Comprehensive Modeling of Long-Term Stock Market Dynamics

Dr. Andrey Sarantsev Michael Reyes

March 1, 2020

1. Introduction

This research is devoted to comprehensive modeling of long-term stock market dynamics in the United States of America. We use as a benchmark the Standard & Poor 500 Index, consisting of 500 large American companies, created in 1957, and its predecessor, the Standard & Poor 90, created in 1926. We will refer to it as the S&P Composite Index. We are particularly interested in the real (inflation-adjusted) earnings growth, which is correlated with the US economic growth.

We model its returns as linear regressions upon the following factors:

1. Short-term (3 month) Treasury bill rate, which is closely correlated with the federal funds rate, controlled by the Federal Reserve Board and measures yield of Risk-Free investment.

2. The spread between 3 month and 10 year Treasury rate which measures long-term investor optimism.

3. A more complicated quantity which measures how overheated the stock market is; this is called "Temperature".

We measure stock market returns by equity premium, which is defined as total returns (including both price appreciation and reinvested dividends) minus Risk-Free annual return (composed by reinvesting in 3-month Treasury bills for each of the 4 quarters). This measures excess return provided by risky equity investments.

Earnings are net income after all taxes and expenses. We sum the annual earnings of all companies in the S&P Composite Index and normalize them to get earnings per share of the S&P Composite Index. Earnings are fundamental for companies: They represent a part of income which companies use pay their dividends, reinvest in business enterprises, or pay debt.

Assume that on average, equity premium exceeds real earnings growth by 2% per year. However, in recent years, this difference was 5%. Then we can claim that the stock market is detached from fundamentals and is in a bubble. We can expect lower future returns than historical returns. To capture this insight, we create a new variable called "temperature". A complete description is in Section 2.

The main question is whether the stock market is now in a bubble. We can present the following arguments, for or against:

(A) Temperature of the market is currently high, thus stock market returns have run away from earnings growth. This implies that we are in a bubble.

(B) Short-term interest rates are historically very low now, the money is cheap, and this fuels the stock market. It is rational for investors to avoid bonds, since they yield so little, and to buy stocks, even if they are expensive relative to their earnings.

(C) The spread between 10-year and 3-month Treasury rates is close to zero, which implies a flat yield curve. This has historically preceded recessions and bear stock markets.

This article is an attempt to formalize and statistically evaluate these arguments.

We note that often researchers use simpler measures of market valuations; various versions of price-earnings-ratio, defined as current index price divided by last year's earnings (or average of last k year's earnings). If this indicator is well above its long-terms average, then the market is considered overvalued. However, companies sometimes deliver returns to their shareholders not only by paying dividends, but by stock buybacks, which tend to raise the stock price. This method of cash distribution became increasingly popular recently. This implies that high price-earnings ratio might be the indicator of an increase in buybacks, not a bubble. Our valuation measures must consider total return, both from price appreciation and dividends.

One major difficulty in financial research is that fluctuations of the index, returns, and financial indicators are not modeled by the normal law. Even in 1-year horizons, regression residuals are not normal. This does not allow us to use standard statistical tests of significance. If we apply Bayesian inference, we cannot use convenient prepackaged formulas for the posterior. We bypass this problem by considering 3-year intervals. Our computations show that 3-year cumulative equity premium has normal iid. residuals for regression upon the three factors stated above.

We model real earnings growth (from which temperature is derived), and the spread between 10-year and 3-month Treasury rates as simple autoregressions, again with time step 3 years. These autoregressions also have normal residuals. We decided not to model 3-month interest rate, as an autoregression, because it is controlled by the Federal Reserve. Instead, we run the simulations under some reasonable assumptions about future actions of the Fed. For example, that in 3-year steps this rate would be iid normal with mean 1.5% and an appropriate standard deviation.

After fitting the model, we apply Bayesian inference. We impose a Jeffrey's non-informative prior on regression and autoregression parameters, as well as a certain normal prior upon short-term interest rate. This allows us to use an explicit posterior, and simulate parameters. We take data from 1935 to 2018 which contains 26 3-year intervals. Thus, we have 25 data points in our regressions. To account for uncertainty in estimating, we shall use Bayesian inference.

The article is organized as follows. In Section 2, we introduce the data, explain its sources, and define notation. In Section 3, we describe our regression and autoregression models. In Section 4, we provide point estimates for parameters and confidence intervals for regression coefficients. In Section 5, we impose Bayesian prior upon parameters, as described above, and infer the posterior. We use those results in Section 6 to simulate the long-term behavior of the stock market over 30 years (10 3-year time steps). Section 7 summaries the article and sketches directions for possible future research.

2. Data and Notation

In progress

3. Model Description

In this section, we will describe our autoregressions and regression models as well as discuss the tests regarding the sufficiency of the models.

