Log number:

Estimating Economic Index Trends Using Statistical and Computational Simulations

Chanyeol Kim, Richard Kyung
CRG-NJ (Choice Research)

Abstract

In order to speculate about the future of an economy, it is crucial to understand the trend of it. The fundamental relationship between the share price of a company's stock and its per-share earnings, also known as the P/E ratio, is a good indicator of the trend of the economy. While investors have used the P/E ratio to determine whether to buy a certain stock or sell it, relatively less attention has been paid to the significance of the P/E ratio in terms of it being a reflection of the status of the economy.

This paper collected about a century's worth of P/E ratio data and attempted to analyze it using the Gumbel distribution. Then, the processed data was plotted in a frequency graph in order to visualize the general trend of the data. In addition, the data was plugged into Matlab, a programming software, and was experimented with multiple fitting models that could offer further insight to the trend of the data.

Using these methods, it has been revealed that lower P/E values are more likely to repeatedly occur, while higher P/E values are much less likely to repeatedly occur, which can be seen as a tendency for the P/E values to remain in the lower ranges. Put in context, this shows that while the economy may fluctuate for a short amount of time, the economy (in the long run) has a tendency to return back to its previous state and maintain normalcy.

Introduction

Because there are many factors that can determine economic index in the US, it is hard to use a single variable linear regression to determine the effectiveness of finding a pattern. In order to accurately predict the effectiveness of such trend, an objective main factor that determines the status of economic status in the United States was chosen in this paper. The task of modeling economic pattern in a focused area and performing data analysis is not easy, especially when the data pattern is complex.

By using statistical and computational simulations including a regression analysis using least square method, this paper investigates the impact of Price Earnings ratio (P/E ratio) on economy and identifies risk factors associated with these outcomes. The Price Earnings ratio (P/E ratio) is the ratio for estimating the valuation of a company. It is calculated by dividing the current market price of the stock of the company by its earning per share. [1] P/E ratio is one of the most widely used tools for economic index and stock selection. [2]

Gumbel distribution is used to model the distribution of the maximum/minimum of a number of P/E ratio samples of various distribution. [3] After finding exceedance probability that exceeds a critical value, return period, which is an estimate of the likelihood of an event occurrence, was found. [4] A statistical measurement based on historic data denoting the average recurrence interval was found accordingly. Also, theoretical probability was calculated; and finally, estimated probability was checked to find if it approaches the theoretical probability as the number of trials gets larger.

I. BACKGROUND AND THEORY

A. *Time Series Analysis and Forecasting*

There have been numerous different kinds of data such as stock prices and interest rates observed and gathered in the past. The sequential nature of these data require us to account for the dynamic nature using special statistical skill and techniques. Time series analysis provide the appropriate methods necessary in order to analyze sequential data.

B. *Smoothing*

It may be problematic to picture the essential, underlying trend of the data if the time series has a lot of noise. To distinguish the signal and the noise from each other, various linear and nonlinear smoothers must be applied.

C. *Curve fitting*

In MATLAB®, best-fit line are available using the least-sum-of-squares line from the data. Also known as linear least squares regression or least squares regression line (LSRL), this type of linear modeling minimizes the sum of the squares of the deviations between the model and the actual data. [6] The deviations are squared in order to reduce the influence of negative or positive signs when added.

In general terms, curve fitting involves either interpolation - in which the fitting model exactly matches the data, but often in piecewise manners - or smoothing - in which the noise of the data is reduced and a function approximates the overall trend of the data. [7] [8] Curve fitting can be used to not only map out the data and render it computable in general terms, but also extrapolate other data points based on the trend provided by the model. [8]

II. Data/Results

A. *Plotting the PE ratio Frequency Curve using Gumbel Distribution*

Background information:

- Gumbel distribution is used to model the distribution of the maximum/minimum of a number of samples of various distribution. [3]
- Exceedance probability is the probability that the event will exceed some critical value (usually far from the mean).
- Return period is an estimate of the likelihood of an event to occur. [4] A statistical measurement based on historic data denoting the average recurrence interval.
- Theoretical probability is the fraction of times we expect the event to occur if we repeat the same experiment over and over (i.e. flipping a coin and getting heads or tails is each 0.50).
- Estimated probability approaches the theoretical probability as the number of trials gets larger. It is an approximation of theoretical probability.

