Long-Term Bayesian Modeling of the American Stock Market

University of Nevada in Reno

Taran Grove, Andrey Sarantsev

January 17, 2020. JMM Denver

Stock Market

We take as benchmark the Standard & Poor 500, the index composed of 500 large publicly traded American companies. It starts from 1926, when the predecessor consisting of 90 stocks was created. We take annual data 1935–2018, since in 1934 Securities & Exchange Commission was created.

This index is capitalization-weighted (in proportion to market capitalization of a company, which is total value of all its shares). The largest companies as of this writing are `AAPL` (Apple) and `MSFT` (Microsoft), with capitalizations 1.39T$ and 1.25T$.

Robert Shiller: Earnings, dividends for S&P, and inflation.
Yahoo Finance: S&P beginning of year index price.

Total Return and Equity Premium

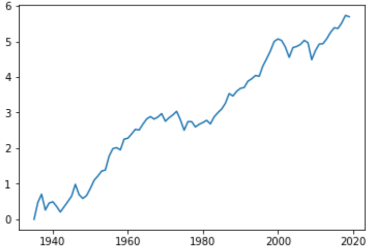
Take a year t. Let $S(t)$ = price at start of year t, $D(t)$ = dividend per share paid in year t. Total return includes reinvested dividends.

Quarterly return: $Q(t) = \ln \frac{S(t+1)+D(t)}{S(t)}$.

Real return: $R(t) = Q(t) - I(t)$, where $I(t)$ = inflation in year t.

Logarithm of Cumulative Wealth

If we invested 1$ in S&P on January 1, 1935, and reinvest dividends at the start of every year, this plot show the logarithm of wealth, adjusted for inflation.



On average, stocks grow faster than inflation, which shows in upward trend. However, there are dips, including recent crashes in 2000–2002 (dotcom bubble) and 2007–2009 (housing bubble).

Net Income

Every stock has net income, or earnings, reported every quarter. Example: 2019Q3 Microsoft earnings were $10.678B$\$.

$I(t)$ = earnings per share of S&P, summed over all component stocks during year t.

Earnings yield: $E(t) = I(t)/S(t)$. Highly volatile.

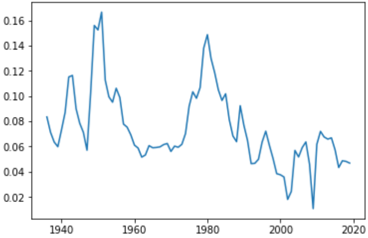


Figure: Graph of $E(t)$ vs t.

Background

Previous earnings are low compared to level of S&P 500 $\Rightarrow$ future return is low: Robert Shiller, Nobel Prize in Economics 2013.

See his book Irrational Exuberance (he uses 10-year income averages).

Example: Earnings yield was very low in 2000 (dotcom bubble), 1929 (eve of the Great Depression), and right now.

Problem: Too volatile earnings yield.

Solution: Average Variables

Compute average income over the last k years. Leads to smoothed version of earnings yield: $E_k(t)$. Correlation between $E_k(t-1)$ and $R(t)$ is strongest for $k = 7$ and equal to 23%.

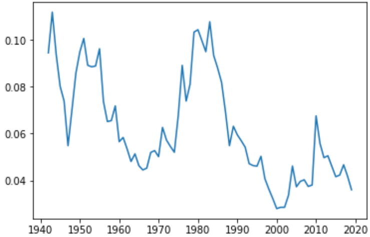


Figure: Graph of $E_7(t)$ vs t.

Regression Model

Model $E_7(t)$ as $\mathrm{AR}(1)$ and $P(t)$ vs $E_7(t-1)$:

$$E_7(t) = c + sE_7(t-1) + \delta(t), \quad \delta(t) \sim \mathcal{N}(0, \sigma_\delta^2),$$
$$R(t) = \alpha + \beta E_7(t-1) + \varepsilon(t), \quad \varepsilon(t) \sim \mathcal{N}(0, \sigma_\varepsilon^2),$$

where all residuals $\delta(t)$ and $\varepsilon(t)$ are independent.