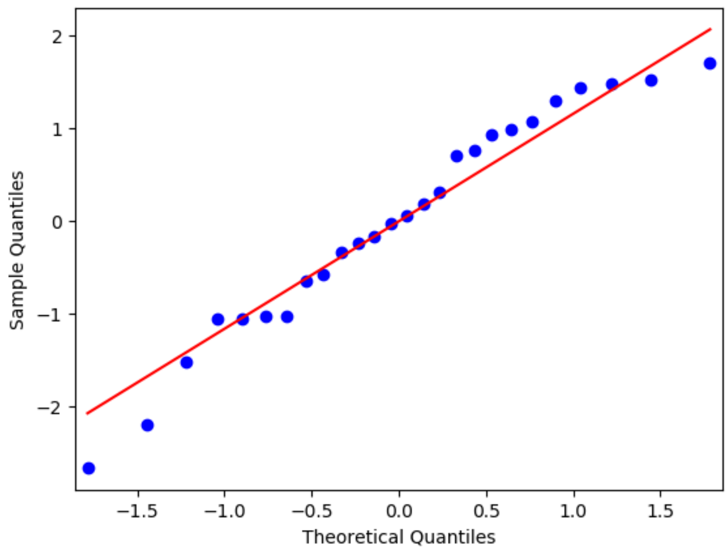
3.1 Spread

The Autoregression model was used on the Spread $S(t_n)$ for 3-year time steps between 1935-2018.

$$S\left(t_n\right) = \beta_0 + \beta_1 \cdot S\left(t_{n-1}\right) + \epsilon_s(t_n) \tag{1}$$

where $\epsilon_s(t_n) \sim \mathcal{N}(0,\, \hat{\sigma_s^2} = 1.452020)$ and the MSE = 1.204998

The following QQ Plot for Spread shows that our model is linearly related.



We used the Sharpiro-Wilk and Jarque-Bera test to test for normality. Since, the Shapiro-Wilk p = 0.353956 and Jarque-Bera p = 0.569580, the data comes from a normal distribution.

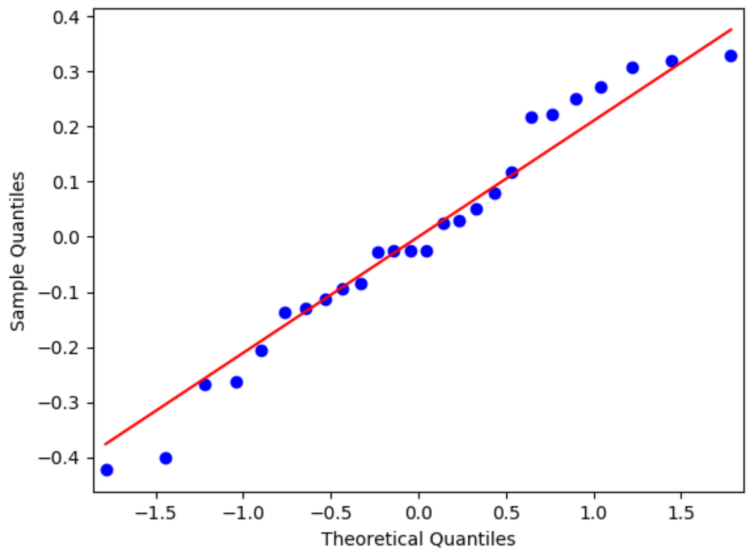
3.2 Real Earnings Growth

The Autoregression model was used on the Real Earning Growth $R(t_n)$ for 3-year time steps between 1935-2018.

$$R\left(t_n\right) = \beta_2 + \beta_3 \cdot R\left(t_{n-1}\right) + \epsilon_r(t_n) \tag{2}$$

where $\epsilon_r(t_n) \sim \mathcal{N}(0, \hat{\sigma_r^2} = 0.048026)$ and the MSE = 0.219147

The following QQ Plot for Real Earnings Growth shows that our model is linearly related.



We used the Sharpiro-Wilk and Jarque-Bera test to test for normality. Since, the Shapiro-Wilk p = 0.393727 and Jarque-Bera p = 0.711620, the data comes from a normal distribution.

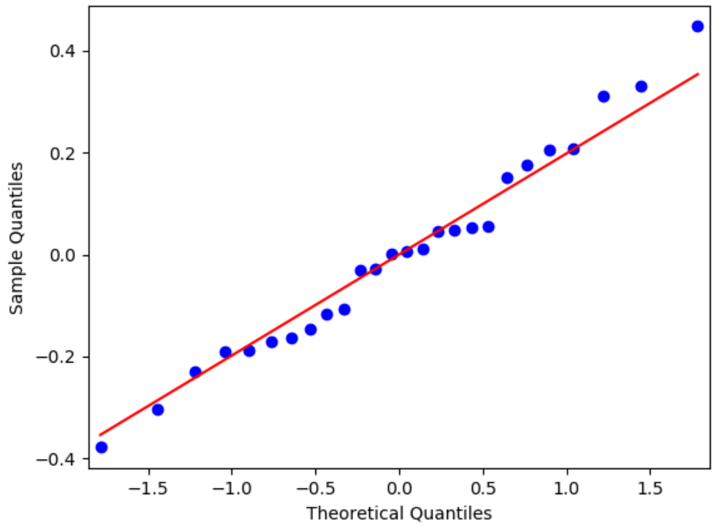
3.3 Cumulative Earnings Premium

A Multivariate regression model was used on the Premium $C(t_n)$ for 3-year time steps between 1935-2018.

