A NOTE ON LONG-TERM BAYESIAN MODELING OF STANDARD & POOR COMPOSITE INDEX RETURNS

TARAN GROVE, ANDREY SARANTSEV

ABSTRACT. This is a short note, which contains a simple dynamic factor model: Total (including dividends) real (inflation-adjusted) annual returns of Standard & Poor Composite Index is regressed upon earnings yield, which is defined as earnings (net income) per share of this index (summed over all companies in this index), divided by the current value of this index. This earnings yield is modeled as a simple autoregression. We use 7-year trailing earnings yield, with earnings averaged over the last 7 years. This version of earnings yield has maximal correlation with next year's total real returns. Regression residuals fail standard normality tests, but we still use this model, since we are interested in long-run modeling. To account for shortage of data (1935–2019 annual), we use Bayesian inference. We confirm the conventional wisdom that future long-run stock market returns are likely to be lower than the historical averages.

1. Introduction

The modeling of future stock market movements, based on current indicators, is a major area of research in financial economics. One such indicator is the current earnings (net income after taxes and other expenses) of a stock compared to its price. If the earnings are low, this implies that the stock market is overpriced and, consequently, in a bubble. Therefore, the future expected returns of the stock market are low.

This research dates back to Benjamin Graham who popularized *value investing:* choosing stocks with low price-earnings ratios (computed as the current price of a stock divided by its earnings over the last year). This is opposed to *growth investing*, which prefers high P/E ratios. The growth investing strategy is so named because stocks with high P/E ratios have, in the minds of investors, a high potential for earnings growth. Stocks with low P/E ratios are called *value stocks*, while stocks with high P/E ratios are called *growth stocks*. Indeed, subsequent research by [4] confirmed that value stocks outperform growth stocks in the long run, relative to their level of risk. Robert Shiller and his collaborators applied this method to the stock market as a whole. When the P/E ratio of the stock market is high, the future expected returns of the stock market (say over the next 10 years) are low. Since the seminal work by Robert Shiller [3, 8], the price-to-earnings (PE) ratio approach to investing has captured the attention of academics and practitioners in finance, and seeped into the popular literature.

However, the traditional P/E ratio can be significantly influenced by the high volatility of net income. For example, in 2008, the P/E ratio was very high because multiple companies in the financial sector reported negative incomes. This, in turn, greatly reduced the total net income for the stock market as a whole. We compute the P/E ratio of the stock market

2010 *Mathematics Subject Classification.* 62J05, 62M10, 62P05.

Key words and phrases. Linear regression, Jeffrey's prior, quantile-quantile plots, Bayesian inference, earnings yield, autoregression.

(measured by a benchmark such as the Standard & Poor 500 index) by summing market capitalization (market value) of all stocks in this index and dividing it by the sum of the net incomes of all these companies in the last year. This, however, suffers from a drawback. Namely, the negative income of one company does not cancel the positive income of another company. This is called *aggregation bias*.

Shiller considered the trailing 10-year P/E ratio, where the earnings are averaged over the last 10 years. This quantity is called the Shiller P/E ratio. Shiller found that future 10-year and 20-year returns are highly negatively correlated with the Shiller P/E ratio. However, he was unable to use standard linear regression tools since these 10-year periods overlap and the corresponding residuals are not independent. Instead, he devised a new statistical approach, which we shall not utilize here. It is impossible to survey all existing literature; instead, we refer the reader to the articles cited above, as well as [7, 5, 1, 6].

In this article, we consider the inverse of the P/E ratio: *earnings yield*, defined as last year's total net income of all stocks in the benchmark index, divided by the current sum of the market capitalizations of all these companies.

We take data that dates back to 1935, when the SEC was created, and the worst of the Great Depression was in the past. At this time, the U.S. Federal Government adopted an activist monetary and fiscal policy. This represents a crucial turning point, and there were no stock market crashes comparable to the Great Depression since then. Indeed, consider the events in 2008–2009: the Federal Reserve and the Federal Goverment stepped in to arrest the collapse of the economy. Yet, the resulting drop in stock prices and incomes was small compared to the Great Depression. Another example is the Black Monday of 1987. This stock market crash was every bit as terrifying as October, 1929. However, the Federal Government cut interest rates, earnings growth did not slow down, and markets soon recovered. The total inflation-adjusted wealth from $1 invested at the beginning of 1935 and dividends reinvested is shown in Figure 1, on the logarithmic scale.

