Prediction in Econometrics: Towards Mathematical Justification of Simple (and Successful) Heuristics

Vladik Kreinovich[1,2], Hung T. Nguyen[3,2], and Songsak Sriboonchitta[2]

[1]Department of Computer Science
University of Texas at El Paso
El Paso, TX 79968, USA
vladik@utep.edu

[2]Faculty of Economics
Chiang Mai University
Chiang Mai 50200 Thailand
songsak@econ.cmu.ac.th

[3]Department of Mathematical Sciences
New Mexico State University
Las Cruces, NM 88003, USA
hunguyen@nmsu.edu

Abstract

Many heuristic and semi-heuristic methods have been proposed to predict economic and financial processes. Some of these heuristic processes are intuitively reasonable, some seemingly contradict to our intuition. The success of these heuristics leads to a reasonable conjecture that these heuristic methods must have a more fundamental justification. In this paper, we provide such a justification for two simple (and successful) prediction heuristics: of an intuitive exponential smoothing that provides a reasonable prediction for slowly changing processes, and of a seemingly counter-intuitive idea of an increase in volatility as a predictor of trend reversal. As a possible application of these ideas, we consider a new explanation of the price transmission phenomenon.

1 Introduction

Prediction is important. Prediction (forecasting) is of upmost importance in economics and finance. If we can accurately predict the future prices, then we can get the largest return on investment – whether we invest in stocks or in industry. Vice versa, if we make decisions based on the wrong predictions, then our financial investments collapse, and the manufacturing plants that we built are non-profitable and thus idle.

There exist many heuristic prediction techniques. Because the prediction problem is so important, many heuristic and semi-heuristic methods have been proposed to predict economic and financial processes. Many proposed heuristic methods turn out to be very successful.

Some of these methods are *semi-heuristic* in the sense that some of their features have known explanations. Often, many of these features are intuitively reasonable. Other methods are purely heuristic – in the sense that they are justified solely by their empirical success. Some of these empirically efficient methods are very difficult to explain and understand, because their success seems to contradict our intuitive understanding of economical and financial phenomena.

Need to justify heuristic strategies. The success of prediction heuristics leads to a reasonable conjecture that these heuristic methods must have a more fundamental justification. In general, when we have a theoretical justification, it helps:

- we can use the corresponding theory to fine-tune the method,
- we can also use this theory to get a clearer understanding in what situations the method is efficient and in what situations it is not efficient.

What we do. In this paper, we provide such a justification for two simple (and successful) prediction heuristics:

- of an intuitive exponential smoothing procedure, that provides a reasonable prediction for slowly changing processes, and
- of a seemingly counter-intuitive idea of an increase in volatility as a predictor of trend reversal.

As a possible application of these ideas, we consider a new simple explanation of the known phenomenon of asymmetric price transmission – when:

- an increase in raw component prices leads to an immediate increase in consumer prices, but
- a following decrease in raw component prices leads to a much slower decrease in consumer prices.

2 First Result: Exponential Smoothing is the Only Predictor for Which the Effect of Noise Always Decreases with Time

Prediction: case of slowly changing processes. In many economic and financial situations, we have a sequence of observations of a certain quantity x at different moments of time. Based on these observations, we would like to predict the future value of this quantity. Let x_1 be the result of the most current observation, x_2 the result of the second recent observation, etc. Based on these values $x_1, \ldots, x_T$, we would like to compute the estimate X_0 for the future value x_0 of the observed quantity; see, e.g., [2, 6].

To describe such an estimate, we need to describe a function $F(x_1, \ldots, x_T)$ that takes, as input, the values x_t and returns the desired estimate X_0. What are the reasonable properties of this prediction function?

Continuity. The first property is related to the fact that in many cases, the values x_t are only approximate. For example, if we are interested in predicting GDP or unemployment rate, we have to take into account that the existing methods of measuring these characteristics are approximate. Thus, the actual values $x_t^{\rm act}$ of the corresponding variables are, in general, slightly different from the observed values x_t. It is therefore desirable that the actual result $F(x_1, \ldots, x_T)$ of applying the function F to these slightly modified values x_t be only slightly different from the desired result $F(x_1^{\rm act}, \ldots, x_T^{\rm act})$ of applying F to the (unknown) actual values $x_t^{\rm act}$.

In other words, if the inputs to the function F change slightly, the output should also change slightly. In precise terms, we want the function F to be continuous.