$$C\left(t_n\right) = \beta_4 + \beta_5 \cdot H\left(t_{n-1}\right) + \beta_6 \cdot t_{n-1} + \beta_7 \cdot S\left(t_{n-1}\right) + \epsilon_c(t_n) \tag{3}$$

where $H = \sum_{t=0}^{t_{n-1}}$ Deviation(t), Deviation $(t) = C(t) - R(t)$, $\epsilon_c(t_n) \sim \mathcal{N}(0,\, \hat{\sigma_p^2} = 0.046361)$, and the MSE = 0.215316

The following QQ Plot for Cumulative Earnings Premium shows that our model is linearly related.



We used the Sharpiro-Wilk and Jarque-Bera test to test for normality. Since, the Shapiro-Wilk p = 0.884902 and Jarque-Bera p = 0.749887, the data comes from a normal distribution.

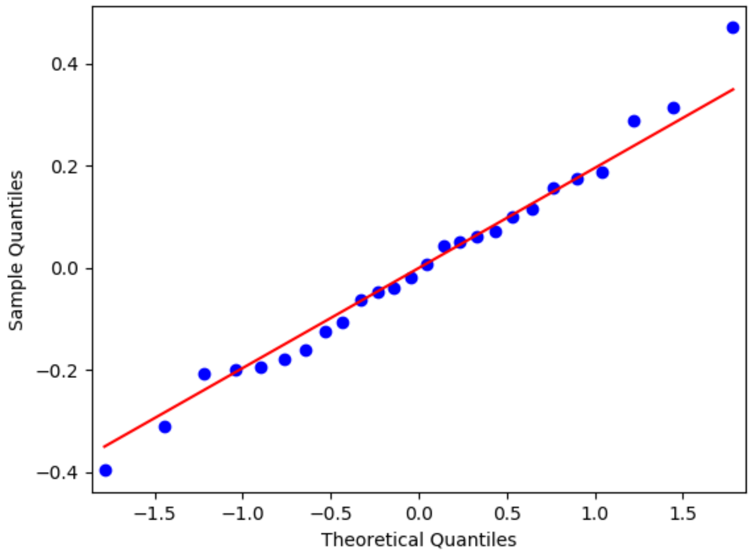
3.4 Cumulative Earnings Premium with Short-Term Interest Rate

A Multivariate regression model including Short-Term Interest Rate $I(t_n)$ was used on the Cumulative Earnings Premium $C(t_n)$ for 3-year time steps between 1935-2018.

$$C\left(t_n\right) = \beta_8 + \beta_9 \cdot H\left(t_{n-1}\right) + \beta_{10} \cdot t_{n-1} + \beta_{11} \cdot S\left(t_{n-1}\right) + \beta_{12} \cdot I(t_n) + \epsilon_c(t_n) \quad (4)$$

where $\epsilon_c(t_n) \sim \mathcal{N}(0,\, \hat{\sigma_c^2} = 0.048569)$ and the MSE = 0.217926

The following QQ Plot for Cumulative Earnings Premium with the Short-Term interest rate shows that our model is linearly related.



We used the Sharpiro-Wilk and Jarque-Bera test to test for normality. Since, the Shapiro-Wilk p = 0.993658 and Jarque-Bera p = 0.864589, the data comes from a normal distribution.

4. Point Estimates and Confidence Intervals

In this section, we state the point estimates and the confidence intervals of the coefficients for each model.

4.1 Spread

The following table details the analysis of the coefficients of the Spread autoregression model:

Table 1: Spread Autoregerssion Model Coefficients Summary

Coefficient	Estimated Value	Standard Error	t-value	p-value	[0.025	0.975]
β_0	1.8927	0.351	5.385	0.000	1.167	2.618
β_1	-0.3743	0.187	-2.001	0.057	-0.760	0.012

4.2 Real Earnings Growth

The following table details the analysis of the coefficients of the Real Earnings Growth auto-regression model:

Table 2: Real Earnings Growth Autoregerssion Model Coefficient Results

Coefficient	Estimated Value	Standard Error	t-value	p-value	[0.025	0.975]
β_2	0.1204	0.044	2.719	0.012	0.029	0.212
β_3	-0.6043	0.158	-3.821	0.001	-0.931	-0.278

4.3 Cumulative Earnings Premium

The following table details the analysis of the coefficients of the Cumulative Earnings Premium multivariable regression model:

Table 3: Cumulative Earnings Premium Model Results

Coefficient	Estimated Value	Standard Error	t-value	p-value	[0.025	0.975]
β_4	0.3850	0.118	3.253	0.004	0.140	0.631
β_5	-0.2387	0.115	-2.076	0.050	-0.477	0.000
β_6	0.0212	0.015	1.457	0.159	-0.009	0.051
β_7	0.0301	0.034	0.893	0.382	-0.040	0.100