Equivalently, we can define earnings yield as the 12-month trailing earnings per share of this index, divided by the current index level. The time step is equal to 1 year and we model this quantity as a simple autoregression of order 1. Next, we compute the annual total returns of the index, which includes dividends. We regress next year's total returns over the current earnings yield. Since the 1-year earnings yield is too volatile, as mentioned above, we consider versions of this indicator with net incomes averaged over the last k years. The strongest correlation between the k-year earnings yield and next year's total returns is when $k = 7$. We use this instead of the 10-year Shiller P/E ratio.

We find that this model does not fit well, since the residuals in both regressions are not normally distributed, and the correlation between earnings yield and total returns are still just weakly correlated. Indeed, even if the residuals could be modeled as i.i.d. normal, we could not reject the null hypotheses (zero slope) based on standard statistical significance tests for a simple linear regressoin. In other words, the annual returns are very volatile. Another difficulty that arises is that there are only a few years of data (less than 100).

However, we can account for the uncertainty in the parameter estimate using a Bayesian framework. We put a non-informative Jeffrey's prior on all the regression parameters, and utilize well-known, explicit formulae for the posterior. Then, we simulate 20-year future returns, starting both from the historical average of this 7-year earnings yield and from its current (start of 2019) value. Since we are interested in 20-year returns (the average of 20 1-year returns), it is arguably fair to use a model with normal residuals, even when they are

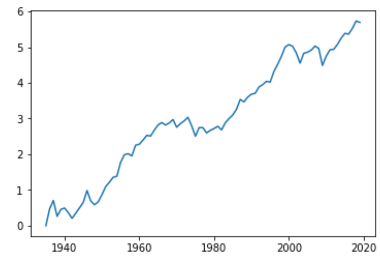
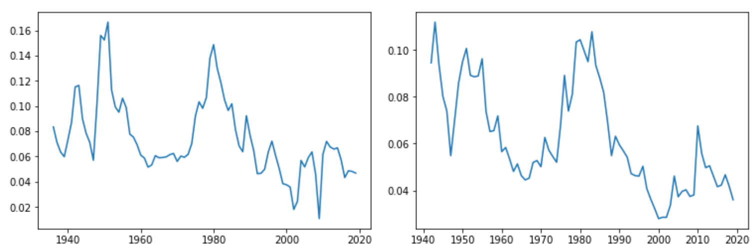


FIGURE 1. Inflation-adjusted wealth for S&P index, including dividends



(A) $E_1(t)$ (B) $E_7(t)$

FIGURE 2. Earnings yield for 1-year and 7-year windows

not actually normal. Indeed, by the Central Limit Theorem, a sum of many independent (but not normally distributed) random variables is close to normal, under some conditions.

We simulate each of the two scenarios 10000 times. We are interested in the follow quantities: mean, standard deviation, and 5% value at risk (meaning the value v such that 95% of the simulation results are greater than v). Our simulations showed that future returns in 2019-2038 have much a much lower mean, a higher standard deviation, and a lower 5% value at risk than returns starting from the historical average of 7 year earnings yield. This is consistent with the conventional wisdom, sometimes expressed in popular finance books and journals: future returns are not likely to be as good as they were in the past.

The article is organized as follows. In Section 2, we present the data and its sources. In Section 3, we describe the model, estimate the parameters, and evaluate goodness of fit. In Section 4, we apply the Bayesian prior, get explicit formulae for the posterior, and provide the simulation results. Section 5 deals with possible extensions and improvements that could be made. The Appendix contains well-known results about Bayesian simple linear regression with a Jeffrey's prior, which we will use in this article. Our code and data can be found at the `GitHub` repository `asarantsev/BayesianLT`

2. DATA

Our benchmark index is the Standard & Poor 500, created in 1957, which contains the 500 largest companies in the United States of America. Our benchmark also contains its predecessor — the Standard & Poor 90, created in 1926. Below, we simply refer to this index as S&P. We take the beginning-of-year level of this index from Yahoo Finance. To adjust for inflation, we use the annual Consumer Price Index (CPI) January data. We also use annual earnings and dividends data for the S&P share. The data can be found in the book [10] and (with updates) Robert Shiller's Yale University web site.

