Thai Journal of Mathematics : (2017) 29–34
Special Issue on Entropy in Econometrics
http://thaijmath.in.cmu.ac.th
ISSN 1686-0209

MaxEnt-Based Explanation of Why Financial Analysts Systematically Under-Predict Companies' Performance

Vladik Kreinovich[†,1] **and Songsak Sriboonchitta**[‡]

[†]Department of Computer Science, University of Texas at El Paso
El Paso, TX 79968, USA
e-mail : vladik@utep.edu

[‡]Faculty of Economics, Chiang Mai University
Chiang Mai 50200 Thailand
e-mail : songsakecon@gmail.com

Abstract : Several studies have shown that financial analysts systematically under-predict the companies' performance, so that quarter after the quarter, 70-75% of the companies outperform these predictions. This percentage remains the same where the economy is in a boom or in a recession, whether we are in a period of strong or weak regulations. In this paper, we provide a possible Maximum Entropy-based explanation for this empirical phenomenon – an explanation rooted in the fact that financial analysts mostly analyze financial data, while to get a more accurate prediction, it is important to go deeper, into the technical issues underlying the companies functioning.

Keywords : maximum entropy; interval uncertainty; financial analysis.

2010 Mathematics Subject Classification : 94A17; 91B50; 90B50.

We acknowledge the partial support of the Center of Excellence in Econometrics, Faculty of Economics, Chiang Mai University, Thailand. This work was also supported in part by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721, and by an award "UTEP and Prudential Actuarial Science Academy and Pipeline Initiative" from Prudential Foundation.

[1]Corresponding author.

1 Formulation of the Problem

Financial predictions: what seem natural to expect. Every quarter, financial analysts try their best to predict the companies' performance as accurately as possible. One would expect that with this strive for accuracy, the estimates will be close to the actual performance, with no special preference for positive or negative prediction errors. In other words, we expect that:

- in approximately half of the cases, a company would perform slightly better than the analysts predicted, and
- in about half of the cases, the company would perform slightly worse than the financial analysts predicted.

What we actually observe. Surprisingly, quarter after quarter, the companies beat expectations in 70-75% of the cases; see, e.g., [1] and references therein. This percentage is the same whether the economy as a whole is good or bad, whether the analysts predict profits or losses, etc.

How can we explain this phenomenon? Some of this may be caused by a collusion between analysts and companies – since companies are interested in beating expectations, it increases the value of their stock. This collusion is a known problem, and the regulatory agencies all over the world are trying their best to fight against it.

Sometimes, the regulatory agencies are more successful, they penalize some companies and some analysts, scaring everyone else into obedience. Sometimes, budget cuts led to serious cuts in the regulatory activities, as a result of which collusion becomes more widely spread.

Collusion is a real phenomenon, no doubt about it, but if it was the main source of underestimation, then the underestimation rate would fluctuate with changes in regulatory activity:

- when the regulators are more active, we would expect the percentage of the companies beating financial expectations to drop to close to 50%, while
- in the periods when the regulatory activity decreases, we would expect this percentage to rise again.

And we do not observe any such fluctuations. No matter what is the current level of regulatory activity, the percentage of companies that beat expectations remain approximately the same – 70-75%.

So, we must look beyond collusion to explain this empirical phenomenon.

2 Analysis of the Problem and the Resulting Explanation

There should be a fundamental explanation. Estimation methods change, algorithms change, new technologies appear – but the percentage of companies beating the analysts' expectations remains the same. This makes us think that this percentage should have some deep fundamental explanation. To find such an explanation, let us consider this problem from the viewpoint of uncertainty.

Predictions in general and their uncertainty: a brief reminder. Predictions are ubiquitous, they are ubiquitous in science, they are ubiquitous in medicine, they are ubiquitous in finance. Traditionally, when we want to predict the future value of a quantity, we come up with a single number. For example, a meteorologist may predict that tomorrow's temperature will be 21 degrees, a financial analyst may predict that the company's profit will be 1.3 billion dollars, etc.

These prediction numbers sound exact, but, of course, everyone understands that predictions are approximate: the temperature may be slightly higher or slightly lower than predicted, the company's profit may be slightly higher or slightly lower than predicted, etc. For a prediction to be meaningful, we need to know not only the number, but also how accurate is the number. If the prediction is not accurate at all, then it is useless: for example, if the predicted temperature is 20 degree but the accuracy is ± 20 degrees, then we may have a hear wave of 40 degrees, or we may have snow on the ground at the zero temperature.

