ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εθνικόν και Καποδιστριακόν
Πανεπιστήμιον Αθηνών
ΤΜΗΜΑ ΟΙΚΟΝΟΜΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

Διδακτορική Διατριβή
Τμήμα Οικονομικών Επιστημών

The Profit-Based Approach
Αποτίμηση Κεφαλαιακών Τίτλων από τα Θεμελιώδη Μεγέθη: Η Προσέγγιση των Κερδών

Νικόλαος Κ. Στραβελάκης

Αθήνα, Ιανουάριος 2020

Chapter 4
The empirical testing of the profits-based approach

Introduction

The fourth and final chapter of this project refers to the empirical testing of certain aspects of the profit-based approach. A full test of all aspects of the theory requires a good deal of development and application of empirical techniques that needs more than a dissertation. This will become evident from the present chapter although it is confined to the empirical estimation of the profit-based approach on stocks. The reason is that additional econometric techniques are required to assess certain parts of the theory even on stocks alone.

The exposition begins by replicating and extending to the present time the empirical tests of the theory conducted by Anwar Shaikh (1997, 2016). This is done in section 4a. Shaikh (1997, 2016) estimated stock prices and stock returns against the incremental rate of profit. Specifically, he estimated the S&P 500 rate of return from the Shiller data base and calculated the incremental rate of profit from the BEA tables. All estimations are described and explained in section 4a and in Appendix 4.1. The National Income and Product Accounts (NIPE) of the U.S. Bureau of Economic Analysis (BEA) data on profitability goes back to 1947 and for this reason this is the starting year for the calculations. As it will become evident in section 4a the results are supportive of the profit-based approach.

Nevertheless, the time series of both and S&P 500 rate of return and the incremental rate of profit are not stationary. This raises estimation issues. Specifically, it was impossible for Shaikh to estimate the model parameters and draw statistical inference on them. For this reason, he restricted himself in drawing inference coming from the simulation of the warranted against the actual S&P 500 prices presented below in Figure 4.3 (Shaikh 2016) and the calculation of the R^2 from detrended data as presented in Figure 4.2 (Shaikh 1997). Although, the results are strongly supportive. no specific inference could be drawn on the influence of the incremental rate of profit on actual prices and returns of the S&P 500.

In order to address these issues, I have modified the empirical model. This is elaborated in section 4b. Specifically, I have used the rate of growth of the Earnings Per Share (EPS) of the companies comprising the S&P 500 (the data is recorded in the Shiller data base) as a proxy for the incremental rate of profit.[1] I argue that the change in the EPS of these companies (between the previous and the current year) is a good proxy of changes in the profitability of the 'regulating capital' (see section 3a for the definitions). Assuming further that investment is a linear function of the last period

---

[1] The EPS is calculated from net corporate earnings divided by the number of shares (excluding from the denominator the number of shares comprising corporate buybacks). It is the key measure of corporate profitability for bankers and fund managers as indicated by George Soros (1994) and elaborated in section 3.6.

corporate profit, the rate of growth of the EPS becomes an approximation of the incremental rate of profit ($Irop_t$).[2] This permits us to regress the S&P 500 detrended logarithmic values against the logarithm of the EPS for a period going back to 1900. Following Shiller (1989b: 78-82), and Shaikh (1997: 398), I detrended the data using the 30-year moving average. The results are impressive and enable us to draw statistical inference for the model parameters.

Yet, the most important contribution of this chapter has to do with the estimation of the dynamics of the correlation between the incremental rate of profit and the rate of return of the S&P 500. Shaikh (1997) correctly pointed out that the two variables are not linearly related and for this reason, the Pearson statistic takes a small value. Nevertheless, the incremental rate of profit and the stock market rate of return have roughly the same mean and standard deviation. This implies among other issues that for the profit-based approach there is no unexplained volatility of stock returns since stock market volatility reflects the variability of the underlying corporate fundamentals. In section 4c.6 this finding is elaborated further. I will apply a non-parametric statistic named 'Mutual Information' (MI). It measures the reduction of the uncertainty about stock returns from knowing the corporate fundamentals. It is a non-linear correlation statistic originally applied by Shannon (1948) that has been incorporated and developed in the context of Transfer Entropy (TE). The application will reveal important patterns on the dynamics of the relation between stock market returns and corporate fundamentals especially in the transition from normal times to market crash. As it has been argued analytically in Chapter 3, for the profit-based approach financial turmoil is the trigger rather than the cause of economic crises. The idea is that corporate fundamentals deteriorate before the stock market crashes and not the other way around. Using data going back to 1880 I will show empirically that this is the case, and that 'phase transition' in the stock exchange reflects a pattern explained in Soros' 'reflexivity theory' elaborated in the previous chapter.