We take the data 1935–2018 for earnings and dividends, and 1935–2019 for S&P index and CPI levels. Although Robert Shiller's data goes back to 1871, we believe that the events of the 1930's fundamentally altered the U.S. financial system, as explained in the Introduction.

Denote by $S(t)$, the S&P index level at the beginning of the year t, with $t = 0$ corresponding to 1935. Similarly, denote by $D(t)$ and $I(t)$, the dividends and net income per S&P share, respectively, with $t = 1$ corresponding to 1935. Also, denote by $C(t)$, the CPI level in January of the year t, with $t = 0$ corresponding to 1935. Then, we define total real returns during the year t by

$$R(t) = \ln\frac{S(t) + D(t)}{S(t-1)} - \ln\frac{C(t)}{C(t-1)},\ \text{for } t = 1, \dots, 84, \tag{1}$$

and the k-year earnings yield by

$$E_k(t) = \frac{I(t) + \dots + I(t-k+1)}{kS(t)},\ \text{for } t = k, \dots, 84.$$

We know $E_k(t)$ by the beginning of the year $t+1$. Thus, we wish to use $E_k(t)$ to predict $R(t+1)$, returns during the year $t+1$. We use logarithmic (geometric) returns in (1), not standard arithmetic returns, to avoid the complication of compound interest. The same remark applies for the logarithmic rate of inflation in (1). Earnings yields for $k = 1$ and $k = 7$ (we will use the latter smoothed version below) are shown in Figure 2.

3. Model Fitting

Our model is given by the following two independent regressions, for $t = 2, \dots, 84$:

$$\begin{aligned} R(t) &= \alpha + \beta E_k(t-1) + \delta(t), \delta(t) \sim \mathcal{N}(0, \sigma_\delta^2) \text{ i.i.d.} \\ E_k(t) &= a + bE_k(t-1) + \varepsilon(t), \varepsilon(t) \sim \mathcal{N}(0, \sigma_\varepsilon^2) \text{ i.i.d.} \end{aligned} \tag{2}$$

where ε and δ are independent. In total, we have $T = 83$ time steps. We choose k to make the correlation between $R(t)$ and $E_k(t-1)$ as large as possible: The result is $k = 7$. However, the correlation is still only 23%, see Figure 3. Then, we use standard linear regression theory (least squares method) to get point estimates of $a, b, \alpha, \beta, \sigma_\varepsilon, \sigma_\delta$:

$$\begin{aligned} R(t) &= -0.031 + 1.63E_7(t-1) + 0.153\varepsilon(t), \\ E_7(t) &= 0.0056 + 0.898E_7(t-1) + 0.0094\delta(t), \end{aligned} \tag{3}$$

However, when we try quantile-quantile plots of either $\varepsilon(t), t = 2, \dots, 84$ or $\delta(t), t = 2, \dots, 84$ from (3) versus the standard normal distribution, the result is not a good fit.

Alternatively, we can apply normality tests to $\varepsilon(t)$ and $\delta(t)$. Namely, we can use the Jarque-Bera test (which analalyzes the skewness and kurtosis) or the Shapiro-Wilk test (a more powerfult test), for each of the two regressions in (2). All four tests give very low p-values. This means we must reject the null hypothesis (that the residuals are i.i.d. normal).

As explained in the Appendix, the non-informative Bayesian prior on $a, b, \alpha, \beta, \sigma_\varepsilon, \sigma_\delta$ is

$$\begin{aligned} \pi(\sigma_\varepsilon^2) &\sim \sigma_\varepsilon^{-2}, \pi(a, b \big| \sigma_\varepsilon^2) = 1, \\ \pi(\sigma_\delta^2) &\sim \sigma_\delta^{-2}, \pi(\alpha, \beta \big| \sigma_\delta^2) = 1. \end{aligned} \tag{4}$$

Thus, we have no prior knowledge about these parameters. We compute the *posterior*:

$$p(a, b, \alpha, \beta, \sigma_\varepsilon, \sigma_\delta \big| E, R) \sim L(E, R \big| a, b, \alpha, \beta, \sigma_\varepsilon, \sigma_\delta)\pi(a, b, \alpha, \beta, \sigma_\varepsilon, \sigma_\delta)$$

