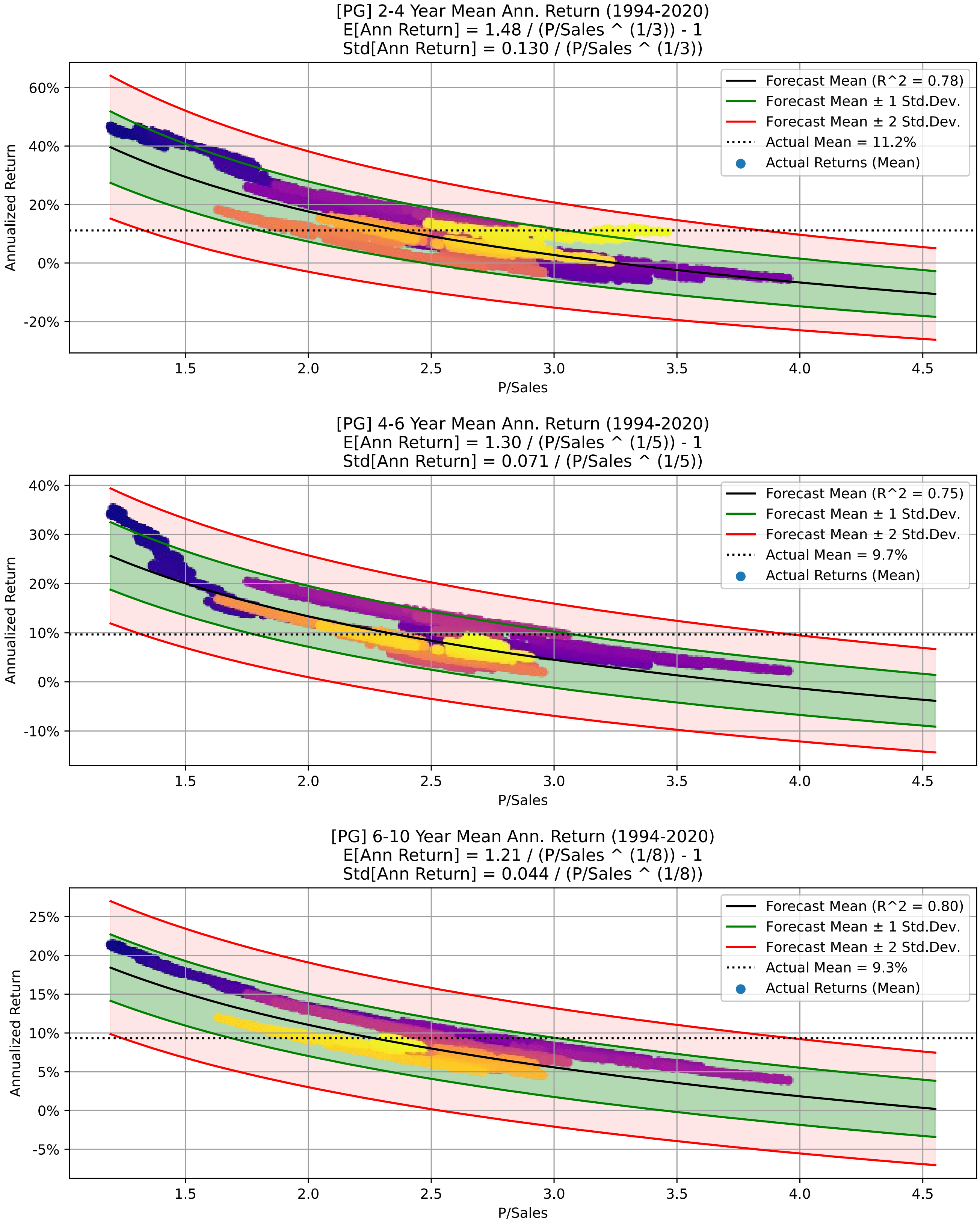


Figure 4: Historical <u>mean</u> annualized returns for the company Procter & Gamble (PG).

5 Case Study: Walmart

In the previous case study we saw an example of how the forecasting model fit the historical data exceptionally well for the Procter & Gamble (PG) stock. This case study is an example of the opposite, where the forecasting model has a horrible fit on the historical data of the company Walmart (WMT), which is a gigantic retailer founded in 1962 in USA and now has 2.2 million employees worldwide.

5.1 Basic Data

Figure 5 shows the basic financial data for the WMT stock between mid 1975 and early 2020. The top plot shows the growth in the Share-Price and Sales Per Share, which have followed each other closely between 1975 and 2020, but with the Share-Price being far more volatile than the Sales Per Share. Note that the y-axis is log-scaled so this has been nearly exponential growth for several decades.

The second plot in Figure 5 shows the P/Sales ratio which is just the Share-Price divided by the Sales Per Share from the previous plot. The P/Sales ratio has ranged between 0.3 and 1.9 with a mean of 0.8. The P/Sales ratio was low between 1975 and 1982, and then low again from 2005 until today in 2020. Between these two periods, roughly from 1982 to 2005, the P/Sales ratio was mean-reverting around a value of 1. So already here we might expect that the forecasting model will have trouble fitting the entire 45-year period, because the P/Sales ratio has behaved differently for different periods.

The third plot in Figure 5 shows the annual growth in Sales Per Share, which was around 40% in 1975 and has gradually decreased to around 5% in 2020. This is a good example of a company with explosive sales-growth that did not always trade at high valuation ratios. The Walmart company had average annual growth in Sales Per Share around 35% between 1975 and 1990, but for the early part of this period between 1975 and 1982, the WMT stock was trading at a very low valuation ratio. It is possible that the entire stock-market was trading at very low valuations at that time, but unfortunately our data for the S&P 500 index in Section 6 only goes back to 1991, so we cannot check that. Furthermore, because the annual growth in Sales Per Share was not mean-reverting, we might expect that the forecasting model will also have trouble fitting this data for the entire 45-year period.

The last plot in Figure 5 shows the Dividend Yield for the WMT stock, which had a very strange spike at 17.5% around the year 1985. We would expect this to show as a corresponding drop in the P/Sales ratio for that time, but that is not the case. This Share-Price data has been obtained from the free internet website called Yahoo Finance and I have noticed similar problems with early Share-Prices and Dividend Yields for other stocks. I think it is a data-problem, possibly related to missing stock-split adjustments for the dividends, or something like that. Fortunately it is not a big problem for us, as it just means that the average Dividend Yield is slightly over-estimated for this 45-year period, but that is not the reason why the forecasting model cannot fit the historical data, as we will see next.

5.2 Annualized Returns

Figure 6 shows the historical annualized returns for the WMT stock. This is calculated similarly to the previous case study for the PG stock that was explained in greater detail in Section 4.3. The top plot in Figure 6 shows 1-year returns with the forecasting model overlaid. As we can see, the forecasting model has a horrible fit to the historical 1-year stock-returns with $R^2 = -2.18$, as explained previously it can be negative because the forecasting model is non-linear.

The middle plot in Figure 6 shows the historical 5-year annualized returns for the WMT stock and now the forecasting model fits slightly better. But even though the mean forecast is in the centre of the historical data-points, the data-points are so dispersed that the fit only has $R^2 = 0.11$. Also note that the forecasting model's standard deviation really does not cover the historical data very well, as some of the data-points are beyond 3 standard deviations from the forecast mean.

The bottom plot in Figure 6 shows the historical 5-year annualized returns for the WMT stock and now the forecasting model fits slightly better again. Once again the mean forecast is in the centre of the historical data-points, but again the data-points are so dispersed that the fit only has $R^2 = 0.21$, and the forecasting model's standard deviation also does not cover the historical data-points very well.

Why does the forecasting model fit the historical data so poorly, even when the forecasting model's parameters *a* and *b* have been calculated from Eq. (15) and Eq. (19) using the "future" data for the entire 45-year period? We already hinted at the reason in the previous section, namely that the P/Sales ratio and annual growth in Sales Per Share were not mean-reverting during this 45-year period. The forecasting model was fitted to the mean of the historical data, which had an average annual growth in Sales Per Share of around 20%, but in reality the growth-rate was around 40% in 1975 and it gradually decreased to about 5% in 2020. Similarly for the P/Sales ratio which was also not mean-reverting during this entire 45-year period. So in the plots in Figure 6 the forecasting model's so-called Return Curves are shown roughly at the centre of the historical data-points, because the forecasting model has made an average compromise of the entire 45-year period.

To further demonstrate this point, notice the colour-coding of the historical data-points in Figure 6, where dots of similar colour are close in time, and dots of different colours are far apart in time. It is especially clear in the bottom plot for 10-year annualized returns, that the different "levels" or "sub-curves" in the historical data belong to different time-periods because they have very different colours.

Let this be a warning, that when we are doing long-term investing, it has a great impact on our investment return whether a company grows its Sales Per Share at 40% per year or 5% per year, and whether the average valuation ratio for the past will also continue in the future. We cannot blindly take the averages of historical data and assume they will continue to be valid in the future.

5.3 Annualized Returns (Mean)

Figure 7 shows the historical *average* annualized returns for the WMT stock, which is calculated similarly to the previous case study for the PG stock that was explained in greater detail in Section 4.4.

Comparing this to Figure 6 which shows the historical annualized returns for fixed durations of 1, 5, or 10 years, it is obvious that the only difference is that the scatter-plots in Figure 7 are more smooth because the annualized returns are averaged over several years. The forecasting model still fits the historical data-points very poorly, for the same reason explained in the previous section, namely that the WMT stock did not really have a mean-reverting P/Sales ratio, and the annual growth in Sales Per Share decreased gradually during this 45-year period from about 40% in 1975 down to 5% in 2020. The forecasting model was fitted to the average P/Sales ratio and annual growth in Sales Per Share for this entire 45-year period, so naturally the mean forecast is in the centre of the historical data-points, which are otherwise very dispersed, as shown in the plots in Figure 7.

5.4 Summary & Discussion

In this section we saw an example of a stock, where the forecasting model had a very poor fit to the historical annualized returns, even when the model was fitted using “cheating” data for the entire 45-year period. It also did not help to average and smoothen the historical returns over multi-year periods.

The reason is that the WMT stock did not really have a mean-reverting P/Sales ratio during this period, and more importantly, its annual growth in Sales Per Share gradually decreased from about 40% in 1975 to only about 5% in 2020. Because the forecasting model was fitted with the averages for the entire 45-year period, its Return Curves were merely located in the centre of the historical data-points.

The lesson here is that you should not blindly use the historical averages for the Dividend Yield, P/Sales ratio and annual growth in Sales Per Share. For some stocks these tend to be mean-reverting over time, but for other stocks such as WMT they have changed dramatically over time.

In this particular case with the WMT stock, if we were to make a forecasting formula for use today, we would not be using the average annual growth in Sales Per Share for the past 45 years which was around 20%, because it is highly unlikely that the Walmart company can grow its sales by 20% per year at this point, as it is already an incredibly big retail company. So we would probably use the average growth-rate for the past 5-10 years instead, or maybe just make a conservative assumption that the future sales growth will match inflation, and similarly we might use the average P/Sales ratio and Dividend Yield for the past 5-10 years. This would probably give us a more realistic forecast for the future returns on the WMT stock.

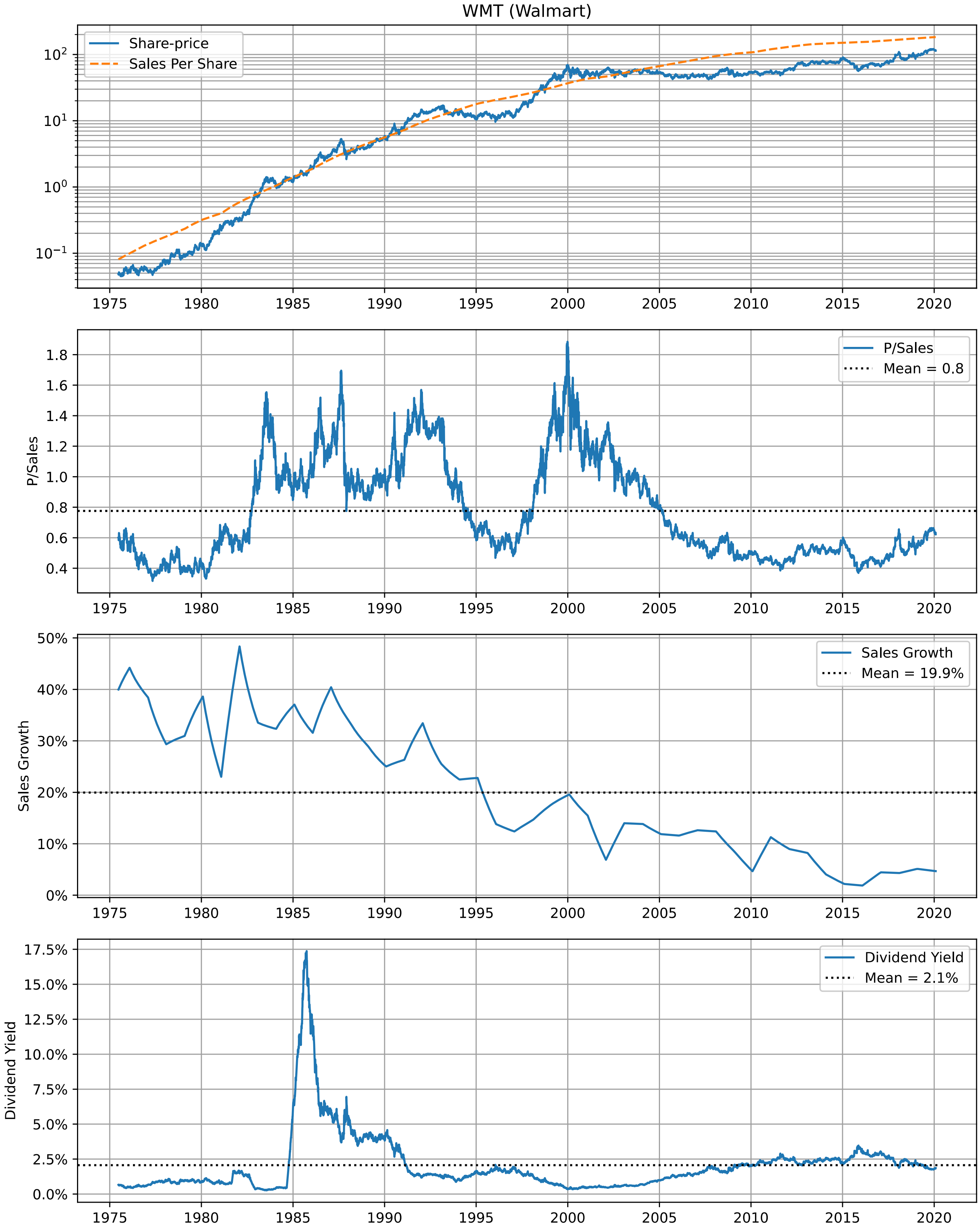


Figure 5: Basic financial data for the company Walmart (WMT).

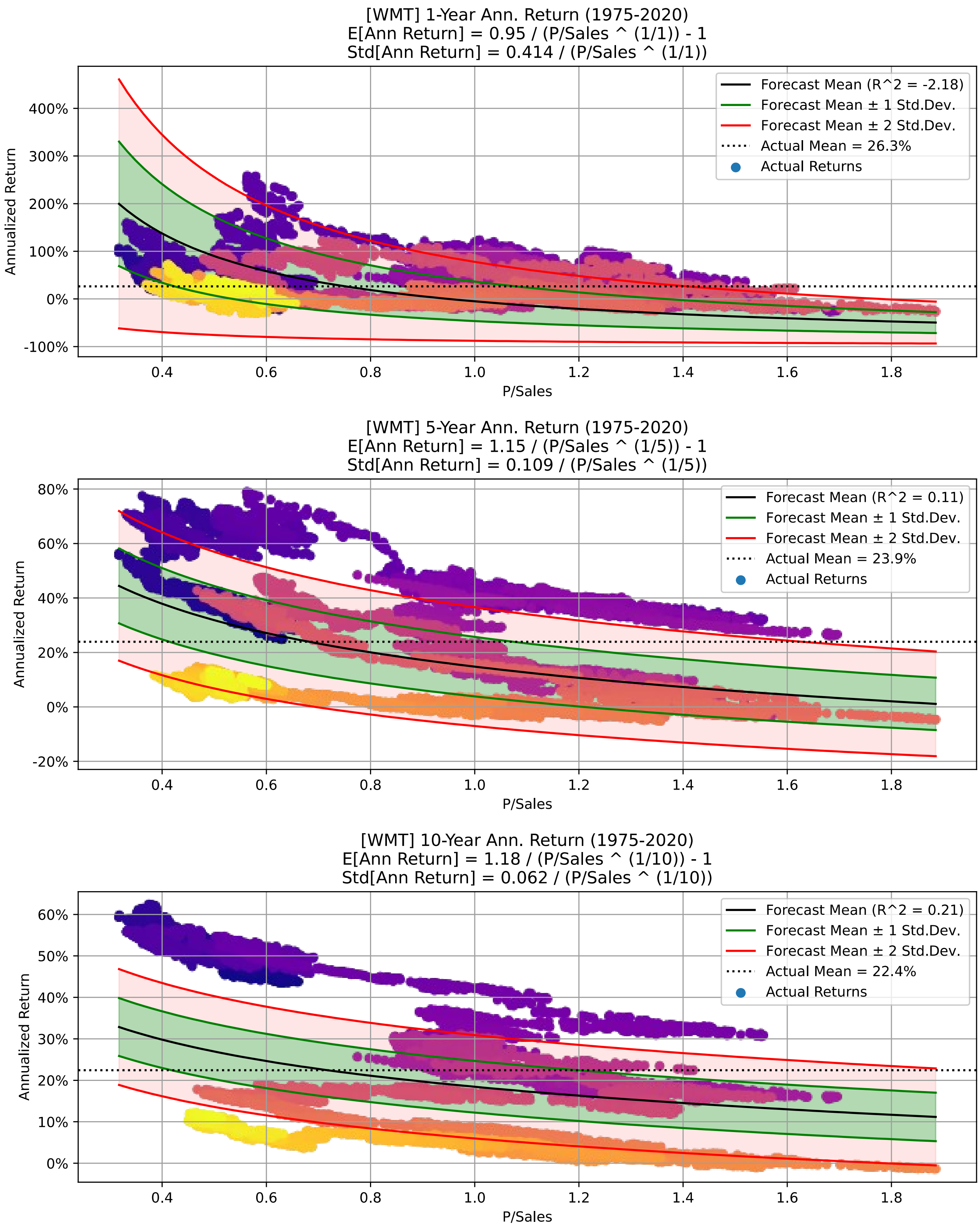
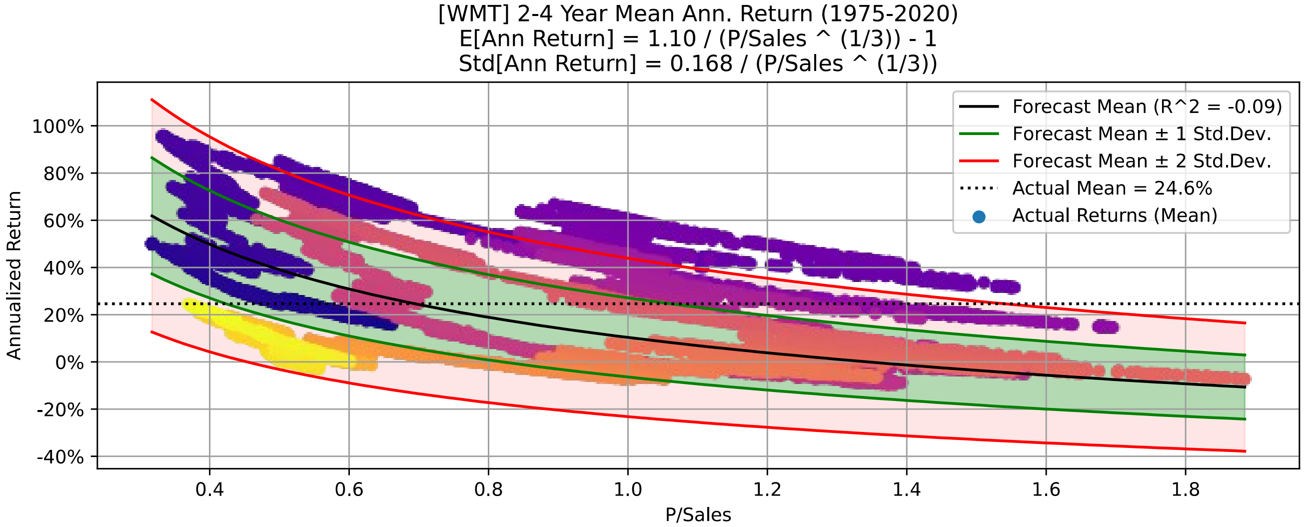
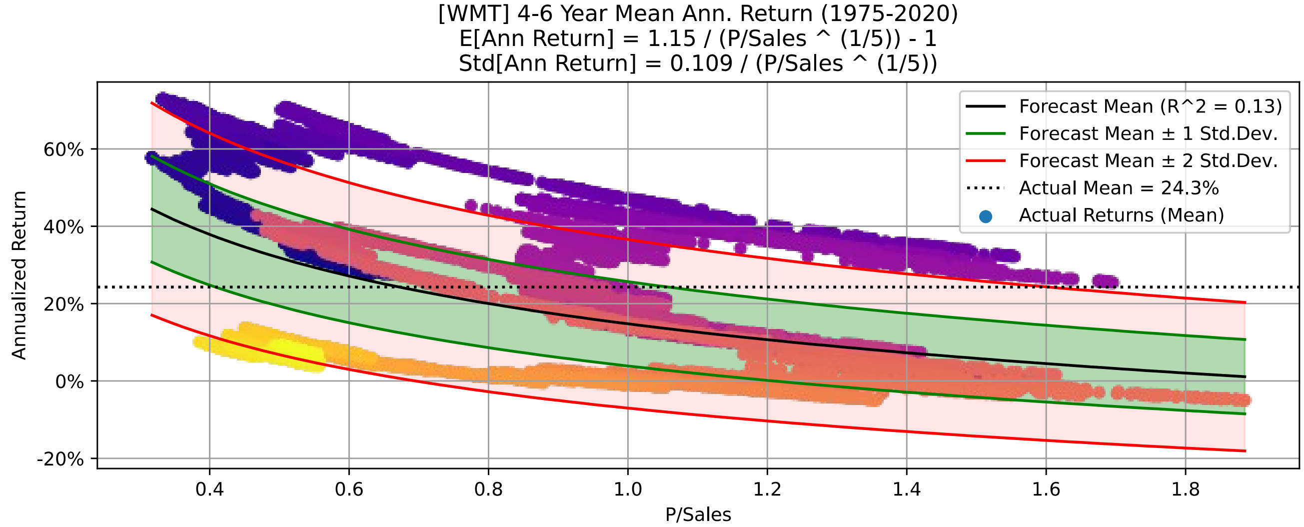


Figure 6: Historical annualized returns for the company Walmart (WMT).





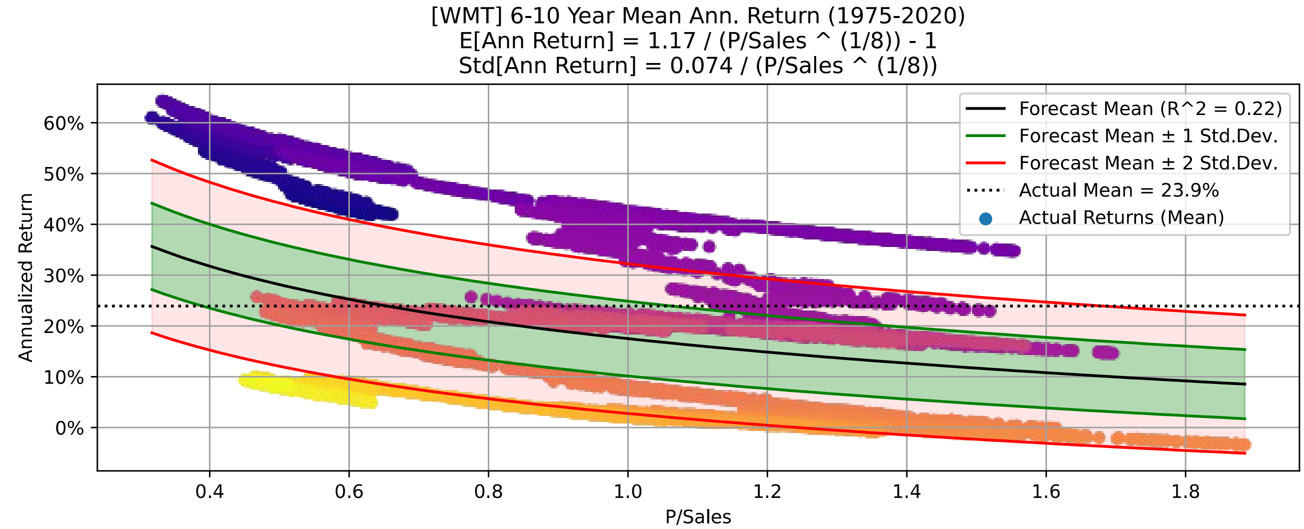


Figure 7: Historical mean annualized returns for the company Walmart (WMT).

6 Case Study: S&P 500 (U.S. Large-Cap)

In the previous two case studies, we first saw an example with the PG stock whose historical annualized returns could be fit very well by the forecasting model, and we then saw another example with the WMT stock whose historical annualized returns could not be fit by the forecasting model. The difference was that the PG stock had somewhat mean-reverting P/Sales ratio and annual growth in Sales Per Share so its historical averages could be used to forecast the future, while the WMT stock did not really have a mean-reverting P/Sales ratio and its annual growth in Sales Per Share decreased gradually during 45 years from 40% in the beginning down to only 5% towards the end. If we had been able to foresee these changes in the P/Sales ratio and Sales Per Share for the WMT stock, then we could have made much better predictions, of course. But if we are just using the historical averages, then the forecasting model will not work if the future is radically different from the historical averages.

That is why it might be better to use the forecasting model on a broadly diversified index of stocks, where the historical averages for the P/Sales ratio and annual growth in Sales Per Share may be a better predictor for the future.

One of the most common stock-indices is the so-called S&P 500 which consists of the stocks of 500 of the largest companies in USA, which operate in a wide variety of industries including energy and utility, financial services, health care, information technology, heavy industry, manufacturers of consumer products, etc. The S&P 500 covers about 80% of the whole U.S. stock-market in terms of size, so it is useful as a gauge for the entire U.S. stock-market.

6.1 Basic Data

Figure 8 shows the basic financial data for the S&P 500 index. The individual stocks are weighted in the index by their market-capitalizations, so the largest companies also have the largest weights in the index. There is a specific methodology used to calculate financial data for the entire index from the weights of the individual stocks in the index, and you are referred to the official website for details.[7] For our purposes we will just consider the S&P 500 index as if it was a single stock and you can indeed invest in shares of so-called Exchange Traded Funds (ETF) that closely track the S&P 500 index, but you trade the ETF shares as if they were normal shares of any other company.

The top plot in Figure 8 shows the Share-Price and annual growth in Sales Per Share for the S&P 500 index, which show similar upwards trends during this period from late 1989 to early 2020. Note that

7 https://us.spindices.com/indices/equity/sp-500

the y-axis is log-scaled so this shows roughly exponential growth. The Share-Price is much more volatile than the Sales Per Share, and usually the Share-Price is highest but sometimes it is lowest.

The ratio between the Share-Price and Sales Per Share is the P/Sales ratio which is shown in the second plot in Figure 8. During this 30-year period it looks like there has been 1.5 cycles for the P/Sales ratio, which was very low at around 0.6 in year 1990, and then it peaked at around 2.2 in year 2000 at the so-called “Dot-Com bubble”, then it was low again in 2009 during the so-called “Financial Crisis”, and then it peaked again in early 2020 right before the “Corona-Virus Panic”. On average the P/Sales ratio was around 1.4 for this 30-year period. If and when the P/Sales ratio of the S&P 500 will crash again is impossible to predict. Instead we want to try and predict the future average P/Sales ratio and a reasonable margin of error, and use these to estimate the future long-term annualized returns, without trying to predict exactly what is going to happen to the P/Sales ratio.