Additivity. In many practical situations, we observe a joint effect of two different signals $x_t = x_t^{(1)} + x_t^{(2)}$. For example, the varying price of the financial portfolio can be represented as a sum of the prices corresponding to different parts of this portfolio: e.g., stocks and bonds. In this case, the desired future value x_0 also consists of two components: $x_0 = x_0^{(1)} + x_0^{(2)}$. Thus, if we separately predict the first component and predict the second component, then the sum of these predicted values can serve as a predictor for the time series as a whole.

Because of this, it is reasonable to require that the result of applying our predictor to the sum $x_0^{(1)} + x_0^{(2)}$ of the two time series should be equal to the sum of the predictions based on individual components $x_0^{(1)}$ and $x_0^{(2)}$ of this sum:

$$F(x_1^{(1)} + x_1^{(2)}, \ldots, x_n^{(1)} + x_T^{(2)}) = F(x_1^{(1)}, \ldots, x_T^{(1)}) + F(x_1^{(2)}, \ldots, x_T^{(2)}).$$

In mathematical terms, this means that the predictor function should be *additive.*

Conclusion: we must consider linear predictors. It is known (see, e.g., [1]) that every continuous additive function is a homogeneous linear function, i.e., it has the form $F(x_1, \ldots, x_T) = \sum\limits_{t=1}^{T} f_t \cdot x_t$ for some values f_t. Thus, we conclude that we must consider linear predictors.

How to naturally describe long time series. The actual number of observed values is always finite. However, in many cases, we have very long time series. For example, we have many decades of daily records of prices of a certain stock. In this case, it is reasonable to assume that we have an infinite number of records, and use the formula $X_0 = \sum\limits_{t=1}^{\infty} f_t \cdot x_t$. In real life, the influence of remote events is small, so if we do not know only the values $x_1, \ldots, x_T$, we simply ignore the remaining (unknown) terms $x_{T+1}, \ldots$ in the above formula, and use an approximate formula $X_0 \approx \sum\limits_{t=1}^{T} f_t \cdot x_t$.

Case of a constant signal. If the observed value x_t does not change at all, i.e., if for some constant c, we have $x_t = c$ for all t, then it is reasonable to predict that the same value $x_0 = c$ will remain in the next moment of time. Thus, it is reasonable to require that in this case, we have $X_0 = F(x_1, \ldots, x_n) = F(c, \ldots, c) = c$. For the above expression, this equality leads to $\sum\limits_{t=1}^{\infty} f_t \cdot c = c \cdot \sum\limits_{t=1}^{\infty} f_t = c$, i.e., to $\sum\limits_{t=1}^{\infty} f_t = 1$.

Which predictors should we use? The prediction quality depends on the choice of the predictor, i.e., on the choice of the coefficients f_t that describe the predictor function. The only requirement that we have described so far is that $\sum\limits_{t=1}^{\infty} f_t = 1$. There are many possible predictor functions with this property.

Exponential smoothing: a brief reminder. Empirically, it was found that an *exponential smoothing* predictor (also known as *exponential moving average*), in which $f_t = \alpha \cdot (1 - \alpha)^{t-1}$ for some $\alpha < 1$, works very well. Exponential smoothing, originally proposed in [3, 8], has become one of the main econometric tools. It is described one of the basic methods described in textbooks (see, e.g., [7]), it is used in many serious econometric studies; see, e.g., [5] and references therein.

When $\alpha < 1$, exponential smoothing provides a weighted average of all previous values of the quantity. When $\alpha > 1$ and $1 - \alpha < 0$, the above formula leads to the estimate $X_0 = \alpha \cdot f_1 - \alpha \cdot (\alpha - 1) \cdot f_2 + \ldots$ which can be represented as $X_0 = \text{const} \cdot f_1 + \alpha \cdot (\alpha - 1) \cdot (f_1 - f_2) + \ldots$, i.e., as attempt to add the expected trend (approximately estimated as $f_1 - f_2$) to the last observed value.

Why exponential smoothing is empirically efficient: what is known. There exist many explanations for the usefulness of exponential smoothing. However, these explanations are usually based on complex, not very intuitively clear statistical models; see, e.g., [9].

What we do in this section. In this section, we provide a new (and rather simple) theoretical explanation for the empirical success of exponential smoothing.

Monotonicity relative to the influence of noise. We have already considered the situation when the original signal x_t is a constant: $x_t = c$ for all moments of time t. In this case, we should have $X_0 = c$.

