Long-Term Bayesian Modeling of the American Stock Market

University of Nevada in Reno

Andrey Sarantsev

January 27, 2020

University of Montana in Missoula

Stock Market

We take as benchmark the Standard & Poor 500, the index composed of 500 large publicly traded American companies. It 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 $1.3T\$$ and $1.2T\$$.

`VFINX`: Vanguard 500 Index Fund, first index fund in the world!

Let $S(t) =$ price at end of quarter t, $D(t) =$ dividend per share paid during quarter t.

Total return includes reinvested dividends.

Total Return and Equity Premium

Take a quarter t.

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

Risk-free rate: 3-month Treasury bills $r(t)$ at end of quarter t.

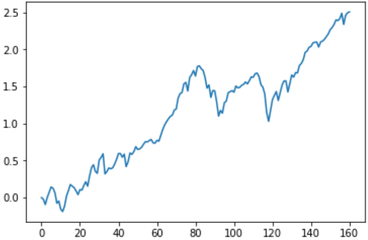
Risk-free return: $R(t) = \ln\left(1 + \frac{r(t-1)}{4}\right)$.

Equity premium (excess return): $P(t) = Q(t) - P(t)$.

160 quarters: 1979Q3–2019Q3.

Logarithm of Cumulative Equity Premium

This is cumulative excess return vs quarter. If we invested 1$ in S&P 500 on June 30, 1979, how much will it grow compared to investment in 3-month Treasury bills.



On average, stocks grow faster than Treasury bills. However, there are two dips, corresponding to market 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 $10.678B$\$.

Usually, we sum all earnings to find net income for S&P 500. But profit for one company is not related to loss of another company. Let $I(t)$ be the average of only positive incomes.

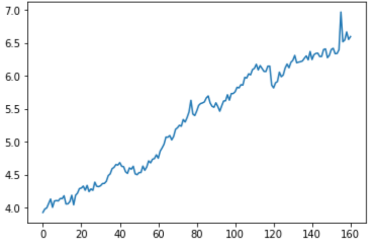


Figure: $(t, \ln I(t))$, $t = 0, \dots, 160$.

Earnings Yield

$C(t)$ = Average market capitalization of a company in S&P 500.

Earnings yield: $I(t)/C(t+1)$. We use $C(t+1)$ because net income for quarter is not known at the end of this quarter.

LEY (log earnings yield): $E(t) = \ln(I(t)/C(t+1))$. We can do this because $I(t) > 0$: we count only positive net income.

Data

Siblis Research: 1979Q1–2019Q1 earnings and market cap quarterly data for S&P 500 companies, including delisted stocks. For annual sum of all incomes (positive and negative), the data is available from Robert Shiller's financial data web page.

YCharts: Prices and dividends for `VFINX` for 1979Q1–2019Q4. (Starting from 1980, it is available on Yahoo Finance.)

Federal Reserve Economic Data: 3-Month Treasury rate.

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 trailing earnings including negative net incomes).

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

Problem: Regress $P(t)$ upon $E(t-1)$, residuals are not normal.

Solution: Average Variables

Compute average of $E(kt-1), \ldots, E(kt-k)$ and get $E'(t)$. Compute $P'(t) = P(kt) + \ldots + P(kt+k-1)$. We now have $160/k$ data points.

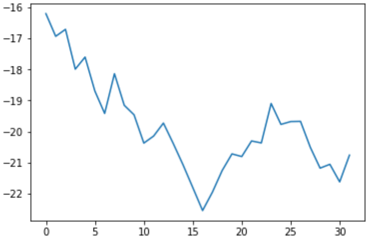


Figure: Graph of $E'(t)$ vs t, for $k = 5$.

The dip in the middle: 2000 dotcom bubble. The peak later: the Great Recession, when prices collapsed.

Regression Model

Model $E'(t)$ as $\mathrm{AR}(1)$ and P' vs E':

$$E'(t) = c + sE'(t-1) + \delta(t),$$
$$P'(t) = \alpha + \beta E'(t-1) + \varepsilon(t).$$

Goal: each residuals $\delta(t), \varepsilon(t)$ i.i.d. normal. Minimal $k = 5$ for this (from QQ plots, Shapiro-Wilk and Jarque-Bera normality tests).

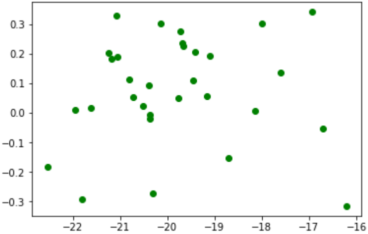


Figure: $(E'(t), P'(t)), t = 1, \ldots, 32$, for $k = 5$.