4.4 Cumulative Earnings Premium with Short-Term Interest Rate

The following table details the analysis of the coefficients of the Cumulative Earnings Premium with Short-Term Interest Rate multivariable regression model:

Table 4: Cumulative Earnings Premium Model Results

Coefficient	Estimated Value	Standard Error	t-value	p-value	[0.025	0.975]
β_8	0.4371	0.142	3.087	0.006	0.143	0.731
β_9	-0.2441	0.117	-2.092	0.049	-0.487	-0.001
β_{10}	0.0236	0.015	1.560	0.134	-0.008	0.055
β_{11}	0.0130	0.042	0.309	0.761	-0.075	0.101
β_{12}	-0.0126	0.018	-0.690	0.498	-0.050	0.025

5. Bayesian Inference

We have 25 data points in our regressions. To account for uncertainty in estimating, we shall use Bayesian inference.

5.1 Spread

Given the autoregression model for Spread (1) and $v = \sigma_s^2$, for β_0, β_1, and v put a non-informative Jeffrey's prior:

$$\pi(v) = v^{-1},\ \pi(\beta_0, \beta_1 | v) = 1$$

The posterior will be:

(A) v has inverse χ^2 law, that is $\dfrac{(n-2)s^2}{v} \sim \chi^2_{n-2}$, where $s^2 = MSE^2$

(B) $\beta_0, \beta_1 | v \sim \mathcal{N}_2\left(\left[\hat{\beta}_0, \hat{\beta}_1\right], M^{-1}v\right)$, where M is the Gram matrix for $(\vec{i}, S(\vec{t_{n-1}}))$.

Thus, $\beta_0, \beta_1 | v \sim \mathcal{N}_2\left([1.8927, -0.3743], M^{-1}v\right)$ with $M = \begin{bmatrix} 26 & 36.1625 \\ 36.1625 & 91.78605 \end{bmatrix}$.

5.2 Real Earnings Growth

Given the autoregression model for Real Earnings Growth (2) and $v = \sigma_r^2$, for β_2, β_3, and v put a non-informative Jeffrey's prior:

$$\pi(v) = v^{-1},\ \pi(\beta_2, \beta_3 | v) = 1$$

The posterior will be:

(A) v has inverse χ^2 law, that is $\dfrac{(n-2)s^2}{v} \sim \chi^2_{n-2}$, where $s^2 = MSE^2$

(B) $\beta_2, \beta_3 | v \sim \mathcal{N}_2\left(\left[\hat{\beta}_2, \hat{\beta}_3\right], M^{-1}v\right)$, where M is the Gram matrix for $(\vec{i}, R(\vec{t_{n-1}}))$.

Thus, $\beta_2, \beta_3 | v \sim \mathcal{N}_2\left([0.1204, -0.6043], M^{-1}v\right)$ with $M = \begin{bmatrix} 26 & 1.7573 \\ 1.7573 & 2.0387 \end{bmatrix}$.

5.3 Cumulative Earnings Premium

Given the multivariable regression model for Cumulative Earnings Premium (3) and $v = \sigma_c^2$, for β_4, β_5, β_6, β_7 and v put a non-informative Jeffrey's prior:

$$\pi(v) = v^{-1},\ \pi(\beta_4, \beta_5, \beta_6, \beta_7 | v) = 1$$

The posterior will be:

(A) v has inverse χ^2 law, that is $\dfrac{(n-4)s^2}{v} \sim \chi^2_{n-4}$, where $s^2 = MSE^2$

(B) $\beta_4, \beta_5, \beta_6, \beta_7 | v \sim \mathcal{N}_4 \left(\left[\hat{\beta_4}, \hat{\beta_5}, \hat{\beta_6}, \hat{\beta_7} \right], M^{-1} v \right)$, where M is the Gram matrix for $(\vec{i}, H(\vec{t_{n-1}}), \vec{t_{n-1}}, S(\vec{t_{n-1}}))$.

Thus, $\beta_4, \beta_5, \beta_6, \beta_7 | v \sim \mathcal{N}_4 \left([0.3850, -0.2387, 0.0212, 0.0301], M^{-1} v \right)$

$$\text{with } M = \begin{bmatrix} 26 & 53.3062 & 325.0 & 36.1625 \\ 53.3062 & 132.5983 & 836.2912 & 74.2474 \\ 325.0 & 836.2912 & 5525 & 465.2792 \\ 36.1625 & 74.2475 & 465.2792 & 91.7861 \end{bmatrix}.$$

5.4 Cumulative Earnings Premium with Short-Term Interest Rate

Given the multivariable regression model for Cumulative Earnings Premium with Short-Term Interest Rate (4) and $v = \sigma_c^2$, for β_8, β_9, β_{10}, β_{11}, β_{12} and v put a non-informative Jeffrey's prior:

$$\pi(v) = v^{-1}, \ \pi(\beta_8, \beta_9, \beta_{10}, \beta_{11}, \beta_{12} | v) = 1$$

The posterior will be:

(A) v has inverse χ^2 law, that is $\dfrac{(n-5)s^2}{v} \sim \chi^2_{n-5}$, where $s^2 = MSE^2$

(B) $\beta_8, \beta_9, \beta_{10}, \beta_{11}, \beta_{12} | v \sim \mathcal{N}_5 \left(\left[\hat{\beta_8}, \hat{\beta_9}, \hat{\beta_{10}}, \hat{\beta_{11}}, \hat{\beta_{12}} \right], M^{-1} v \right)$, where M is the Gram matrix for $(\vec{i}, H(\vec{t_{n-1}}), \vec{t_{n-1}}, S(\vec{t_{n-1}}), I(\vec{t_{n-1}}))$.