The natural next case if when the actual signal is constant $x_t^{\rm act} = c$, but the observed signal x_t also contains noise: $x_t = x_t^{\rm act} + n_t = c + n_t$, for some values n_t. Let us consider one specific noise pattern, i.e., a specific sequence of noise values $p_1, \ldots, p_k$. This pattern may occur at the end of the observation period, in which case we have $x_i = c + p_i$ for $i = 1, \ldots, k$ and $x_t = c$ for all other values t. This pattern may have ended right before the m-th moment, in which case $x_{m+i} = c + p_i$ for $i = 1, \ldots, k$ and $x_t = c$ for all other values t.

It is reasonable to require that the farther in the past is this noise pattern, the smaller the the smaller the effect of this noise value on our estimate Y_0, i.e., the smaller the absolute value $|X_0 - c|$ of the difference between the estimate X_0 and the no-noise estimate c. Thus, we arrive at the following definitions.

Definition 1.

- *By a* time series x, *we mean an infinite sequence of real numbers* $x_1, \ldots, x_n, \ldots$
- *By a* predictor function f, *we mean an infinite sequence of real numbers* $f_1, \ldots, f_n, \ldots$ *for which* $\sum\limits_{t=1}^{\infty} f_t = 1$.
- *By the* prediction $X_0(f, x)$ *made by the predictor function* f_t *for the time series* x_t, *we mean the value* $\sum\limits_{t=1}^{\infty} f_t \cdot x_t$.
- *By a* noise pattern p, *we mean a finite sequence of real numbers* $p_1, \ldots, p_k$.
- *Let* c *be a real number, and let* m *be a natural number. By* $x(p, c, m)$, *we mean a time series for which* $x_{m+i} = p_i$ *for* $i = 1, \ldots, k$, *and* $x_t = c$ *for all other* t. *We say that this time series* $x(p, c, m)$ *corresponds to a* a constant signal plus a noise pattern p before moment m.
- *We say that for a predictor function* f_t, the effect of noise always decreases with time *if for every noise pattern* p, *for every real number* c *and for every two natural numbers* $m > m'$, *we have*

$$|X_0(f, x(p, c, m)) - c| \leq |X_0(f, x(p, c, m')) - c|.$$

Proposition 1.

- *For every $\alpha \in (0,2)$, for the predictor function $f_t = \alpha \cdot (1-\alpha)^{t-1}$, the effect of noise always decreases with time.*
- *If for a predictor function f, the effect of noise always decreases with time, then there exists a constant $\alpha \in (0,2)$ for which $f_t = \alpha \cdot (1-\alpha)^{t-1}$.*

Comment. This result shows that exponential smoothing is the only predictor for which the effect of noise always decreases with time. Thus, the need to satisfy this natural property explains the efficiency of exponential smoothing.

Proof.

1°. The above condition – that the effect of noise always decreases with time – is formulated in terms of the differences $X_0(f, x(p, c, m)) - c$. To simplify the analysis of this condition, let us first find an explicit expression for this difference.

Due to the definitions of $X_0(f, x)$ and of $x(p, c, m)$, we have

$$X_0(f, x(p, c, m)) = \sum_{t=1}^{\infty} f_t \cdot x_t = \sum_{i=1}^{k} f_{m+i} \cdot (c + p_i) + \sum_{t \neq m+i} f_t \cdot c.$$

Here, $\sum\limits_{i=1}^{k} (c + p_i) = \sum\limits_{i=1}^{k} f_{m+i} \cdot c + \sum\limits_{i=1}^{k} f_{m+i} \cdot p_i$, so

$$X_0(f, x(p, c, m)) = \sum_{i=1}^{k} f_{m+i} \cdot c + \sum_{i=1}^{k} f_{m+i} \cdot p_i + \sum_{t \neq m+i} f_t \cdot c.$$

By re-arranging this sum, we get

$$X_0(f, x(p, c, m)) = \sum_{i=1}^{k} f_{m+i} \cdot c + \sum_{t \neq m+i} f_t \cdot c + \sum_{i=1}^{k} f_{m+i} \cdot p_i.$$

The first two sums in the right-hand side form the sum $\sum\limits_{t=1}^{\infty} f_t \cdot c$, so

$$X_0(f, x(p, c, m)) = \sum_{t=1}^{\infty} f_{m+i} \cdot c + \sum_{i=1}^{k} f_{m+i} \cdot p_i.$$