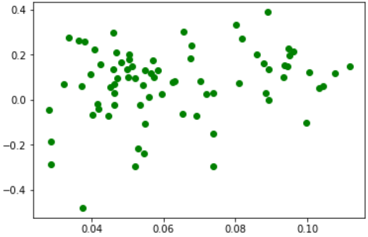


Figure: $(E_7(t), R(t)), t = 7, \ldots, 84.$

Bayesian Regression

We use Bayesian framework to account for errors in estimation.

Bayes' formula: $\mathbf{P}(F_1 \mid A) = \frac{\mathbf{P}(A|F_1)\mathbf{P}(F_1)}{\mathbf{P}(A)}$.

$\mathbf{P}(F_1) =$ Prior, $\mathbf{P}(F_1 \mid A) =$ Posterior, $\mathbf{P}(A \mid F_1) =$ Likelihood.

Bayesian Regression

Consider a simple linear regression:

$$y_i = \alpha + \beta x_i + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2) \quad \text{i.i.d.}$$

Continuous analogy of the above Bayes' theorem:

$$p(\alpha, \beta, \sigma \mid y) \propto L(y \mid \alpha, \beta, \sigma)\pi(\alpha, \beta, \sigma).$$

π = Prior, p = Posterior, L = Likelihood.

Jeffrey's non-informative prior: $\pi(\alpha, \beta, \sigma) \propto \sigma^{-1}$. Although this π is an infinite measure, the posterior is finite and explicit.

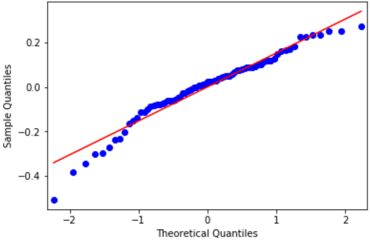
Regression Results

$$R(t) = -0.031 + 1.63E_7(t-1) + 0.153\varepsilon(t),$$
$$E_7(t) = 0.0056 + 0.898E_7(t-1) + 0.0094\delta(t),$$
$$\varepsilon(t) \sim \mathcal{N}(0,1), \quad \delta(t) \sim \mathcal{N}(0,1), \quad \text{i.i.d.}$$

But the QQ plots for residuals for each regression and normality tests (Shapiro-Wilk, Jarque-Bera) show that residuals are not i.i.d. normal.

QQ plots

This is a QQ plot for residuals of $P(t)$ versus $E_7(t-1)$.



Still, arguably it is fine to use normal model for non-normal residuals, since we are interested in long-run simulation, which implies sum of independent variables, which by CLT is normal.

Monte Carlo Simulations

Step 1. Simulate regression parameters from Bayesian posterior.

Step 2. Simulate 10000 times: 20-year process.

Step 3. Compute average return over 20 years for each simulation.

Step 4. From 10000 numbers, find mean, standard deviation, and 5% percentile.

Simulation Results

Start simulations from different values of earnings yield $E_7(t)$.

Current 3.6%: 4.4% mean, 5.2% stdev, -4.4% 5% level.
Average 6.2%: 6.4% mean, 5.0% stdev, -1.8% 5% level.

Currently, earnings yield is low compared to long-term average, and the stock market is overpriced. Thus we cannot count on the same results as in the past. The stock market will likeiy yield smaller but positive results.

3-Year Earnings

If we use $E_3(t)$ instead of $E_7(t)$, the correlation between $E_3(t-1)$ and $R(t)$ is 19%, not far from correlation 22% between $E_7(t-1)$ and $R(t)$. Repeat:

Current 4.0%: 5.0% mean, 5.0% stdev, -3.3% 5% level.
Average 7.0%: 6.6% mean, 4.8% stdev, -1.1% 5% level.

Same conclusion: Currently, earnings yield is low compared to long-term average, and the stock market is overpriced. Thus we cannot count on the same results as in the past. The stock market will likeiy yield smaller but positive results.

Historical Returns 1935–2018

During these 84 years, statistics of S&P actual returns?

Mean 6.8%, stdev 17%, 5% level -28%.

Future returns are likely smaller than historical returns.