Thus, $\beta_4, \beta_5, \beta_6, \beta_7 | v \sim \mathcal{N}_5 \left([0.4371, -0.2441, 0.0236, 0.0130, -0.0126], M^{-1} v \right)$

$$\text{with } M = \begin{bmatrix} 26 & 53.3062 & 325.0 & 36.1625 & 98.39 \\ 53.3062 & 132.5983 & 836.2912 & 74.2474 & 224.3771 \\ 325.0 & 836.2912 & 5525 & 465.2792 & 1420.9299 \\ 36.1625 & 74.2475 & 465.2792 & 91.7861 & 82.9035 \\ 98.39 & 224.3771 & 420.9299 & 82.9035 & 616.1005 \end{bmatrix}$$

6. Simulation Results

To simulate the random errors, we will use the multivariate normal. However, we need to find the covariance matrix between the three models (1,2,3) We can do so using the correlation matrix and the formula $Cov(A, B) = Cor(A, B) \cdot \sigma_A \cdot \sigma_B$

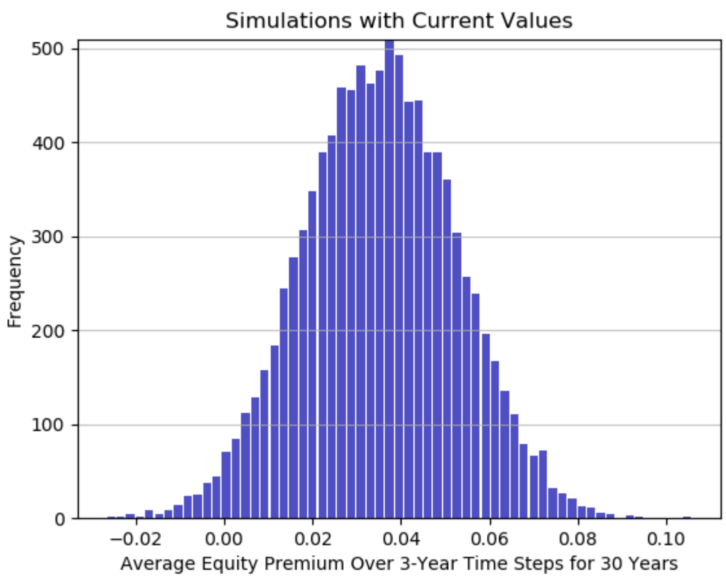
$$\text{Correlation Matrix} = \begin{array}{c} \\ Spread \\ REG \\ CEP \end{array} \begin{array}{c} \begin{array}{ccc} Spread & REG & CEP \end{array} \\ \begin{bmatrix} 1 & 0.346147 & 0.417229 \\ 0.346147 & 1 & 0.545088 \\ 0.417229 & 0.545088 & 1 \end{bmatrix} \end{array}$$

$$\text{Covariance Matrix} = \begin{array}{c c} & \begin{array}{ccc} Spread & REG & CEP \end{array} \\ \begin{array}{c} Spread \\ REG \\ CEP \end{array} & \begin{bmatrix} 1.452020 & 0.091408 & 0.108252 \\ 0.091408 & 0.048026 & 0.025721 \\ 0.108252 & 0.025721 & 0.046361 \end{bmatrix} \end{array}$$

We can now simulate the models. We will simulate the next 3-year time steps over 30 years 10000 times using the following starting points: Long Terms Averages and Current Values of Spread, Real Earnings Growth, and Heat.

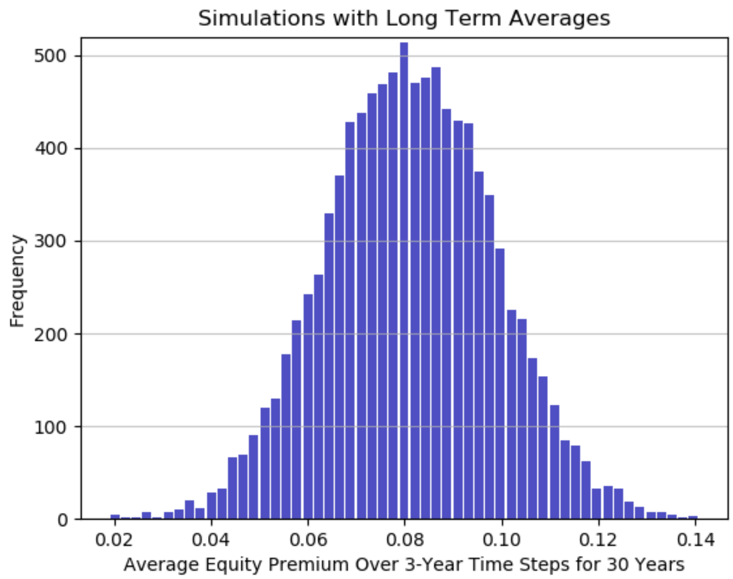
6.1 Cumulative Earning Premium Simulations using Covariance Matrix