In the sum $\sum\limits_{t=1}^{\infty} f_t \cdot c$, the value c is a constant factor, so it can moved outside the sum. As a result, we get $\sum\limits_{t=1}^{\infty} f_t \cdot c \cdot \sum\limits_{t=1}^{\infty} f_t$. By definition of a prediction function,

$\sum_{t=1}^{\infty} f_t = 1$, hence $\sum_{t=1}^{\infty} f_t \cdot c = c$, and

$$X_0(f, x(p, c, m)) = c + \sum_{i=1}^{k} f_{m+i} \cdot p_i,$$

thus,

$$X_0(f, x(p, c, m)) - c = \sum_{i=1}^{k} f_{m+i} \cdot p_i.$$

2°. Let us now prove that for the predictor function $f_t = \alpha \cdot (1 - \alpha)^{t-1}$, the effect of noise always decreases with time.

Indeed, for this predictor function, we have

$$X_0(f, x(p, c, m)) - c = \sum_{i=1}^{k} f_{m+i} \cdot p_i = \sum_{i=1}^{k} \alpha \cdot (1 - \alpha)^{m+i-1} \cdot p_i.$$

Here, $(1 - \alpha)^{m+i} = (1 - \alpha)^m \cdot (1 - \alpha)^{i-1}$ and thus,

$$X_0(f, x(p, c, m)) - c = \sum_{i=1}^{k} \alpha \cdot (1 - \alpha)^{m-1} \cdot (1 - \alpha)^{i-1} \cdot p_i.$$

The term $(1 - \alpha)^m$ is a constant factor, so it can taken out of summation:

$$X_0(f, x(p, c, m)) - c = (1 - \alpha)^m \cdot \sum_{i=1}^{k} \alpha \cdot (1 - \alpha)^{i-1} \cdot p_i.$$

For m', we similarly have

$$X_0(f, x(p, c, m')) - c = (1 - \alpha)^{m'} \cdot \sum_{i=1}^{k} \alpha \cdot (1 - \alpha)^{i-1} \cdot p_i.$$

Thus, the ratio of these two difference has the form

$$\frac{X_0(f, x(p, c, m)) - c}{X_0(f, x(p, c, m')) - c} = \frac{(1 - \alpha)^m}{(1 - \alpha)^{m'}} = (1 - \alpha)^{m-m'}.$$

Thus,

$$\left| \frac{X_0(f, x(p, c, m)) - c}{X_0(f, x(p, c, m')) - c} \right| = \frac{|X_0(f, x(p, c, m)) - c|}{|X_0(f, x(p, c, m')) - c|} = |1 - \alpha|^{m-m'}.$$

Since $\alpha \in (0, 2)$, we have $-1 < 1 - \alpha < 1$, i.e., $|1 - \alpha| < 1$. Since $m > m'$, we have $|1 - \alpha|^{m-m'} < 1$, hence $\frac{|X_0(f, x(p, c, m)) - c|}{|X_0(f, x(p, c, m')) - c|} < 1$, and thus indeed $|X_0(f, x(p, c, m)) - c| \le |X_0(f, x(p, c, m')) - c|$. The statement is proven.

3°. To complete the proof, let us prove that if the effect of noise always decreases with time, then the predictor function has the form $f_t = \alpha \cdot (1-\alpha)^{t-1}$ for some $\alpha \in (0,2)$.

3.1°. Let us first consider a one-value pattern $p = (p_1)$, with $p_1 = 1$ and $m' = m-1$. In this case,

$$X_0(f, x(p,c,m)) - c = \sum_{i=1}^{k} f_{m+i} \cdot p_i = f_{m+1},$$

and similarly,

$$X_0(f, x(p,c,m')) - c = X_0(f, x(p,c,m-1)) - c = \sum_{i=1}^{k} f_{m-1+i} \cdot p_i = f_m.$$

For this example, the requirement that the effect of noise decreases with time means that

$$|f_{m+1}| \le |f_m|.$$

This means, in particular, that if $f_m = 0$, then $f_{m+1} = 0$.

3.2°. We can now conclude that $f_1 \ne 0$.

Indeed, if we had $f_1 = 0$, then, due to Part 3.1 of this proof, we would have $f_2 = 0$, then $f_3 = 0$, etc. – i.e., $f_t = 0$ for all t, which contradicts to our condition that $\sum\limits_{t=1}^{\infty} f_t = 1$.