The next plot in Figure 8 shows the annual growth in Sales Per Share, which seems to show smaller but also more rapid cycles of maybe 4-6 years. It looks like it has a tendency to revert to its mean of around 4.0% annual growth in Sales Per Share. This confirms our intuition that an index of many broadly diversified stocks would probably have more stable sales growth than individual companies, and it gives us some confidence that we might be able to use the historical average growth in Sales Per Share as a good estimate for the future.

The last plot in Figure 8 shows the Dividend Yield of the S&P 500 index, which shows roughly the inverse pattern of the P/Sales ratio. That is because the dividend payouts are generally more stable than the share-price, so when the stock-index is expensive and has a high P/Sales ratio, the dividend payouts are smaller compared to the share-price, and vice versa. For the entire 30-year period the Dividend Yield was about 2.1% on average.

6.2 Annualized Returns

Figure 9 shows the historical annualized returns of the S&P 500 index overlaid with the forecasting model which is calculated exactly the same way as explained for a single stock in Section 4.3, that is, the parameters a and b are calculated using Eq. (15) and Eq. (19) with the historical Dividend Yield, P/Sales ratio and annual growth in the Sales Per Share of the S&P 500 index, and these parameters are then used in Eq. (14) and Eq. (18) to create the forecasting formulas, just as we did for individual stocks. The forecasting formulas for the S&P 500 index are shown above each of the plots in Figure 9.

Also shown in the plots are the R^2 values when fitting the mean forecast to the historical annualized returns. The forecast has a horrible fit to 1-year returns on the S&P 500 index with $R^2 = -5.1$, which can be negative because the forecasting model is non-linear. We can also see from the scatter-plot that there is no clear relation between the P/Sales ratio and the 1-year returns on the S&P 500 index.

The middle plot in Figure 9 shows the historical annualized returns on the S&P 500 index for 5-year investment periods as multi-coloured dots, which begin to follow the Return Curves of the forecasting model, but there are several periods with outliers in the returns, so the fit only has $R^2 = 0.54$.

The bottom plot in Figure 9 shows the historical annualized returns on the S&P 500 index for 10-year investment periods. This has a much better fit to the forecasting model with $R^2 = 0.75$, but there are still periods with outliers, which is why we also want to consider the average annualized returns over a range of investment periods.

6.3 Annualized Returns (Mean)

Figure 10 shows the *average* annualized returns on the S&P 500 index for a range of investment periods. The top plot shows it for 2-4 year investment periods which only has $R^2 = 0.07$ so the forecasting model is barely better than just using the average of the values on the y-axis. But if you look closely at the plot, you will notice that the multi-coloured dots run somewhat parallel to the black curve for the mean forecast. The explanation is probably that the S&P 500 has a fairly long cycle of maybe 10 years for its P/Sales ratio to revert to its mean, as we saw in Figure 8. So the historical data-points often show as curves that are somewhat parallel to the mean forecast, but shifted up or down relative to the mean forecast.

The middle plot in Figure 10 shows the *average* annualized returns on the S&P 500 index for 4-6 year investment periods. This can be compared directly to the middle plot in Figure 9 that used fixed 5-year investment periods, which are calculated using the same forecasting formulas, as explained in Section 4.4. The only difference is that we are now using *average* annualized returns so the historical data-points are more smooth, and this gives a slightly higher $R^2 = 0.60$ for the forecasting model. When we look at the middle plot in Figure 10 we see that the forecasting model is mostly a pretty good fit to the historical annualized returns, but there are some outliers especially for low P/Sales ratios. But the combination of using both the forecast mean and standard deviation would give a pretty good estimate for the range of possible outcomes, so let us try and demonstrate how to use that here.

At the time of this writing in September 2020 the P/Sales ratio for the S&P 500 index is around 2.40 which is the highest it has been for the last 30 years. We will insert this into the forecasting formulas.

The forecasting formula for the mean annualized return is shown above the middle plot in Figure 10 and it is calculated using Eq. (14) with the parameter *a* being calculated using Eq. (15) as usual. This is the same formula we would use whether we want to forecast for fixed 5-year investment periods or for the average of 4-6 year investment periods. Inserting today's P/Sales ratio of 2.40 gives:

$$E[\mathit{Ann\,Return}] = \frac{1.13}{P/Sales^{1/5}} - 1 = \frac{1.13}{2.40^{1/5}} - 1 \simeq -5.1\%$$

The standard deviation measures the degree of uncertainty around the forecast mean, and it is calculated using Eq. (18) where the parameter *b* has been calculated by the computer using Eq. (19). The resulting formula is also shown above the middle plot in Figure 10 and inserting today's P/Sales ratio gives:

$$Std[Ann\,Return] = \frac{0.076}{P/Sales^{1/5}} = \frac{0.076}{2.40^{1/5}} \simeq 6.4\,\%$$

So the forecasting model tells us that if we invest in the S&P 500 today at a P/Sales ratio of around 2.40, and we hold the investment for 4-6 years, then we can expect an annualized loss of -5.1% with a standard deviation of 6.4%. This is because the forecasting model assumes the currently very high P/Sales ratio of 2.4 will decrease towards its historical average of 1.4, which by itself would give an annualized loss of about -10.2%, and the forecasting model assumes this will be partially made up for by annual growth in the Sales Per Share of around 4.0%, as well as a Dividend Yield of 2.1%. Note that these numbers do not add up perfectly because the calculations in the forecasting formulas are multiplicative instead of additive, as explained in Section 2.3.

Finally the bottom plot in Figure 10 shows the *average* annualized returns on the S&P 500 index for 6-10 year periods, which has a really good fit to the forecasting model's mean with $R^2 = 0.84$. If we again use a P/Sales ratio of 2.40 as we did in the example above, the forecasting formulas for 6-10 year periods, which are shown above the bottom plot in Figure 10, gives a forecast mean annualized return of about -1.4% with standard deviation 4.4%, if we were to invest in the S&P 500 at a P/Sales ratio of 2.40 and hold the investment for 6-10 years. These are different from the forecasts for 4-6 year periods because the loss originating from the revaluation of the P/Sales ratio are spread out over more years.

Whether these will be accurate forecasts for the future return on the S&P 500 index depends especially on whether the future P/Sales ratio will revert towards its historical 30-year average, or if the growth in Sales Per Share will be so high in the coming years, so as to justify the currently high valuation ratio. There is also the possibility that the large technology companies have now become such a big part of the S&P 500 index that their higher profit margins justify a higher P/Sales ratio for the index as a whole, which is something you are encouraged to investigate. You can easily make adjustments to the forecasting formula, if you believe the future P/Sales ratio should indeed by higher than its 30-year historical average.

6.4 Summary

In this section we studied the S&P 500 index of large companies in USA. We found that its P/Sales ratio could not be used to forecast the S&P 500 returns for 1-year periods, and instead we need to consider 5-year and preferably 10-year investment periods, where the forecasting model had a very good fit to the historical data. The reason seems to be that the P/Sales ratio for the S&P 500 index goes through long cycles before it reverts to its mean, while the sales growth cycles are much shorter.

Whether these long-term forecasting models for the S&P 500 index will also work in the future, depends entirely on how accurately you have estimated the model parameters, namely the future average P/Sales ratio, the future growth in Sales Per Share, and the future Dividend Yield.

Because the S&P 500 is an index of 500 broadly diversified stocks from many industries, it seems reasonable to assume that the future long-term averages for these parameters will be somewhat similar to the past 30 years of historical data. But if you believe the future will be different somehow, then you can simply change the parameters of the forecasting model accordingly.

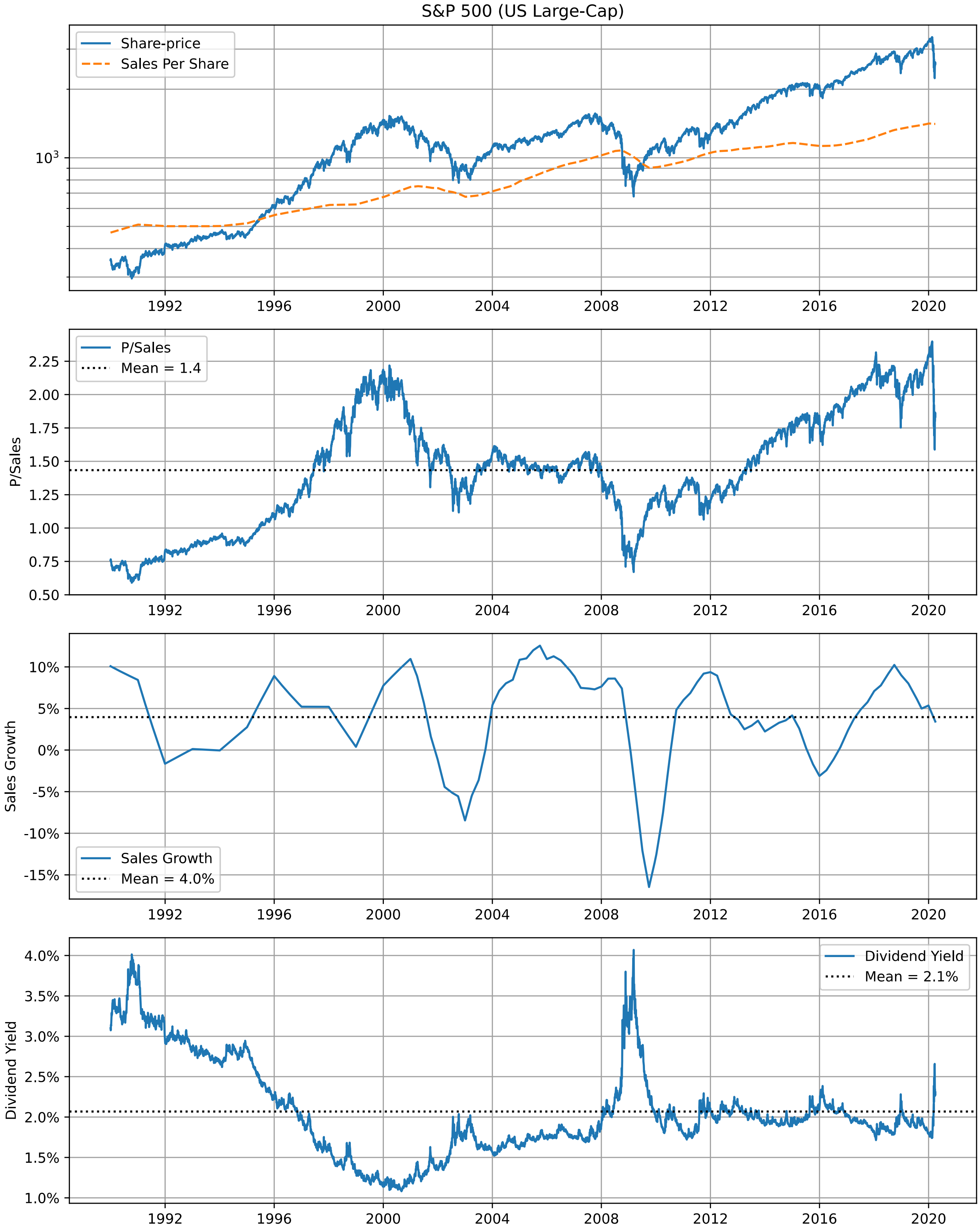


Figure 8: Basic financial data for the S&P 500 stock-index.

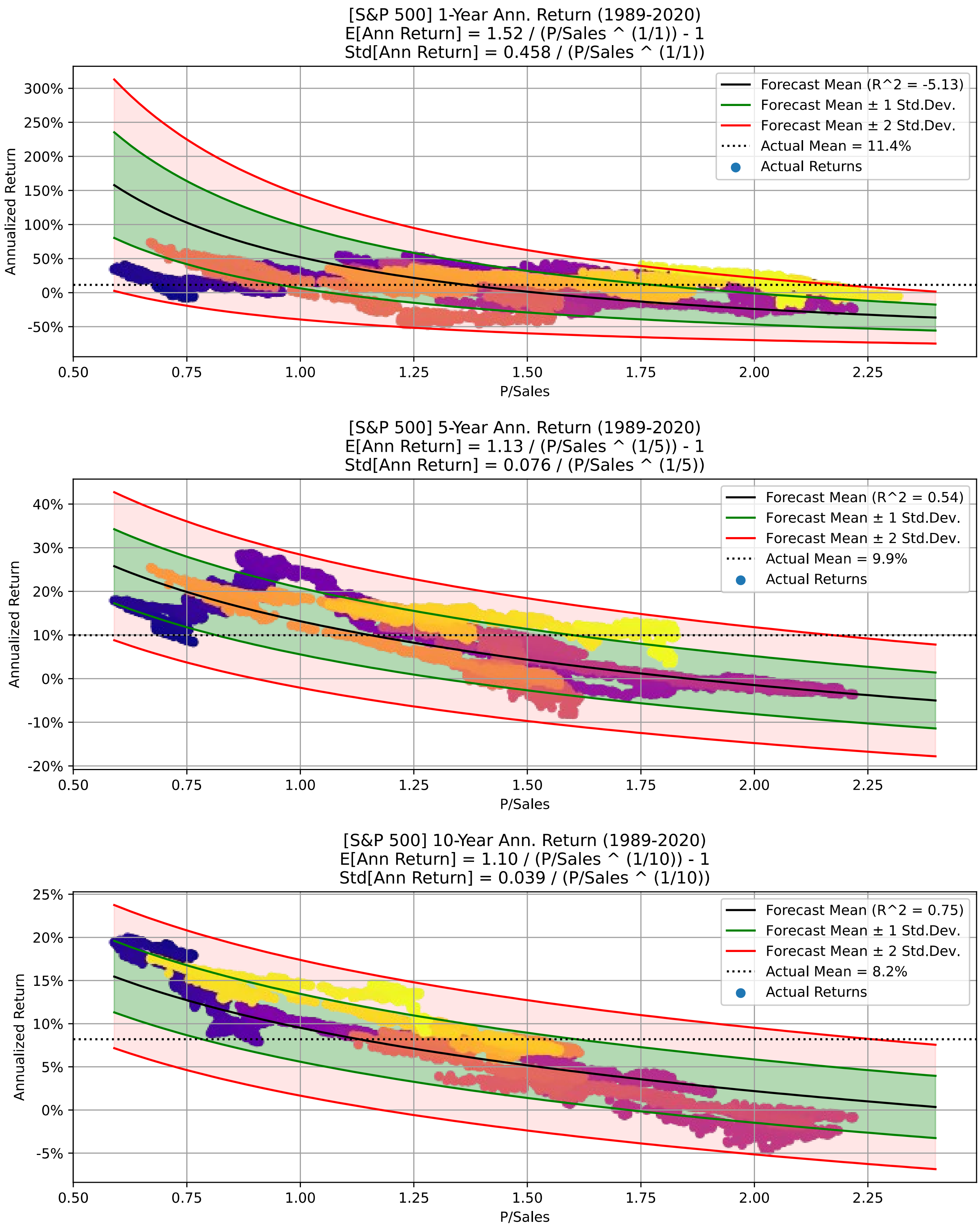


Figure 9: Historical annualized returns for the S&P 500 stock-index.

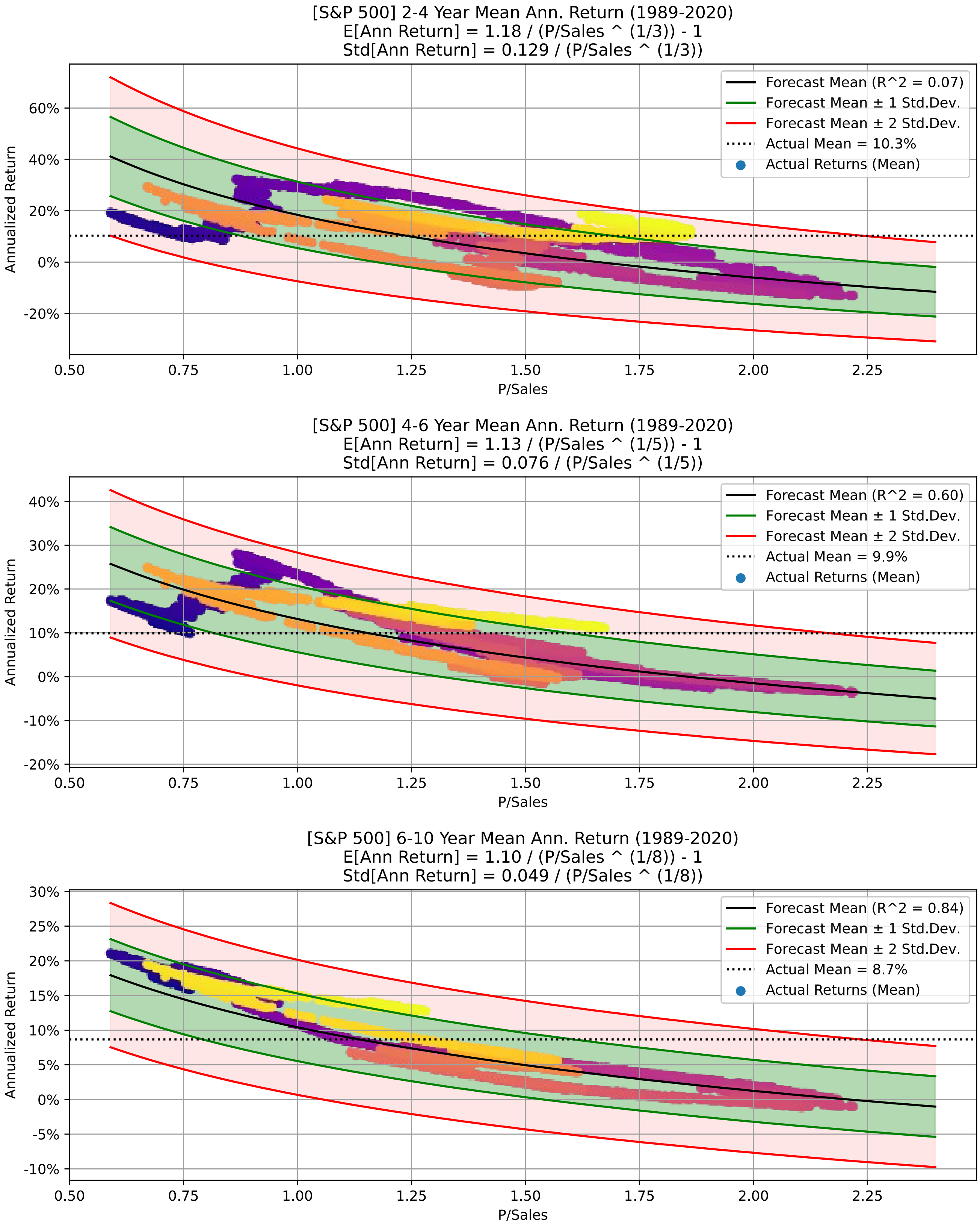


Figure 10: Historical mean annualized returns for the S&P 500 stock-index.

7 Case Study: S&P 400 (U.S. Mid-Cap)

The S&P 400 is a stock-index for 400 U.S. companies that have mid-sized market capitalizations. It is broadly diversified over many different industries. The most comprehensive statistical comparison of the S&P 500 (Large-Cap), S&P 400 (Mid-Cap) and S&P 600 (Small-Cap) stock-indices is probably my own previous paper [7], which concluded that historically the S&P 400 Mid-Cap index generally performed better than the S&P 600 Small-Cap index, and both of these generally performed better than the S&P 500 Large-Cap index. But the paper did not explain *why* that was the case, whether it can be expected to continue in the future, and how we can assess which stock-index might do best in the future. All these questions can be at least partially answered using the forecasting model.

7.1 Basic Data

Figure 11 shows the basic financial data for the S&P 400 index. The top plot shows that the Share-Price and Sales Per Share have grown similarly between late 1996 and mid 2019, with the Share-Price being far more volatile in the short-term than the Sales Per Share.

The second plot in Figure 11 shows the P/Sales ratio between the Share-Price and Sales Per Share, which seems to be mean-reverting with a historical average of 1.1 for this nearly 23-year period.

The third plot in Figure 11 shows the annual growth in Sales Per Share which was 8.8% on average, with its highest growth in 2001, and only having a decline in Sales Per Share in years 2002 and 2009.

The final plot in Figure 11 shows the Dividend Yield which was 1.3% on average. The biggest outlier is during the “Financial Crisis” in 2009 where the Dividend Yield reached 2.8% because the Share-Price became so low, which can also be seen from the low P/Sales ratio in the second plot in Figure 11.

7.2 Annualized Returns

Figure 12 shows the historical annualized returns for the S&P 400 index. The top plot shows the 1-year returns overlaid with the forecasting model that has been calculated using the historical averages for the P/Sales ratio, annual growth in Sales Per Share, and Dividend Yield. The P/Sales ratio could not be used to predict the 1-year returns on the S&P 400 index, even when the forecasting model was “cheating” by having been fitted with the average data for the entire period.

The middle plot in Figure 12 shows the historical annualized returns for 5-year investment periods on the S&P 400 index. The forecasting model has a reasonable fit with $R^2 = 0.53$ which would probably have been better if it were not for a certain part of the data-points that were more than 2 standard deviations from the forecasted mean.

The bottom plot in Figure 12 shows the historical annualized returns for 10-year investment periods on the S&P 400 index, where the forecasting model has a poor fit with only $R^2=0.11$. It is interesting that this has a much worse fit than for the shorter 5-year investment periods which had $R^2=0.53$. The explanation could be that we are considering investment periods that are exactly 10 years, and there were particular 10-year investment periods that were abnormal. We would then expect the plots in the following section to show a better fit because the average of many investment periods are considered.

7.3 Annualized Returns (Mean)

Figure 13 shows the *mean* annualized returns on the S&P 400 stock-index. The top plots shows it for 2-4 year investment periods, where the historical returns fit the forecasting model with $R^2=0.39$. The middle plot in Figure 13 shows the mean annualized returns for 4-6 year investment periods, which has a better fit to the forecasting model with $R^2=0.58$. But then again we see in the bottom plot in Figure 13 that the mean annualized returns for 6-10 year investment periods has a much worse fit to the forecasting model with only $R^2=0.12$.

Several things should be noted here. First note that the forecasting model produces both a mean and standard deviation for the stock-returns, but the R^2 is only calculated from the forecasted mean, so the R^2 does not tell us how well the forecasted distribution (i.e. both the mean and standard deviation) fits the historical data. If we look at the bottom plot in Figure 13 it is clear that most of the data-points are within 1 standard deviation of the forecasted mean and all data-points are within 2 standard deviations, so the forecasting model does fit the historical data-points reasonably well. Also note that the dots are coloured according to their points in time and therefore show that it is particular time-periods that have a somewhat poor fit to the forecasted mean.

In general, we get a poor fit of the forecasting model whenever one of the three components of the historical returns were significantly different than their mean values, that is, if either the change in P/Sales ratio, or the growth in Sales Per Share, or the Dividend Yield were significantly different from the values that were used in the forecasting model.

If we look carefully at the plots in Figure 11 with the basic financial data for the S&P 400 index, we see that the P/Sales, annual growth in Sales Per Share, and Dividend Yield are not completely mean-reverting, but seems to have slight trends over time. This could be shown better by plotting their moving averages for e.g. 10-year periods,[8] which would show that the moving averages have indeed changed somewhat over the past 23 years, and that is why their averages for the entire 23-year period are not accurate for all sub-periods, and therefore gives a slightly poor fit of the forecasting model – similar to what we saw with the more extreme Walmart case-study in Section 5.

8 This is done in the computer code that accompanies this paper, see Section 16, but has been omitted here to save space.

Another reason that we sometimes see a lower R^2 for longer investment periods compared to shorter investment periods, is that we have fewer data-points as the investment periods become longer, and this might remove some of the data-points that had a good fit to the forecasting model when considering shorter investment periods. This can be hard to see in scatter-plots such as Figure 13 because there might be a lot of data-points shown on top of each other, so a small region of the scatter-plot with poorly fitting data-points could actually hold a substantial number of all the data-points available, which would result in a lower R^2 because of how the R^2 formula is defined mathematically.

This naturally raises the question whether we can trust and use the forecasting model for making future predictions when the R^2 is not very high. The answer is as usual, that the future accuracy of the forecasting model depends entirely on how accurately you predict its three components: The future P/Sales ratio, the future growth in Sales Per Share, and the future Dividend Yield. Their historical averages are only useful if they are also good estimates for the future, and you should generally take the R^2 values with a grain of salt.

For example, if we believe the next 10 years of the S&P 400 index is going to be more like the past 10 years than the past 23 years, then we can use the last 10-year averages for the P/Sales ratio, annual growth in Sales Per Share, and Dividend Yield as parameters in the forecasting model, instead of using their full 23-year averages. We then get an average P/Sales ratio of 1.18 compared to 1.12 for the entire 23-year period, and an average annual growth in Sales Per Share of 6.6% compared to 8.8% for the entire 23-year period, and an average Dividend Yield of 1.5% compared to 1.3% for the entire 23-year period.

Using only the past 10 year of financial data to calculate the parameters of the forecasting model in Eq. (15) and Eq. (19), we get the parameters $a \simeq 1.10$ and $b \simeq 0.027$ compared to the parameters $a \simeq 1.12$ and $b \simeq 0.032$ when using the full 23 years of historical data.

At the time of this writing in October 2020 the P/Sales ratio for the S&P 400 index is around 1.13 which is below the average of 1.18 for the past 10 years and only slightly above its average of 1.12 for the entire 23-year data-period. We insert this along with the parameter $a \simeq 1.10$ into Eq. (14) to estimate the future mean annualized return on the S&P 400 index for 6-10 year investment periods:

$$E[Ann\,Return] = \frac{a}{P/Sales^{1/Years}} - 1 = \frac{1.10}{1.13^{1/8}} - 1 \simeq 8.3\,\%$$

And similarly we insert today’s P/Sales ratio with the parameter $b \simeq 0.027$ into Eq. (18):

$$Std[Ann\,Return] = \frac{b}{P/Sales^{1/Years}} = \frac{0.027}{1.13^{1/8}} \simeq 2.7\,\%$$

So if we are using the last 10 years of financial data to forecast the future 6-10 year returns on the S&P 400 index starting in October 2020 where the P/Sales ratio was about 1.13, then the model forecasts a mean annualized return of 8.3% with standard deviation of 2.7%. The majority of this return comes from the assumed growth in Sales Per Share of 6.6% per year and an assumed Dividend Yield of 1.5% per year. The small revaluation from the current P/Sales ratio of 1.13 to the assumed future P/Sales ratio of 1.18 corresponds to an annualized return of only 0.5%. Note that these components do not add up perfectly to the forecasted mean of 8.3%, because the formula is actually multiplicative instead of additive as explained in Section 2.3.

7.4 Summary

In this section we studied the S&P 400 index of mid-sized companies in USA. As usual, we found that its P/Sales ratio could not be used to forecast its returns for 1-year periods, but the forecasting model worked reasonably well for 5-year investment periods. Interestingly the forecasting model's R^2, which measures how well the model fits the historical data, was much worse for longer investment periods such as 6-10 years. There can be several explanations for this phenomenon.

One likely explanation is that we are removing a significant amount of data-points when considering longer investment periods, and this can remove some of the data-points where the forecasting model had a good fit for shorter investment periods, so the R^2 is being calculated with fewer data-points that have a good fit, and the data-points with a poor fit then dominate the so-called "sums of squared errors" that go into calculating the R^2 value.

A more general explanation why we sometimes see a forecasting model that has a poor fit to historical stock-returns, is because the three components of the forecasting model were not completely mean-reverting as we assume when using their mean values in the forecasting model. In this case the S&P 400 had slightly different returns for some 6-10 year periods than the 23-year averages because its return-components were not completely mean-reverting over long periods.

But most of the historical returns for the S&P 400 were still within one standard deviation of the forecasted mean, and all returns were within two standard deviations. So the forecasting model fit the historical data reasonably well. This is not reflected properly in the R^2 which only measures how well the forecasted mean fits the data, and not whether the combination of forecasted mean and standard deviation fits the data well.