Normality of Residuals

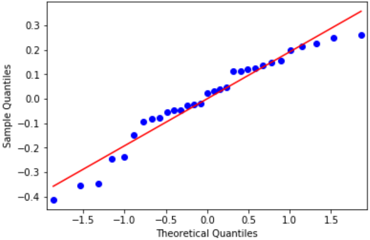


Figure: QQ plot for residuals of regression of $P'(t)$ vs $E'(t-1)$

Jarque-Bera normality test: $p = 0.34$.
Shapiro-Wilk normality test: $p = 0.10$.

Bayesian Regression

Since we have only $32 = 160/5$ observations for $E'(t)$ and $P'(t)$, 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

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

Continuous analogy:

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

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

$$
\begin{aligned}
P'(t) &= 0.191 + 0.0058E'(t-1) + \varepsilon'(t), \\
E'(t) &= -3.84 + 0.814E'(t-1) + \delta'(t), \\
&(\varepsilon'(t), \delta'(t)) \sim \mathcal{N}_2(0, \Sigma) \quad \text{i.i.d.}
\end{aligned}
$$

with covariance matrix of residuals

$$
\Sigma = \begin{bmatrix} 0.0327 & 0.0173 \\ 0.0173 & 0.4099 \end{bmatrix}
$$

Correlation between $P'(t)$ and $E'(t-1)$ is only 0.05. Standard regression tests would not reject null hypothesis for this regression (zero slope).

Monte Carlo Simulations

Step 1. Simulate regression parameters from Bayesian posterior.

Step 2. Simulate 10 years (8 time steps) 10000 times.

Step 3. Compute average annual premia for each simulation.

Step 4. From 10000 numbers, find mean, 5% and 95% percentiles.

Simulation Results

Start simulations from different values of earnings yield $E'(t)$.

Current value -20.76: 5.69% mean, -4.31% 5p, 15.61% 95p.
Long-term average -19.84: 5.96% mean, -3.83% 5p, 15.66% 95p.
Initial value -16.21: 7.03% mean, -3.55% 5p, 18.09% 95p.

Currently, earnings yield is low compared to long-term average, and the stock market is overpriced. In 1980, it was high, and the market was underpriced.

Equity premia per year 1979Q3–2019Q2 were 6.26%. This included both fundamental return (earnings growth and dividends paid) and speculative return (increase in price compared to earnings).

Modification

We can regress upon earnings yield $E(t) = I(t)/C(t+1)$ itself instead of its logarithm.

Need averaging window of at least 8 to make residuals in both regressions normal, $160/8 = 20$ data points.

Regression Results

$$P'(t) = 0.038 + 0.595E'(t-1) + \varepsilon'(t),$$
$$E'(t) = 0.029 + 0.766E'(t-1) + \delta'(t),$$
$$(\varepsilon'(t), \delta'(t)) \sim \mathcal{N}_2(0, \Sigma) \quad \text{i.i.d.}$$

with covariance matrix of residuals

$$\Sigma = \begin{bmatrix} 0.0357 & -0.00012 \\ -0.00012 & 0.000312 \end{bmatrix}$$

Correlation between $P'(t)$ and $E'(t-1)$ is only 0.17. Standard regression tests would not reject null hypothesis for this regression (zero slope).

Earnings Yield Smoothed

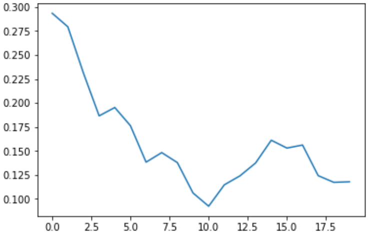


Figure: Graph of $E'(t)$ vs t, for $k = 8$.

Plot of Equity Premium vs Earnings Yield

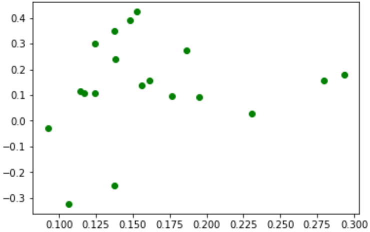


Figure: $(E'(t), P'(t)), t = 1, \ldots, 32$, for $k = 5$.

Simulation Results

Start simulations from different values of earnings yield $E'(t)$.

Current value 11.8%: 5.38% mean, −2.89% 5p, 13.85% 95p.
Long-term average 16.0%: 6.19% mean, −1.81% 5p, 13.9% 95p.
Initial value 29.3%: 8.78% mean, −0.27% 5p, 17.88% 95p.

Again, we see that earnings yield is low compared to long-term average, and the stock market is overpriced. In 1980, it was high, and the market was underpriced.