3.3°. Let us now consider the case when $f_1 \ne 0$ and $f_2 = 0$. In this case, due to Part 3.1 of this proof, we conclude that $f_3 = \ldots = 0$, hence the condition $\sum\limits_{t=1}^{\infty} f_t = 1$ implies that $f_1 = 1$. This is a particular case of exponential smoothing corresponding to $\alpha = 1$.

3.4°. To complete the proof, it is thus sufficient to consider the remaining case $f_1 \ne 0$ and $f_2 \ne 0$.

3.5°. To analyze this case, let us consider a general situation when we have $f_m \ne 0$ and $f_{m+1} \ne 0$ for some integer m. Let us consider a two-value pattern $p = (p_1, p_2)$, with $p_2 = 1$ and $m' = m-1$. In this case,

$$X_0(f, x(p,c,m)) - c = \sum_{i=1}^{k} f_{m+i} \cdot p_i = f_{m+1} \cdot p_1 + f_{m+2},$$

and

$$X_0(f, x(p,c,m')) - c = X_0(f, x(p,c,m-1)) - c = \sum_{i=1}^{k} f_{m-1+i} \cdot p_i = f_m \cdot p_1 + f_{m+1}.$$

For this example, the requirement that the effect of noise decreases with time means that

$$|f_{m+1}\cdot p_1+f_{m+2}|\le |f_m\cdot p_1+f_{m+1}|.$$

For $p_1=-\dfrac{f_{m+1}}{f_m}$, we have $f_{m+1}=(-p_1)\cdot f_m$, hence $f_m\cdot p_1+f_{m+1}=0$ and thus, we must have $f_{m+1}\cdot p_1+f_{m+2}=0$ as well, i.e., $f_{m+2}=(-p_1)\cdot f_{m+1}$. From $f_{m+1}\ne 0$, we conclude that

$$\frac{f_{m+1}}{f_m}=\frac{f_{m+2}}{f_{m+1}}=-p_1.$$

3.6°. We start with the values $x_1\ne 0$ and $x_2\ne 0$. Then, by induction, from using Part 3.5 of this proof, we get

$$\frac{f_1}{f_2}=\frac{f_3}{f_2}=\ldots=\frac{f_{m+1}}{f_m}=\ldots$$

for all m. Thus, the values f_m form an arithmetic progression. If we take α such that $1-\alpha=-p_1$ (i.e., $\alpha=1+p_1$), then we conclude that $f_t=f_1\cdot(1-\alpha)^{t-1}$. The requirement that $\sum\limits_{t=1}^{\infty}f_t=1$ implies that this geometric progression converges, i.e., that $-1<1-\alpha<1$ and $0<\alpha<2$. It is known that the sum of the geometric progression z^m is equal to

$$\sum_{m=0}^{\infty}z^m=\frac{1}{1-z},$$

hence

$$\sum_{t=1}^{\infty}f_t=f_1\cdot\sum_{t=1}^{\infty}(1-\alpha)^{t-1}=f_1\cdot\sum_{m=0}^{\infty}(1-\alpha)^m=f_1\cdot\frac{1}{1-(1-\alpha)}=f_1\cdot\frac{1}{\alpha}=1,$$

hence $f_1=\alpha$. Thus, $f_t=\alpha\cdot(1-\alpha)^{t-1}$. The proposition is proven.

3 Application: A Simple Explanation of Asymmetric Price Transmission

As an example of an application of this idea, let us consider the phenomenon of asymmetric price transmission.

What is price transmission. The price of a manufacturing product is determined by the price of the components and the price of the labor. If one of the component prices changes, this change affects the product's price; this is called *price transmission.*

For example, when the oil price changes, the gasoline prices change as well; when the gasoline prices change, the transportation prices change as well; when the transportation prices change, the price of all the goods that need to be transported change as well, etc.

What is asymmetric price transmission. In many cases, price transmission is *asymmetric* in the following sense: when the component (input) price increases, the final product (output) price starts increasing right away. On the other hand, when the input price starts decreasing, the output price sometimes stays high and does not decrease at all – sometimes, decreases somewhat but much slower than it was decreasing. As a result, when the input price falls to the original lower level, the output price remains much higher than the original one; see, e.g., [11, 12, 17].

Why this is a problem. This phenomenon seems to contradict to the usual economic assumption that markets are efficient, and that the price of each product is determined by the equilibrium of supply and demand. Under this assumption, when the demand remains fixed and the labor costs remain fixed, the price of the final product should be uniquely determined by the price of the components. However, in reality, when the component prices goes back to the original level, the final product price does not get back to the same level, it remains higher.