Although, the profit-based approach has not been tested extensively even for stocks, in all of the few empirical tests the findings are highly supportive. This will become evident from the presentation of Shaikh's (1997, 2016) estimations, as well as my contribution that follows. Overall, empirical testing provides a strong initiative for further elaboration on the analytical and empirical findings of this project.

---

[2] The idea is kind of simple as outlined in the equations that follow. I start from the definition of the incremental rate of profit and arrive at the approximation. A full explanation is provided in section 4b.

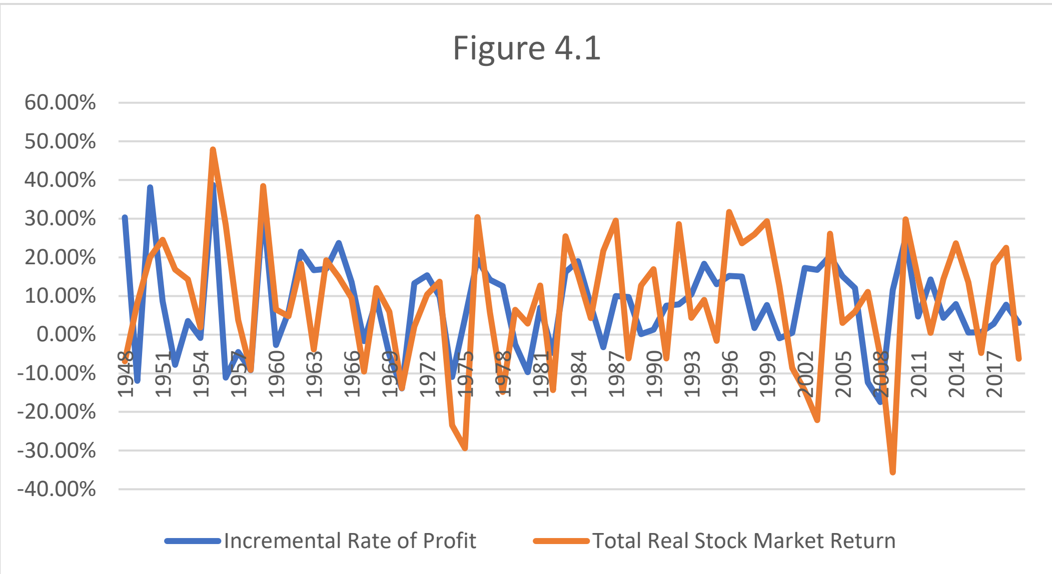
$$Irop_t = \frac{P_t - P_{t-1}}{I_{t-1}} \; if \; P_t - P_{t-1} \approx EPS_t - EPS_{t-1} \; and \; I_{t-1} = \rho \cdot P_{t-1} \; and \; \rho > 0 \; \rightarrow Irop_t = \frac{EPS_t - EPS_{t-1}}{EPS_{t-1}}$$

$$and \; P_t = corp.\ profit, I_{t-1} = Investment, EPS = Earnings\ per\ Share, Irop_t = Incr.Profit\ Rate$$

4a. The main assumptions of the profit-based approach on stocks - Summary and Empirical Handling