How can we describe the prediction's accuracy? In the first approximation, we want to know which future values are possible and which are not. In most applications, if two values are possible, this means that all intermediate values are possible as well. Thus, the set of all possible future values of a quantity x is a connected set – i.e., an interval $[\underline{x}, \overline{x}]$.

Ideally, for each of the possible values $x \in [\underline{x}, \overline{x}]$, we would also like to know how possible it is, e.g., what is the probability that this value will occur. Coming up with such probabilities – in addition to the intervals – is a very complex task, requiring a large amount of effort. Indeed, instead of generating two numbers $\underline{x}$ and $\overline{x}$, we now need to generate probabilities corresponding to several subintervals of this interval.

For example, if we want 10 probabilities – by the way, a very crude description of a probability distribution – we need to generate 5 times more values than simply two endpoints – so, in general, we need 5 times more efforts.

In these terms, what is the accuracy of financial predictions? A local meteorologist team tries to predict a few numbers – tomorrow's temperature, wind velocity, etc. They can therefore afford to spend some time trying to come up with probabilities of different possible values of these quantities.

In contrast, a financial analyst predicts many financial characteristics describing the future performance of numerous companies. There is simply no way that an overworked analyst will perform five times more work to come up with the probabilities.

So, internally, what the analyst does is comes up, for each financial characteristic, with a range $[\underline{x}, \overline{x}]$ of its possible values. If he or she is a good analyst, this range should more or less accurately reflect the actual performance of the company.

So, why is there a systematic underestimation?

How an analyst transforms an internal interval estimate into a numerical prediction. To answer this question, let us recall that while internally, an analyst understands very well that his/her prediction is uncertain, that he/she can only predict the interval of possible values – what the analyst needs to generate is a number, not an interval.

So how should a rational person go from an interval to a number? If we know the probability distribution $\rho(x)$ on the interval $[\underline{x}, \overline{x}]$, then a reasonable idea is to select a number $\widetilde{x}$ for which the mean square deviation from the actual value is the smallest possible:

$$\int (x - \widetilde{x})^2 \cdot \rho(x)\, dx \to \min_{\widetilde{x}}.$$

According to calculus, we can find the minimum of the function if we look for points where its derivative is equal to 0. Differentiating the above minimized expression and equating the derivative to 0, we conclude that the optimal estimate is

$$\widetilde{x} = \int x \cdot \rho(x)\, dx,$$

i.e., the mean value of x.

What if we do not know the distribution $\rho(x)$? Since we have no reason to believe that some values x from the interval $[\underline{x}, \overline{x}]$ are more probable than others, it is reasonable to assume that all these values are equally probable – i.e., that we have a uniform distribution on the interval. In the continuous case, this idea – known as the *Laplace Indeterminacy Principle* – is a particular case of the *Maximum Entropy Principle*, according to which in situations when we several candidates $\rho(x)$ for the actual probability distribution, we should select a one for which the entropy $S = -\int \rho(x) \cdot \log_2(\rho(x))\, dx$ is the largest possible; see, e.g., [2].

For the uniform distribution on an interval, the mean value is the midpoint $\widetilde{x} = \dfrac{\underline{x} + \overline{x}}{2}$ of this interval. This midpoint is what the analysts return.

But what if we look beyond financial data? Financial analysts look mostly at the financial data. This makes sense, they are specialists in finance, they are not specialists in specifical technical areas of the companies that they analyze. For example, they can analyze trends in sales records of smart phones, but they cannot meaningfully analyze the possible effect of different technical decisions. We are not intending this as a criticism: it is amazing that, without the technical knowledge, based only on the financial data, the analysts can make reasonably accurate predictions.

However, clearly, any additional information would, in general, make predictions more accurate – in particular, additional technical information. How would such an analysis affect the predictions?

Let us consider the simplest case, when the financial result x depends on the selection of some technical parameter a. Companies do their best to find optimal values of the corresponding technical parameters – i.e., values for which the company would get the largest possible profit. But, of course, the companies also cannot predict everything perfectly, and they cannot implement everything perfectly. As a result, the actual value a that the company uses will be, in general, slightly different from the ideal optimal value a_0.

Let Δ denote the largest possible deviation of the actual value a from the ideal value a. In this case, all we can say about the actual parameter a is that it will be somewhere between $a_0 - \Delta$ and $a_0 + \Delta$. A general value a from this interval $[a_0 - \Delta, a_0 + \Delta]$ can be represented as $a_0 + \Delta a$, where $\Delta a \stackrel{\text{def}}{=} a - a_0$ can take any value from $-\Delta$ to Δ.