How this problem is solved now. There exist explanations of this phenomenon; see, e.g., [4, 10, 16] and references therein. However, these explanations are based on complex models and are far from intuitive clarity.

What we do in this section. In this section, we provide a simple explanation for the phenomenon of asymmetric price transfer.

Our explanation. To describe the main idea behind our explanation, let us consider an often-cited example of price transmission from crude oil prices to gasoline prices. A gasoline company periodically buys crude oil, transforms it into gasoline, and sells this gasoline to the customers. When the company's supply of oil decreases below its re-order threshold, it orders a new oil supply.

This ordering of new oil has to be in the future. Even if the company suspects that the price will increase, it is still usually unable to buy needed oil now, at when the price is still lower: the company may simply not have enough storage for all this oil which will only be needed in the future.

As a result, for the company to stay even, it should determine the price it charges to the customers *not* based on the current price of oil, but rather on its *future* price – at the time when it will re-supply it.

Of course, we do not actually know the exact future price, so we have to use a *predicted* price. As we have seen in the previous section, to predict the price, we should use not only the current price, but also prices at previous moments of time. Let us show, on a simplified example, that this simple idea indeed leads to what we observe as asymmetric phase transmission.

For simplicity, let us assume that as a predictor of the future price, we use the arithmetic average of the two previous observations.

- Let us assume that in the past, the oil price was steady at $20 per barrel. During this trend, the predicted price is also equal to $\frac{20+20}{2} = 20$.
- Then, one year, the oil price shoots to $100. At this moment, the predicted price, based on the two previous values of $20 and $100, becomes equal to $\frac{20+100}{2} = 60$, a threefold increase.
- In the next year, the price falls back to the old amount $20. Here, the predicted price, based on the two previous values $100 and $20, is equal to $\frac{100+20}{2} = 60$. In other words, the oil price decreased back to the original level, but the gasoline price did not decrease at all!
- Finally, in the year after that, when the oil stays at its low price of $20, the average becomes again equal to $\frac{20+20}{2} = 20$, i.e., finally, the gasoline price goes down.

We see that when the oil price increased, the gas price increased right away, but when the oil price decreased to the old level, it took the whole year for the gas price to go down to the old level. This is exactly what is called the asymmetric price transmission.

Comment. A general idea that many problems can be resolved if we take into account the dynamic, time-changing character of many real-life phenomena, has been advocated, on numerous examples, by L. Perlovsky; see, e.g., [13, 14, 15].

Additional intuitive arguments in favor of our explanation. To explain the observed phenomenon of asymmetric price transmission, we considered a situation in which the price of the component remains stable and then experiences a sudden increase. In this case, we have an abrupt increase in the customer price of the final product – both empirically and in accordance with our explanation.

We can also consider an opposite situation: when the price of the component remains stable and then experiences a sudden decrease. In this case, our explanation predicts a sudden decrease in the customer price of a final product as well. We indeed observe such a phenomenon on the example of consumer electronics: when the computer chips become cheaper, many electronic products become cheaper as well.

Comment. One might argue that still, in real life, we observe more cases when consumer prices go up fast but go down slower, and similarly cases when consumer prices go down fast but go up slower are rarer. This would indeed somewhat contradict our justification if prices, on average, went down as frequently as they go up – thus remaining, on average, stable. In reality, the average prices go up, there is a small but steady inflation, which means that indeed, prices going up are more frequent than prices going down. Thus, the observed phenomenon of asymmetric price transmission (when the prices suddenly go up and then go somewhat down) is indeed more frequent than the opposite phenomenon: when consumer prices suddenly go down, and then somewhat go back up.

4 Second Result: A Simple Explanation of Why Volatility Increase is a Good Predictor of Trend Reversal

Formulation of the problem. In the previous sections, we described the problem of predicting the new value when we are within a certain trend. Another important problem is predicting when a trend will reverse, e.g., when recession will end – or, vice versa, when an economic boom is expected to end.

Volatility increase as a good predictor of trend reversal: an empirical fact. It is a known empirical fact that volatility – measured, e.g., the standard deviation of the value from its local average – tends to increase before trend reversals. Thus, such volatility increases are a known predictor of trend reversals.

This empirical fact is somewhat counter-intuitive. Economists and econometricians have used volatility increase as a predictor of trend reversal for so long that to most of them, this relation is natural and intuitive. However, as well see, from the purely mathematical viewpoint, this predictor is somewhat counter-intuitive.