The main assumption of the profit-based approach when it comes to stock pricing is that stock market returns tend to become equal to the returns in the corporate sector. This is not something new, since mainstream theory also assumes equalization of risk-adjusted rates of return (Chapter 2). However, in the profit-based approach, the equalization of returns takes place around the volatile incremental and not the constant or slowly varying average rate of return (profit) of mainstream theory. The application of the incremental rate of profit is based on the Classical/Marxian theory of competition. In this context, equalization takes place between the incremental profit rates of the regulating capitals of each industry (Chapter 3 Section 3a). Given that in competitive economies corporations are expected to constantly bring to the market new products and apply new techniques, the incremental rate of profit is expected to be a highly volatile measure. Its volatility is enhanced further from the structure of expectations (Chapter 3 section 3f). We saw that in the 'reflexivity theory', expectations affect prices which, in turn, affect fundamentals that reflect upon prices, and so on. This means that stock price investments are inherently short term, since persistent variations in the incremental rate of profit bring forward new positions of risky arbitrage, or 'turbulent arbitrage', as Shaikh (1997) calls it. In other words, equalization is a dynamic and evolving process around an equilibrium path.

For this reason, when it comes to empirical testing the reasonable thing to do is to directly associate the incremental rate of profit with stock prices and returns. Although the assumption is that the incremental rate of profit of regulating capitals tends to become equalized, following Shaikh 1997, we will begin by calculating the time series of the average incremental rate of profit. the latter is the independent variable in our calculations. The dependent variable is the rate of return of the S&P 500 as presented in the publicly available database of the Nobel prize laureate Robert Shiller that can be found online (**Shiller 1**). The time series are pictured in the chart that follows:



Source: author's calculations

Figure 4.1 extends previous calculations of the time series of the (adjusted) incremental rate of profit (blue line) to the real total return (including the dividend yield) of the S&P 500 as recorded in the Shiller database. The incremental rate of profit is the ratio of the yearly change in real corporate profits (including depreciation) divided by real investment. Data sources (BEA tables), deflators, and formulas applied in the calculations are detailed in Appendix 4.1. The calculation covers the period from 1948-2019. It is an extension of previous calculations of the same series in Shaikh (1997) covering the period from 1948 – 1992 and in Shaikh (2016: 470-471) for the period 1948-2011.
In all three calculations, the time series retains the same properties. They have almost the same mean, standard deviation, and coefficient of variation. The prices of these descriptive statistics are detailed in table 4.1 that follows:

Table 4.1

	IROP	**Total Return S&P500**
Average	7.77%	8.46%
Standard deviation	11.90%	16.11%
Coefficient of Variation	1.53	1.90

Although both the time series and the descriptive statistics indicate that the two variables are strongly associated the R^2 between the two is only 10% (the Pearson correlation coefficient is 0.31). The reason is that the correlation between them is non-linear. Nevertheless, the data summarized in table 4.1 prove that for the profit-based approach there is no 'unexplained volatility' in stock market prices. You will recall that this whole project began (Chapter 1 section 1b) by presenting Shiller's (1989 a, b) work on the matter. The latter had shown that the variability of stock prices cannot be justified by the variations of dividends discounted by a constant, or almost constant, required rate of return that underlies the efficient market hypothesis. Here, using the incremental rate of profit as the required rate of return, it is shown that the volatility you get in the stock market is the one you should expect. It results from the rough return equalization process between the stock market and the corporate sector.
There are, however, additional elaborations in the initial empirical evaluations of the profit-based approach for stocks that are worth mentioning. The initial reaction of Shaikh to the empirical findings outlined above was a calculation of an equation like 3.45. Nevertheless, to apply the traditional correlation statistics, the data needs to be detrended to avoid spurious correlation. Following Shiller (1989b), this was done by dividing stock market prices by the 30-year moving average of the earnings per share. Shaikh followed this practice to make his results comparable to those of Shiller. Specifically, in his (1989b: 78-82) book Shiller compares the detrended by the 30year moving EPS average S&P 500 prices to the DCF model prices calculated with a constant discount factor. From this comparison Shiller concluded that there exists excess, i.e., unexplained, volatility of actual stock prices as compared to DCF prices. By following

the exact same practice and changing only Shiller's constant discount factor with the incremental rate of profit Shaikh claims that any difference in the results of two models is due to different discount factors applied. The results of Shaikh s' (1997) estimation extended to 2020 is summarized in Figure 4.2