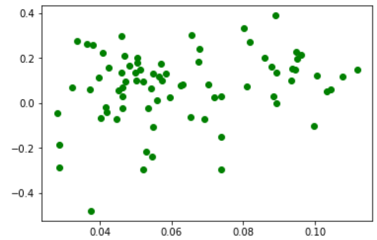
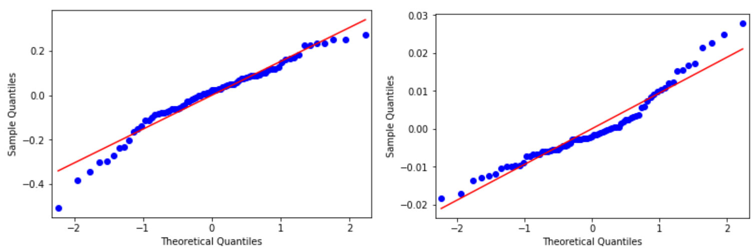


FIGURE 3. $(E_7(t-1), R(t))$, $t = 2, \dots, 84$.



(A) $\varepsilon(t)$, $t = 2, \dots, 84$ (B) $\delta(t)$, $t = 2, \dots, 84$

FIGURE 4. Quantile-quantile plots of residuals versus standard normal law.

where $L(E, R \big| a, b, \alpha, \beta, \sigma_\varepsilon, \sigma_\delta)$ is the *likelihoood* of y in (1): the product of bivariate normal densities. Actually, the prior π is an infinite measure, not a probability distribution, as its integral over all a, b, α, β and positive $\sigma_\varepsilon, \sigma_\delta$ is equal to infinity. However, the posterior for $a, b, \alpha, \beta, \sigma_\varepsilon, \sigma_\delta$ is a probability measure, found in [9, Chapter 4] and provided in the Appendix. Since the residuals of the two regressions in (2) are independent, we can apply Bayesian inference with Jeffrey's non-informative priors independently for each regression.

4. Simulation Results

For both the current value of the 7-year earnings yield and the long-term average, we perform 1000 simulations. For each simulation we do the following steps:

(a) Simulate the values of the regression parameters for both regressions. First, find σ_ε and σ_δ, and then find a, b, α, β.
(b) Simulate 20 steps of these two regressions, starting from the given value of the 7-year earnings yield. This corresponds to 20 years.
(c) Compute the average real return over these 20 years.
(d) Next, for each of the two cases, we have 10,000 results: the average returns for each simulation. Thus, we compute the mean, standard deviation, and 5% value at risk.

For the current (beginning of 2019) value 3.6% of the 7-year earnings yield, the results are: mean 4.4%, standard deviation 5.2%, 5% value-at-risk -4.4%.

For the long-term average value 6.2% of the 7-year earnings yield, the results are: mean 6.4%, standard deviation 5.0%, 5% value-at-risk -1.8%.

5. Conclusion

The current article fits and simulates a very simple model of stock market valuation. Even though it does not fit well, it is able to explain some variation in market returns and support the conventional wisdom that the stock market is currently overvalued. However, there is much room for improvement. One wonders whether the current high valuations are due to very low interest rates — since bonds are unnatractive, investors switch to other classes of assets are are willing to buy them, even at high prices. A better model would incorporate both short-term Treasury rates (highly correlated with the federal funds rate which is controlled by the Federal Government) and long-term Treasury rates, which measure investors' optimism about the economy.

The other drawback of our model was mentional already: the residuals of our regression are not normally distributed. One way to overcome this would be to increase the time window (our preliminary research indicated that switching from 1-year to 3-year returns leads to normality). In 78 years of our observations, we have $26 = \frac{78}{3}$ 3-year windows. This is only four data points. Again, however, we may use a Bayesian framework to account for uncertainty. Then, we will need to consider cumulative earnings and dividends over 3 years in these time steps. Another approach would be to consider heavy-tailed distributions for the residuals, or a non-parametric setting. This is left for future research.

6. Appendix: Bayesian Simple Linear Regression

The following short review section is taken from [9, pp.46-47]. Assume we have a simple linear regression

$$y_i = m + kx_i + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \rho^2) \quad \text{i.i.d.} \quad i = 1, \ldots, n.$$

Classic statistics gives us the point estimates:

$$\begin{aligned} \hat{k} &:= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad \hat{m} := \bar{y} - \hat{k}\bar{x}, \\ \hat{\sigma}^2 &= s^2 = \frac{1}{n-2} \sum_{i=1}^n \left(y_i - \hat{m} - \hat{k}x_i\right)^2. \end{aligned} \tag{5}$$