Indeed, from the mathematical viewpoint, when the trend changes from decrease to increase, this means that the corresponding quantity reaches its minimum. When the trend changes from increase to decrease, this means that the corresponding quantity reaches its maximum. The problems of detecting minimum and maximum are known in mathematics for a long time. From the fundamental viewpoint, the questions of how to detect minima and maxima have been solved by Newton's and Leibnitz's calculus: at the point where a function attains its minimum or its maximum, the derivative of this function is equal to 0.

The derivative is, crudely speaking, a measure of how much (and in what direction) a function changes locally (in the vicinity of a given point). In these terms, the result about zero derivative means that when a function attains minimum or maximum, then the corresponding measure of local change attains its smallest possible value (value 0). Now, for econometric sequences, the opposite seems to be empirically true: in trend reversal situations (i.e., when the corresponding function attains its minimum or maximum), the corresponding measure of local change (in this case, volatility) attains its largest value!

This is what we meant by saying that the above empirical fact is somewhat counter-intuitive.

Towards an explanation. Let us provide a simple explanation of the empirical fact, an explanation that will hopefully make this fact more intuitively convincing. As an example of a time series, let us consider a stock price. With respect to the given stock, some traders are optimistic, some are pessimistic.

An optimistic trader believes that the stock will rise, so he/she is willing to pay a little extra for this stock – in the expectation of larger gains in the future. A pessimistic trader believes that the stock will go down, so he/she is willing to sell this stock even for a price which is somewhat lower than the last recorded price for this stock – because if the trader does not sell this stock even at this price, he/she may lose more when this stock decreases in value.

The overall price of the stock can be computed as an average over all the transactions. Let x be the last recorded price for the stop, and let δ be the average value of the small increase/decrease in stock in transactions by optimists and pessimists. We are interested in the average behavior of all the traders in the market, and in such an average behavior, individual differences tend to average out. Thus, it seems safe to ignore the individual differences between the corresponding increases and decreases and simply assume that each optimist performs transactions with this stock at the price $x + \delta$, while each pessimist performs transactions at the price $x - \delta$.

Let p be the proportion of optimists, i.e., the probability that a randomly selected trader is an optimist with respect to this stock. To further simplify our description, we will also assume that all the traders are independent from each other. Let n denote the total number of traders. Thus, we arrive at the following model:

- We start with the price x.
- At the next moment of time, we have a price $x' = \dfrac{x_1 + \ldots + x_n}{n}$, where $x_i = x + \eta_i \cdot \delta$ and η_i are independent random variables each of which attains:
  - the value $\eta_i = 1$ with probability p, and
  - the value $\eta_i = -1$ with the remaining probability $1 - p$.

Since the variables x_i are independent and identically distributed, their expected values coincide: $E[x_1] = \ldots = E[x_n]$. Thus,

$$E[x'] = \frac{E[x_1] + \ldots + E[x_n]}{n} = E[x_i],$$

where

$$E[x_i] = x + E[\eta_i] = x + (p \cdot 1 + (-1) \cdot (1 - p)) = x + (2p - 1) \cdot \delta.$$

According to this formula, when $2p - 1 > 0$, i.e., when $p > 1/2$, the price increases; when $2p - 1 < 0$, i.e., when $p < 1/2$, the price decreases. The trend reverses itself when p is equal to the threshold value $p = 1/2$.

For this situation, as a measure of volatility, we will take standard deviation $\sigma = \sqrt{V}$. For each x_i, the variance $V[x_i]$, i.e., the expected value of the square $(x_i - E[x_i])^2$ of the difference $x_i - E[x_i]$. When $x_i = x + \delta$, this difference is equal to

$$(x_i - E[x_i])^2 = (x + \delta) - (x + (2p - 1) \cdot \delta) = 2 \cdot (1 - p) \cdot \delta.$$

When $x_i = x - \delta$, this difference is equal to

$$(x_i - E[x_i])^2 = (x - \delta) - (x + (2p - 1) \cdot \delta) = -2 \cdot p \cdot \delta.$$

Thus, the variance is equal to

$$V[x_i] = p \cdot 4 \cdot (1-p)^2 \cdot \delta^2 + (1-p) \cdot 4 \cdot \delta^2 = 4 \cdot \delta^2 \cdot p \cdot (1-p) \cdot ((1-p)+p) = 4 \cdot \delta^2 \cdot p \cdot (1-p).$$