The financial result $x(a)$ depends on the selection of the parameter a. Substituting $a = a_0 + \Delta$ into this dependence, we conclude that $x = x(a_0 + \Delta a)$. Since the value Δa is small, we can expand this dependence $x(a_0 + \Delta a)$ in Taylor series in Δa and keep only the main terms in this dependence:

$$x(a_0 + \Delta a) = x(a_0) + x'(a_0) \cdot \Delta x + \frac{x''(a_0)}{2} \cdot (\Delta a)^2 + \ldots$$

By definition, the function $x(a)$ attains its maximum at the value a_0, thus $x'(a_0) = 0$ and $x''(a_0) < 0$, so we have

$$x(a_0 + \Delta a) \approx x(a_0) - c \cdot (\Delta a)^2,$$

where we denoted $c \stackrel{\text{def}}{=} -\frac{x''(a_0)}{2}$.

The largest value of this function $x(a)$ on the interval $[a_0 - \Delta, a_0 + \Delta]$ is attained when $a = a_0$, and is thus equal to $\overline{x} = x(a_0)$. One can easily see that the smallest value of the function $x(a)$ is attained when $a = a_0 \pm \Delta$, and is equal to $\underline{x} = x(a_0) - c \cdot \Delta^2$.

Provided that the financial analyst is accurate, these are exactly the bounds that he or she generates. Thus, as we have explained, the analysts returns the following number as his/her prediction

$$\widetilde{x} = \frac{\underline{x} + \overline{x}}{2} = x(a_0) - \frac{c}{2} \cdot \Delta^2.$$

What should we be able to predict if we could go into technical details? All we know about the actual value of the technical parameter a is that this value is somewhere between $a_0 - \Delta$ and $a_0 + \Delta$. We do not know the probabilities of different values of a within this interval. A reasonable idea is thus to use the Maximum Entropy principle and to conclude that the parameter a is uniformly distributed on this interval, and that, equivalently, the difference Δa is uniformly distributed on the interval $[-\Delta, \Delta]$.

Based on this probability distribution for the unknown value a, let us find the probability that the actual performance $x(a)$ will exceed the predicted value $\widetilde{x}$.

The actual performance is equal to $x(a) = x(a_0) - c \cdot (\Delta a)^2$. The predicted value is equal to $\widetilde{x} = x(a_0) - \frac{c}{2} \cdot \Delta^2$. Thus, the condition $x(a) \geq x(a_0)$ is equivalent to

$$x(a_0) - c \cdot (\Delta a)^2 \geq x(a_0) - \frac{c}{2} \cdot \Delta^2.$$

If we subtract $x(a_0)$ from both sides of this inequality and divide both sides of the resulting inequality by $-c$, we get an equivalent inequality $(\Delta a)^2 \le \dfrac{1}{2} \cdot \Delta^2$, i.e., equivalently,

$$-\frac{\sqrt{2}}{2} \cdot \Delta \le \Delta a \le \frac{\sqrt{2}}{2} \cdot \Delta.$$

Thus, the actual company performance outperforms the analyst's estimate if Δa belongs to the subinterval $\left[-\dfrac{\sqrt{2}}{2} \cdot \Delta, \dfrac{\sqrt{2}}{2} \cdot \Delta\right]$ of the original interval $[-\Delta, \Delta]$.

For the uniform distribution, the probability to be in an interval is proportional to the width of this interval. Thus, the probability p for Δa to be in the above interval is equal to the ratio of the width $\sqrt{2} \cdot \Delta$ of the corresponding subinterval to the width 2Δ of the original interval:

$$p = \frac{\sqrt{2} \cdot \Delta}{2\Delta} = \frac{\sqrt{2}}{2}.$$

This probability is approximately equal to 70% (and slightly larger than 70%), in perfect accordance with the empirical observation that in 70-75% of the cases, the companies outperform their predictions.

So, we indeed have an explanation for the above empirical fact – namely, we explain it by the fact that:

- the financial analysts only perform the analysis of the financial data,
- while the analysis of the technical data – which is, unfortunately, beyond the financial analysts' area of expertise – would have led to more accurate estimates.

References

[1] T. Gryta, S. Ng, T. Francis, Companies routinely steer analysts to deliver earnings surprises, Wall Street Journal, August 5, 268 (30) (2016) pp. A1 and A8.

[2] E. T. Jaynes, G. L. Bretthorst, Probability Theory: The Logic of Science, Cambridge University Press, Cambridge, UK, 2003.

(Received 31 August 2017)
(Accepted 30 October 2017)