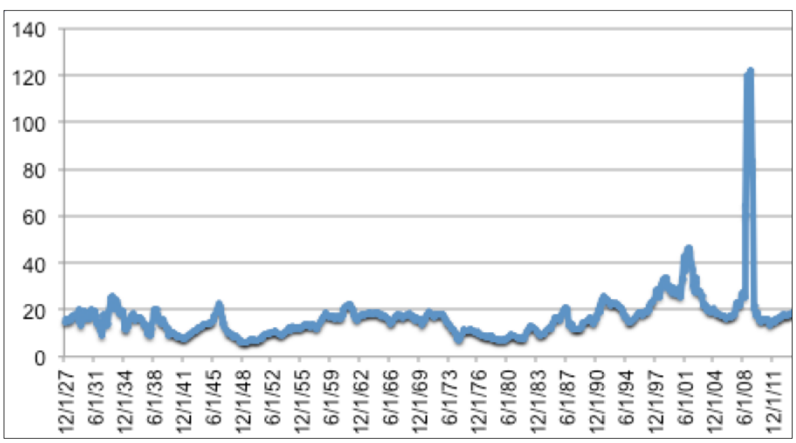

Fig. 1. S&P 500 PE Ratio 90 Year Graph

B. PE Ratio Trend Analysis

Instructions:

1) Label column A as ‘time’ and column B as the ‘PE Ratio’ and enter them in appropriate values in each cell, which is shown in Table A below.

TABLE I. TIME AND P/E RATIO

Time	PE Ratio	Time	PE Ratio
01/12/27	15.9099	01/09/16	24.338
01/01/28	14.8898	01/10/16	22.487
01/02/28	14.6271	01/11/16	23.2555
01/03/28	16.339	01/12/16	23.6788
01/04/28	15.8	01/01/17	22.7228
01/05/28	16	01/02/17	23.5681
01/06/28	15.312	01/03/17	23.5589
01/07/28	14.8321	01/04/17	22.9206
01/08/28	15.9313	01/05/17	23.1859
01/09/28	16.1374	01/06/17	23.2975

2) Select the PE Ratio values in column B then sort the values from smallest to largest by clicking on the ‘Sort and Filter’ tool. Allow expanding the selection. An example can be seen in Table B below.

TABLE II. SORTED P/E RATIO

PE Ratio	continued...
5.9	46.453
5.9125	46.4998
6.1417	60.2312
6.1429	60.7023
6.2929	65.1042
6.3277	78.7464
6.3682	81.3892

6.395	84.2967
6.441	107.156
6.5188	116.2197
6.5588	116.3076
6.6376	120.3907
6.6769	122.3888
	122.4128

3) Label column C as ‘Rank (i)’ and rank the data in decreasing order (from N to 1).

4) Create a fourth column called q_i. Gringorten’s plotting position formula will be used to calculate the estimated exceedance probabilities relevant to past observations. An example can be seen in Table C below.

$$q_i = \frac{i-a}{N+1-2a} \qquad (1)$$

q_i = exceedance probability associated with a specific observation

$N =$ number of annual maxima observations

$i =$ Rank of specific observation (i=1 is the largest and i=N is the smallest)

$a =$ constant for estimation (0.44)

TABLE III. q_i OF PE RATIO VALUES

Time	PE Ratio	rank	q_i
01/06/49	5.9	1075	0.999479128
01/05/49	5.9125	1074	0.998548999
01/04/49	6.1417	1073	0.99761887
01/02/49	6.1429	1072	0.996688742
01/07/49	6.2929	1071	0.995758613
01/03/49	6.3277	1070	0.994828484
01/08/49	6.3682	1069	0.993898356
01/01/49	6.395	1068	0.992968227
01/11/48	6.441	1067	0.992038098

01/09/49	6.5188	1066	0.991107969
01/07/50	6.5588	1065	0.990177841
01/12/48	6.6376	1064	0.989247712
01/03/80	6.6769	1063	0.988317583
01/08/50	6.7721	1062	0.987387454
01/10/79	6.852	1061	0.986457326

5) Make another column and label it p_i. Then make it equal to $1 - q_i$. p_i refers to the non-exceedance probability.