In general, we cannot conclude from a low or high R^2 for the historical data, whether the forecasting model will work in the future. This depends entirely on how well we can predict the future P/Sales ratio, annual growth in Sales Per Share, and Dividend Yield. If you can predict these precisely, then you can also predict the future stock-returns precisely, because it is just made up of these components. When you cannot predict these three components precisely, perhaps you can predict a reasonable range of possibilities using their mean and standard deviation.

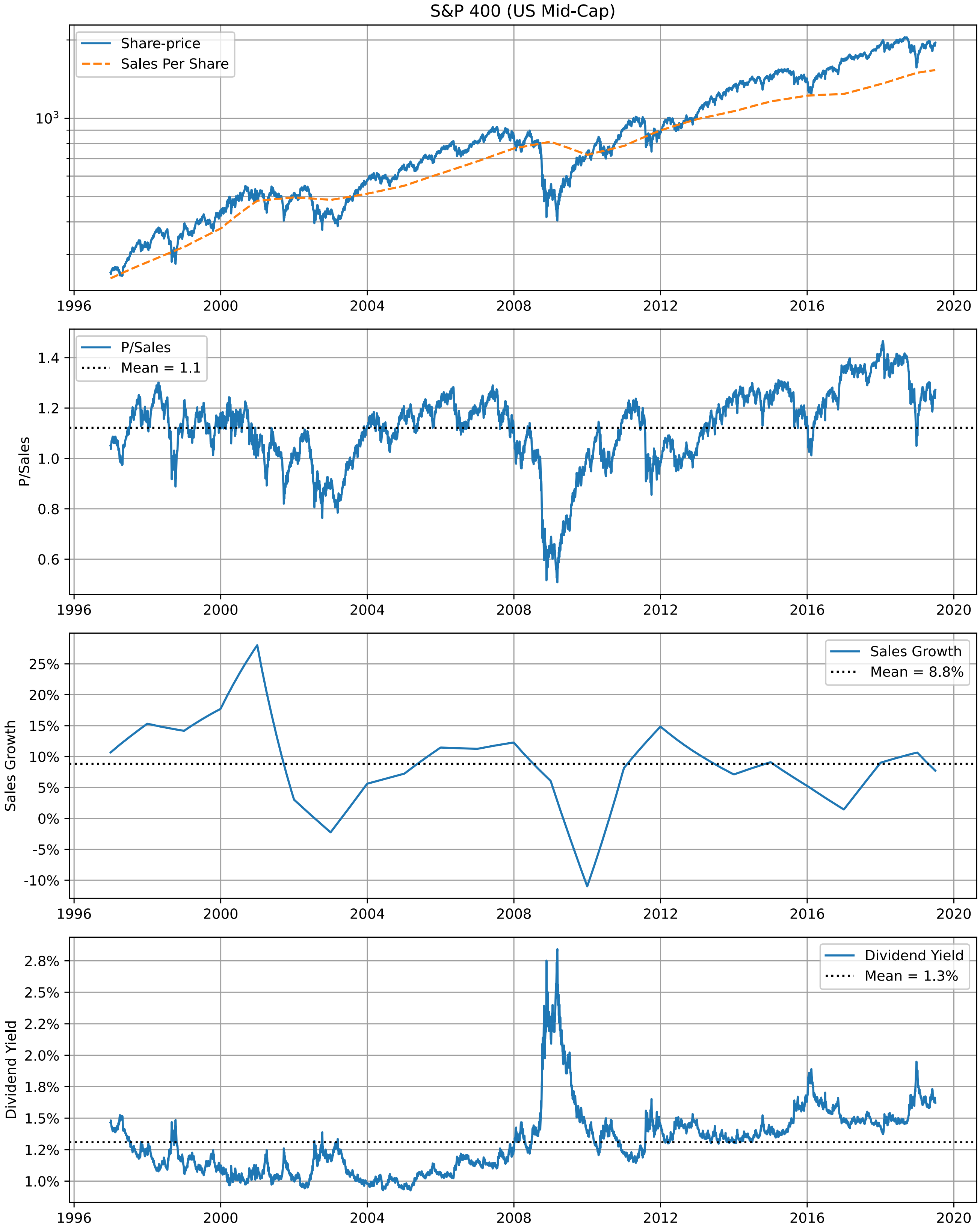


Figure 11: Basic financial data for the S&P 400 stock-index.

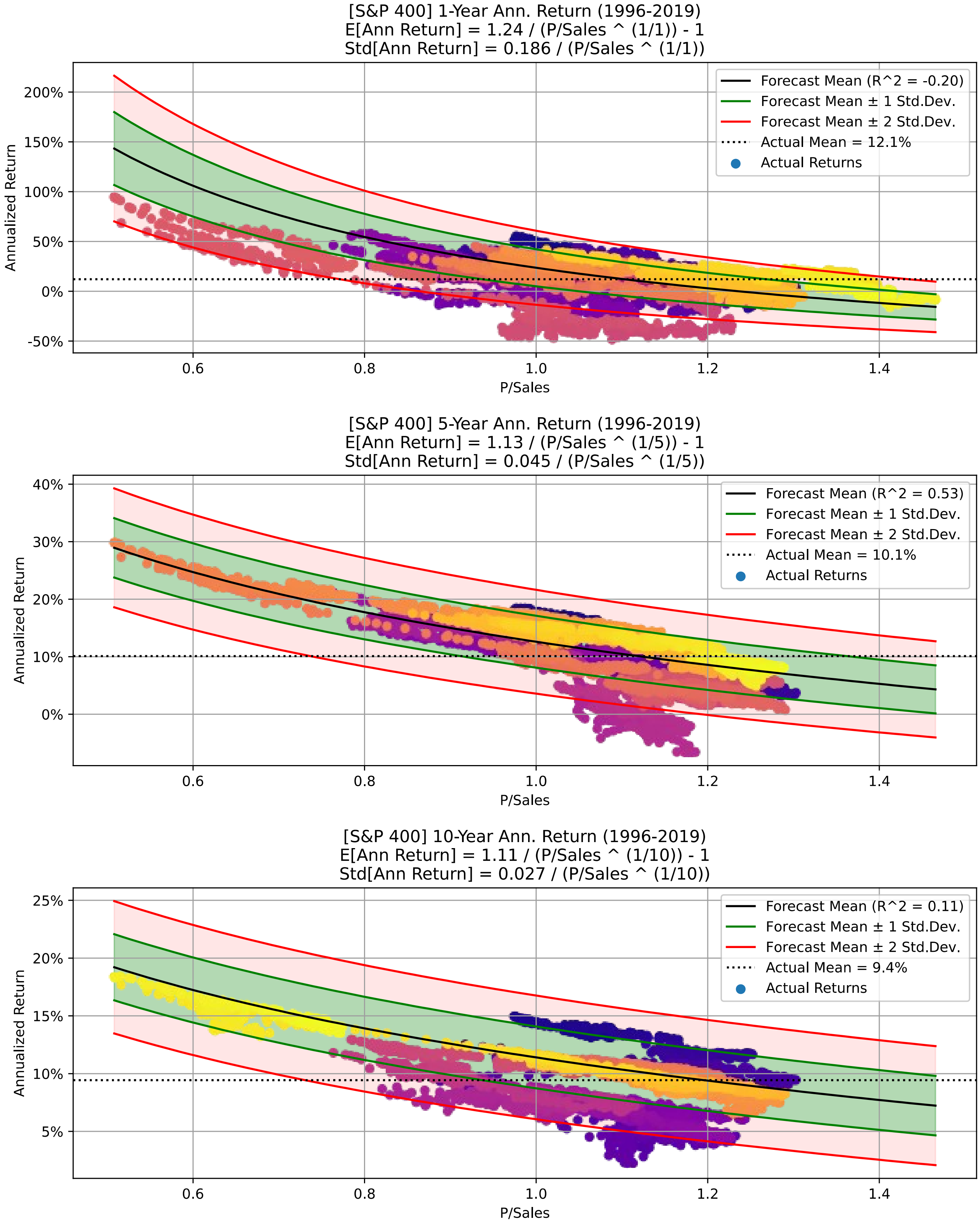


Figure 12: Historical annualized returns for the S&P 400 stock-index.

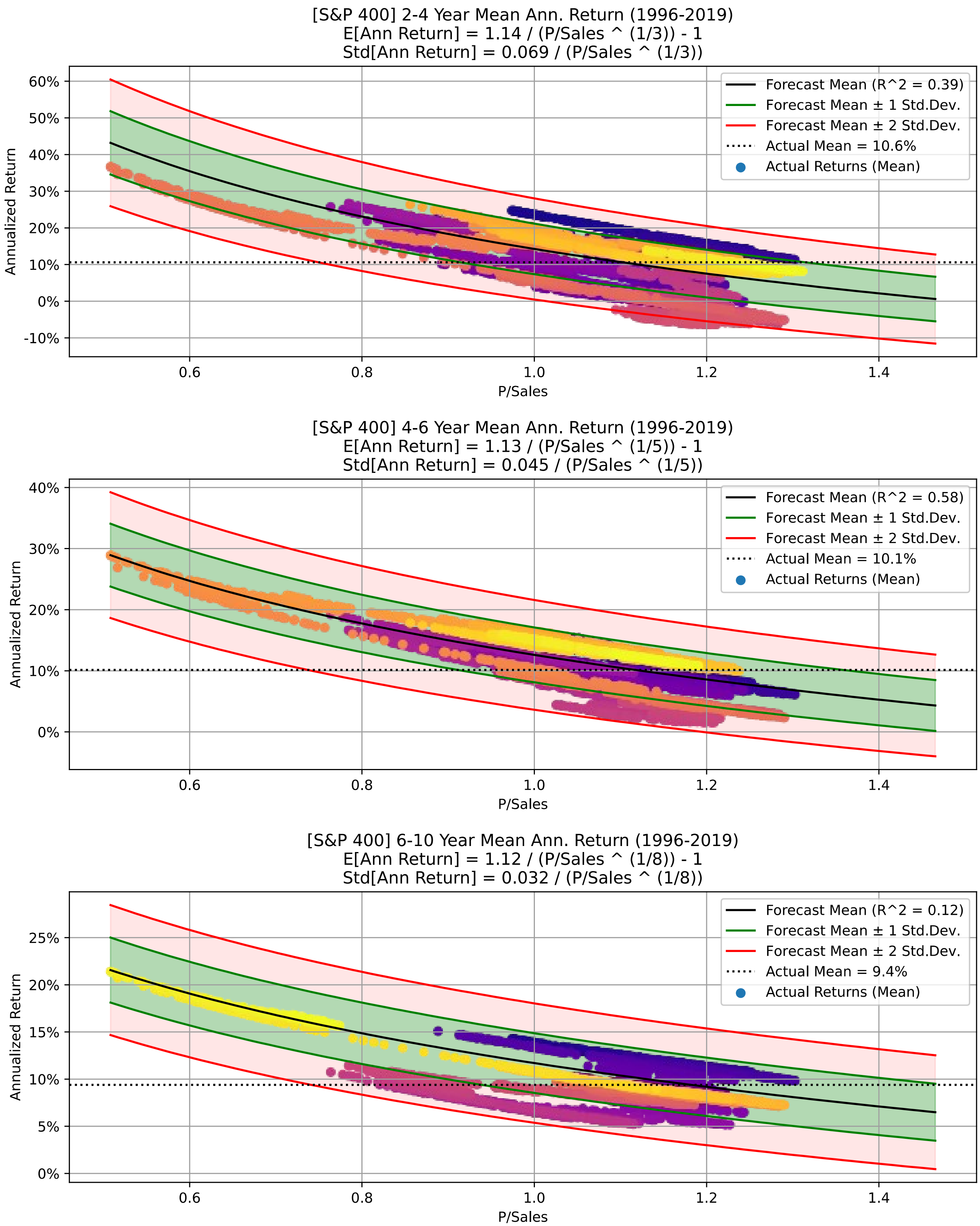


Figure 13: Historical mean annualized returns for the S&P 400 stock-index.

8 Case Study: S&P 600 (U.S. Small-Cap)

The S&P 600 is a stock-index for 600 U.S. companies with small market capitalizations. It is broadly diversified across many different industries. As previously mentioned, the most comprehensive statistical comparison of the S&P 500 (Large-Cap), S&P 400 (Mid-Cap) and S&P 600 (Small-Cap) stock-indices is probably my paper [7], which concluded that historically the S&P 400 Mid-Cap index generally performed better than the S&P 600 Small-Cap index, and both of these generally performed better than the S&P 500 Large-Cap index. But the paper did not explain *why* that was the case, whether it can be expected to continue in the future, and how we can assess which stock-index might do best in the future. All these questions can be at least partially answered using the forecasting model.

8.1 Basic Data

Figure 14 shows the basic financial data for the S&P 600 index. The top plot shows that the Share-Price and Sales Per Share have grown very similarly between late 1996 and mid 2019, with the Share-Price being much more volatile in the short-term than the Sales Per Share.

The second plot in Figure 14 shows the P/Sales ratio between the Share-Price and Sales Per Share, which seems to be mean-reverting with a historical average of 0.9 for this nearly 23-year period, although the cycles are quite long of perhaps 8 years or more.

The third plot in Figure 14 shows the annual growth in Sales Per Share which was 8.7% on average, with the highest growth being about 23% in both years 2000 and 2017, and only having a decline in Sales Per Share in years 2002 and 2009. The annual growth in Sales Per Share seems to be somewhat mean-reverting.

The final plot in Figure 14 shows the Dividend Yield which was 1.0% on average. The biggest outlier is during the “Financial Crisis” in 2009 where the Dividend Yield reached 2.5% because the Share-Price became so low, which can also be seen from the low P/Sales ratio in the second plot in Figure 14. The Dividend Yield has been gradually trending upwards from 0.8% in late 1996 and nearly doubled to 1.5% in mid 2019.

8.2 Annualized Returns

Figure 15 shows the historical annualized returns for the S&P 600 index. The top plot shows the 1-year returns overlaid with the forecasting model that has been calculated using the historical averages for the P/Sales ratio, annual growth in Sales Per Share, and Dividend Yield. As we have seen for the other stocks and indices, the P/Sales ratio could not be used to predict the 1-year returns on the

S&P 600 index either, even when the forecasting model was "cheating" by using the average data for the entire 23-year period.

The middle plot in Figure 15 shows the historical annualized returns on the S&P 600 index for 5-year investment periods, which has a decent fit to the forecasting model with $R^2=0.44$. There is one large cluster of outliers and the colour-coding of the dots shows that these data-points are for the same period. A guess would be that it is the 5-year period that ended around the "Financial Crisis" in 2009 where the share-price for the S&P 600 index was very low along with most other stocks.

The bottom plot in Figure 15 shows the historical annualized returns on the S&P 600 index for 10-year investment periods. The forecasted mean has a very poor fit to the historical data with $R^2=-0.24$. This is similar to what we saw for the S&P 400 index in the previous section. When we get such a poor fit for long investment periods, the first thing to try is to use an average of many investment periods, which we will do in the following section.

8.3 Annualized Returns (Mean)

Figure 16 shows the *mean* annualized returns on the S&P 600 index. The top plots shows it for 2-4 year investment periods, where the historical returns fit the forecasting model with a somewhat low $R^2=0.22$. The middle plot in Figure 16 shows the mean annualized returns for 4-6 year investment periods, which has a decent fit to the forecasting model with $R^2=0.50$. But then again we see in the bottom plot in Figure 16 that the mean annualized returns for 6-10 year investment periods has a much worse fit to the forecasting model with negative $R^2=-0.09$.

This is similar to what we saw for the S&P 400 index in the Section 7.3. Whenever we see historical annualized returns that are far from the mean of the forecasting model, it is because at least one of the three components of the stock-return were far from the mean that was used in the forecasting model. For example, the investment periods that ended during the bottom of the "Financial Crisis" in 2009 would have P/Sales ratios that were much lower than their average for the entire 23-year period.

In the case of the S&P 600 index, both the Dividend Yield and P/Sales ratio have trended somewhat upwards during this 23-year period. This can perhaps be seen from their raw plots in Figure 14, but it becomes much clearer when plotting e.g. 10-year moving averages of the Dividend Yield and P/Sales ratio, which has been omitted here but is available in the computer code, see Section 16.

When we have a stock or a stock-index whose return-components (that is, the P/Sales ratio, annual growth in Sales Per Share, and Dividend Yield) are not exactly mean-reverting but have a slight trend over time, yet we still use the mean of these return-components to calculate the parameters of the forecasting model, then the forecasted mean will be in the middle of the historical data-points as we see in Figure 15 and Figure 16, and the data-points may be more or less dispersed around the forecasted mean, depending on how much the return-components trend over time. We saw a more extreme example of this with the Walmart stock in Section 5, which had explosive sales growth in its early years that gradually declined towards low growth as Walmart became one of the biggest companies in the world.

Although a highly diversified stock-index is probably much more likely to have predictable P/Sales ratios, annual growth in Sales Per Share, and Dividend Yield than individual stocks, both the S&P 600 and S&P 400 indices are good examples that highly diversified stock-indices can also have return-components that are not completely mean-reverting, at least not in short cycles of just a few years, but sometimes seem to have trending return-components for several years.

As we did for the S&P 400 index in the previous section, we could also assume that the next 10 years of the S&P 600 index are going to be more like the previous 10 years, rather than its previous 23 years for the full data-period available. We then get an average P/Sales ratio of 1.01 compared to 0.93 for the entire 23-year period, and an average annual growth in Sales Per Share of 8.7% which was the same for the entire 23-year period, and an average Dividend Yield of 1.2% compared to 1.0% for the entire 23-year period.

Using only the past 10 years of financial data to calculate the parameters of the forecasting model in Eq. (15) and Eq. (19), we get the parameters $a \simeq 1.10$ and $b \simeq 0.034$ compared to the parameters $a \simeq 1.09$ and $b \simeq 0.035$ when using the full 23 years of historical data, and when we want to forecast for 8-year investment periods (which corresponds to 6-10 year average investment periods).

At the time of this writing in October 2020 the P/Sales ratio for the S&P 600 index is around 0.74 which is far below the average of 1.01 for the past 10 years and also well below its average of 0.93 for the entire 23-year data-period. We insert this along with the parameter $a \simeq 1.09$ into Eq. (14) to estimate the future mean annualized return on the S&P 600 index for 6-10 year investment periods:

$$E[Ann\,Return] = \frac{a}{P/Sales^{1/Years}} - 1 = \frac{1.09}{0.74^{1/8}} - 1 \simeq 13.2\%$$

And similarly we insert today's P/Sales ratio with the parameter $b \simeq 0.035$ into Eq. (18):

$$Std[Ann\,Return] = \frac{b}{P/Sales^{1/Years}} = \frac{0.035}{0.74^{1/8}} \simeq 3.6\%$$

So if we are using the last 10 years of financial data to forecast the future 6-10 year returns on the S&P 600 index, starting in October 2020 where the P/Sales ratio was about 0.74, then the model forecasts a mean annualized return of 13.2% with standard deviation of 3.6%. The majority of this return comes from the assumed growth in Sales Per Share of 8.7% per year and an assumed Dividend Yield of 1.2% per year. The revaluation from the current P/Sales ratio of 0.74 to the assumed future P/Sales ratio of 1.01 corresponds to an annualized return of about 4.0%. Note that these components do not add up perfectly to the forecasted mean of 13.2%, because the formula is actually multiplicative instead of additive as explained in Section 2.3.

8.4 Comparing Stock Indices

Now let us compare the forecasted means for the three stock-indices: Section 6.3 gave a forecasted mean annualized loss of -1.4% for the S&P 500 index, while Section 7.3 gave a forecasted mean return of 8.3% for the S&P 400 index, and in the previous section we calculated a forecasted mean return of 13.2% for the S&P 600 index.

If we believe the assumptions that went into these forecasting models are reasonable, then it would seem better to invest more in the S&P 600 index than the S&P 400 index, and only invest little or nothing in the S&P 500 index.

But because all of this is very uncertain and our assumptions for the future could be very wrong, you should be careful not to make decisive bets and invest everything in a single stock-index such as the S&P 600 index in this example, even though its forecasted return was far better than for the S&P 500 and S&P 400 indices. It would probably be wiser to have investments in all three indices, but weigh them according to the forecasted returns as well as how much you trust your assumptions for the future that went into calculating the parameters of the forecasting models for these stock-indices.

An example portfolio allocation could have 10% of your portfolio invested in the S&P 500, 30% invested in the S&P 400, 50% invested in the S&P 600, and the remaining 10% invested in a low-risk bond-fund. You could then rebalance as the valuation ratios of the three indices change relative to each other, and the forecasting models would give other predictions for the future returns.

8.5 Summary

In this section we studied the S&P 600 index consisting of 600 small and diversified companies in USA. The result was somewhat similar to the S&P 400 index in the previous section, namely that the forecasting model did not work for 1-year investment periods, but it worked reasonably well for 4-6 year investment periods – but then the forecasting model had a very poor fit to the historical returns for 6-10 year investment periods. A likely explanation is that we remove a lot of data-points when considering longer investment periods, as explained in more detail for the S&P 400 index in Section 7.3. Another explanation is that the three return-components were not exactly mean-reverting for these two stock-indices: The P/Sales ratio, annual growth in Sales Per Share, and Dividend Yield can trend slightly for extended periods of time, and this means the forecasting model may not fit all sub-periods of the entire 23-year period that were used for the model's parameters. Once again it must be stressed that the forecasting model is only as good as your assumptions for the future P/Sales ratio, the future annual growth in Sales Per Share, and the future Dividend Yield.

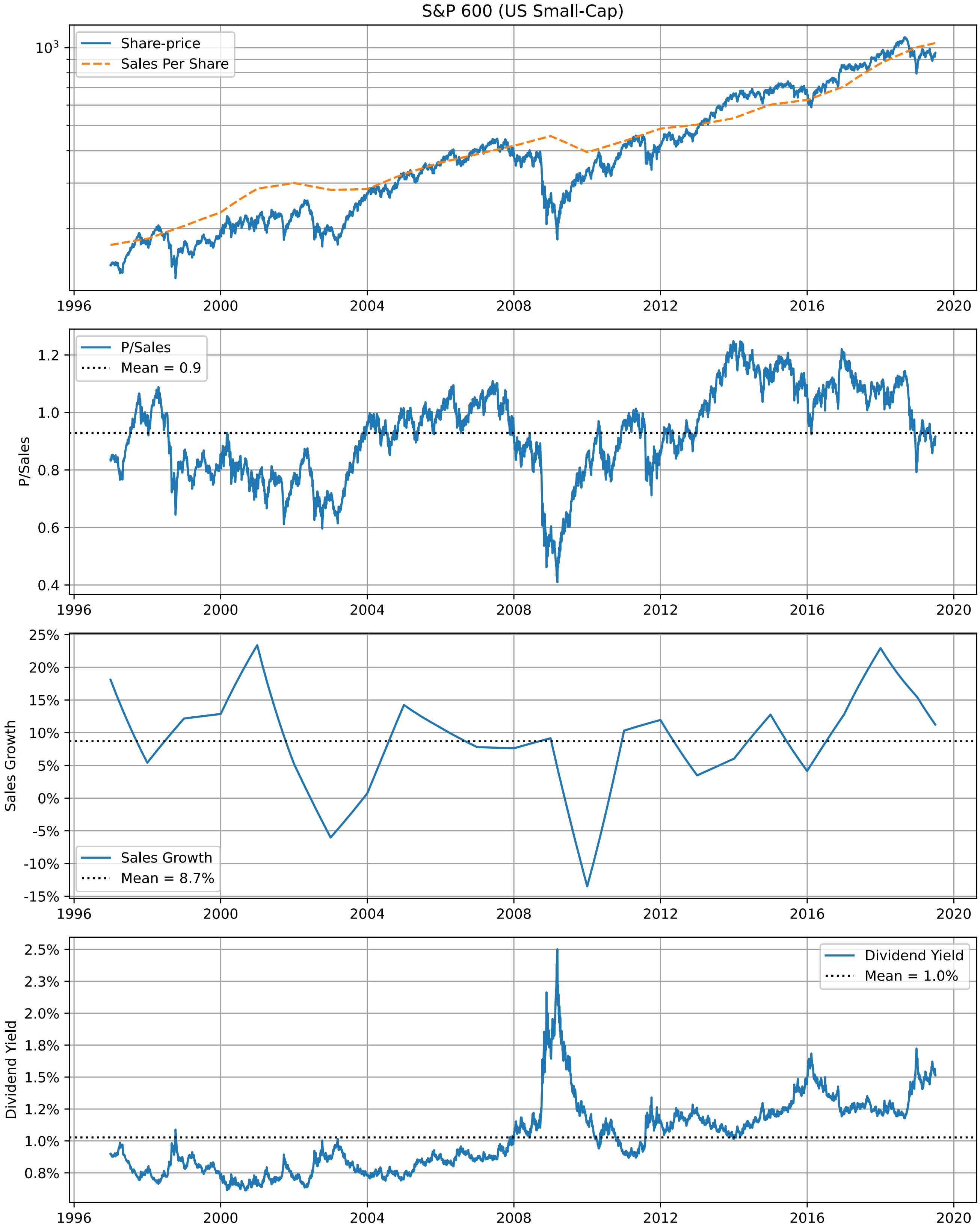


Figure 14: Basic financial data for the S&P 600 stock-index.

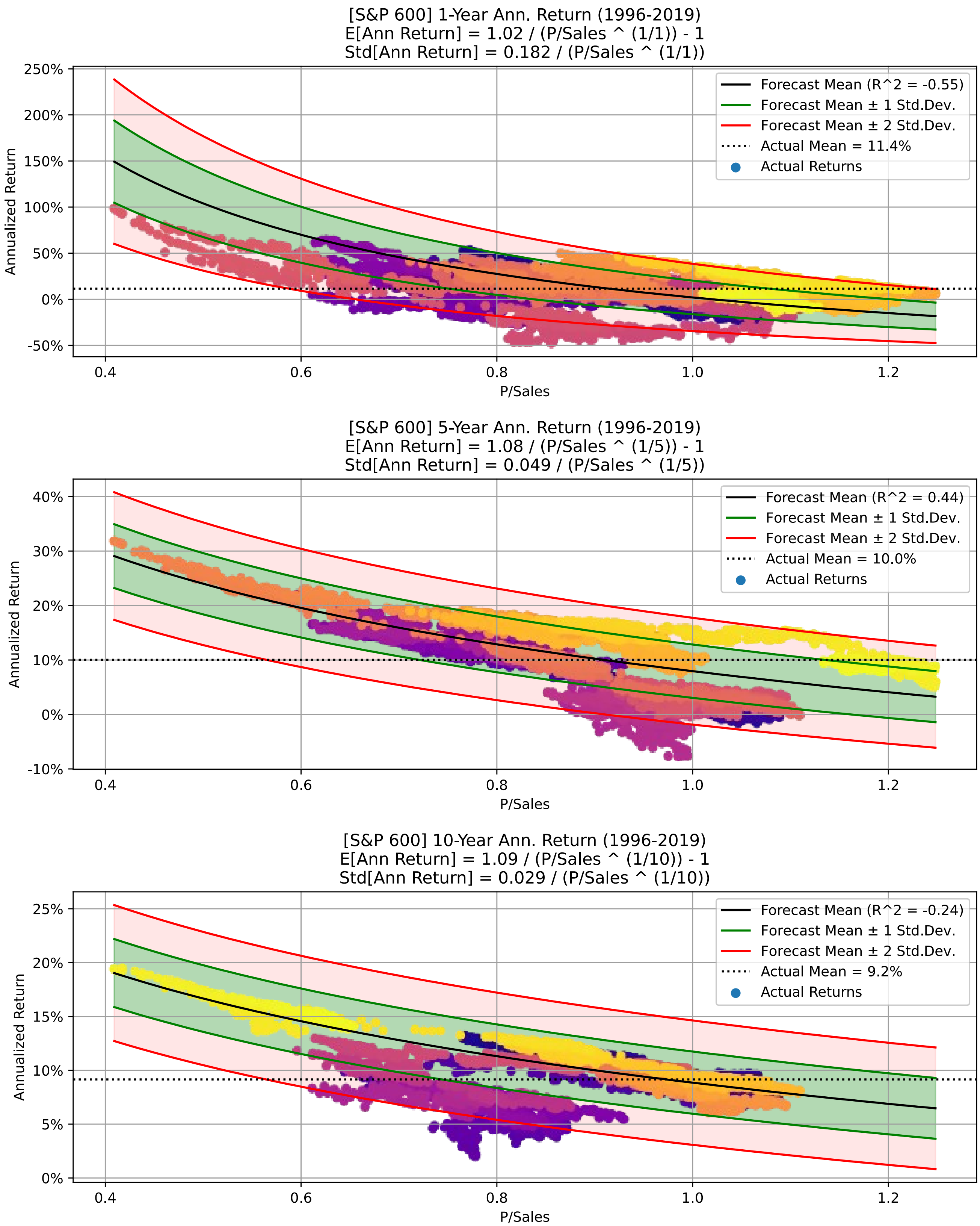


Figure 15: Historical annualized returns for the S&P 600 stock-index.