We ran simulations with current values, below are the results:



The Average Equity Premium Over 3-year time steps annualized for all the simulations was: 0.035311. The standard deviation of Average Equity Premium over 3-year time steps annualized for all the simulations was 0.017266. We found that the 90% Value at Risk was 0.133742 and the 95% Value at Risk was 0.006743.

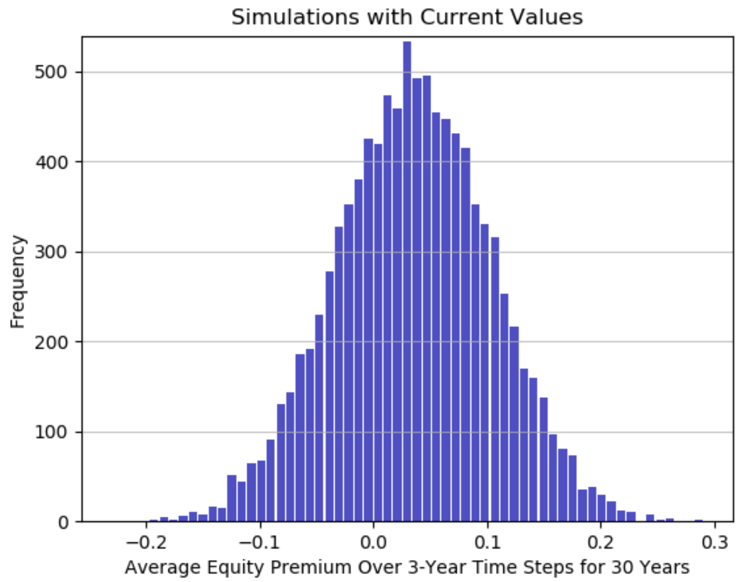
We also ran simulations with Long Term averages, below are the results:



The Average Equity Premium over 3-year time steps annualized for all the simulations was: 0.081234. The standard deviation of Average Equity Premium Over 3-Year Time Steps annualized for all the simulations was 0.017389. We found that the 90% Value at Risk was 0.059105 and the 95% Value at Risk was 0.052681.

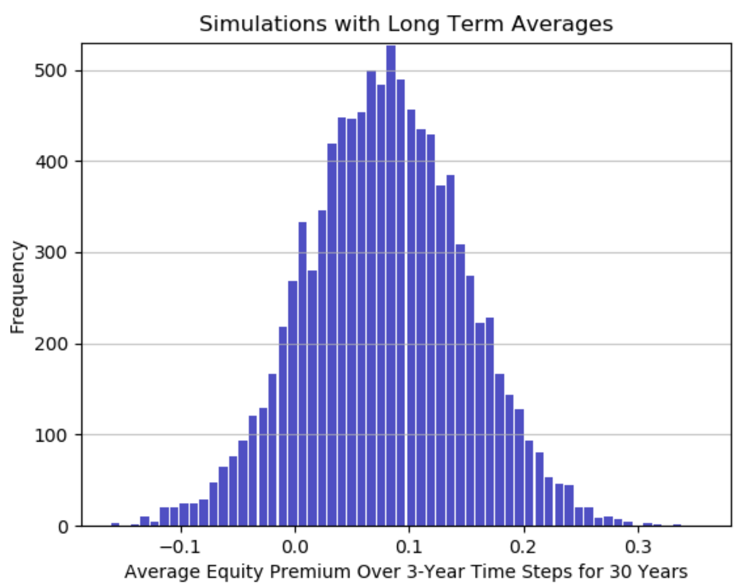
6.2 Cumulative Earning Premium Simulations using Identity Matrix

We also wanted to simulate the models with the covariance matrix being the identity matrix for the random errors. Those simulations gave the following results using current values:



The Average Equity Premium Over 3-year time steps annualized for all the simulations was: 0.035742. The standard deviation of Average Equity Premium over 3-year time steps annualized for all the simulations was 0.068646. We found that the 90% Value at Risk was -0.054352 and the 95% Value at Risk was -0.079388.