C. *Statistical Definition of Return Period*

- If X is a random variable with a cumulative distribution function $F_x(x)$, the probability that X is less than equal (not exceeding) to a given event x_p is:

$$F_x(x) = P(X \le x_p) = p \qquad (2)$$

- The probability that this event will be exceeded is now $1 - p$, and the percent exceedance would be $100(1 - p)\%$.
- For an event x_p, the return period corresponding to this exceedance probability is denoted by T.

$$T = \frac{1}{(1-p)} \qquad (3)$$

- For example, a 100-year return period is an event with a probability of exceedance $1 - p = 0.01$ or a non-exceedance probability $p = 0.99$. There is a 99% chance that this event will not be exceeded within a given year.

6) Create one more column and label it 'T_p estimated' and evaluate the values in p_i using the equation for the return period. An example can be seen in Table D below.

TABLE IV. q_i, p_i, T_p OF PE RATIO VALUES

Time	PE Ratio	Rank	q_i	p_i	T_p estimated
01/06/49	5.9	1075	0.999479	0.000520	1.000521
01/05/49	5.9125	1074	0.998548	0.001451	1.001453
01/04/49	6.1417	1073	0.997618	0.002381	1.002386
01/02/49	6.1429	1072	0.996688	0.003311	1.003322

01/07/49	6.2929	1071	0.995758	0.004241	1.004259
01/03/49	6.3277	1070	0.994828	0.005171	1.005198
01/08/49	6.3682	1069	0.993898	0.006101	1.006139
01/01/49	6.395	1068	0.992968	0.007031	1.007081
01/11/48	6.441	1067	0.992038	0.007961	1.008025

continued...

01/12/08	60.7023	11	0.009822	0.990177	101.81060
01/10/08	65.1042	10	0.008892	0.991107	112.46025
01/07/09	78.7464	9	0.007961	0.992038	125.59813
01/08/09	81.3892	8	0.007031	0.992968	142.21164
01/09/09	84.2967	7	0.006101	0.993898	163.89024
01/02/09	107.156	6	0.005171	0.994828	193.36690
01/04/09	116.2197	5	0.004241	0.995758	235.77192
01/03/09	116.3076	4	0.003311	0.996688	302
01/01/09	120.3907	3	0.002381	0.997618	419.96875
01/05/09	122.3888	2	0.001451	0.998548	689.17948
01/06/09	122.4128	1	0.000520	0.999479	1919.857143

'T_p estimated' is the estimated distribution of 91 years of data. We assume that the data follows a specific distribution to estimate the parameters.

7) We will follow the 'Gumbel' or 'Extreme Value Type 1' distribution. The CDF (Cumulative distribution of function) of the Gumbel distribution is the following:

$$F_x(x) = exp\,[-exp\,(-\frac{x-u}{\alpha})\,] \;=\; p \qquad (4)$$

x is the observed P/E data; u and α are the calculated parameters of the distribution. This distribution will allow us to calculate the theoretical estimate of p.

8) Create two columns labeled '$(x-u)/\alpha$' and 'p-theoretical. Using the following equations, calculate x̄, s_x, u and α. Table E (below) shows such values that result from the existing data.

$$\underline{x} = \sum_{i=1}^{n} \frac{x_i}{n} \qquad (5)$$

$$s_x{}^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (x_i - \underline{x})^2 \qquad (6)$$

$$u = \underline{x} - 0.5772\alpha \qquad (7)$$

$$\alpha = \frac{\sqrt{6}s_x}{\pi} \qquad (8)$$

TABLE V. SPECIAL VALUES

$\underline{x}$
17.07033098
${s_x}^2$
10.61780738
u
12.29188239
α
8.278670447

9) Use the P/E Ratio values (x) and populate the column $(x-u)/\alpha$ as shown in Table F below:

TABLE VI. q_i, p_i, T_p ESTIMATED, AND $(x-u)/\alpha$ OF PE RATIO VALUES

time	PE Ratio	rank	T_p estimated	$(x-u)/\alpha$
01/06/49	5.9	1075	1.000521144	-0.77209045
01/05/49	5.9125	1074	1.001453109	-0.770580546
01/04/49	6.1417	1073	1.002386813	-0.742894941
01/02/49	6.1429	1072	1.003322259	-0.74274999
01/07/49	6.2929	1071	1.004259453	-0.724631139
01/03/49	6.3277	1070	1.005198399	-0.720427565
01/08/49	6.3682	1069	1.006139103	-0.715535476
01/01/49	6.395	1068	1.007081569	-0.712298241
01/11/48	6.441	1067	1.008025803	-0.706741793
01/09/49	6.5188	1066	1.008971808	-0.697344149
01/07/50	6.5588	1065	1.009919591	-0.692512455
01/12/48	6.6376	1064	1.010869156	-0.682994018
01/03/80	6.6769	1063	1.011820509	-0.678246879
01/08/50	6.7721	1062	1.012773654	-0.666747448
01/03/02	46.453	14	79.28613569	4.126401434
01/12/01	46.4998	13	85.59872611	4.132054516

01/11/08	60.2312	12	93.00346021	5.790702494
01/12/08	60.7023	11	101.8106061	5.847607767
01/10/08	65.1042	10	112.460251	6.379323581
01/07/09	78.7464	9	125.5981308	8.027196883
01/08/09	81.3892	8	142.2116402	8.346426887
01/09/09	84.2967	7	163.8902439	8.697630624
01/02/09	107.156	6	193.3669065	11.45885903
01/04/09	116.2197	5	235.7719298	12.55368459
01/03/09	116.3076	4	302	12.56430224
01/01/09	120.3907	3	419.96875	13.05750945
01/05/09	122.3888	2	689.1794872	13.29886463
01/06/09	122.4128	1	1919.857143	13.30176365

10) Use the CDF equation from step 7 to calculate the value of p-theoretical.

11) Use the equation used to calculate 'T_p estimated' and use it to calculate 'T_p theoretical' using the p theoretical values, as seen in Table G below.

TABLE VII. $(x-u)/\alpha$, P THEORETICAL, AND T_p THEORETICAL OF PE RATIO VALUES

(x-u)/a	p theoretical	T_p theoretical
-0.77209045	0.114831911	1.129728933
-0.770580546	0.115207496	1.13020849
-0.742894941	0.122210304	1.139225038
-0.74274999	0.122247543	1.13927337
-0.724631139	0.126949026	1.145408493
-0.720427565	0.128052902	1.146858568
-0.715535476	0.129343763	1.148558935
-0.712298241	0.130201603	1.149691703
-0.706741793	0.131680746	1.151650151
-0.697344149	0.134201675	1.155003389

-0.692512455	0.135507139	1.156747552
-0.682994018	0.138097349	1.160223836
-0.678246879	0.139398252	1.161977654
-0.666747448	0.142574389	1.166281934

continued...

4.126401434	0.983988707	62.45591876
4.132054516	0.984078243	62.80713726
5.790702494	0.99694883	327.7430845
5.847607767	0.997117366	346.9049483
6.379323581	0.998305168	590.0289459
8.027196883	0.999673591	3063.643305
8.346426887	0.999762785	4215.59282
8.697630624	0.999833033	5989.205912
11.45885903	0.999989445	94737.41657
12.55368459	0.999996468	283137.0556
12.56430224	0.999996505	286159.3158
13.05750945	0.999997866	468602.6772
13.29886463	0.999998324	596518.4617
13.30176365	0.999998328	598250.286

D. *Graphing the P/E Ratio Frequency Curve*

6) Go to ‘insert’ tab and select charts. Plot $'T_p$ estimated’ vs Ratio Value. On the same graph, also plot $'T_p$ theoretical’ vs Ratio Value. Label the chart title and the axes on the obtained graph. Figure 2 (below) shows how the graph may look like.