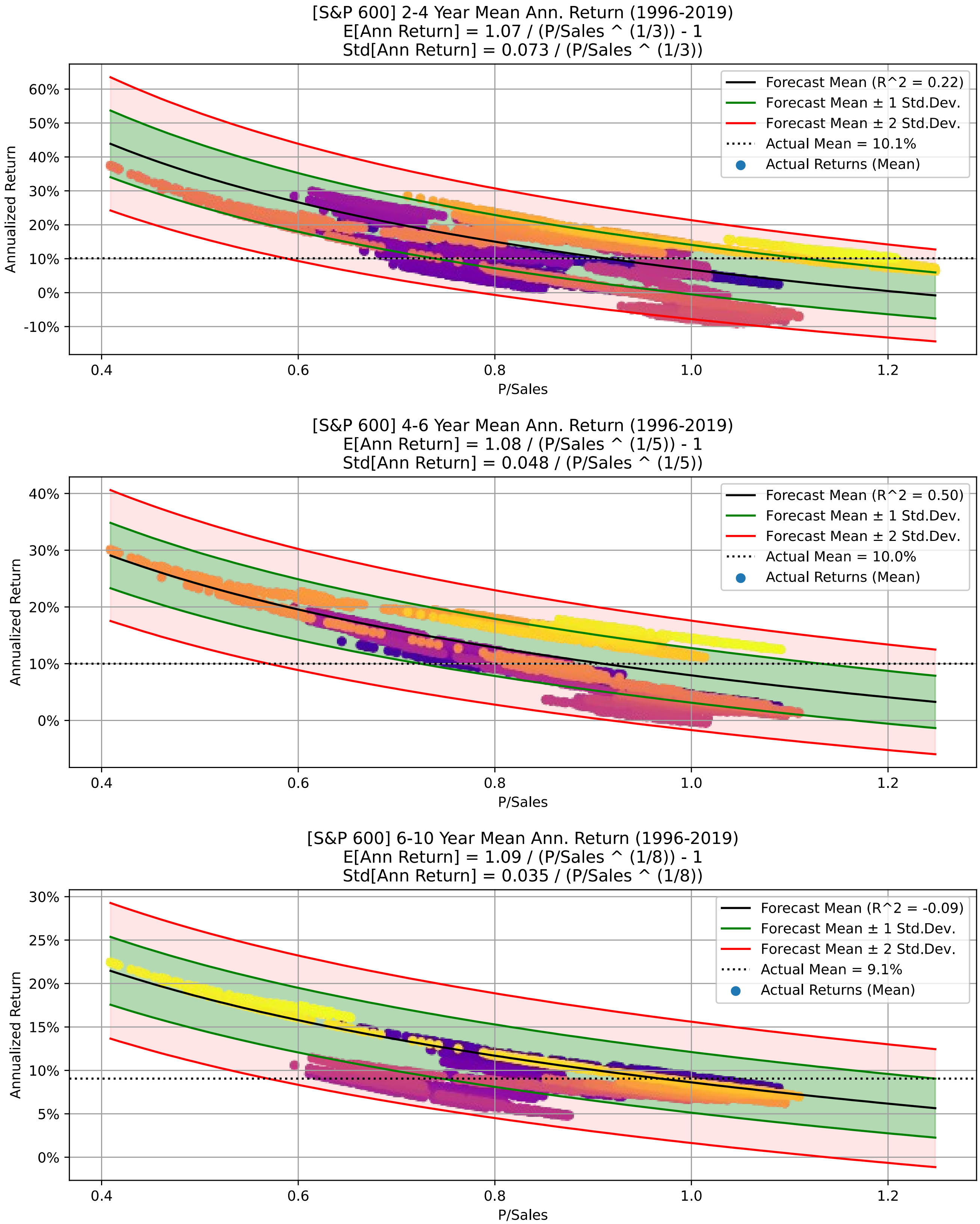


Figure 16: Historical mean annualized returns for the S&P 600 stock-index.

9 Case Study: NASDAQ 100 (U.S. Tech & More)

The NASDAQ 100 index contains 100 of the largest non-financial stocks traded on the NASDAQ exchange. Many of these stocks are for large technology companies such as Alphabet / Google, Apple, Facebook, Microsoft, etc., but the NASDAQ 100 index also contains non-technology stocks such as Starbucks and PepsiCo. We do not have data for the NASDAQ 100 index itself, so we use data for an Exchange Traded Fund (ETF) with the ticker symbol QQQ that tracks the NASDAQ 100 index.[9] Unfortunately we only have this data until February 2019 and there are some other problematic issues relating to its calculation, as explained in Section 16.

9.1 Basic Data

Figure 17 shows the basic financial data for the NASDAQ 100 index (that is, the ETF with ticker QQQ). The top plot shows that the Share-Price and Sales Per Share have had similar upwards trends between 2004 and 2019, but the Share-Price is much more volatile than the Sales Per Share.

The second plot shows the P/Sales ratio between the Share-Price and Sales Per Share which was around 4.0 at the beginning of the data in 2004. The lowest P/Sales ratio was around 1.4 during the “Financial Crisis” in 2009. The average P/Sales ratio was around 2.8 for this 15-year period, and for large parts of the period the P/Sales ratio was fairly close to its average, with only a few shorter periods having more extreme outliers for the P/Sales ratio.

The third plot in Figure 17 shows the annual growth in Sales Per Share which looks quite cyclical with a full cycle lasting maybe 3-4 years. The lowest growth-rate was actually a sales decline of only about -2% around year 2010 in the aftermath of the “Financial Crisis”. This was followed by tremendous sales growth that reached 33% in 2011. The peaks of the growth-cycles often reached sales growth of 20-30%, which are so high that it makes you wonder if there are errors in the data. On average the annual growth in Sales Per Share was 11.7%.

The bottom plot in Figure 17 shows that the Dividend Yield was nearly 0% in 2004 and reached its maximum of nearly 1.6% in 2014. On average it was 0.8% during this 15-year period.

9.2 Annualized Returns

Figure 18 shows the historical annualized returns on the NASDAQ 100 stock-index (i.e. the QQQ ETF). The top plot shows the 1-year returns overlaid with the forecasting model that has been fitted to the historical averages for the P/Sales ratio, annual growth in Sales Per Share, and Dividend Yield.

9 https://www.invesco.com/us/financial-products/etfs/product-detail?productId=QQQ
http://portfolios.morningstar.com/fund/holdings?t=QQQ&region=usa&culture=en-US

Even though the forecasting model was fitted using the “future” data which is a form of cheating, the model still could not predict the 1-year returns. This is because the P/Sales ratio has no predictive power on the NASDAQ 100 returns for such short investment periods of only 1 year.

The middle plot in Figure 18 shows the historical annualized returns for 5-year investment periods of the NASDAQ 100 index. This has a quite good fit with $R^2=0.69$ and the historical returns shown as multi-coloured dots in the plot also clearly follow the Return Curves of the forecasting model.

The bottom plot in Figure 18 shows the historical annualized returns for 10-year investment periods, which has a very good fit with $R^2=0.81$. However, it should be noted that there is not a lot of data in this plot, because the data starts in late 2003 and only goes to early 2019, so the last possible start-date for a 10-year investment period is in early 2009. Therefore the start-dates shown in this plot are only between late 2003 and early 2009, which is only a bit more than 5 years of data-points. Fortunately this short period has the full range of historical P/Sales ratios, and their relation to the future 10-year annualized returns can be seen to fit the forecasting model very well.

9.3 Annualized Returns (Mean)

Figure 19 shows the *mean* annualized returns on the NASDAQ 100 stock-index. The top plots shows it for 2-4 year investment periods, whose historical returns have a reasonable fit to the forecasting model with $R^2=0.54$. The middle plot in Figure 19 shows the mean annualized returns for 4-6 year investment periods, which has an even better fit to the forecasting model with $R^2=0.72$.

The bottom plot shows the mean annualized returns for 6-10 year investment periods, which has an exceptionally good fit to the forecasting model with $R^2=0.87$. But as noted above for fixed 10-year investment periods, this plot also does not have a lot of data-points, because the last valid 10-year period begins in early 2009 and ends in early 2019, which means that we are only showing data-points that start between late 2003 and early 2009, so there is only a bit more than 5 years of data-points.

The QQQ ETF which tracks the NASDAQ 100 index has a P/Sales ratio of about 4.9 at the time of this writing in September 2020. So let us try and use these three forecasting models with this P/Sales ratio. The forecasting formulas are written above the plots in Figure 19.

For 2-4 year investment periods the forecasted mean annualized return is actually a loss of -7.0%. So if we invest in the ETF with ticker QQQ for the NASDAQ 100 index in September 2020 and hold the investment until somewhere between September 2022 and September 2024, then the forecasting model expects a loss of about -7.0% per year. The forecasting model is not certain about this projected loss, so it also provides a standard deviation to give a spread of possible outcomes. That formula is also found above the top plot in Figure 19 and using the current P/Sales ratio of 4.9 gives a standard deviation of 6.8%. The reason the forecasting model predicts an overall loss, is that the model has been fitted with the historical average P/Sales ratio which was only 2.8, so the model expects a significant loss of about -17% per year from this revaluation alone, which is only partially made up by assuming the future annual growth in Sales Per Share is going to be the same as its historical average of 11.7%, and similarly that the future Dividend Yield is going to be the same as its historical average of 0.8%.

Now let us use the forecasting model from the middle plot in Figure 19 for 4-6 year investment periods with the P/Sales ratio of 4.9 from September 2020. This gives a forecasted mean annualized return of 0.4% with standard deviation 4.7%. So the forecasting model expects an investment in the NASDAQ 100 will only barely break even from September 2020 until somewhere between 2024 and 2026.

The last forecasting model from the bottom plot in Figure 19 is for 6-10 year investment periods, and using it with a P/Sales ratio of 4.9 gives a forecasted mean annualized return of about 4.9% with standard deviation 3.4%.

The forecasting models for longer investment periods have higher forecasted annualized returns, even though they are all using the same assumptions for the future average P/Sales ratios, growth in Sales Per Share, and Dividend Yield. The models give different forecasts because of the revaluation part of the forecasting formula, where the currently very high P/Sales ratio of 4.9 is assumed to decrease to its historical average of only 2.8. This corresponds to an absolute loss of about -43%. For 3-year investment periods this gives a loss of about -17% per year, and for 5-year investment periods the loss is about -10.6% per year, and for 8-year investment periods the loss is about -6.8% per year. This is because the loss is spread over more years for longer investment periods. But the forecasting model still assumes that the average growth in Sales Per Share is 11.7% per year regardless of how long the investment period is, and similarly for the Dividend Yield which is assumed to be 0.8% per year.

9.4 Summary

In this section we studied the forecasting model on the NASDAQ 100 index. As usual we found that the forecasting model did not work for 1-year investment periods, because the P/Sales ratio had no predictive power for such short periods. But for investment periods of 2-4 years the forecasting model started to have a good fit on the historical returns, and the fit became increasingly better for longer investment periods of 6-10 years.

However, it should be noted that we only have data for the NASDAQ 100 index (i.e. the QQQ ETF) for a bit more than 15 years, so for the plots with 10-year investment periods, there are only really data-points for a bit more than 5 years in the plots.

Using the forecasting models with the very high P/Sales ratio in September 2020, suggests that the NASDAQ 100 index will experience a significant loss in the coming years, provided the P/Sales ratio reverts towards its historical mean and the growth in Sales Per Share cannot offset the loss.

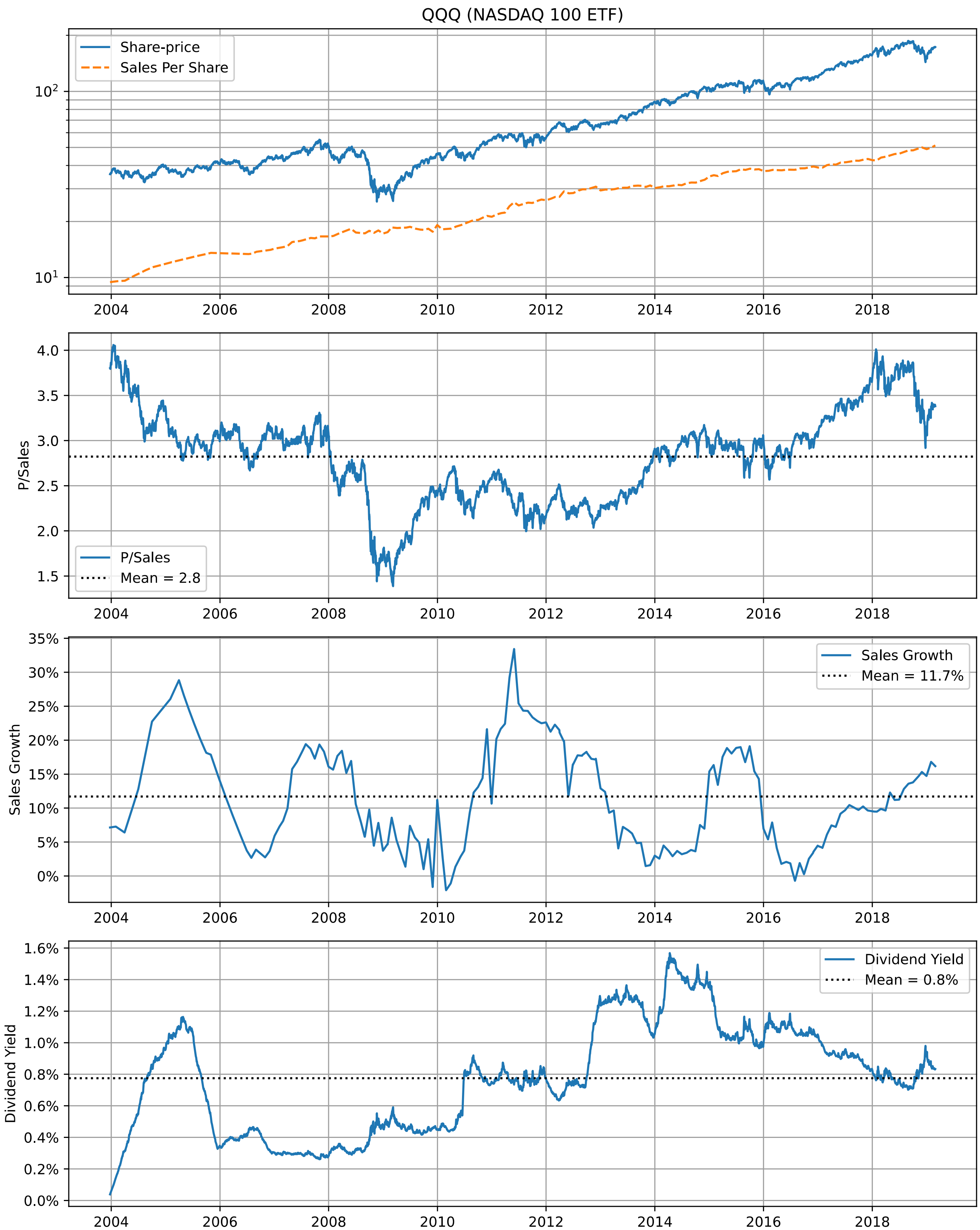


Figure 17: Basic financial data for the NASDAQ 100 stock-index (QQQ ETF).

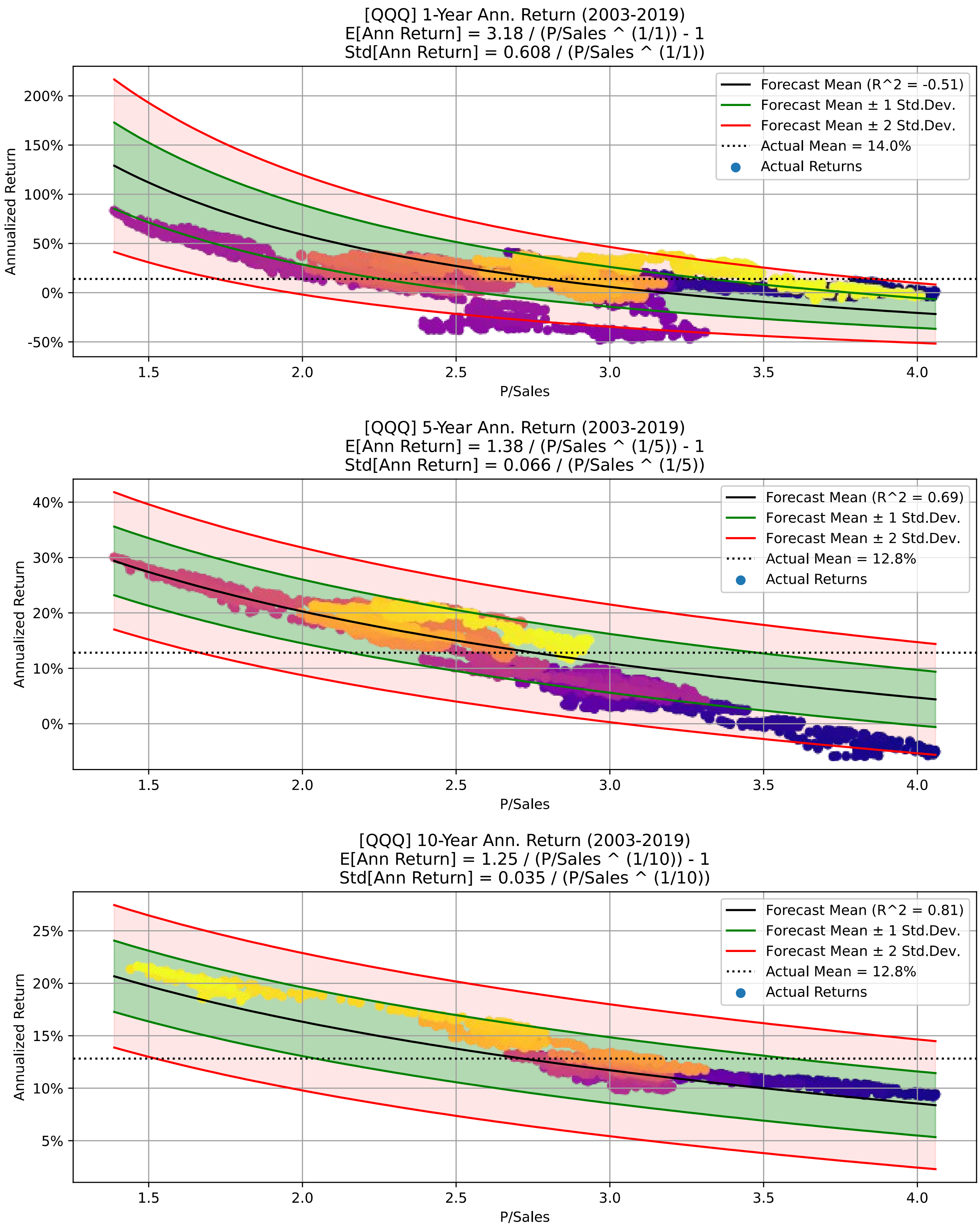


Figure 18: Historical annualized returns for the NASDAQ 100 stock-index (QQQ ETF).

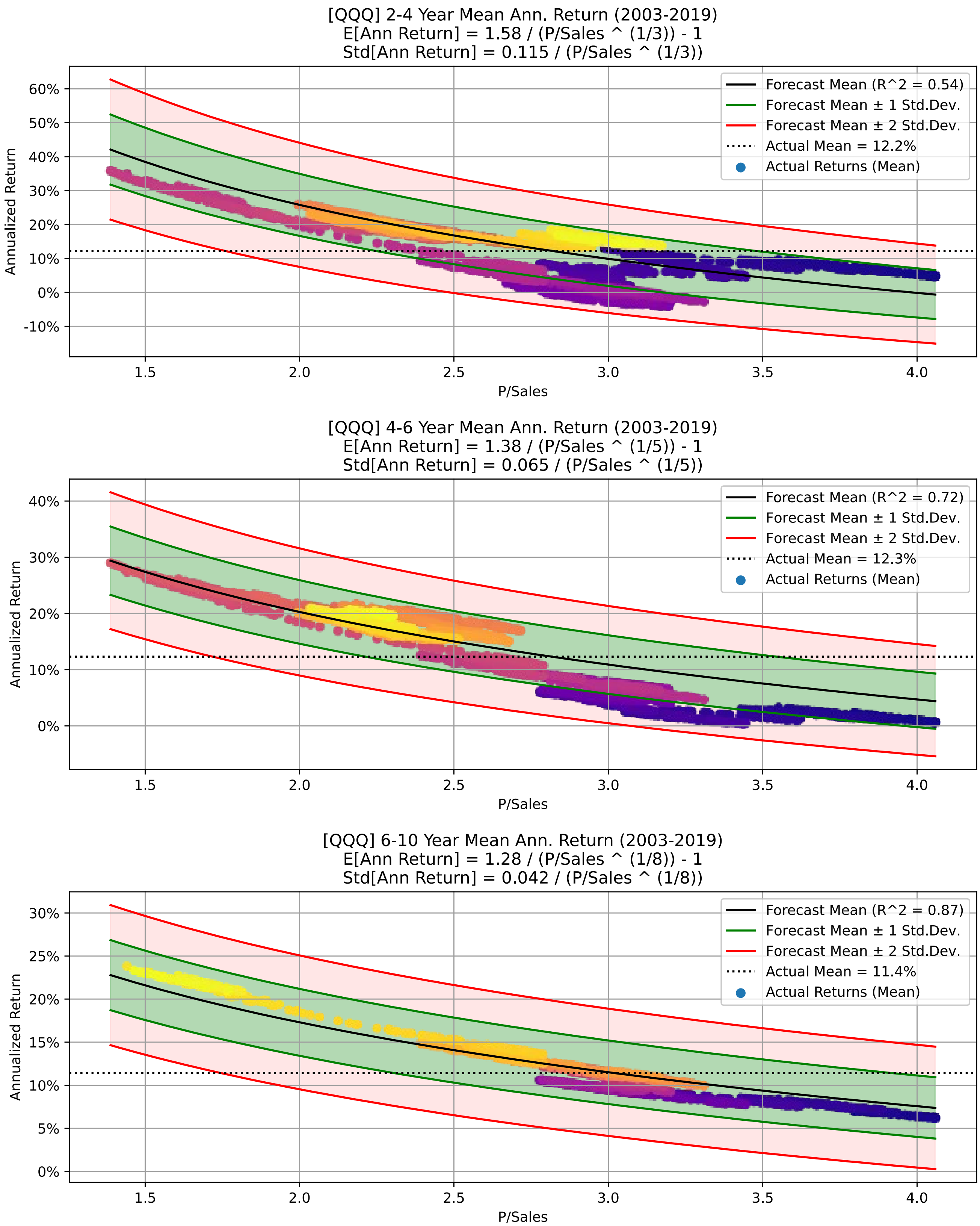


Figure 19: Historical mean annualized returns for the NASDAQ 100 stock-index (QQQ ETF).

10 Case Study: Developed European Markets

The FTSE Developed Europe All Cap index contains nearly 1300 stocks of broadly diversified companies in developed European countries, including Nestle, Novo Nordisk, Siemens, etc. We do not have data for the actual index, so we instead use data for an Exchange Traded Fund (ETF) with the ticker symbol VGK that tracks the index.[10] Unfortunately we only have this data until February 2019 and there are some other problematic issues relating to its calculation, as explained in Section 16.

10.1 Basic Data

Figure 20 shows the basic financial data for the FTSE Developed Europe stock-index (that is, the ETF with ticker VGK). The top plot shows that the Share-Price and Sales Per Share have merely fluctuated around their means between mid 2006 and early 2019, but there is no overall growth as we have seen in all the previous case-studies. Perhaps there is even a slight downwards trend in the Sales Per Share for this stock-index.

The second plot in Figure 20 shows the P/Sales ratio between the Share-Price and Sales Per Share which seems to have fluctuated slowly around its mean value of 1.0.

The third plot in Figure 20 shows the annual growth in Sales Per Share whose mean is only 0.3% but there is great volatility from year to year, with the decline in Sales Per Share reaching -20% in some years and the gain reaching +15% in other years. This data looks strange and it makes you wonder if there are errors in the data. As explained in Section 16, we do not have the raw Sales Per Share data for these ETF's, so it has instead been estimated from the P/Sales ratio and the Share-Price. Furthermore, it is also possible that there is noise from the currency conversion, as also explained in Section 16. However, as we are interested in the long-term stock returns, and we will use the average growth in Sales Per Share over many years, it turns out that the forecasting model still fits the historical stock-returns very well, as shown in the following sections.

The last plot in Figure 20 shows the Dividend Yield which was 3.8% on average with its greatest peak around year 2009 during the global "Financial Crisis" where the world's stock-markets crashed.

10 https://investor.vanguard.com/etf/profile/VGK
http://portfolios.morningstar.com/fund/holdings?t=VGK&region=usa&culture=en-US

10.2 Annualized Returns

Figure 21 shows the historical annualized returns for the Developed Europe stock-index (i.e. the VGK ETF). The top plot shows it for 1-year investment periods, where the historical data has a very poor fit to the forecasting model, as usual. The middle plot shows it for 5-year investment periods, where the historical data has a remarkably good fit to the forecasted mean with $R^2=0.81$. The bottom plot shows it for 10-year investment periods which has an even better fit with $R^2=0.91$, although that plot actually only has less than 3 years of data-points because the data starts in mid 2006 and ends in early 2019, so the last 10-year investment period starts in early 2009 and ends in 2019, thus only giving valid data-points between mid 2006 and early 2009.

10.3 Annualized Returns (Mean)

Figure 22 shows the historical mean annualized returns for the VGK ETF which are averaged over many investment periods. The top plot shows it for 2-4 year investment periods which has a decent fit to the forecasted mean with $R^2=0.59$. The middle plot shows it for 4-6 year investment periods which has a very good fit with $R^2=0.86$. And the bottom plot shows it for 6-10 year investment periods, which has an ever better fit with $R^2=0.89$, but it should again be noted that there is actually less than 3 years of data-points in this plot, because the data starts in mid 2006 and ends in early 2019, so the last valid 10-year investment period starts in early 2009 and ends in 2019, thus only giving valid start-dates between mid 2006 and early 2009.

At the time of this writing in October 2020, the P/Sales ratio was 1.06 for the VGK ETF. If we use this value in the forecasting formulas from the middle plot in Figure 22, we get the forecasted return for 4-6 year investment periods, where the forecasted mean return is 2.8% and the standard deviation is 5.9%. The forecasting model expects a small loss from revaluation from the current P/Sales ratio of 1.06 down to its historical average of 1.0 which corresponds to an annualized loss of about -1.2%. The forecasting model further expects annual growth in Sales Per Share of only 0.3%, and it expects a Dividend Yield of 3.8%. As explained in Section 2.3, these numbers do not add up exactly to the forecasted mean, because the mathematical formula actually uses multiplication instead of addition.

10.4 Summary

We used the forecasting model on data for an Exchange Traded Fund (ETF) tracking a stock-index, instead of using data for the index itself. The ETF data for annual growth in Sales Per Share seemed to be excessively volatile with frequent changes between -20% and +15%, which suggests the data might be erroneous. Even so, the forecasting model still had a very good fit to the historical returns on the stock-index, because the forecasting model uses the average Sales Growth over several years.