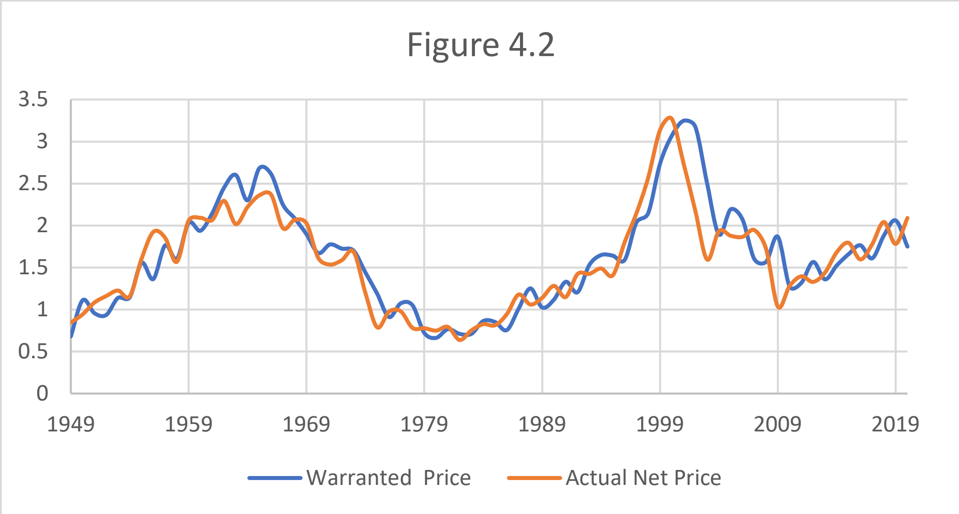


Figure 4.2 compares the warranted price (blue line) calculated from equation 3.45 ($Pr_{rt}{}^{w} = Pr_{t-1} \cdot [1 + (rror_t - divy_t)]$) with the detrended actual net price of the S&P 500 (brown line). The required rate of return is the incremental rate of profit calculated above from which I have deducted the dividend yield $divy_t$. The actual real price of the S&P 500 is taken from the Shiller database divided by the 30-year moving average of the earnings per share. The R^2 between the two variables is 0.8. The strong correlation holds also for subperiods. Specifically, Shaikh (1997) performing the same calculation found an R^2 of 0.875 for the period 1948-1993. These results are extremely strong compared to calculations of DCF models where the R^2 is never greater than 0.09 (Shiller 1989: 81-82, Barsky and De Long 1993). Moreover, the difference in the R^2 can be attributed to the application of the incremental rate of profit instead of the Shiller constant discount factor.

Finally, a third elaboration is presented in Shaikh (2016). It is based on the methodology applied by Shiller in his book *Irrational Exuberance* published in 2000 (Shiller 2009) and the databases he updates and makes available online ever since (**Shiller 1**). The difference with the previous model is that, in this case, the data is not detrended. This time Shaikh (2016) wanted to show that the warranted price calculated from the profit-based approach is the 'gravity center' of the actual price. Therefore, any deviations cannot be attributed to irrationality, as claimed by the behaviorists. As you may recall, in Chapter 2 (section 2d.4) the behaviorist approach was considered in the context of the alternatives offered by orthodox theorists to the empirical failure of mainstream asset pricing models. It suggests that mainstream models fail because agents are 'irrational'. The latter leads to positive or negative extremes in financial asset prices. However, the benchmark of rationality for this theory is the 'efficient market hypothesis' (EMH). Shiller presented this notion empirically using a version of the martingale model (Samuelson 1964), presented in section 2f.2, against the real actual price of the S&P 500. Specifically, he calculated an average interest rate for the period from 1871-1999 (7.6%) and used it as the constant discount factor. Then he discounted dividends and arrived at 'Present Value of Real Dividend Prices'. The latter appears in his database (**Shiller 2**) updated until 2009.

Shaikh performed a simulation of the equation 4.1 (appearing herein below). It is similar in concept to equation 3.45 with the difference that, in the former, warranted prices are calculated based on previous warranted prices ${P_{rt-1}}^{w}$ and not actual prices Pr_{t-1}. However, in order to perform the iteration an initial price must be estimated based on an acceptable criterion. Shaikh reproduced the calculation constructing an initial price that equalizes the warranted and actual price averages for the period 1948-1995. On this ground, he constructed a simulation of the profit-based approach warranted prices using equation 4.1.

$$4.1\ {P_{rt}}^{w} = {P_{rt-1}}^{w} \cdot [1 + \mathrm{rror}_t] - Div_t$$

$$and\ rror_t \approx Irop_t$$

The difference between the two calculations is the required rate of return. In equation 4.1 it is the highly volatile incremental rate of profit $Irop_t$ pictured in Figure 4.1 whereas the Shiller-EMH prices are calculated basis a constant rate of 7.6%. The results are summarized in Figure 4.3 below. The only thing I have added is that the profit-based approach warranted price calculation is extended to 2020.