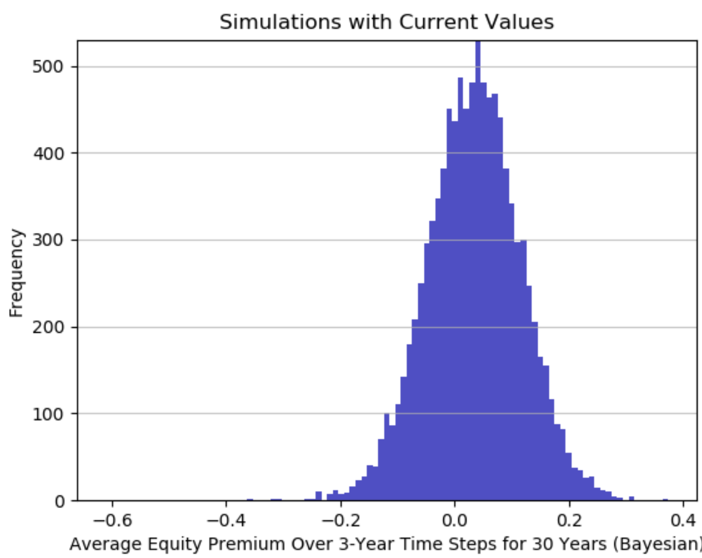
We also ran simulations with Long Term averages, below are the results:



The Average Equity Premium Over 3-year time steps annualized for all the simulations was: 0.079417. The standard deviation of Average Equity Premium over 3-year time steps annualized for all the simulations was 0.069482. We found that the 90% Value at Risk was -0.007992 and the 95% Value at Risk was -0.032936.

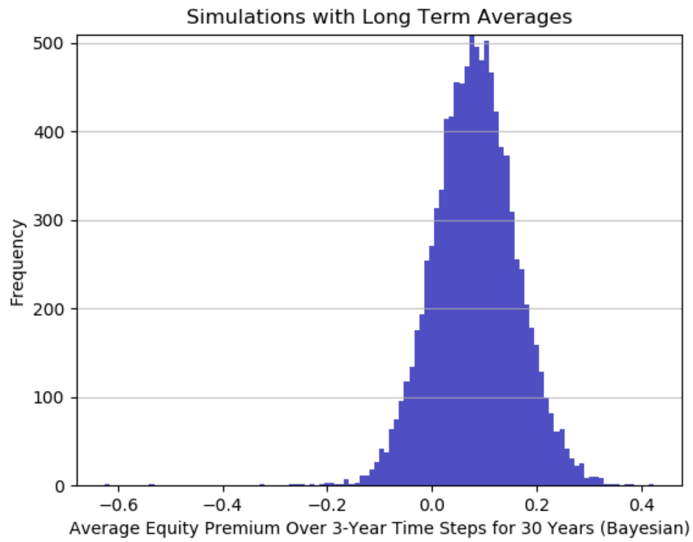
6.3 Cumulative Earning Premium Simulations using Bayesian Inference

We have 25 data points in our regressions. To account for uncertainty in estimating, we shall use Bayesian inference for models 1,2, and 3. We use the identity matrix for the covariance matrix for simulating the random errors. Those simulations gave the following results using current values:



The Average Equity Premium Over 3-year time steps annualized for all the simulations was: 0.03419. The standard deviation of Average Equity Premium over 3-year time steps annualized for all the simulations was 0.08256. We found that the 90% Value at Risk was -0.06876 and the 95% Value at Risk was -0.10064.

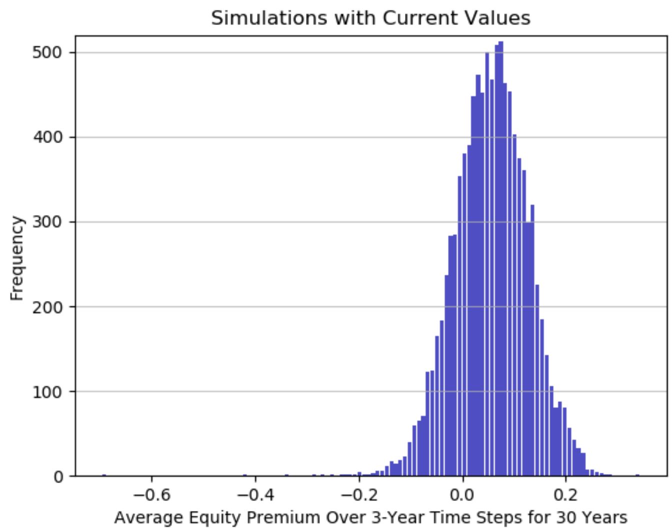
We also ran simulations with Long Term averages, below are the results:



The Average Equity Premium Over 3-year time steps annualized for all the simulations was: 0.0827. The standard deviation of Average Equity Premium over 3-year time steps annualized for all the simulations was 0.079050. We found that the 90% Value at Risk was -0.01598 and the 95% Value at Risk was -0.04569.