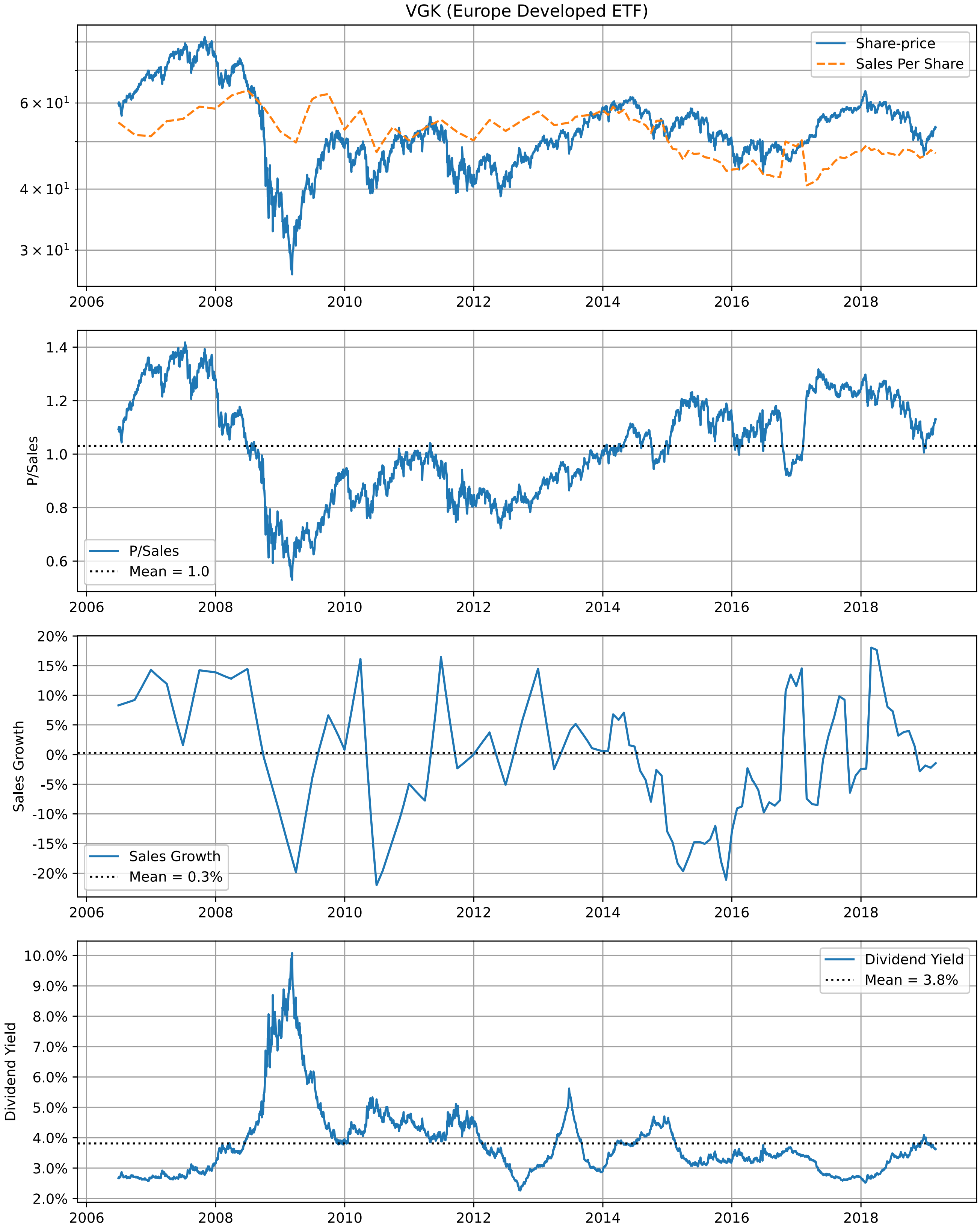


Figure 20: Basic financial data for the FTSE Developed Europe stock-index (VGK ETF).

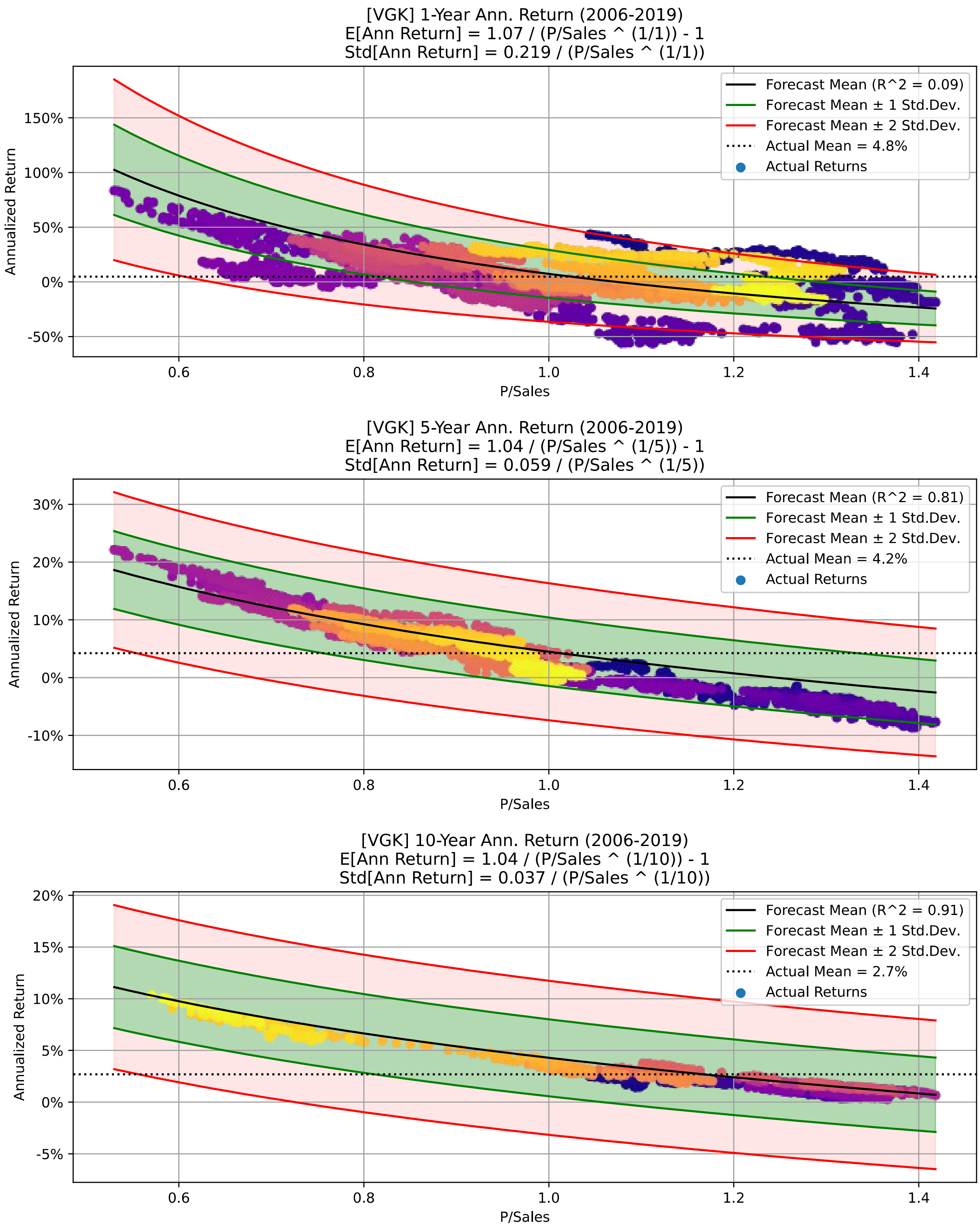


Figure 21: Historical annualized returns for the FTSE Developed Europe stock-index (VGK ETF).

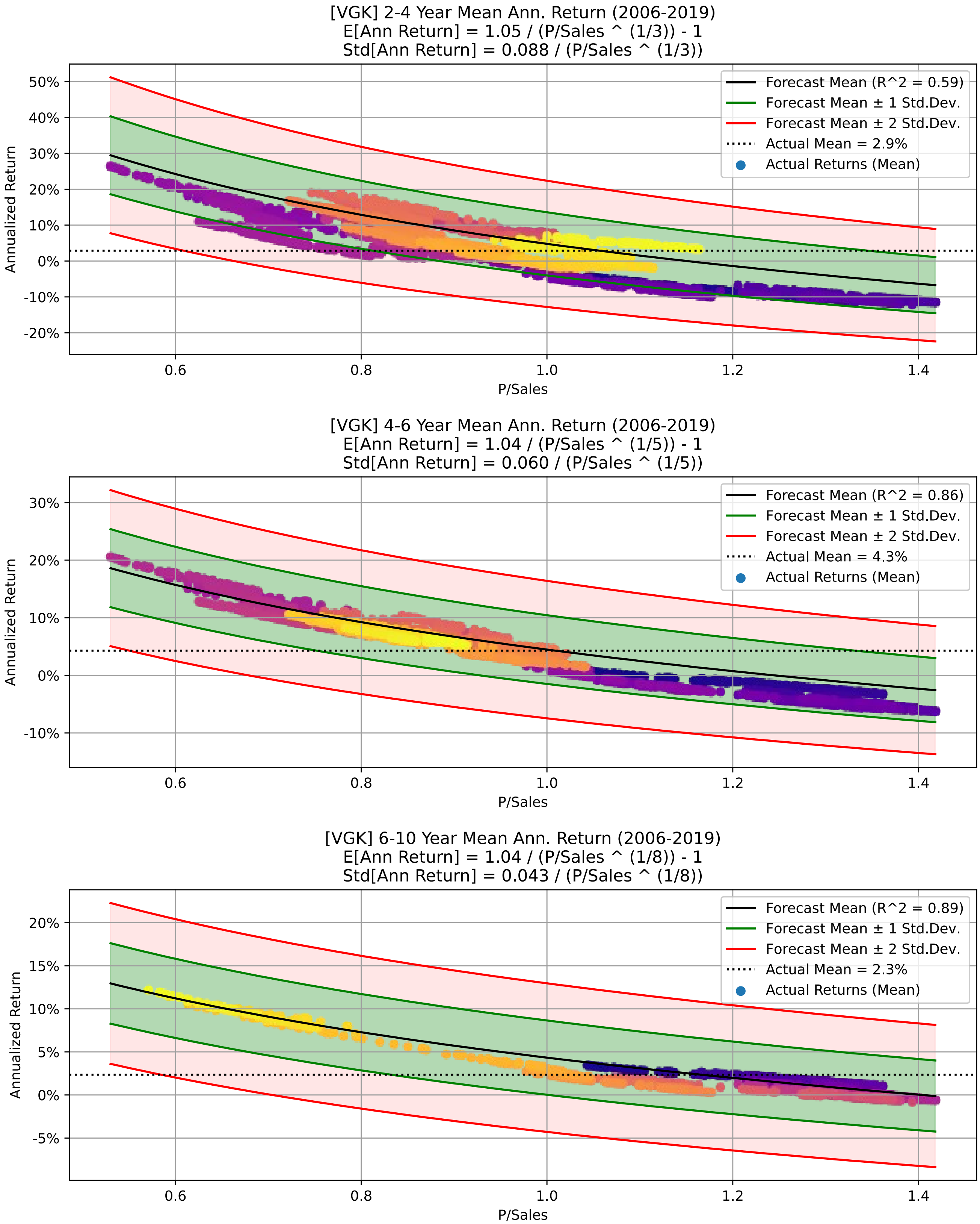


Figure 22: Historical mean annualized returns for the FTSE Developed Europe index (VGK ETF).

11 Case Study: Emerging Markets

The MSCI Emerging Markets index currently contains about 1200 large and mid-sized stocks from Latin America, Emerging Europe, Africa, Middle East, and Asia. The stocks include Alibaba, Tencent, Samsung, Baidu, etc. We do not have data for the actual index, so we instead use data for an Exchange Traded Fund (ETF) with the ticker symbol EEM that tracks the index.[11] Unfortunately we only have this data until February 2019 and there are some other problematic issues relating to its calculation, as explained in Section 16.

11.1 Basic Data

Figure 23 shows the basic financial data for the Emerging Markets stock-index (that is, the ETF with ticker EEM). The top plot shows the Share-Price and Sales Per Share, which seems to have had two distinct periods separated by the global “Financial Crisis” around year 2009. Prior to year 2009 the EEM had significantly higher Share-Price relative to the Sales Per Share, and after year 2009 they have been more equal. This is shown more clearly as the P/Sales ratio in the second plot in Figure 23.

The third plot in Figure 23 shows the annual growth in Sales Per Share which has some very large annual gains of 50-60%, as well as some declines greater than -20%. Between 2012 and 2017 the annual change in Sales Per Share was mostly negative. As with the other ETF’s we have studied, you should probably view this data with some skepticism because of the way it has been estimated, as explained in Section 16. But we will only use the long-term average growth in Sales Per Share, so the short-term volatility is not so relevant for us, regardless of whether it is accurate or erroneous.

The bottom plot in Figure 23 shows the Dividend Yield which appears to have generally trended upwards, with a large spike during the “Financial Crisis” in 2009 where global stock-markets crashed.

11.2 Annualized Returns

Figure 24 shows the historical annualized returns for the Emerging Markets stock-index (i.e. the EEM ETF). The top plot shows it for 1-year investment periods which has a very poor fit to the forecasting model, as usual. The middle plot shows it for 5-year investment periods, which also has a very poor fit with negative $R^2=-0.56$. The bottom plot shows it for 10-year investment periods, whose fit looks decent with $R^2=0.54$, but the full data-period is only from mid 2004 to early 2019, so the last valid 10-year investment period starts in early 2009, and the start-dates for the data-points therefore only range between mid 2004 and early 2009, which is less than 5 years of data-points.

11 https://www.ishares.com/us/products/239637/ishares-msci-emerging-markets-etf
http://portfolios.morningstar.com/fund/holdings?t=EEM&region=usa&culture=en-US

11.3 Annualized Returns (Mean)

Figure 25 shows the mean annualized returns for the Emerging Markets index (i.e. the EEM ETF), which have been averaged over many investment periods so as to smoothen the effect of outliers. Even so, the top and middle plots still have very poor fits of the historical data to the forecasted mean, for both 2-4 and 4-6 year investment periods.

Whenever we see a poor fit between the forecasted mean and the actual historical stock-returns, it is because one or more of the actual historical return components is very different from the means used in the forecasting model. Figure 23 shows that neither the P/Sales ratio nor the annual growth in Sales Per Share seem to be mean-reverting over the entire 15-year period.

The computer code described in Section 16 also provides a function that makes it easy to plot and compare the actual versus forecasted stock-returns, to see which periods had poor fits, and what the P/Sales ratios and Sales Growth were for those historical periods.

The bottom plot in Figure 25 for 6-10 year investment periods has a good fit with $R^2=0.76$, but once again it should be noted, that this actually only has data-points between mid 2004 and early 2009, because the last valid 10-year investment period starts in early 2009 and ends in 2019.

What should we do, if we want to forecast the future returns on the Emerging Markets index? For shorter investment periods the forecasting model had very poor fits to the historical data. For longer investment periods the forecasting model had a good fit to the historical returns, but the dataset was quite small. As always, the answer is that the historical data is only relevant if it also represents the future. For the Emerging Markets index, we saw in Figure 23 that there seems to be two distinct periods before and after the global “Financial Crisis” in year 2009 regarding the P/Sales ratio and Sales Growth. So let us try and use the financial data from year 2010 onwards, but there is no guarantee that this will also be representative of the future.

Let us say we are interested in forecasting 5-year annualized returns for the EEM ETF. Using the formulas from Section 2 and the financial data between 2010 and 2019, we get the parameters $a=1.10$ and $b=0.072$ for the forecasting model, and using these in Eq. (14) and Eq. (18) with the current P/Sales ratio of 1.17 at the time of this writing in October 2020, gives a mean annualized return of 6.9% with standard deviation 7.0% for the EEM ETF.

11.4 Summary

The Emerging Markets stock-index did not seem to have had mean-reverting historical financial data, so it is hard to say what the future P/Sales ratio and growth in Sales Per Share might be. In such cases it might be better to make a reasonable guess at what the future could bring in terms of valuation ratio and growth, rather than assuming the historical averages will just continue in the future.

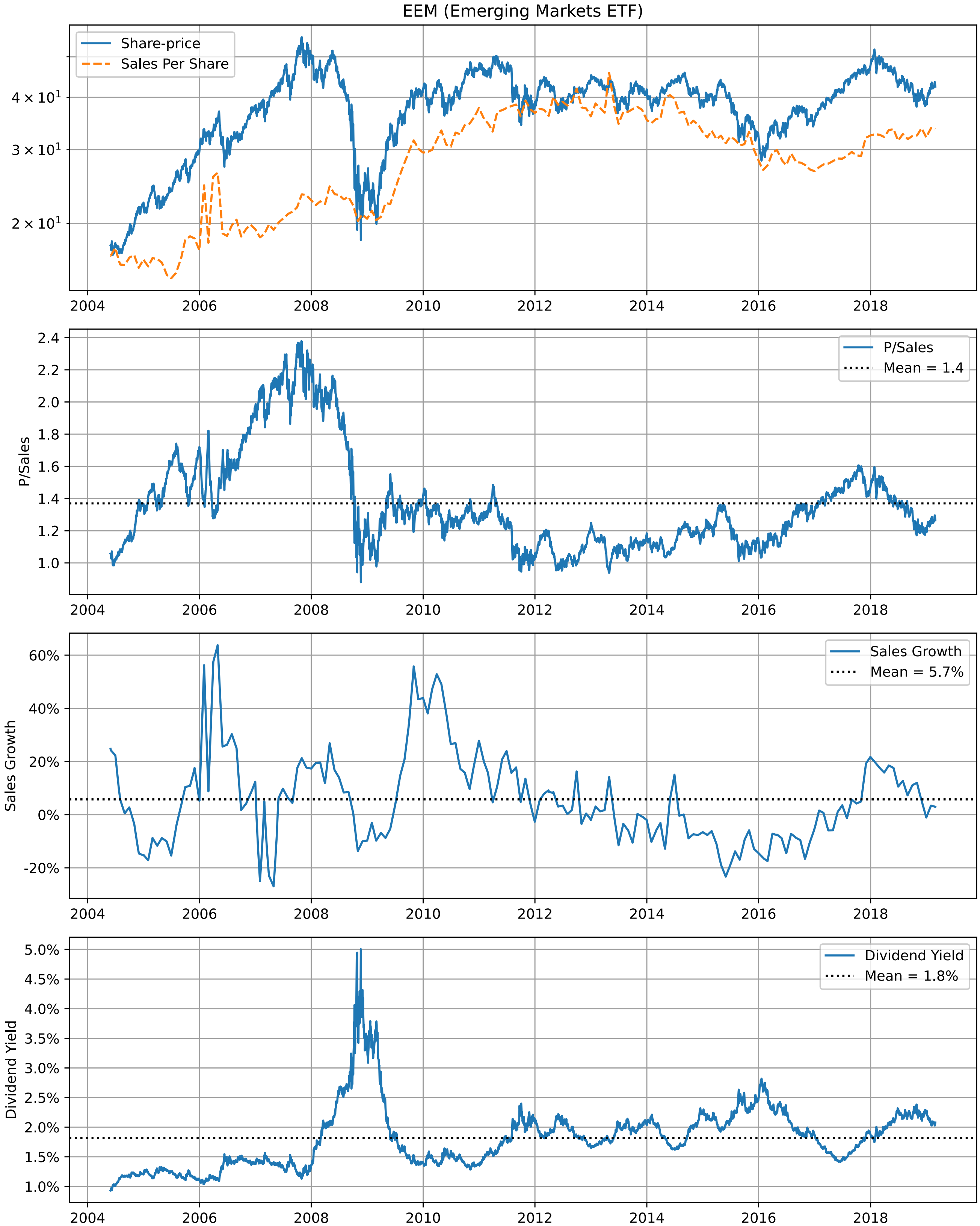


Figure 23: Basic financial data for the MSCI Emerging Markets stock-index (EEM ETF).

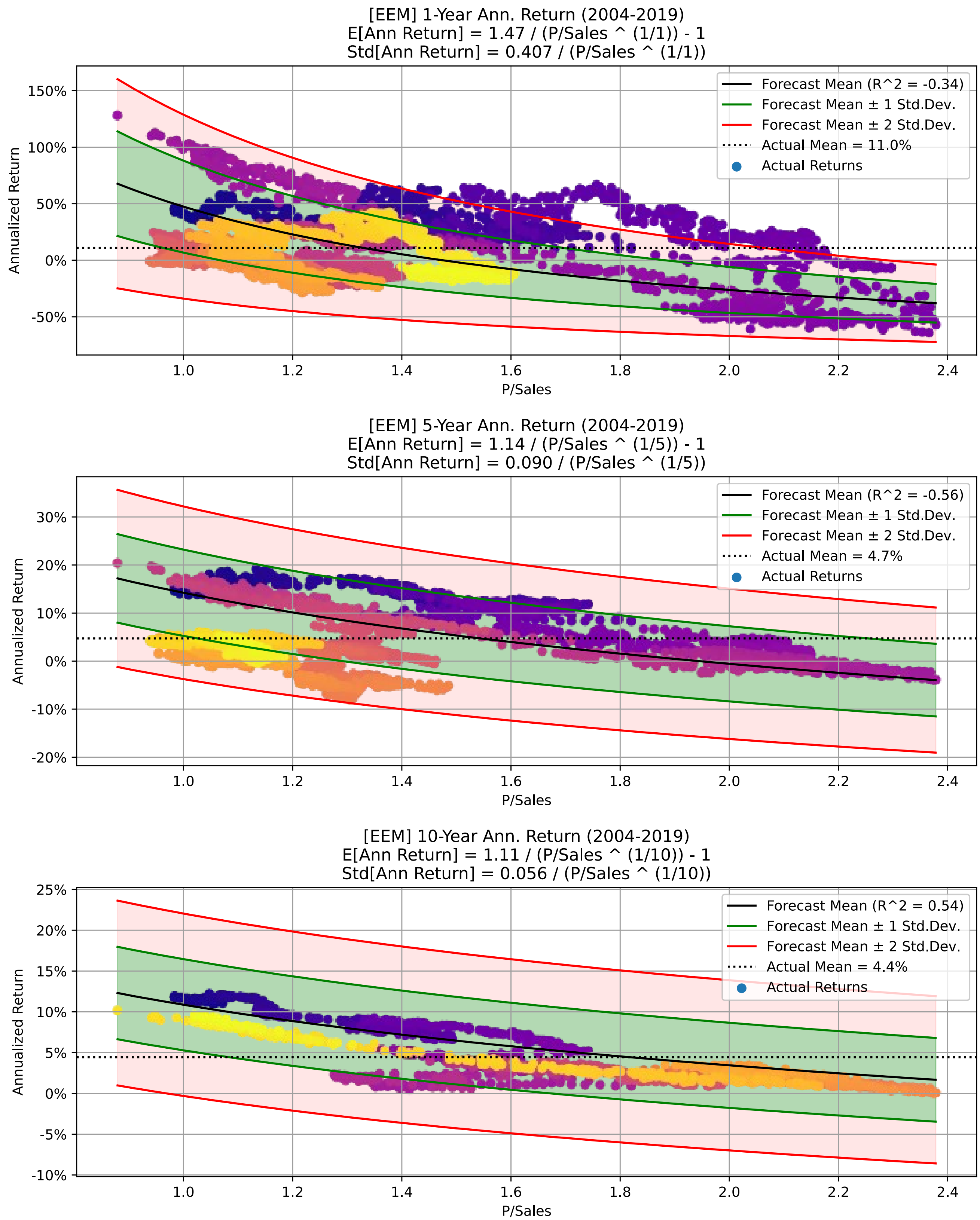


Figure 24: Historical annualized returns for the MSCI Emerging Markets stock-index (EEM ETF).

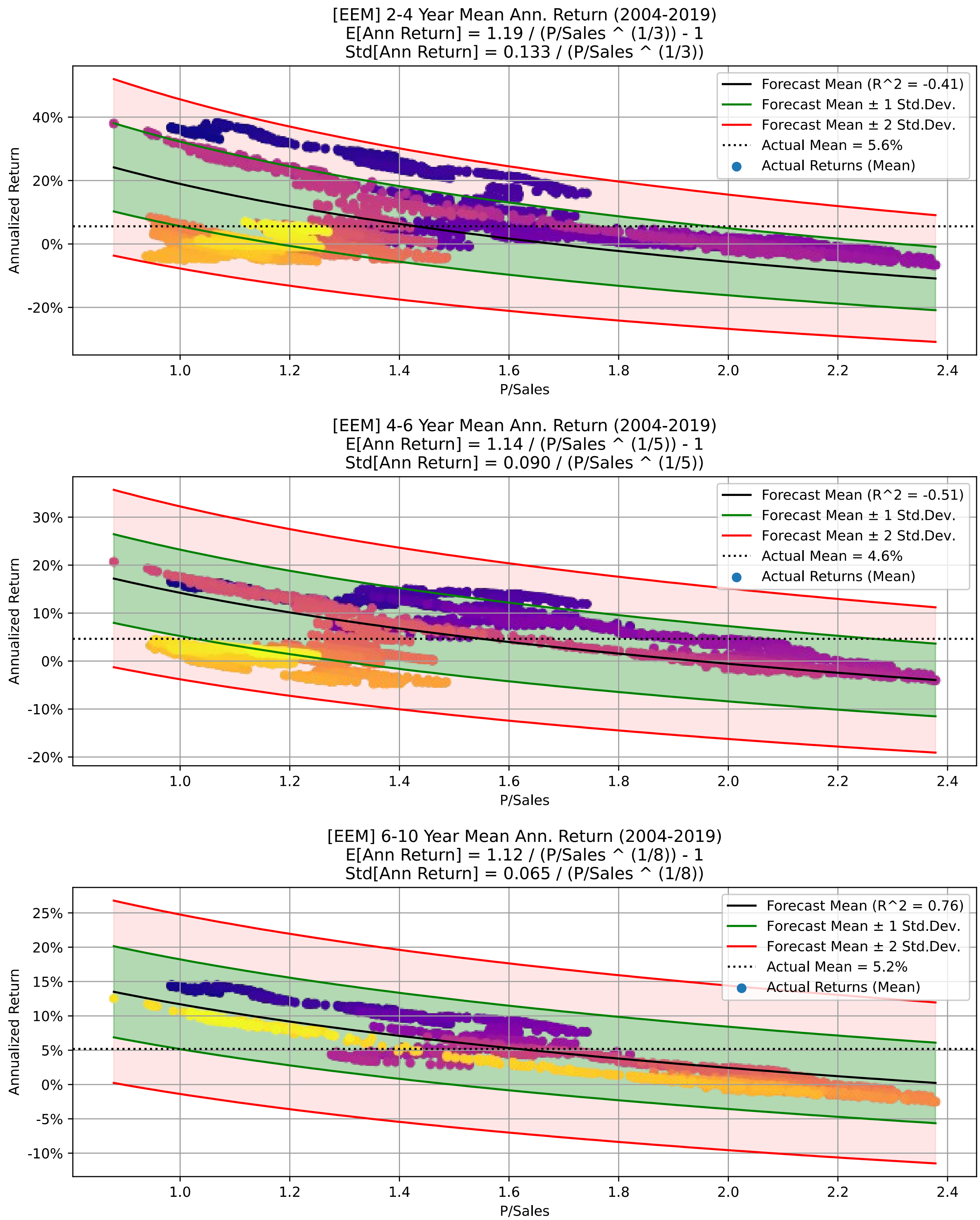


Figure 25: Historical mean annualized returns for the MSCI Emerging Markets index (EEM ETF).

12 Case Study: Austria

The MSCI Austria index contains the stocks of 27 Austrian companies. Although the companies are in diverse industries such as gun manufacturing, cow milking, and yodeling, the Austrian stock-index is obviously much less diversified than the previous stock-indices we have studied, such as the S&P 500, Developed Europe, and Emerging Markets, each of which contain several hundreds or even more than a thousand companies. The Austrian stock-index may therefore provide an interesting case-study.

We do not have data for the MSCI Austria index itself, so we use data for an Exchange Traded Fund (ETF) with the ticker symbol EWO that tracks the index.[12] Unfortunately we only have this data until February 2019 and there are also other problematic issues with this data, as explained in Section 16.