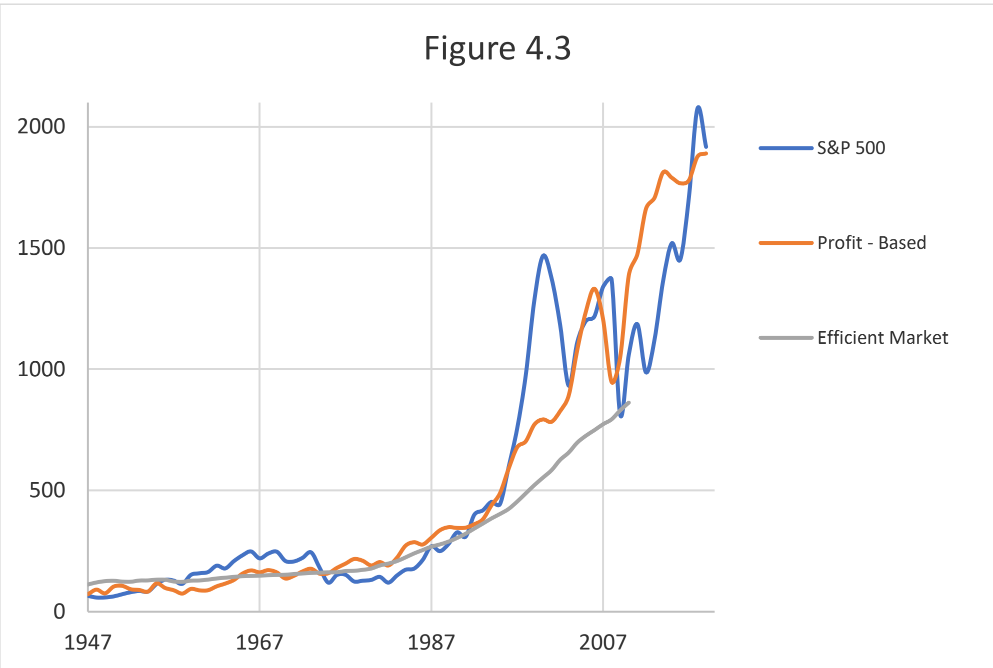


Figure 4.3 is indicative in many respects. The blue line which pictures the real price of the S&P 500 is the benchmark for the warranted price calculated from the profit-based approach (brown line) and the Efficient Market Hypothesis (grey line). For Shiller, the differences in actual prices from the grey line are indications of irrationality. The same is true of the volatility of actual prices compared to the smooth EMH prices. For the profit-based approach bubbles or underpricing can appear as indicated by Soros' reflexivity theory presented in Chapter 3 (section 3f). However, as Soros' reflexivity theory indicates, fundamentals rule in the end. This is confirmed in the simulation pictured above and it holds both in normal times as well as during financial and economic crises. The chart pictures the boom of the 1950s and the 1960s where actual prices exceed fundamental prices, the opposite happens during the times of the great stagflation (1970-1981) when actual prices underscore warranted prices. This pattern persists in the first years of the neoliberal era. However, following 1985 actual prices overshoot the underlying fundamentals reaching a climax in the dot com bubble (around 2000). The correction that follows was steep but short. Actual prices gravitated around fundamental prices and collapsed shortly after the collapse in the underlying fundamentals that preceded the 2007 crisis. A finding that supports the analytical approach of section 3d of Chapter 3 which argues that a financial crisis is only a trigger of major depressions. The irony is that Shiller's EMH estimated prices lose any association to the actual prices in the years of neoliberalism when the idea of 'self-regulated markets' was at its peak.