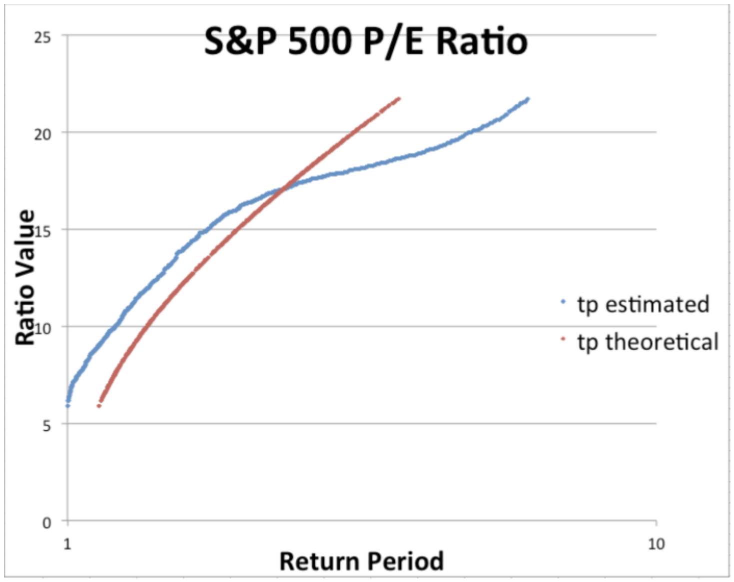

Fig. 2. T_p estimated, T_p theoretical vs P/E Ratio Value

7) Right click on the curve and select 'change chart type' and click on 'scatter with smooth lines' for theoretical and 'scatter' for estimated.

8) Right click again and select 'format chart area' to use the axis options command for the X-axis and select logarithmic scale. It will then contain return periods displayed from 1 to 100 in log scale. Figure 3 (below) shows what the graph may look like.

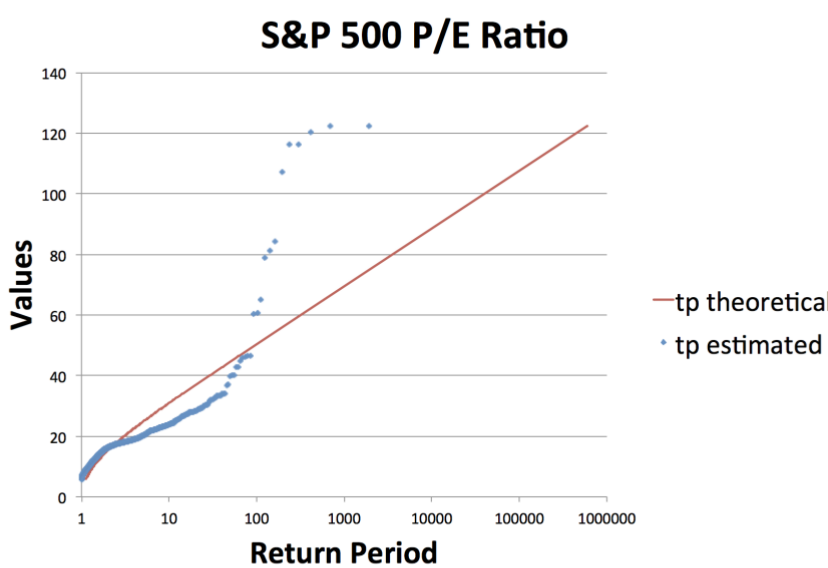

Fig. 3. T_p estimated, T_p theoretical vs P/E Ratio Value, Logarithmic Scale

The red line in Fig. 3 shows the theoretical distribution, while the blue dots stand for the fit of the annual P/E data with respect to a Gumbel distribution. You can predict P/E values corresponding to any return period from 1 to 100. The curve follows the distribution very well for low P/E values, but starts to drift away from the theoretical at higher values.

We can see that lower P/E values have much shorter return periods (meaning that they are more likely to repeatedly occur), while for higher P/E values, the length of the return periods exponentially increase as the values increase (meaning that higher P/E values are much less likely to repeatedly occur). This can be seen as a tendency for the P/E values to remain in the lower ranges. The higher the P/E values, the more over speculation and overestimation the stock market is, so it is natural for the P/E values to come down after some time. [1] [2] This trend is expected to repeat for some time in the future, since stock market bubbles are always bound to pop at some point, bringing the high P/E values down to normalcy. [1] [2]