12.1 Basic Data

Figure 26 shows the basic financial data for the Austrian stock-index (that is, the ETF with the ticker EWO). The top plot shows that the Sales Per Share has merely fluctuated a bit up and down, but not really trended upwards or downwards between mid 2003 and early 2019. But the Share-Price peaked before the "Financial Crisis" in 2009 where it crashed and has not reached the same level since. This is shown more clearly in the second plot, where the P/Sales ratio peaked around 2 in the year 2008, but since 2010 the P/Sales ratio has only been around 0.7 on average.

The third plot in Figure 26 shows the annual growth in Sales Per Share, which has been incredibly volatile and often exceeding ±20% growth per year. In the previous case studies for ETF's, we also observed strange growth-rates, but not quite as extreme as this. As discussed in Section 16, it is unclear exactly what causes this, so you should view this data with some skepticism.

The bottom plot in Figure 26 shows the Dividend Yield which was 2.4% on average and peaked around 8.0% during the "Financial Crisis" in 2009 where the global stock-markets crashed.

12.2 Annualized Returns

Figure 27 shows the annualized returns on the Austrian stock-index (i.e. the EWO ETF). The top plot shows it for 1-year investment periods, where the forecasting model had a very poor fit as usual. The middle plot shows it for 5-year investment periods, where the forecasting model had a decent fit with $R^2=0.49$. The bottom plot is for 10-year investment periods, where the forecasting model had a very good fit with $R^2=0.87$, although it should be noted that it actually has less than 6 years of data-points from mid 2003 to early 2009, because the last 10-year period starts in early 2009 and ends in 2019.

12 https://www.ishares.com/us/products/239609/ishares-msci-austria-capped-etf
http://portfolios.morningstar.com/fund/holdings?t=EWO&region=usa&culture=en-US

12.3 Annualized Returns (Mean)

Figure 28 shows the mean annualized returns on the Austrian stock-index (i.e. the EWO ETF), which uses the average over many investment periods so as to smoothen the short-term volatility. The top plot shows it for 2-4 year investment periods, which seems to have two different levels in the data-points, probably corresponding to the two periods before and after the market-crash in 2009, where the Austrian stock-index traded at very different P/Sales levels.

The middle plot in Figure 28 shows the mean annualized returns for 4-6 year investment periods, where the forecasting model has a good fit to the historical data-points with $R^2=0.62$. The bottom plot shows it for 6-10 year investment periods, where the forecasting model has a very good fit with $R^2=0.85$.

This is an interesting case study, because the annual growth in Sales Per Share from Figure 26 was so volatile that the data seems to be erroneous, so why does the forecasting model still have such a good fit to the historical long-term returns? That is because the annual growth in Sales Per Share is apparently mean-reverting with fairly short cycles of only a few years, so as long as the average growth-rate is approximately right, it can be used in the forecasting model.

If we use the historical averages in the forecasting model, we assume that those averages will continue in the future. We should always assess whether the historical averages are reasonable. In this case the Austrian stock-index had a much lower P/Sales ratio from 2010 onwards. So let us try and use only the financial data from year 2010 onwards. This gives an average P/Sales ratio of 0.78, an average annual decline in Sales Per Share of -1.3%, and average Dividend Yield of 2.6%. Using this data with the formulas from Section 2 gives us the parameters $a=0.96$ and $b=0.071$ for use in the forecasting model's Eq. (14) and Eq. (18), and then using a P/Sales ratio of 0.62 from the time of this writing in October 2020, gives a mean annualized return of 5.7% with standard deviation 7.8%. The forecasting model expects the P/Sales ratio to revert to its historical average of 0.78 which would give an annualized return of about 4.7%. The forecasting model also expects an annual *decrease* in Sales Per Share of -1.3%, and a Dividend Yield of 2.6%. These do not add up exactly to the forecasted mean of 5.7% because the forecasting formula is multiplicative and not additive, as explained in Section 2.3.

12.4 Summary

The Austrian stock-index is an interesting case-study because it only contains 27 companies, and the (possibly erroneous) financial data has highly volatile Sales Growth from year to year. Even so, the forecasting model could fit the historical returns very well because the Sales Growth was mean-reverting with short cycles of only a few years.

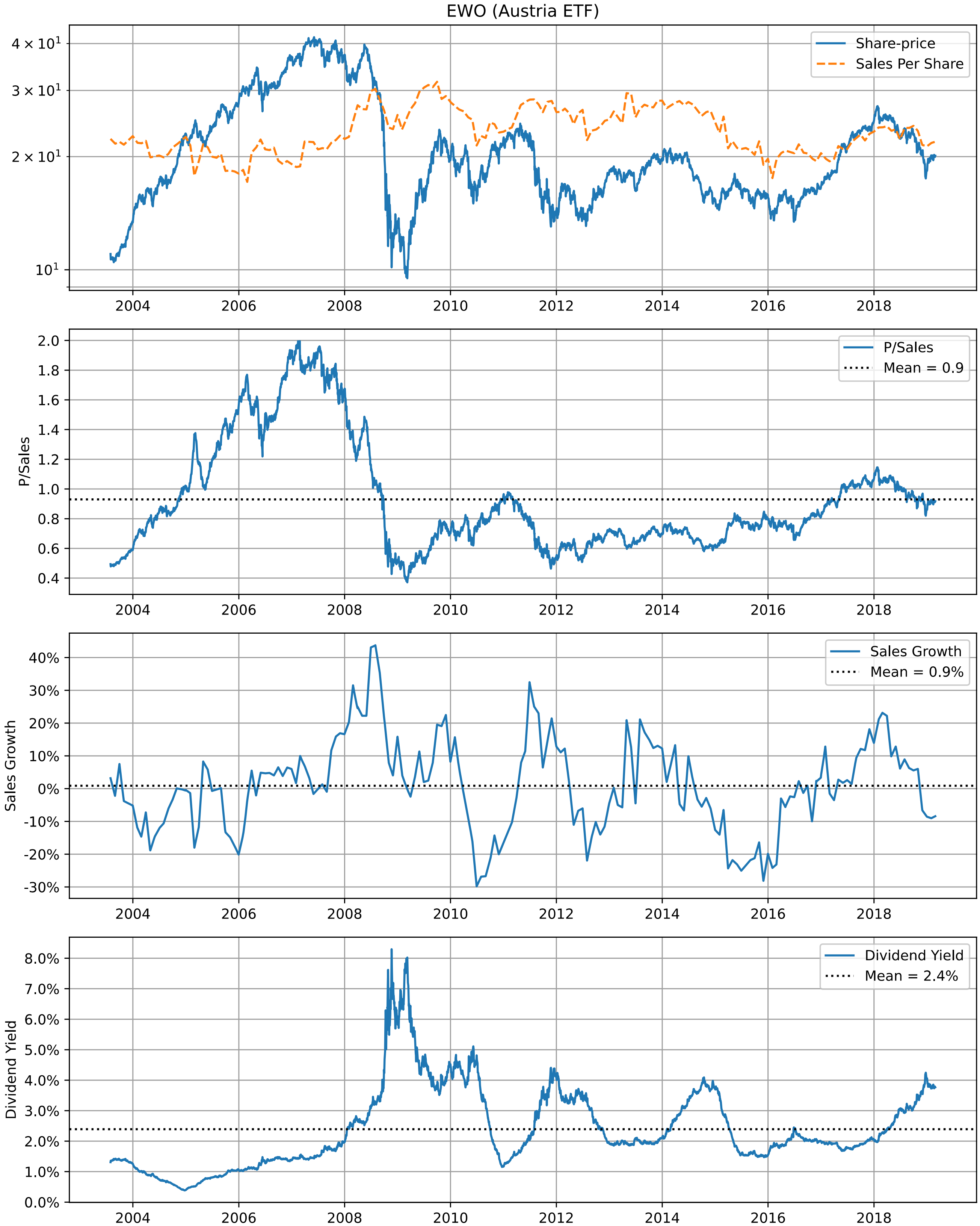


Figure 26: Basic financial data for the Austrian stock-index (EWO ETF).

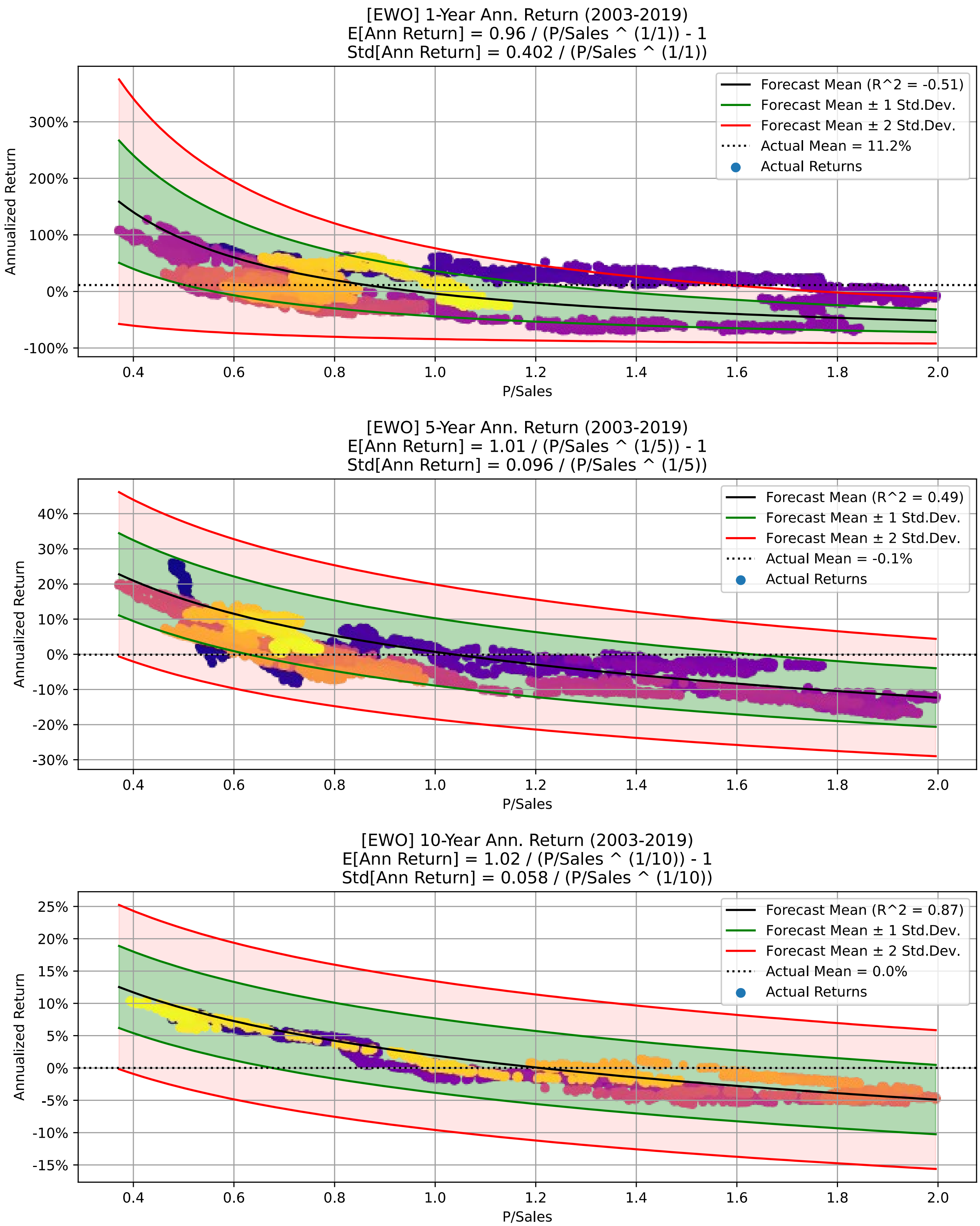


Figure 27: Historical annualized returns for the Austrian stock-index (EWO ETF).

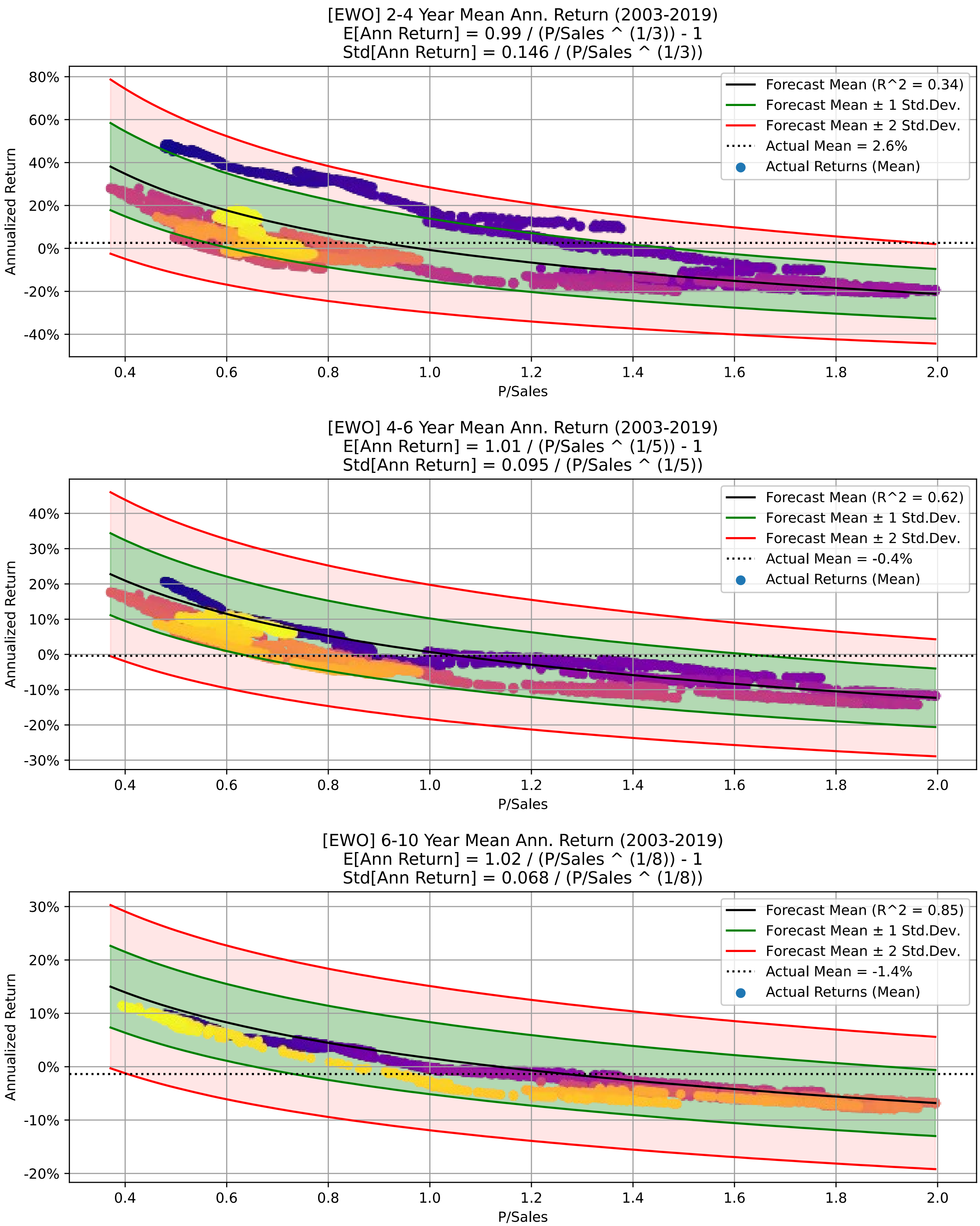


Figure 28: Historical <u>mean</u> annualized returns for the Austrian stock-index (EWO ETF).

13 Synthetic Data

In the previous sections with real-world case studies, we used historical data for several stocks and stock-indices. Sometimes the forecasting model had a poor fit to the historical stock-returns or there were other and more peculiar phenomena. A very useful tool for helping us to better understand how the forecasting model behaves in particular circumstances, is to artificially generate the financial data. This gives us complete control over the financial data so we can better understand what causes the phenomena we observe in the real-world data.

13.1 Data Generator

The computer code for this paper also provides a basic method for generating synthetic data, see Section 16. You can modify the computer code to test other phenomena with the forecasting model.

The basic method sets the Dividend Yield to zero so we focus entirely on the effect of varying the P/Sales ratio and the annual growth in Sales Per Share. The P/Sales ratio is generated from three components: 1) A simple linear component so we can simulate a general trend over time, 2) a cyclical component generated from the sinus-function, and 3) a random component for adding unpredictable noise. Each of these components can be controlled and scaled in the experiments. The annual growth in Sales Per Share is generated from only two components: 1) A linear component so we can simulate a general trend in Sales Growth over time, and 2) a random component for adding unpredictable noise.

In the following experiments we will show how the forecasting model behaves for different kinds of synthetic data that has been generated using this simple method.

13.2 Clean & Simple Data

Figure 29 shows the synthetic financial data for the first experiment, where the P/Sales ratio is a purely cyclical sinus-curve with a period of 4 years. The Sales Per Share grows at a constant rate of 5% per year. The P/Sales and Sales Per Share are multiplied to get the Share-Price which equals the Total Return because the Dividend Yield is set to zero.

Figure 30 shows the annualized returns for the synthetic data in Figure 29. The top plot is for 1-year investment periods where the scatter-plot of the stock-returns is shown as an elongated oval-shape. This is because the annual growth in Sales Per Share is fixed at 5%, and the synthetic P/Sales ratio is a purely cyclical sinus-curve without any noise, where the frequency of the sinus-curve is 4 years, so the 1-year stock-returns are also cyclical.

The middle plot in Figure 30 is for 4-year investment periods which is the same frequency as the P/Sales sinus-curve, so the stock-returns are all identical because the P/Sales ratios are the same at the beginning and end of all 4-year periods. The bottom plot in Figure 30 is for 8-year investment periods which is twice the frequency of the cyclical P/Sales ratio. In both these plots the stock-return is therefore only due to growth in the Sales Per Share, which is set to 5% per year when generating this synthetic data. So the annualized stock-returns should all be exactly 5%. But as shown by the "Actual Mean" in the middle and bottom plots in Figure 30, the average annualized stock-return for all 4 and 8-year investment periods is only 4.7%. This small discrepancy is because the synthetic data for the Sales Growth and Sales Per Share is generated from yearly data-points which are then linearly interpolated to obtain daily data-points. This is the same method used for the real-world data as described in Section 16.3. This data interpolation has the advantage of creating smooth plots for the Sales Per Share as shown in Figure 29, which would otherwise have a stair-case pattern where the Sales Per Share would jump 5% after each year. But the linear interpolation has the disadvantage of slightly distorting the intermediate Sales Per Share numbers, which in turn gives a slightly incorrect growth in the Share-Price, which then shows up as an "Actual Mean" of only 4.7% for the annualized stock-returns, where it should have been exactly equal to the 5% annual growth in Sales Per Share.

Although it is possible to generate the synthetic stock-data using other methods, it can be difficult to get rid of all artefacts in the data and plots, mostly because of the occasional leap-years which can be challenging to handle correctly. For our small experiments here, the slight distortion that arises from the linear interpolation of Sales Per Share is irrelevant, and you would probably not have noticed the small discrepancy if it had not been mentioned. In these experiments we are primarily interested in the overall shape of the plots. But if you want to make very detailed experiments with precise synthetic data, you may want to completely rewrite the computer code and set an investment year to always contain e.g. exactly 200 trading-days without leap-years. This would make it much easier to precisely generate synthetic data, but it would require substantial changes to the computer code used here, which was made specifically for the normal calendar dates.

Figure 31 shows the mean annualized return for a range of investment periods. The top plot shows it for all 1-3 year investment periods, which has an elongated oval shape, somewhat similar to the top plot in Figure 30 for 1-year investment periods. Once again this is because the Share-Price consists of a cyclical sinus-curve from the synthetic P/Sales ratio with a frequency of 4 years, and it also grows 5% per year from the annual increase in the Sales Per Share. Because there is a mismatch in the frequencies, this creates the elongated oval shape in the scatter-plot of the 1-3 year annualized returns.

The middle plot in Figure 31 shows the mean annualized return for 2-6 year investment periods which corresponds to a midpoint of 4 years and is the same as the frequency of the sinus-curve for the P/Sales ratio. The bottom plot shows the mean annualized return for 6-10 year investment periods where the midpoint of 8 years corresponds to twice the cyclical frequency of the P/Sales ratio. Both these plots have nearly perfect fits to the mean of the forecasting model with $R^2=0.96$ and $R^2=0.99$.

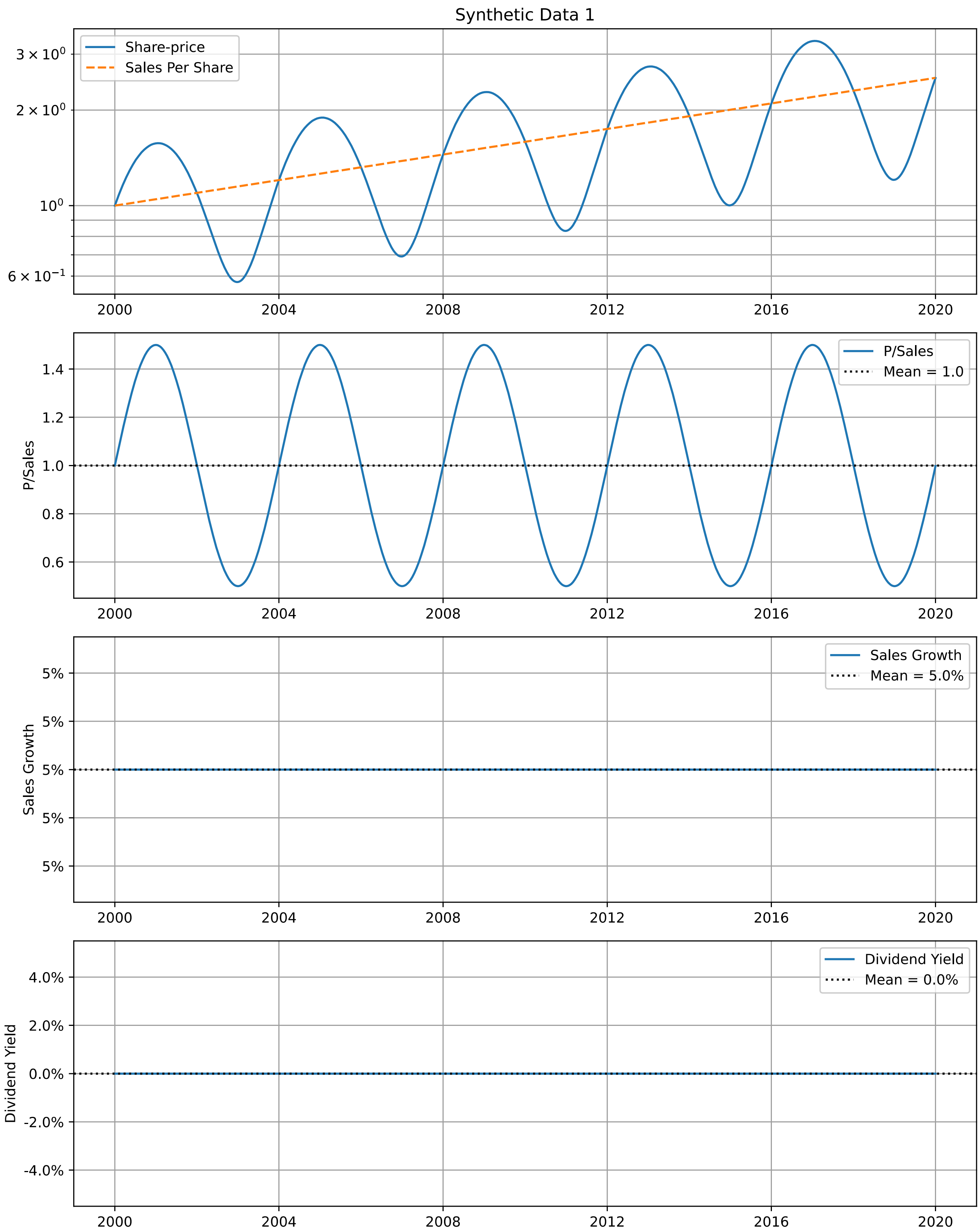


Figure 29: Synthetic financial data. The P/Sales ratio is a sinus-curve with a full cycle of 4 years. The annual growth in Sales Per Share is 5%. The Dividend Yield is zero. There is no random noise.

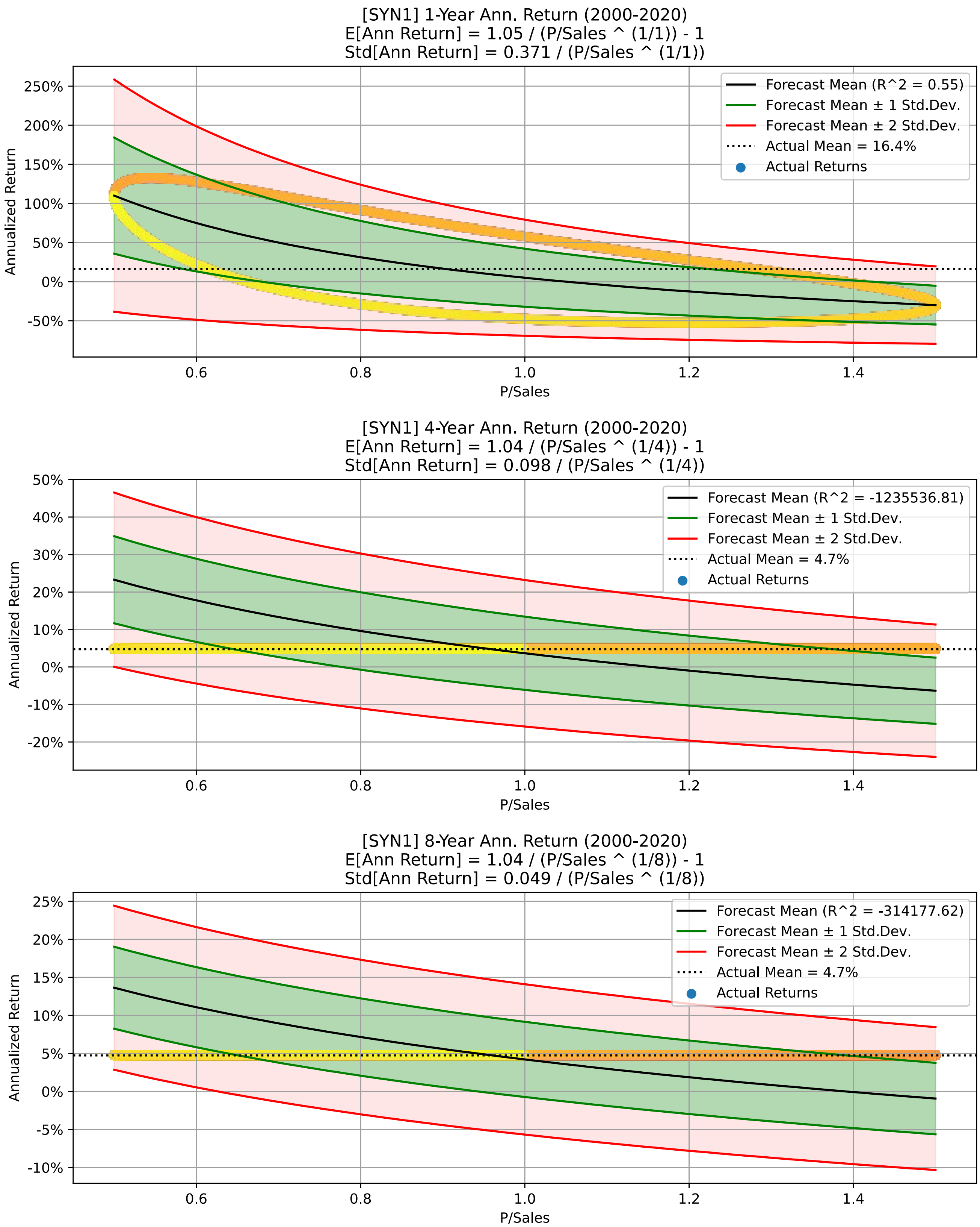


Figure 30: Annualized returns for the synthetic financial data in Figure 29, which has a purely cyclical P/Sales ratio and a constant annual growth of 5% in Sales Per Share, there is no noise.