Nevertheless, the most striking part of the simulation is that when extended to 2019 it does not picture a bubble as most of us would expect (at least I did). Of course, we should keep in mind that the time interval for the calculation of the initial price was picked arbitrarily. For a different initial price, we would end with a different warranted price. For this type of calculation, an 'unobserved component model' identifying the relation between stock market returns and the incremental rate of profit is required. Here I will present a different (non-parametric) statistic to identify this relation. Although we cannot derive a warranted price by applying this method, it will prove a step forward for the empirical evaluation of the profit-based approach. But before we move to this, we can draw interesting statistical inference by applying traditional econometric methods to the detrended stock returns and EPS data. This will be presented in the next section.

4b. Linear Regression - Statistical Inference for Detrended Prices – Using EPS Data

One of the missing points in the empirical analysis of the profit-based approach for stocks is the absence of any direct statistical inference for the explanatory variables. However, the correlation between detrended data and corporate fundamentals pictured in Figure 4.2 is a good starting point for an econometric model from which we can draw statistical inference.

The first step for this is to approximate the Incremental Rate of Profit with the rate of growth of the Earnings Per Share (EPS). This will provide access to data going back to

1871 and perform econometric calculations with time series from 1900 to 2019 even when we detrend the data using the 30year moving average.[3] This will permit us to take advantage of the properties of large samples. Moreover, this handling of the data does not contradict the profit-based approach. George Soros (1994), who's reflexivity theory is an integral part of the profit-based approach, uses the EPS as the key fundamental in his stock valuations and investment decisions.

The EPS growth, when associated with the companies of the S&P 500, can be considered as a closer proxy of the incremental rate of profit of regulating capitals, rather than the average incremental rate of profit we have used so far. If one simply looks at the companies that comprised the index through the years, he will realize that most corporations that reshaped and created markets for more than a century were at their peak in the S&P 500 index. For example, companies like Apple, Microsoft, Dupont, and General Mills are currently members of the index. To put it differently, the 500 corporations with the greatest market capitalization (this is the basic criterion for the construction of the S&P 500 since 1988[4]) in the NY stock exchanges and maybe in the world are probably (although not necessarily) regulating capitals.

As far as the numerator of the incremental rate of profit formula $Irop_t = \frac{P_t - P_{t-1}}{I_{t-1}}$ is concerned, substituting profit differentials with the change in earnings per share $EPS_t - EPS_{t-1}$ is a good proxy for the change in profitability. The EPS is the ratio of net corporate profits to the number of shares adjusted for any share buyback.[5] Things are more complicated for the denominator, since we cannot directly construct time series of investment for the S&P 500 companies from their balance sheets. Nevertheless, we can find an approximation for this measure by making certain restrictive, but plausible, assumptions. I elaborate on this by using equation 3.14 as follows:

$$4.2 \quad \frac{K_t - K_{t-1}}{K_{t-1}} = s_t \cdot \left(r_{1_t} - i_t\right)$$

$$K_t = capital\ advanced, s_t = r.o.savings, r_{1_t} = regulating\ r.o.profit, i_t = r.o.interest$$

$$let\ i_t = \delta \cdot r_{1_t}\ and\ \delta > 0, I_t = K_t - K_{t-1} and\ s_t = s$$

$$\rightarrow I_t = s \cdot (1 - \delta) \cdot r_{1_t} \cdot K_{t-1}$$

[3] I keep this assumption so that the calculation will be comparable to Shiller (1989b) and Shaikh (1997).

[4] The composition criteria of the S&P 500 are not uniform throughout its history. To start with it did not always include 500 stocks. Originally it tracked only 233 stocks. In 1957 when it included 425 industrial, 50 utilities and 15 railway stocks it represented 90% of the total capitalization of the stock exchange. It is surprising that Financial companies were first included in the index in 1970. However, the objective throughout its history was to create a gauge for the market if not for the economy. The impact of the index is so great that some mainstream economists wonder if there exists an S&P 500 index effect on the prices of the stock that comprises the index (Kasch and Sarkar 2012).