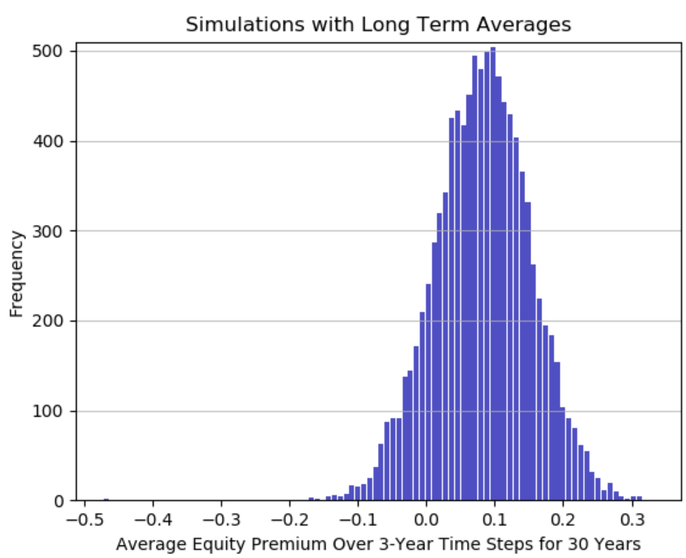
6.4 Cumulative Earning Premium with Short-Term Interest Rate Simulations using Identity Matrix

We also wanted to simulate the models (1,2,4). We use the identity matrix for the covariance matrix for simulating the random errors. Those simulations gave the following results using current values:



The Average Equity Premium Over 3-year time steps annualized for all the simulations was: 0.05631. The standard deviation of Average Equity Premium over 3-year time steps annualized for all the simulations was 0.07125. We found that the 90% Value at Risk was -0.0161 and the 95% Value at Risk was -0.0315.

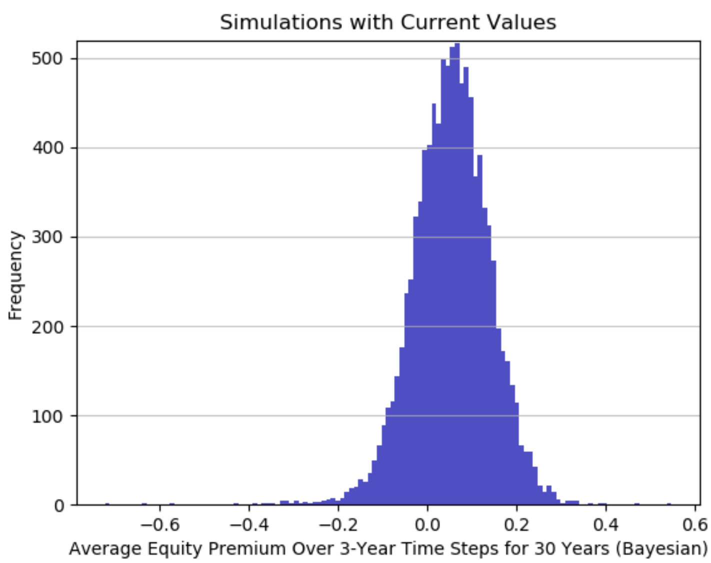
We also ran simulations with Long Term averages, below are the results:



The Average Equity Premium Over 3-year time steps annualized for all the simulations was: 0.08217. The standard deviation of Average Equity Premium over 3-year time steps annualized for all the simulations was 0.07046. We found that the 90% Value at Risk was 0.06759 and the 95% Value at Risk was 0.07208.

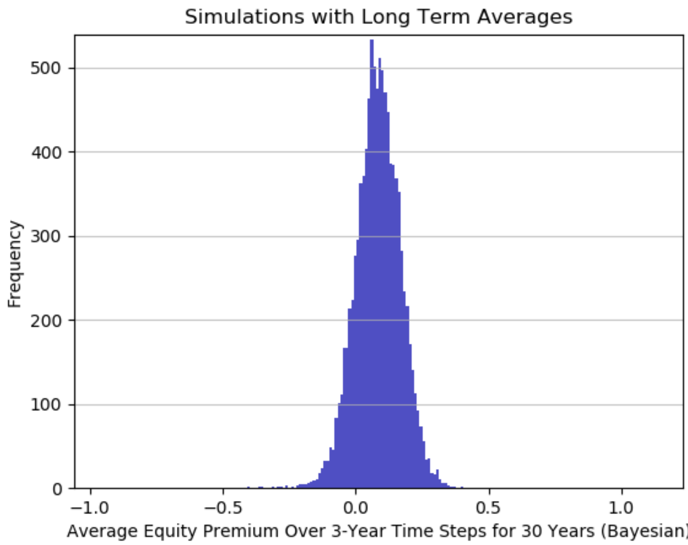
6.5 Cumulative Earning Premium with Short-Term Interest Rate Simulations using Bayesian Inference

We have 25 data points in our regressions. To account for uncertainty in estimating, we shall use Bayesian inference for models 1, 2, and 4. We use the identity matrix for the covariance matrix for simulating the random errors. Those simulations gave the following results using current values:



The Average Equity Premium Over 3-year time steps annualized for all the simulations was: 0.05511. The standard deviation of Average Equity Premium over 3-year time steps annualized for all the simulations was 0.08682. We found that the 90% Value at Risk was -0.04888 and the 95% Value at Risk was -0.08438.

We also ran simulations with Long Term averages, below are the results:



The Average Equity Premium Over 3-year time steps annualized for all the simulations was: 0.08426. The standard deviation of Average Equity Premium over 3-year time steps annualized for all the simulations was 0.08936. We found that the 90% Value at Risk was -0.02237 and the 95% Value at Risk was -0.05454.

7. Conclusions and Future Research

In progress