Impose a non-informative Jeffrey's prior, with the following Lebesgue density:

$$\pi(k, m, \sigma) = \sigma^{-1}.$$

This means we do not have any pre-existing information about k, m, and σ. Conditioned upon σ, it is uniform (Lebesgue measure) with respect to k, m. This prior is actually *improper*, since the integral of this density with respect to $k, m \in \mathbb{R}$ and $\sigma > 0$ is infinite. Thus π is not a true probability measure. However, the posterior p is a proper probability measure. It is computed using the *likelihood* $L(x, y \mid k, m, \sigma)$, a product of normal densities:

$$\begin{aligned} L(x, y \mid k, m, \sigma) &= \prod_{i=1}^n \varphi(y_i - kx_i - m, \sigma), \\ \varphi(z, \sigma) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-z^2/2\sigma^2\right), \\ p(k, m, \sigma) &\propto L(x, y \mid k, m, \sigma)\pi(k, m, \sigma), \end{aligned} \tag{6}$$

The marginal distribution of σ^2 is the inverse χ^2:

$$p(\sigma^2) \sim \Xi(n-2, s^2), \tag{7}$$

where s^2 is taken from (5). Here, the inverse χ^2 distribution $\Xi(\nu, c)$ is defined as the distribution on $(0, \infty)$ with Lebesgue density

$$\frac{1}{\Gamma(\nu/2)}\left(\frac{\nu}{2}\right)^{\nu/2} c^{\nu} x^{-\nu/2-1} \exp\left(-\frac{\nu c}{2x}\right).$$

Next, the conditional distribution of (k, m) given σ is given by the bivariate normal distribution (with $\hat{k}$ and $\hat{m}$ taken from (5)):

$$\begin{gathered} p(k, m \mid \sigma) \sim \mathcal{N}_2([\hat{k}, \hat{m}], M^{-1}\sigma^2), \\ M := \begin{bmatrix} n & x_1 + \ldots + x_n \\ x_1 + \ldots + x_n & x_1y_1 + \ldots + x_ny_n \end{bmatrix} \end{gathered} \tag{8}$$

6.1. **Acknowledgements.** We are thankful to other students at the University of Nevada, Reno, who took part in the work: Chyna Metz-Bannister (who collected the data), Lissa Callahan (who tried fitting factors using AR(p) and ARMA(p, q), but this did not lead to normally distributed residuals); Akram Reshad for help with coding. The second author thanks Tran Nhat for useful discussion.

References

[1] Robert D. Arnott, Denis B. Chaves, Tzee-man Chow (2017). King of the Mountain: The Shiller P/E and Macroeconomic Conditions. *The Journal of Portfolio Management* **44** (1), 55–68.
[2] Alicia Barrett, Peter Rappoport (2011). Price-Earnings Investing. *JP Morgan Asset Management, Reality in Returns* **November 2011** (1), 1–12.
[3] John Y. Campbell, Robert J. Shiller (1998). Valuation Ratios and the Long-Run Stock Market Outlook. *The Journal of Portfolio Management* **24** (2), 11–26.
[4] Eugene F. Fama, Kenneth R. French (1993). Common Risk Factors in the Returns on Stocks and Bonds. *J. Fin. Econ.* **33** (1), 3–56.
[5] Jane A. Ou, Stephen H. Penman (1989). Accounting Measurement, Price-Earnings Ratio, and the Information Content of Security Prices. *Journal of Accounting Research* **27**, 111–144.
[6] Thomas Philips, Cenk Ural (2016). Uncloaking Campbell and Shiller's CAPE: A Comprehensive Guide to Its Construction and Use. *The Journal of Portfolio Management* **43** (1), 109–125.
[7] Pu Shen (2000). The P/E Ratio and Stock Market Performance. *Federal Reserve Bank of Kansas City Ecomonic Review* **Q** (4), 23–36.
[8] Robert J. Shiller (2015). *Irrational Exuberance.* Princeton University Press, 3^{rd} edition.
[9] Svetlozar T. Rachev, John S. J. Hsu, Biliana S. Bagasheva, Frank J. Fabozzi (2008). *Bayesian Methods in Finance.* Wiley.
[10] Robert J. Shiller (1992). *Market Volatility.* MIT University Press.

University of Nevada in Reno, Department of Mathematics & Statistics
Email address: tarang@nevada.unr.edu, asarantsev@unr.edu