The variance of the sum of n independence random variables is equal to the sum of their variances, so

$$V[x_1 + \ldots + x_n] = \sum_{i=1}^{n} V[x_i] = 4 \cdot n \cdot \delta^2 \cdot p \cdot (1 - p).$$

Thus, for the standard deviation $\sigma = \sqrt{V}$, we get

$$\sigma[x_1 + \ldots + x_n] = 2 \cdot \delta \cdot \sqrt{n} \cdot \sqrt{p \cdot (1 - p)}.$$

When we divide a random variable by a constant (in this case, by n), its standard deviation is divided by the same constant, so we get

$$\sigma[x'] = \frac{\sigma[x_1 + \ldots + x_n]}{n} = 2 \cdot \delta \cdot \frac{1}{\sqrt{n}} \cdot \sqrt{p \cdot (1 - p)} = \text{const} \cdot \sqrt{p \cdot (1 - p)}.$$

The value $p \cdot (1 - p)$ is the smallest when $p = 0$ and $p = 1$ and attains its largest value when $p = 1/2$. Thus, in this simple model, *volatility is indeed the largest when the trend reverses – exactly as empirically observed.*

Acknowledgments

This work was supported in part by the Project DAR 1M0572 from MŠMT of Czech Republic, by the National Science Foundation grants HRD-0734825 and DUE-0926721, and by Grant 1 T36 GM078000-01 from the National Institutes of Health. The authors are thankful to the anonymous referees for their help.

References

[1] J. Aczél and J. Dhombres, *Functional Equations in Several Variables*, Cambridge University Press, 2008.

[2] A. Bovas and J. Ledolter, *Statistical Methods for Forecasting*, Wiley, New York, 2003.

[3] R. G. Brown, *Statistical Forecasting for Inventory Control*, McGraw Hill, New York, 1959.

[4] G. Frei and M. Manera, *Econometric Models of Asymmetric Price Transmission*, Fondazione Eni Enrico Mattei Working Paper 100, Milan, Italy, 2005.

[5] E. S. Gardner, "Exponential smoothing: state of the art. Part II", *International Journal of Forecasting*, 2006, Vol. 22, pp. 637–666.

[6] J. D. Hamilton, *Time Series Analysis*, Princeton University Press, Princeton, New Jersey, 1999.

[7] M. Hirschey, *Fundamentals of Managerial Economics*, Southwestern Cengage Publ., Mason, Ohio, 2009.

[8] C. C. Holt at al, *Planning Production, Inventories, and Work Force*, Prentice Hall, Englewood Cliffs, New Jersey, 1960, Chapter 14.

[9] A. Koehler, K. Ord, and R. D. Snyder, *Rationalization of Exponential Smoothing in Terms of a Statistical Framework with Multiple Disturbances*, Monash University, Australia, Department of Econometrics, 1992, Paper No. 92/7.

[10] E. Krivonos *The impact of Coffee Market Reforms on Producer Prices and Price Transmission*, World Bank Policy Research Working Paper 3358, 2004.

[11] J. Meyer and S. von Cramon-Taubadel, "Asymmetric price transmission: a survey", *Journal of Agricultural Economics*, 2004, Vol. 55, No. 3, pp. 581–611.

[12] S. Peltzman, "Prices rise faster than they fall", *Journal of Political Economy*, 2000, Vol. 108, No. 3, pp. 466–501.

[13] L. I. Perlovsky, "Neural Network with Fuzzy Dynamic Logic", In: *Proceedings of the In- ternational IEEE and INNS Joint Conference on Neural Networks IJCNN'05*, Montreal, Quebec, Canada, 2005.

[14] L. I. Perlovsky, "Fuzzy Dynamic Logic", *New Math. and Natural Computation*, 2006, Vol. 2, No. 1, pp. 43–55.

[15] L. I. Perlovsky, "Neural Networks, Fuzzy Models and Dynamic Logic", In: R. Köhler and A. Mehler, eds., *Aspects of Automatic Text Analysis: Festschrift in Honor of Burghard Rieger*, Springer, Germany, 2007, pp. 363–386.

[16] M. E. Tappata, “Rockets and feathers: understanding assumetric price transmission”, *RAND Journal of Economics*, 2009, Vol. 40, No. 4, pp. 673–687.

[17] S. S. Wlazlowski, “Petrol and crude oil prices: asymmetric price transmission”, *Ekonomia / Economics*, 2003, Vol. 11, pp. 1–25.