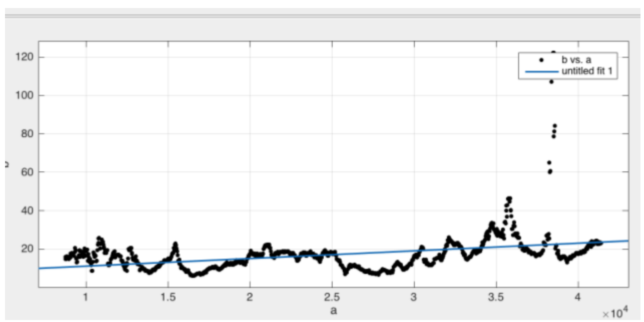

Fig. 4. P/E Ratio with Polynomial Fit to 1 Degree

Figure 4 above shows the data with a polynomial model. When plotted in Matlab with the polynomial fit with 1 degree, the model was not a great fit to the actual data, since the R-squared value was nowhere close to 1 - in fact, it was closer to 0. Even if the degree was put to the maximum available in the software, the R-squared value was still much less than 1, let alone 0.5. This showed that a polynomial model would not suffice.

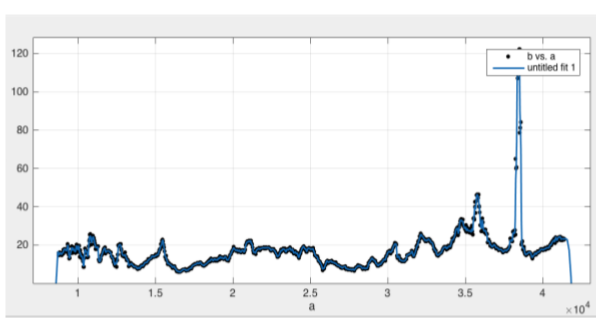

Fig. 5. P/E Ratio with Smoothing Spline Fit

Figure 5 above shows the data with a smoothing spline fit model. The smoothing spline fit was definitely a much better model than the polynomial model (and almost all models in general), as the R-squared value was very close to 1 (being at 0.9889). This can also be seen directly from the graph, since the blue lines pass through almost all of the data points. However, there is no way to find the actual equation of the model per se, since the model is composed of piecewise components, as seen in the image above.

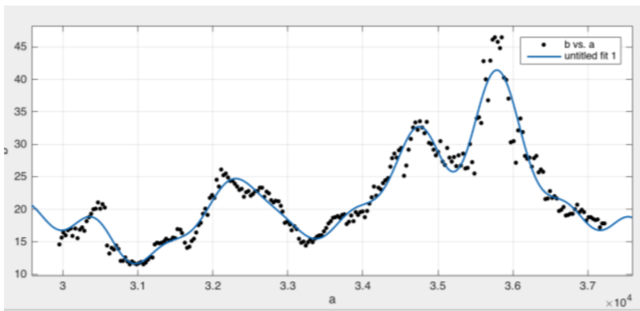

Fig. 6. P/E Ratio with Fourier Fit to 8 Terms

Figure 6 shows the Fourier fit to the data. [5] When the scope of the data was reduced to about two decades' worth of data, the Fourier model provided a great fit. [5] Looking at the graph, it can be seen that the blue model passes through a lot of the data points and is able to map out the trend of the model in general. Although it isn't seen in the figure, the R-squared value is at 0.9377, which shows that the model is a very strong representation of the data.

III. Discussion

This paper collected a century's worth of P/E ratio data and used the Gumbel distribution to map out the overall trend of the P/E ratio in terms of its return period. Also, the data was plotted in Matlab, and multiple fitting models were tested out to see which one fit the data the best. The P/E ratio was chosen due to its significance in the evaluation of stocks' values, and the Gumbel distribution due to its ability to incorporate rapidly fluctuating data into statistical analysis.

References

[1] https://www.investopedia.com/terms/p/price-earningsratio.asphttps://www.investopedia.com/terms/p/price-earningsratio.asp

[2] https://en.wikipedia.org/wiki/Price%E2%80%93earnings_ratio#In_business_culture

[3] https://en.wikipedia.org/wiki/Gumbel_distribution

[4] https://en.wikipedia.org/wiki/Return_period

[5] https://www.usna.edu/Users/oceano/pguth/md_help/html/time6h9j.htm

[6] https://www.itl.nist.gov/div898/handbook/pmd/section1/pmd141.htm

[7] http://web.iitd.ac.in/~pmvs/courses/mel705/curvefitting.pdf

[8] https://en.wikipedia.org/wiki/Curve_fitting