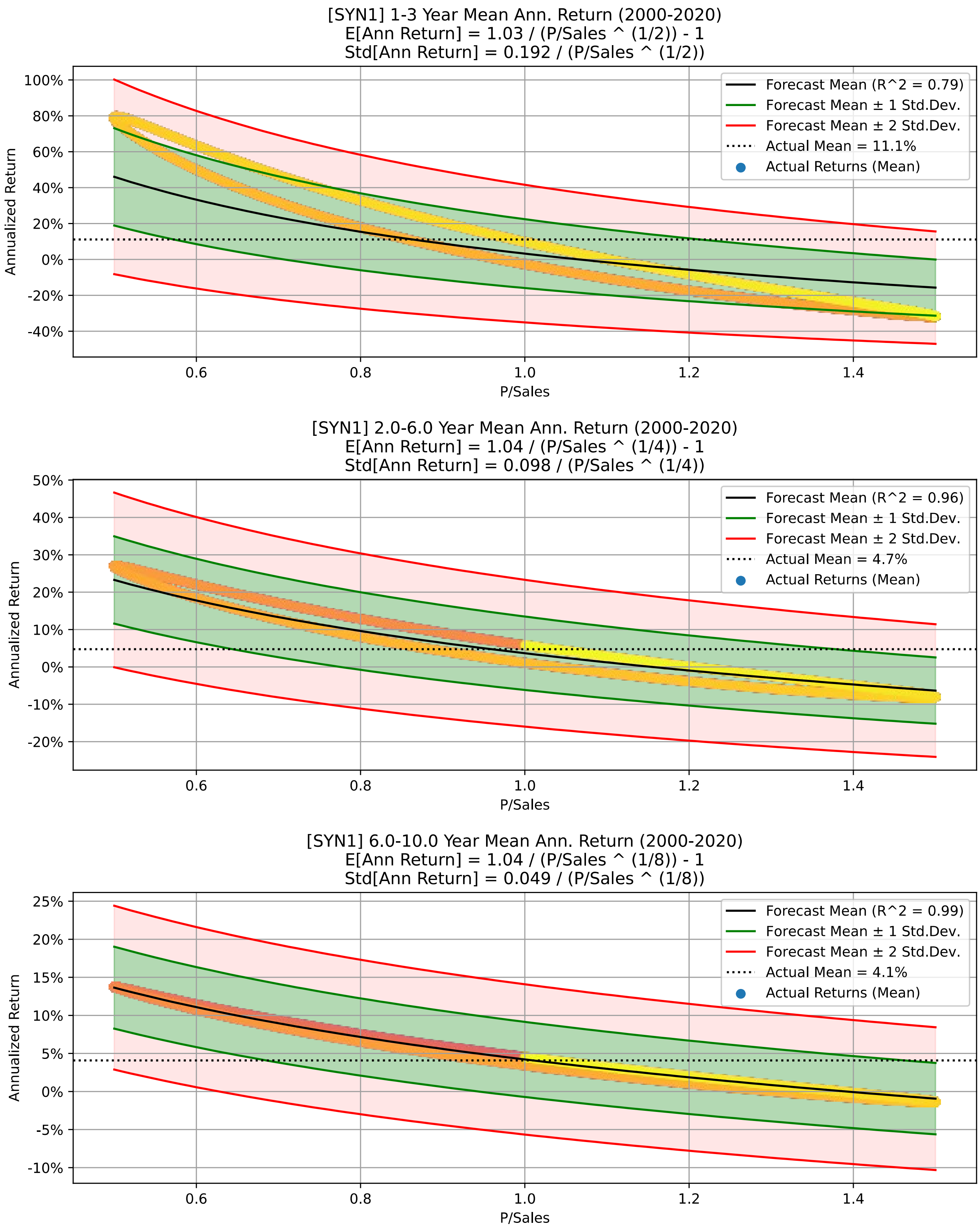


Figure 31: <u>Mean</u> annualized returns for the synthetic financial data in Figure 29, which has a purely cyclical P/Sales ratio and a constant annual growth of 5% in Sales Per Share, there is no noise.

13.3 Random Noise

Figure 32 shows the synthetic financial data for the second experiment. It is generated in the same way as in the previous section, but random noise is now added to both the cyclical P/Sales ratio and the annual growth in Sales Per Share.

Figure 33 shows the annualized returns for the synthetic stock-data. The forecasting model has a very poor fit for 1-year investment periods. This is similar to what we saw in all the real-world case-studies and it is because the P/Sales ratio usually takes more than 1 year to revert to its historical mean.

The middle plot in Figure 33 shows the annualized returns for 4-year investment periods which is exactly the same frequency as the cyclical component of the synthetic P/Sales ratio. But because of the random noise added to the P/Sales ratio and Sales Growth, there is also significant noise in these annualized returns, so the forecasted mean is no longer a great fit for this synthetic data.

The bottom plot in Figure 33 shows the annualized returns for 8-year investment periods which is twice the frequency of the cyclical component of the synthetic P/Sales ratio. But once again, because of the random noise added to the P/Sales ratio and Sales Growth, the data-points are spread around the forecasted mean so the fit is still not great at only $R^2 = 0.48$.

The effect of the random noise on the synthetic P/Sales ratio is to create artificial bubbles and crashes in the stock-prices as can be seen from the second plot in Figure 32. When we are plotting the annualized stock-returns for e.g. exactly 8-year periods, this creates outliers in the scatter-plot for when the start or end-points are in a stock-bubble or crash. Our standard approach to smoothen these outliers is to consider the average stock-returns over a range of investment periods.

This is done in Figure 34 whose top plot shows the mean annualized returns for 1-3 year investment periods. Although the data-points are scattered around the forecasted mean, the colours of the dots show that different periods actually have had a tendency to follow the typical downwards slope of the forecasted mean. Some periods are slightly above the forecasted mean and others are slightly below, depending on the random noise in the Sales Growth and P/Sales ratio for those periods.

The middle plot in Figure 34 shows the mean annualized returns for 2-6 year investment periods, which corresponds to a midpoint of 4 years and is exactly the same as the cyclical component of the synthetic P/Sales ratio. This gives a slightly tighter fit to the forecasting model.

The bottom plot in Figure 34 shows the mean annualized returns for 6-10 year investment periods, so the midpoint is 8 years which is twice the frequency of the cyclical component of the P/Sales ratio. Now the fit is quite good at $R^2 = 0.80$ and the plot looks similar to some of the real-world case-studies from the previous sections.

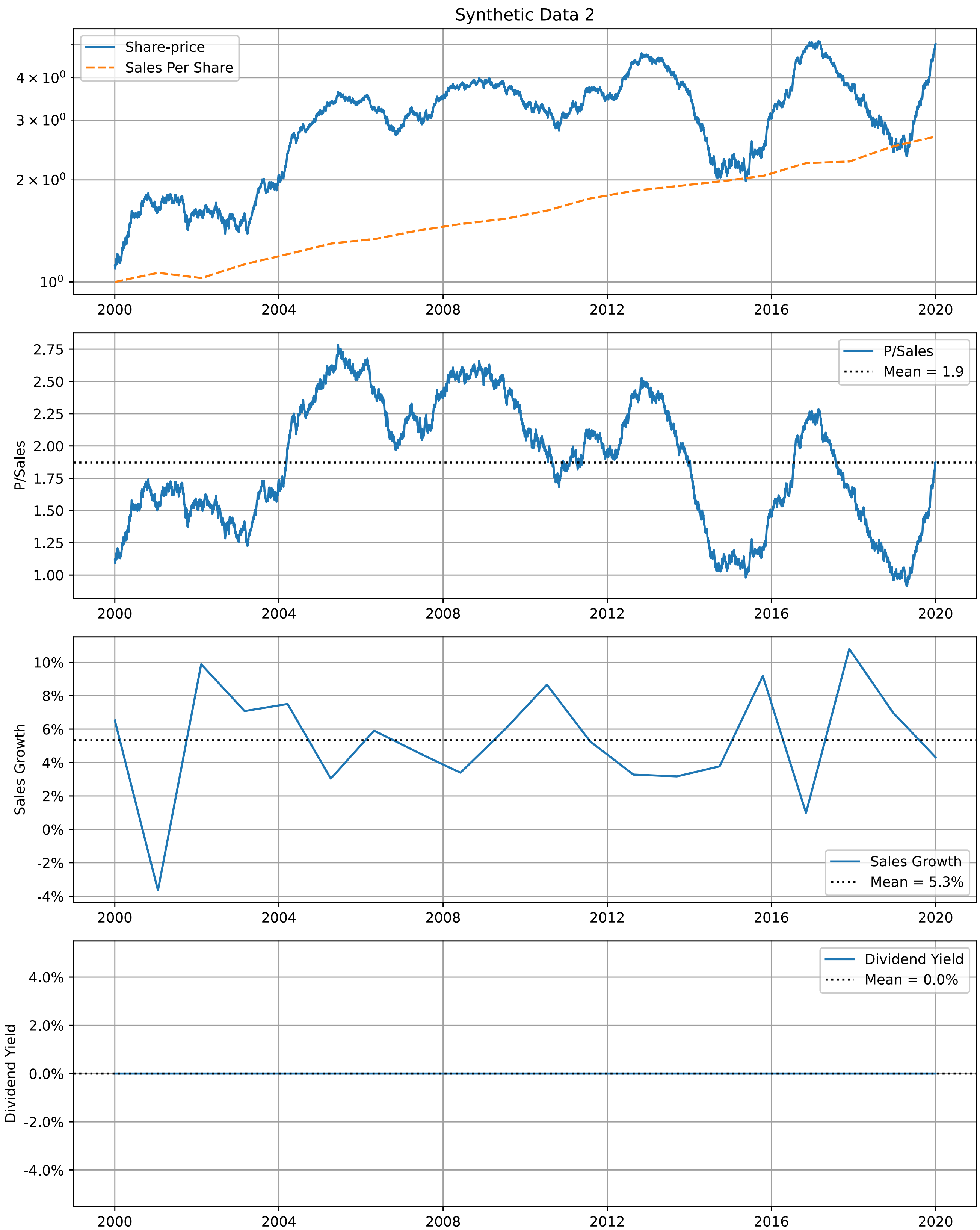


Figure 32: Synthetic financial data. The P/Sales ratio is generated as a sinus-curve with a full cycle of 4 years. The annual growth in Sales Per Share is 5%. Both have random noise. Dividend Yield is zero.

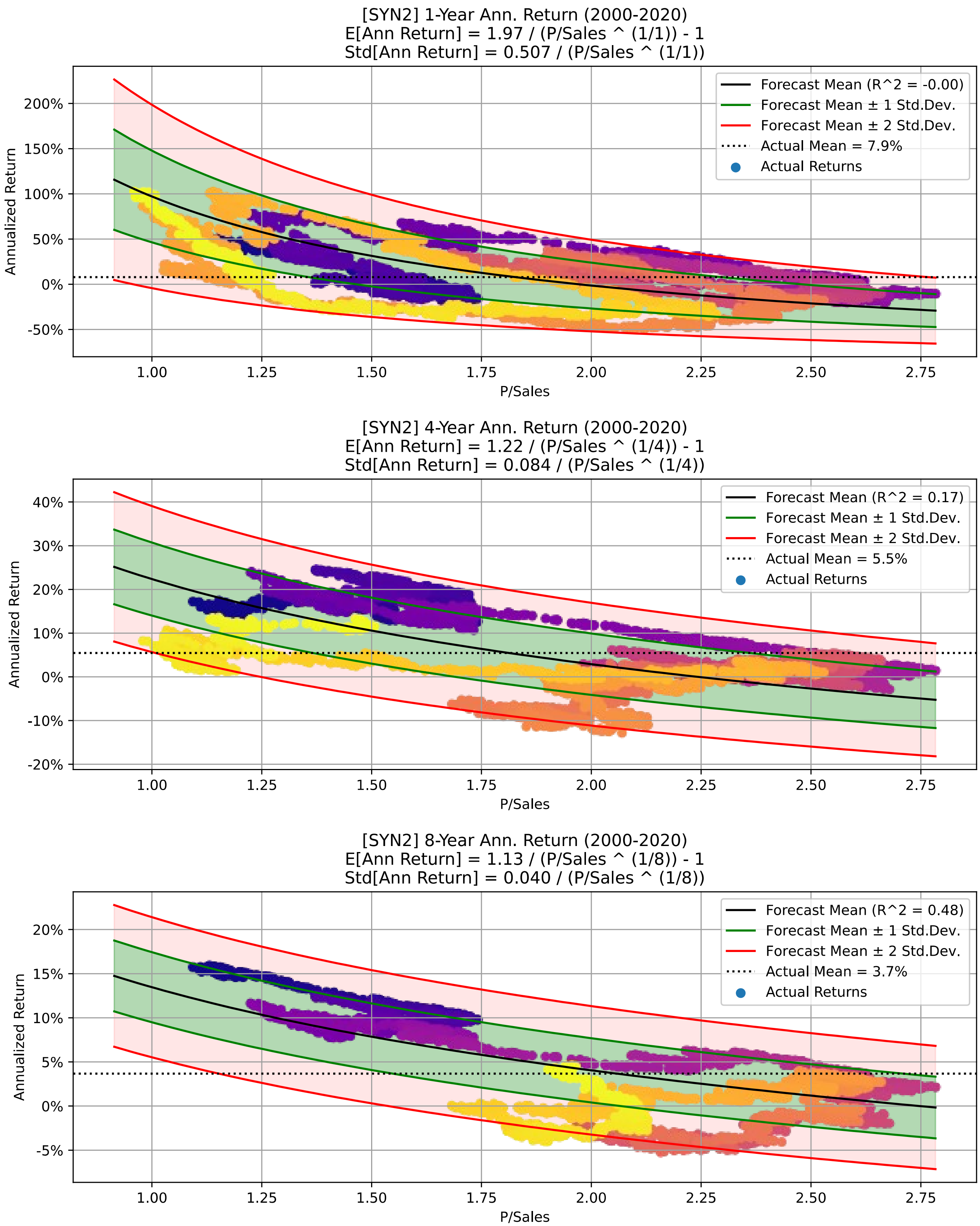


Figure 33: Annualized returns for the synthetic financial data in Figure 32, which has a cyclical P/Sales ratio and annual growth of 5% in Sales Per Share, both with random noise.

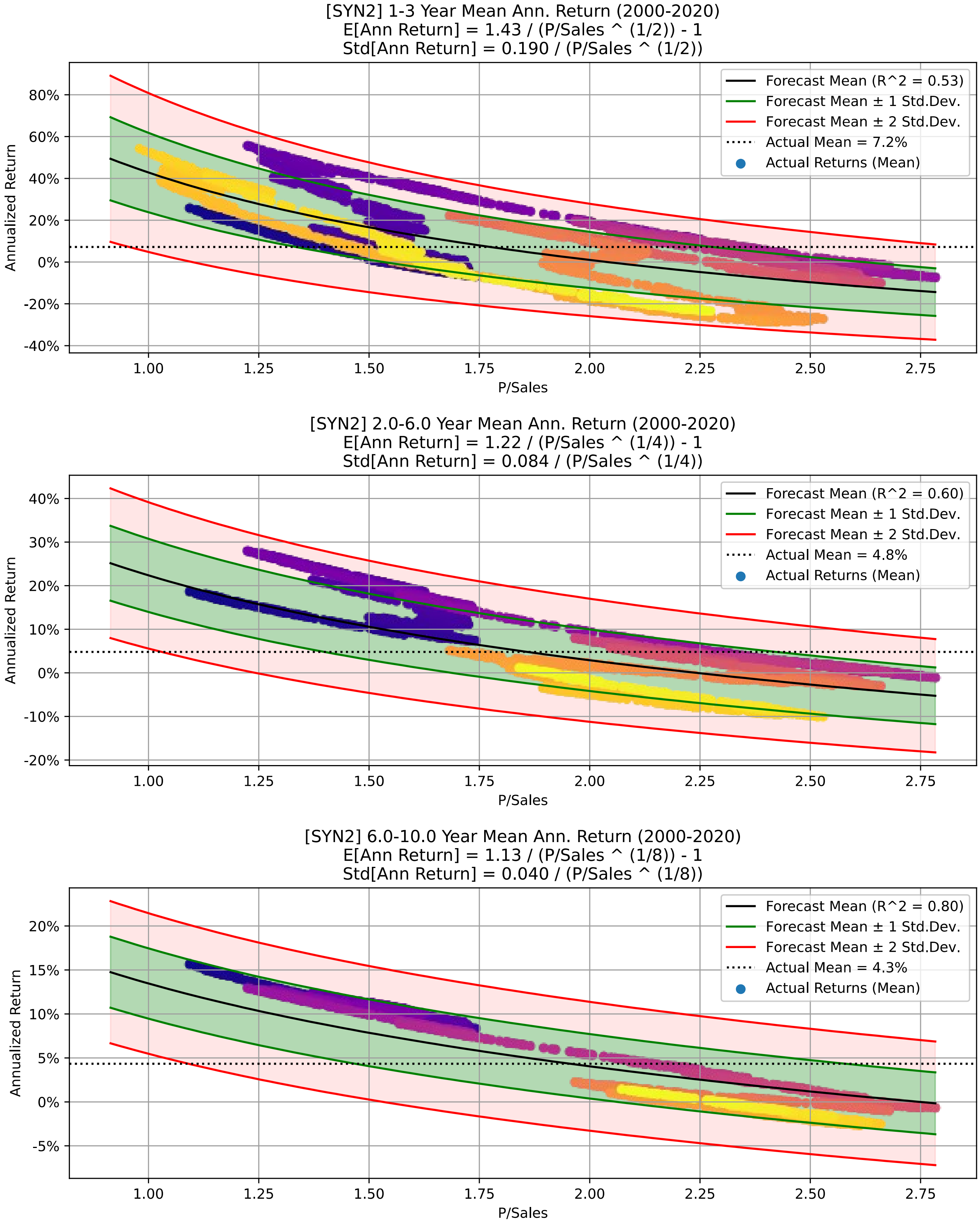


Figure 34: Mean annualized returns for the synthetic financial data in Figure 32, which has a cyclical P/Sales ratio and annual growth of 5% in Sales Per Share, both with random noise.

13.4 Trending Sales Growth

In Section 5 we had a real-world case-study for the company Walmart, which showed a peculiar pattern in the scatter-plots of annualized returns, that we argued was caused by Walmart having had explosive growth of 40% or more per year in its Sales Per Share, that gradually decreased to only a few percent annual growth after 30 years. We can now test this hypothesis using synthetic data.

Figure 35 shows the synthetic financial data that was generated similarly to the previous experiment, so the P/Sales ratio has a cyclical component with a full cycle of 4 years with random noise added. But now the annual growth in Sales Per Share decreases from 40% to 5%, also with random noise added. Also note that we generate 30 years of data instead of the 20 years of data that we generated in the previous two experiments, which is done so we can better see patterns in the scatter-plots similar to what we saw for the Walmart stock.

Figure 36 shows the annualized returns for the synthetic stock-data. The top plot shows it for 1-year investment periods. The middle plot shows it for 4-year investment periods which corresponds to a full period of the cyclical component of the P/Sales ratio. And the bottom plot shows it for 8-year investment periods which corresponds to two cycles of the P/Sales ratio. All these three plots have very poor fits of the forecasted mean to the annualized returns of the synthetic stock-data. Compare these plots to Figure 6 for the Walmart stock which looks quite similar.

Figure 37 shows the mean annualized returns for the synthetic stock-data. The top plot shows it for 1-3 year investment periods. The middle plot shows it for 2-6 year investment periods. And the bottom plot shows it for 6-10 year investment periods. These are averaging the annualized returns over several years and the ranges are specifically chosen to match the frequency of the cyclical component of the P/Sales ratio, but there is still a very poor fit of the forecasting model to the synthetic stock-returns. Although the forecasted mean is in the centre of the data-points, the fit is very poor because the data-points are so dispersed. If we look at the colour-coding of the data-points which represent different periods in time, it seems that individual time-periods have almost the same shape as the forecasted mean, it is just shifted up or down relative to the forecasted mean. This is because the annual growth in Sales Per Share starts at about 40% and gradually decreases to only about 5% at the end of the 30-year period. So when we only have a single forecasting model which uses the average annual growth in Sales Per Share of about 23%, we get a forecasted mean that is in the middle of the scatter-plot, which does not match the periods with much higher or much lower Sales Growth. This is the same problem we sometimes see when fitting real-world data, e.g. for the Walmart stock whose plots in Figure 7 look quite similar to these plots for synthetic data in Figure 37.

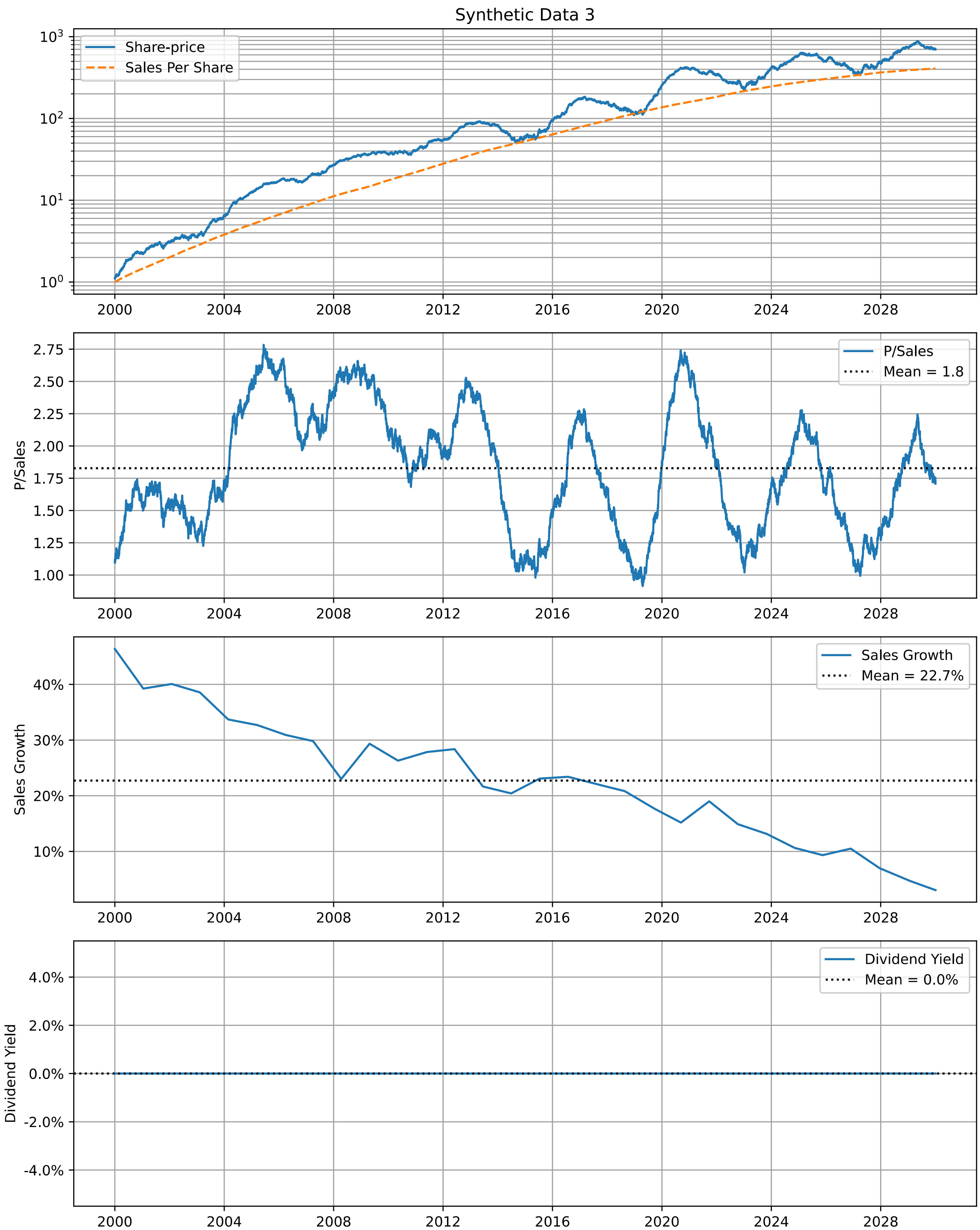


Figure 35: Synthetic financial data. P/Sales is a sinus-curve with frequency 4 years. Annual growth in Sales Per Share goes from 40% down to 5%. Both have random noise. Dividend Yield is zero.

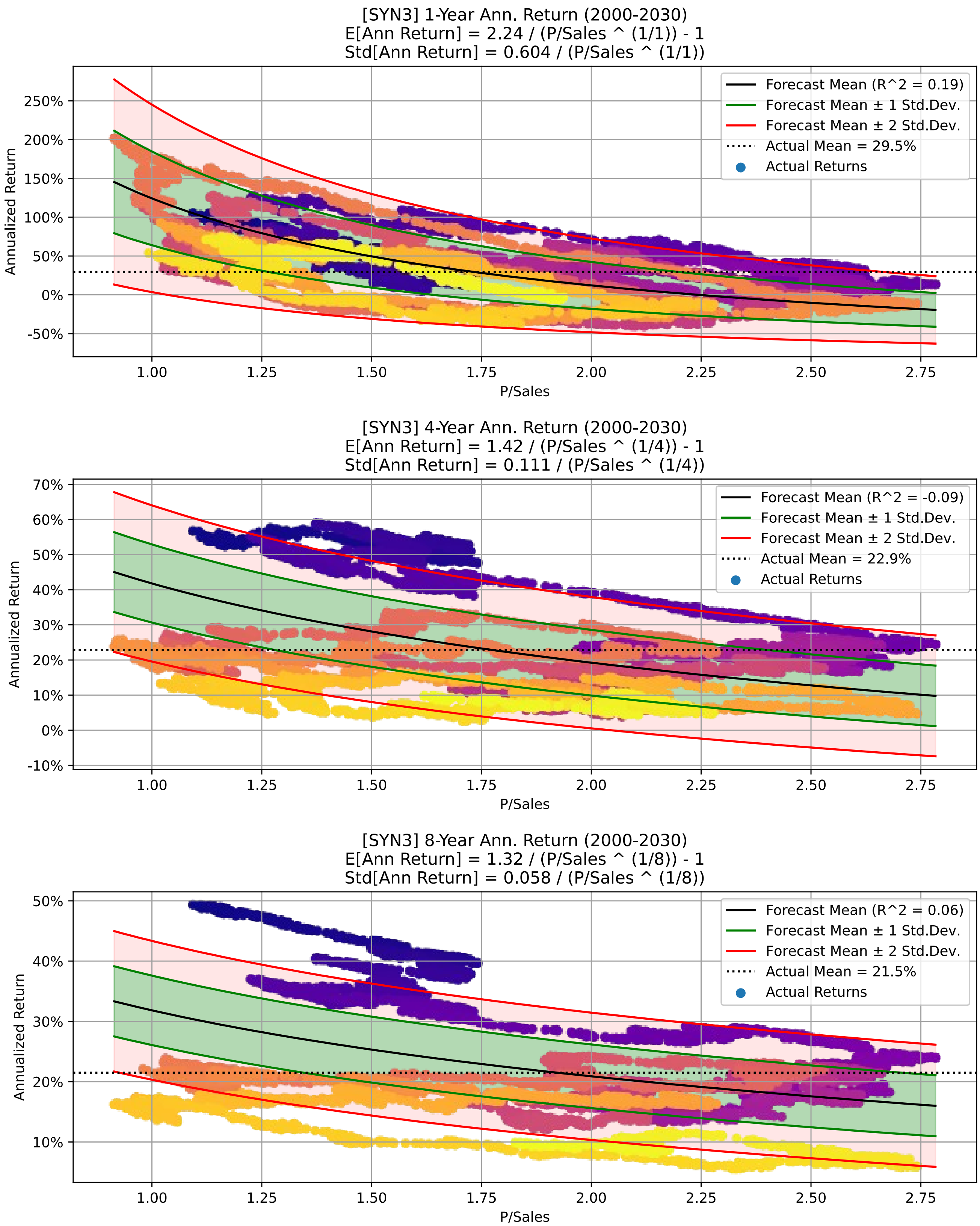


Figure 36: Annualized returns for the synthetic financial data in Figure 35, which has a cyclical P/Sales ratio and the annual growth in Sales Per Share goes from 40% down to 5%, both with noise.

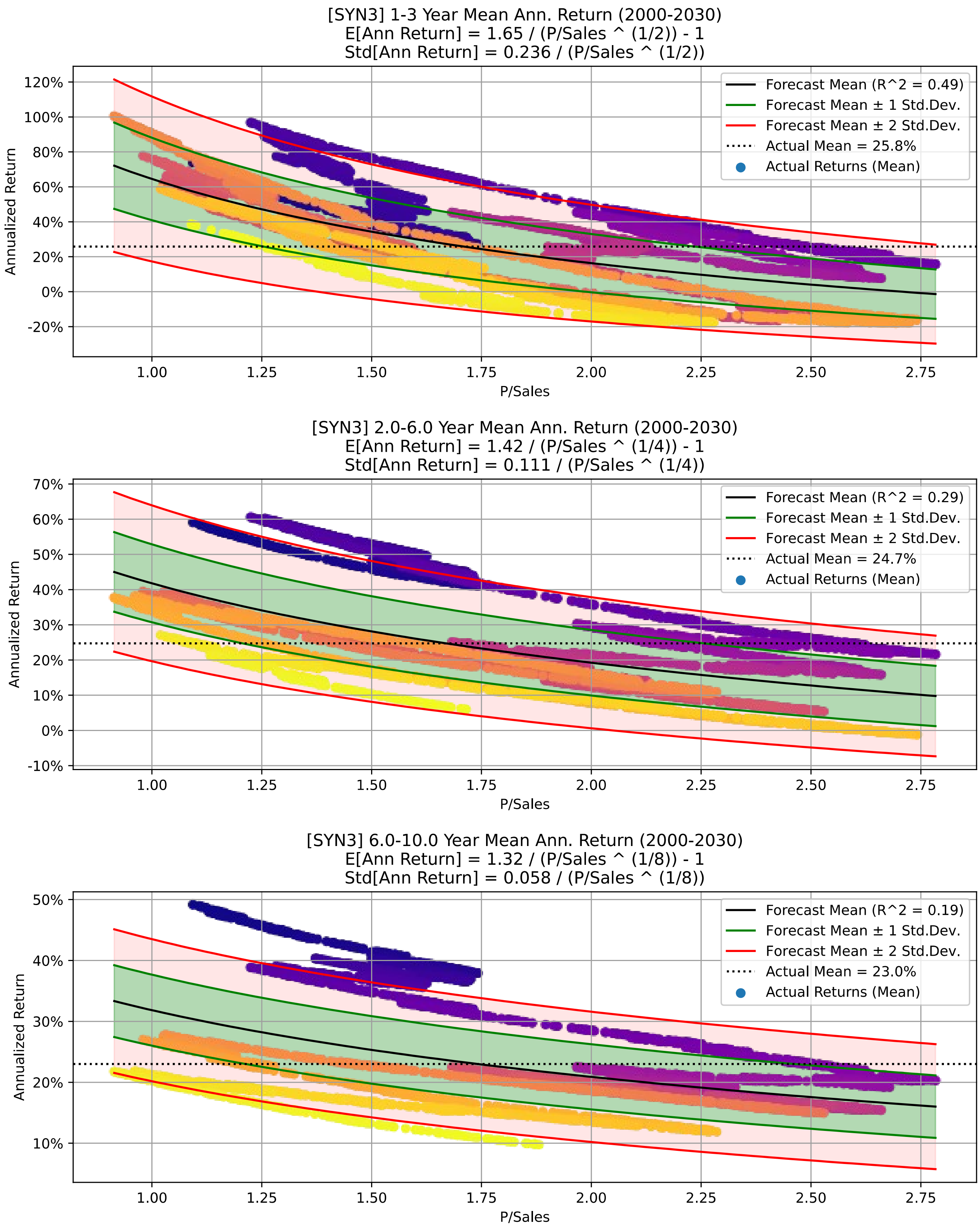


Figure 37: <u>Mean</u> annualized returns for the synthetic financial data in Figure 35, which has a cyclical P/Sales ratio and the annual growth in Sales Per Share goes from 40% down to 5%, both with noise.

14 Discussion

This section contains informal discussions on various topics.

14.1 Out-of-Sample Testing

When fitting statistical or Machine Learning models to data, it is common to split the dataset into training- and test-sets. The model is fitted to the data in the training-set and the model's performance is then evaluated on the test-set. This is also called "out-of-sample testing" and is done to evaluate how the model might perform on previously unseen data. This is necessary when the model does not know anything about the underlying system that has generated the data. Sometimes a model under-fits the data so essential features of the data are not discovered, and at other times a model over-fits the data so it believes that random noise carries essential information for making predictions. In either case the model will perform poorly on out-of-sample data, so we need to test for this when using general statistical or Machine Learning models.

We have not used out-of-sample testing in this paper, because the forecasting model is highly specialized, as it is derived directly from the mathematical definition of annualized stock-returns. This allows us to explain the conditions under which the model will fail, namely if the stock's future valuation ratio such as the P/Sales ratio, or its future growth in Sales Per Share, or its future Dividend Yield is significantly different from those values used in the forecasting model. We made several real-world case-studies, some of which did not have mean-reverting values and some of which did, but even so, there is no guarantee this will continue in the future.

So out-of-sample testing is not useful for this forecasting model, because we already know exactly when it will perform poorly, which is when the future valuation ratio such as P/Sales, or the future growth in Sales Per Share, or the future Dividend Yield are very different from our assumptions used in the forecasting model.

14.2 Growth vs. Revaluation

It is common to hear people talk about "Growth vs. Value stocks" but the exact definition is different amongst people. Some people think of growth-stocks as having high growth in Sales and Earnings Per Share which are accompanied by correspondingly higher valuation ratios. But that is not always the case, as some companies with high growth have stocks that sometimes trade at low valuation ratios. Conversely, some people think of value-stocks as having low valuation ratios because there is something inherently wrong, or highly uncertain or "risky" about the company. But that is not always the case either. Sometimes the stock-market is simply wrong!

It makes much more sense to talk about "Growth vs. Revaluation" because we know that the stock's return will be determined completely by the future change in e.g. Sales or Earnings Per Share (this is the growth-component), and the future change in valuation ratio such as P/Sales or P/E (this is the revaluation-component), as well as the future Dividend Yield (which is often but not always the smallest part of the return).

The question is then if we should prefer stocks based on the expected return from either the growth in Sales or Earnings Per Share, or from its expected return from revaluation, or from its expected return from dividends. But this is not an either-or question, as the stock-return depends on all three components. A company can have very high growth over the next 10 years, but if its shares are currently trading at a very high valuation ratio that will be revalued downwards, then we may still lose money when investing in that stock, if the growth in Sales or Earnings Per Share is not enough to offset the loss from revaluation.

The legendary investor Warren Buffett and his business partner Charlie Munger often say that: *"It is far better to buy a wonderful company at a fair price, than a fair company at a wonderful price."* They never explain exactly what they mean by these terms and how to measure them, but one way to look at it is through this lens of "Growth vs. Revaluation".

For example, let us say a stock is currently trading at a quite low valuation ratio that we believe will eventually double, but the exact timing is unpredictable and we might have to wait 10 years until this revaluation fully materializes. This corresponds to an annualized return of about 7.2% from the doubling through revaluation. If there is no growth in the Sales or Earnings Per Share during those 10 years, then we just get the 7.2% return per year from the revaluation, and whatever the Dividend Yield might be. If this is only a "fair company", or perhaps even worse, a "cigar-butt company" as Warren Buffett sometimes calls them, then we are not even certain about the revaluation, so the upside is only a return of 7.2% per year, while the downside could be a significant loss on our investment.

Now consider another example where the stock of a "wonderful company" is trading at a "fair price" which is taken to mean a "fair valuation ratio", so we do not expect the valuation ratio to change significantly in the future, and because the company is "wonderful" we also do not expect a significant decrease in its Sales Per Share, although the Earnings Per Share could naturally fluctuate a bit from year to year. If the company cannot grow its Sales and Earnings Per Share in the future, and the valuation ratio does not really change either, then the only source of investor-return would be the future Dividend Yield. But if we believe the company still has room to grow but we cannot quantify exactly how much, then this "wonderful company at a fair valuation ratio" with its low risk of long-term loss and its potential for future growth, might be preferable to a "fair company at a cheap valuation ratio" whose future is much more uncertain and the stock-return is perhaps limited to just the revaluation, which is not very much on an annualized basis if it takes 10 years to fully materialize.

14.3 Valuation Ratio

We have already discussed the choice of P/Sales as our valuation ratio in Section 2.5, with the main reason being its stability compared to the raw P/E ratio, which is ill-defined when the Earnings Per Share is zero or negative.

The problem with the P/Sales ratio is that it is difficult to interpret and compare across different stocks. When using the forecasting model, we need to make a reasonable guess about its future valuation ratio, but how can we do that if we don't have an intuitive understanding of it? There is no guarantee that the stock's historical average P/Sales ratio was correct, so we need to understand the valuation ratio better, in order to make a better guess at what it might be in the future.

Section 2.5 suggested using e.g. the moving 10-year average of the company's Net Profit Margin and multiplying this with the Sales Per Share for the Trailing Twelve Months, so as to get the "expected" Earnings Per Share based on the last year's Sales Per Share and the last 10 years of Net Profit Margin. This would give a more stable P/E ratio that would be easier to interpret and compare between stocks.

Alternatively we could try and understand the P/Sales ratio better by researching how it is related e.g. to a company's Net and Gross Profit Margins, growth in Sales Per Share, etc. Some initial research has already been done on this topic in the same collection of computer code mentioned in Section 16.4.

Regardless of which valuation ratio we use, we really need a way to estimate a "fair valuation ratio" that we might expect the future share-price to fluctuate around, that is not just based on historical data, but sound reasoning as to what constitutes a "fair valuation ratio". If share-prices were truly just random walks, then there is nothing to prevent the share-prices and valuation ratios to approach both zero and infinity at different times. Sometimes certain stocks do trade at either very low or very high valuation ratios, which may occur for only a short while or perhaps even a few years in some cases. But over longer periods of time, we nearly always observe that the participants in the stock-market tend to keep valuation ratios within certain ranges. The reason is probably that once the valuation ratio of a stock becomes very low, investors can make a very high return from the dividends alone, and conversely, when the valuation ratio of a stock becomes very high, investors can get a higher and safer return from low-risk bonds. These two extremes probably serve as boundaries that tend to keep valuation ratios within certain ranges, at least over longer periods of time.

14.4 Risk

Investment risk is very poorly understood in academia and seems to go through different fads. Some years ago it was popular to say that a stock's *beta* was a good risk-measure, because it measures the stock's correlation with the entire stock-market. But there are several problems when using this as a risk-measure. Firstly, how exactly should the correlation be calculated? For daily, weekly, monthly or yearly stock-returns? How about 5-year returns? Secondly, we know from the case-studies in this paper, that the valuation ratio is a strong predictor for long-term stock-returns. So we could have a stock trading at a low valuation ratio but also having a very high beta. The low valuation ratio tells us that there is good probability of large gains in the long-term, but that is allegedly contradicted by the high "risk" as indicated by the high beta. Thirdly, how do you measure the risk of the stock-market itself using beta? The reality is that beta only measures what it measures – and that is a stock's volatility relative to the overall stock-market. Beta is generally not a useful measure of investment risk.

Currently it seems popular to say that the variance of stock-returns is a good risk-measure. This gives an elegant framework for maximizing returns and minimizing "risk" in entire portfolios of stocks, which is known as Markowitz mean-variance portfolio optimization. There's just one problem: Variance is not a meaningful measure of investment risk either! Although it may be desirable to minimize your portfolio's daily volatility and hence its variance, this does not imply that "investment risk" is thereby minimized. It is very easy to give a counter-example where the minimum-variance portfolio always results in a loss.[13] It is also easy to show that the variance is inversely proportional to the probability of loss when the mean of the return-distribution is negative.[14] So once again, variance only measures what it measures – which is the spread of stock-returns, and not the probability and magnitude of any losses.

So what exactly is investment risk? The problem is that it is not just one thing, but more like a many-headed monster. Let us start with the dictionary definition that most people would agree with: "*Risk is the probability of loss or injury.*" In investment terms, it means the probability that the stock-return is below zero. But we are also interested in the magnitude of the losses. For example, an investment with 1% probability of a 100% loss may still be worth it if the potential gains are sufficiently large. On the other hand, an investment with 99% probability of 1% loss may not be worth it if the potential gains are too small. So the entire distribution of the future stock-returns is important.

In this paper we have estimated distributions for the future stock-returns by their mean and standard deviation (which could equivalently have been the variance). The mean is the centre of the distribution of stock-returns, and the standard deviation measures the spread around the mean, so it tells us how uncertain the forecast is. When the standard deviation is small, the forecasting model is very certain about the future stock-returns, and vice versa, when the standard deviation is large, the forecasting model is highly uncertain about the future stock-returns. If we assume that the stock-returns are so-called normal-distributed, then there is a mathematical formula that gives us the probability that some of the stock-returns are below zero and therefore investment losses. We could also calculate the magnitude of those losses.

13 https://youtu.be/wr8NzThfpAE
14 https://youtu.be/DzTlH6ipx98

But what if our estimates for the future stock-returns are wrong? The forecasting model might tell us that there is 10% probability of a loss, but how do we know that probability is accurate? The answer is that we don't know, and that is another type of investment risk: Either our forecasting model or its assumptions could be wrong. In our case we know that the forecasting model is mathematically correct, because it is derived directly from the definition of annualized return and a simple decomposition of the share-price. But the model still uses assumptions about the future P/Sales ratio, growth in Sales Per Share, and Dividend Yield, all of which could be very different from what will actually happen in the future. So this is a type of investment risk that arises from estimation errors.

Even if our estimates for the future stock-returns were absolutely correct, these are for long-term investments of several years, and we can make very bad portfolio allocations if we only use these long-term stock-returns to make our decisions. For example, let us say we invest our entire portfolio in a single stock because we have correctly determined that it will have a very high return if held for 5 years or more. But tomorrow the stock-market crashes and this particular stock drops by 90% while the rest of the market only drops 30%. Because all of our wealth is invested in the stock that crashed 90%, not only has our wealth shrunk tremendously, but we can't even take advantage of the stock's much lower price and buy more shares to get an even greater 5-year return from the lower share-price today. This investment risk is a form of "opportunity cost", where our investment in one stock robs us of the opportunity to invest in another. Even if we think we have a diversified portfolio, if all our stocks crash at the same time, we won't be able to take advantage of new opportunities as they arise.

14.5 Investment Strategy

An obvious investment strategy would be to use the forecasting model for all possible stocks with their historical averages for P/Sales, annual growth in Sales Per Share, and Dividend Yield, and then make a diversified investment in the stocks that have the highest forecasted mean return for the future 5 years or so. The problem with using the forecasting model on individual stocks, is that their historical averages may not continue in the future. That is why the forecasting model may work better for broadly diversified stock-indices, which are probably more predictable in the long-term.

For stock-indices the investment strategy would basically be the same, namely to use the forecasting model with the historical averages for P/Sales, annual growth in Sales Per Share, and Dividend Yield, and then invest more heavily in those stock-indices that have the highest forecasted mean return. This is a fine idea, except that the stock-indices tend to be cheap and expensive at the same time, in terms of their valuation ratios. The correlation is not perfect, however, and sometimes one stock-index is cheap while another is expensive.

The computer code mentioned in Section 16 has a comparison of the P/Sales ratio for different stock-indices, which shows that historically the P/Sales ratio has been strongly correlated for the S&P 500 and S&P 400, and for the S&P 600 and S&P 400, but not so strongly correlated for the S&P 500 and S&P 600. So perhaps you might get lucky and sometimes be in a situation where one stock-index is cheap while another is expensive. But even so, it is probably a bad idea to invest your entire portfolio in the stock-index with the highest forecasted return, because there is still considerable uncertainty in the forecasts, even for broadly diversified stock-indices. This was also discussed in Section 8.4.

14.6 Shorting Stocks

Normal investments in stocks are called “long” investments, in which you buy a stock through a broker and then you are the owner of that stock and therefore a part-owner of the given company. A more speculative strategy is called “shorting” in which you borrow a stock from another investor and then you sell that stock in the market. If the share-price goes down, then you buy the stock back and return it to the original investor for a profit. But if the share-price goes up, then your losses are potentially unlimited when you have to buy back the stock and return it to the original investor.

People have different reasons for shorting stocks. Some strategies employ so-called arbitrage, in which a stock is either statistically or contractually related to another stock or financial instrument, so it is possible to make a profit by combining short and long-positions in different stocks and instruments.

Some people short stocks simply because they believe the stocks are overvalued, e.g. using arguments similar to Section 3 for the Tesla stock. We could also argue that a stock has never before traded at such high valuation ratios in its many-year history, so we believe it must come down again. We could even back our reasoning by using the forecasting model which might predict a significant loss if the stock is held for several years. But there are several problems with this approach.

We have seen numerous examples in this paper, that the valuation ratio seems to be completely unpredictable in the short-term, and it may take several years for the valuation ratio to revert to its historical mean. During this time anything can happen – the stock or stock-index can trade around the same high valuation ratio, or it can go much higher. The stock-market can sometimes defy the “laws of financial gravity” for extended periods of time. It is also possible that a high valuation ratio sometimes correctly predicts high future growth in sales and earnings over the next several years, in which case the stock or stock-index will grow enough to justify the current over-valuation.

The point is, that you should be very careful using these forecasting models when shorting stocks or stock-indices. The forecasting models usually only work for long investment periods of several years or more, while shorting a stock is a highly speculative and short-term strategy.

14.7 Share Buybacks vs. Dividends

The forecasting model uses the Dividend Yield, which is the dividends relative to the share-price. But share buybacks are now significantly greater than dividends in USA. How does that affect the forecasting model? The short answer is that it doesn't, provided we use the historical data correctly.

We are forecasting the Total Return, which assumes dividends are reinvested immediately when they are paid out, and that there are no taxes. So our proportion of the outstanding shares is increased according to the dividend payment and the share-price at that day. If instead the company had used some or all of that dividend money for a share buyback, and assuming this did not move the share-price in any way, then the effect would be an identical decrease in the number of shares following the buyback. So there is no difference in the Total Return from a dividend payout and a share buyback.

In the forecasting model we typically use a historical distribution for the Dividend Yield. If the company then stops making dividend payouts and starts making share buybacks instead, then the contribution to the Total Return is simply moved from the Dividend Yield to the Sales Per Share instead, because the number of outstanding shares decreases from the buyback, so the Sales Per Share merely grows by a corresponding amount. The effect on Total Return should be identical.

However, it is important that the data is selected correctly. The data for the Dividend Yield and Sales Per Share must be selected from the same period in time, so we do not overestimate the future Dividend Yield relative to growth in Sales Per Share from share buybacks.

For example, for the S&P 500 we have data for the Dividend Yield from the 1950's onwards, while we only have data for the Sales Per Share from 1988 onwards. During the 1990's companies in USA started shifting a lot of their dividend payouts to share buybacks. We therefore need to use both Dividend and Sales Per Share data from 1988 onwards, so their contribution to Total Return is mutually consistent. We cannot use the dividend-data between 1950 and 1988 to estimate the future Dividend Yield, because it would then be overestimated relative to the effect of share buybacks on the Sales Per Share from 1988 onwards, if companies continue to buyback a lot of shares.

Although the Total Return is identical for share buybacks and dividend payouts, it does not mean that a company should automatically buy back shares. Just as reinvestment of dividends is optional for the individual investor, a share buyback is optional for the company's management. If you trust the forecasting models we have developed here, you would be foolish to reinvest your dividend if the forecasted return on the stock is very low or perhaps even negative. You would either invest the dividend elsewhere or simply keep the cash. Similarly, the company would be foolish to buy back its shares when the forecasted returns are very low. In that case the company should just make a dividend payout and let investors decide what to do with the money, or the company could keep the cash for other opportunities in the future. See my extensive theory on share buyback valuation [8].

15 Conclusion

This paper presented a simple forecasting model, which was derived from A) the mathematical definition of annualized return, and B) the decomposition of the share-price into the product of a valuation ratio such as the P/Sales ratio and the Sales Per Share. Forecasting the future stock-return is then a matter of forecasting the three return components: The future valuation ratio such as P/Sales, the future growth in Sales Per Share, and the future Dividend Yield; and then inserting these values into the forecasting model. The forecasting model is an exact mathematical identity, so if you can precisely predict these three components of stock-returns, then you can precisely predict the future stock-returns.

Typically you cannot predict the three components of stock-returns with great precision, and you will therefore use a range or distribution of possible future values for the three return components. The forecasting model then gives you a mean and standard deviation for the future stock-returns.

The advantage of splitting the forecasting formula into these three components, is that for some stocks – and especially for some stock-indices – it is much easier to predict their future valuation ratio and growth in Earnings or Sales Per Share, than it is to predict the future stock-returns directly.

It seems almost embarrassingly obvious in hindsight, that stock-returns are determined entirely by these three components, although the exact formulas are perhaps not so obvious, which is probably why this paper is apparently the first time the formulas have been published.

In this paper we also used the forecasting formulas in several real-world case studies, both for individual stocks and for entire stock-indices. We tested how well the forecasting model worked when we “cheated” and used the historical data for the three return components before that data was actually known. Sometimes the forecasting model fitted the historical stock-returns very well with high R^2 values, and sometimes the forecasting model had a very poor fit to the historical stock-returns, even though it was “cheating” by using the averages for all the historical data before it was actually known.

The reason that the forecasting model sometimes fails is very simple: If one or more of the three return components turn out to be very different from that assumed in the forecasting model, then the model will give a very inaccurate forecast for the stock-return. It must therefore be stressed again, that the model’s forecast is only as accurate as your estimates for the three return components: The future valuation ratio such as P/Sales, the future growth in Sales Per Share, and the future Dividend Yield.

Although the forecasting model may have a very high R^2 value on historical stock-returns when using the historical averages for the three return components, this does not automatically imply that the model will be accurate in the future. And vice versa: Although the forecasting model may have had a low R^2 value on historical data, this does not imply that the model will be inaccurate in the future. It all depends on the accuracy of your assumptions for the three components of the future stock-return.

This paper also presented numerous empirical studies of how the P/Sales ratio related to stock-returns over different investment periods. It is clear from all the scatter-plots for the different stocks and stock-indices, that the P/Sales ratio had little to no predictive power on the investment returns for only 1-year periods. For some individual stocks and stock-indices, there was already a strong relation between the P/Sales ratio and 2-4 year annualized returns, while for other stocks and stock-indices, this relation only became clear for longer investment periods of 5 years or more. This is because the P/Sales ratio typically takes several years to revert to its historical mean.

So if you are investing mainly to capture stock-returns from revaluation, then you should be prepared to hold the investments for several years. You should not expect a stock or stock-index to have its valuation ratio revert to its historical mean or any other particular value after just 1 year. Revaluation of stocks and entire stock-indices often takes several years.

When studying the historical data of a stock or stock-index whose three return components have been somewhat mean-reverting, the scatter-plot of the P/Sales ratio versus the annualized returns for longer investment periods shows a particular pattern, where higher P/Sales ratios correspond to lower future stock-returns, and vice versa. Moreover, the scatter-plot has a particular downwards-sloping curve, which arises from the mathematical definition of annualized return, as explained in this paper.

When the parameters of the forecasting model have been calculated using the historical averages for the three return components of the stock or stock-index, and these three components have been mean-reverting historically, then the forecasting model will have a very good fit to the historical data. But once again, just because the forecasting model has an excellent fit to the historical data, this does not guarantee that it will provide excellent predictions for the future stock-returns. This depends entirely on how close the historical averages are to the future return components: The future Dividend Yield, the future growth in Earnings or Sales Per Share, and the future change in valuation ratio.

Now that we understand exactly how stock-returns are related to the three return components, we should focus our research efforts on how to forecast each of these three individual return components.

16 Data & Computer Code

This section briefly describes the different data sources, how the data was processed, and where to obtain both the data and computer code so you can rerun all the experiments in this paper by yourself.

16.1 Data Sources

Daily share-prices for individual stocks, indices and ETF's are from Yahoo Finance, e.g. the S&P 500.

Financial data for the individual companies in Section 4 and 5 was collected manually by the author of this paper from the Form 10-K annual reports filed with the US SEC e.g. for Walmart.

Newer financial data for the S&P 500 is from the S&P Earnings & Estimates Report and older data is from the research staff at S&P and Compustat, with some of the older data having been approximated by their research staff. The S&P 400 and S&P 600 data is also from these sources.

Financial data for Exchange Traded Funds (ETF) is from Morningstar Direct.

16.2 ETF Data Processing

For each stock, index or ETF, the forecasting model needs the annual growth in Sales Per Share, the P/Sales ratio, and the Dividend Yield. The data from Morningstar has the P/Sales ratio but not the Sales Per Share, so that was estimated using the following formula, which is just a simple rewrite of Eq. (5) for the definition of the P/Sales ratio:

$$Sales\,Per\,Share_t = \frac{Share\,Price_t}{P/Sales_t}$$

Because the Morningstar data contains the Net Asset Value Per Share (NAV), but not the daily Share-Price, and these are usually very close to each other, it was actually easier to estimate it from:

$$Sales\,Per\,Share_t = \frac{NAV_t}{P/Sales_t}$$

How Morningstar gathers and processes its data is not entirely clear and was only briefly explained to me by their support staff. The P/Sales ratio is apparently calculated for each stock in the ETF's holdings, and these are then weighted according to the stock-weights in the ETF, so as to create the P/Sales ratio for the ETF. This is calculated every month using the Trailing Twelve Months (TTM) of Sales Per Share for the individual stocks in the ETF holdings, along with each stock's price for the last day in the month. It is unclear how different currencies are treated in this calculation and perhaps that contributes to the sometimes extreme volatility we see in the derived Sales Per Share in some of the case-studies. Because of these issues you should view the ETF case-studies with some skepticism.

16.3 Further Data Processing

Once we have the data for each stock, index and ETF, we then interpolate it to obtain daily data-points including weekends and holidays. This makes the data easier to work with and also gives more smooth data-points, instead of only using the last-known values, which would give a “staircase” pattern in the data. But this is a form of cheating because the linear interpolation uses future data that was not known at the time-steps where it is being used. However, this is deemed acceptable because we are analysing long-term stock-returns for several years or more, while the interpolation is being done for quarterly or annual data-points. Interpolation is generally not acceptable for short-term research and development of trading strategies, because it would be cheating.

16.4 Computer Code

The computer code used to generate all the experiments and plots in this paper is written in a so-called Python Notebook which is available on GitHub. You should download the entire GitHub repository so you will get all the necessary code and data-files, so you can repeat all the experiments yourself, or you can reuse the computer code with your own data for other stocks and indices.

16.5 Spreadsheet

There is also a spreadsheet compatible with Microsoft Excel that summarizes the forecasting models for the stock-indices, including internet links for looking up the current P/Sales ratios, so you can easily evaluate the forecasted returns using updated P/Sales ratios.

17 Bibliography

1: Ben Graham and David Dodd, Security Analysis, 1934.

2: John Y. Campbell and Robert J. Shiller, Valuation Ratios and the Long-Run Stock Market Outlook: An Update, Cowles Foundation Discussion Paper 1295, 2001.

3: Robert Shiller, Irrational Exuberance, 3rd ed., 2015.

4: Norbert Keimling, Predicting Stock Market Returns Using The Shiller CAPE, StarCapital Research, 2016.

5: Adam Zaremba, The Cross Section of Country Equity Returns: A Review of Empirical Literature, Journal of Risk and Financial Management, 2019.

6: Magnus Erik Hvass Pedersen, Strategies for Investing in the S&P 500, 2015.

7: Magnus Erik Hvass Pedersen, Comparison of U.S. Stock Indices, 2015.

8: Magnus Erik Hvass Pedersen, The Value of Share Buybacks, 2012.

My previous papers and books can all be downloaded through SSRN and GitHub.

18 Revision History

- 2021, February 14: Rewrote the abstract and clarified a few things in the introduction, conclusion, and Tesla case-study. (104 pages)
- 2020, December 17: First edition. (103 pages)