[5] For example, if a company has net profits of one million euros and one million shares, its earnings per share will be 1 euro. If it buys back 200,000 shares its EPS will increase to 1.25 euros (1,000,000/800,000).

$$making\ use\ of\ r_t = \frac{P_t}{K_{t-1}}\ and\ asuming\ that\ r_1 = \frac{1}{1+z} \cdot r\ as\ in\ 3.15\ we\ get$$

$$\rightarrow \boldsymbol{I_t = \rho \cdot P_t\ and\ \rho = \frac{s \cdot (1-\delta)}{1+z}}$$

Equation 4.2 tells us that if the interest rate follows the regulating rate of profit r_{1_t} (the loan/ reserve ratio is constant) and that the ratio of the regulating to the average gross rate of profit is constant then investment I_t is a linear function of gross profitability. If we apply this to our modification of the incremental rate of profit it will read as follows:

$$4.3\ Irop_t \approx \frac{P_t - P_{t-1}}{I_{t-1}} = \theta \cdot \frac{EPS_t - EPS_{t-1}}{EPS_{t-1}}, \qquad \theta = \frac{1}{\rho} > 0$$

Equation 4.3 tells us that the incremental rate of profit is a linear function of the rate of growth of the earnings per share (EPS). Where earnings per share is the gross measure that includes dividends.

When it comes down to drawing inferences for the variables, we need to linearize these relations by using logarithms. In this regard we turn to log growth and linearize the following relation:

$$4.4\ ln\frac{Pr_t}{Pr_{t-1}} = \theta \cdot ln\frac{EPS_t}{EPS_{t-1}} \rightarrow \ln Pr_t - \ln Pr_{t-1} = \theta \cdot \ln EPS_t - \theta \cdot \ln EPS_{t-1} \rightarrow$$

$$\ln Pr_t = \ln Pr_{t-1} + \theta \cdot \log EPS_t - \theta \cdot \log EPS_{t-1}$$

$$and\ Pr_t = the\ price\ of\ the\ S\&P\ 500\ at\ time\ t$$

Equation 4.4 can be easily transformed into an econometric model where we can estimate θ using Ordinary Least Squares (OLS) provided that the two variables $\log Pr_t, \log EPS_t$ are stationary. This is achieved by dividing the time series of the S&P 500 and the EPS by their 30year moving average. The detrended variables are pictured in the following chart:

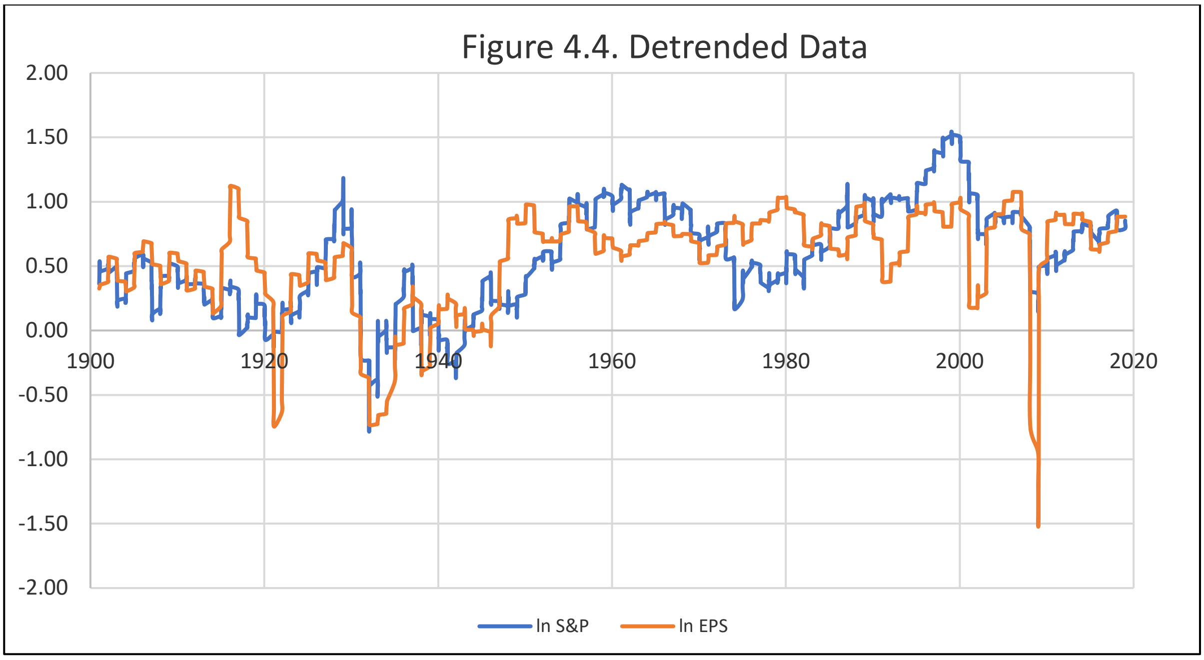


Figure 4.4. Detrended Data

The chart pictures the natural logarithms of the detrended data. In the latter, all the important economic events of the past 120 years can be identified. The great depression of 1929, the great stagflation of the 1970s, the dot.com bubble, the Asian crisis, and the first depression of the new century (2007). Throughout a century the variations of the S&P 500 follow the variations in real earnings per share (EPS). This is confirmed from the regression of the following econometric model:

$$4.5 \;\; \log Pr_t = Const + \log Pr_{t-1} + \theta \cdot \log EPS_t - \theta \cdot \log EPS_{t-1} + \varepsilon_t$$
$$where\ Const = constant, \varepsilon_t = the\ residual\ term$$

The results are summarized in the Table 4.2 that follows.

Table 4.2

N	1418						
Mean of Y	.0.579661034						
Equation	Y = .0.003047 + .0.1269 X - .0.1238 X-1 + .0.9923 Y-1						
R^2	.0.989						
R^2 adjusted	.0.989						
RMSE	.0.042171321						
Parameter	Estimate	95% CI		SE	t	p-value	VIF

Constant	.0.003047	-.0.001108	to .0.007201	2.1179E-03	1.44	.0.1505	-
X	.0.1269	.0.07236	to .0.1815	.0.027810	4.56	<.0.0001	94.39
X-1	-.0.1238	-.0.1783	to -.0.06929	.0.027786	-4.46	<.0.0001	94.20
Y-1	.0.9923	.0.9853	to .0.9993	3.5684E-03	278.07	<.0.0001	1.59

Source	SS	DF	MS	F	p-value
Difference	############	3	73.170552074	41143.57	<0.0001
Error	2.514686267	1414	0.001778420		
Null model	############	1417	0.156687609		

H0: E(Y|X=x) = μ
The model is no better than a null model Y=μ.
H1: E(Y|X=x) = $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots$
The model is better than the null model.

Term	SS	DF	MS	F	p-value
X	0.037034957	1	0.037034957	20.82	<0.0001
X-1	0.035299843	1	0.035299843	19.85	<0.0001
Y-1	############	1	1.375157E+02	77324.64	<0.0001

The impressive thing about this estimation is not the 0.989 R^2. It is that all parameters come at their expected values. Specifically, the constant *Const* is almost equal to zero and statistically insignificant. The parameter of the price in the previous period is positive, significant, and more importantly almost equal to unity (1) as expected. Similarly, the parameter θ is almost equal for $\log EPS_t$, $\log EPS_{t-1}$, statistically significant, and with the appropriate sign for both. All this holds for a model that has been tested for a period over a century during which three major capitalist crises, two world wars, and major financial bubble episodes took place. However, during all these times stock market fluctuations followed closely the variations of the earnings per share. In this context, it is difficult to attribute stock price volatility to externalities or persistent irrationality in investment behavior. In other words, the volatile fundamentals rule.

Nevertheless, this is achieved through significant manipulation in the data, through which the nonlinear relation between stock market returns and the incremental rate of profit is linearized. Therefore, the question of estimating and drawing inference from the original data remains. If we want to evaluate empirically the overall relationship between the incremental rate of profit and the stock market return, we need to apply the appropriate non-parametric statistics because of nonlinearity. This is attempted in the next section.

4c. Non-parametric statistics-The case of transfer entropy

The empirical evaluation of the profit-based approach for stocks is limited because both the stock market (S&P 500) returns, and the Incremental Rate of Profit time series are not stationary. For this reason, direct statistical inference can be drawn from models like equation 4.5 or by applying cointegration techniques (the latter are not implemented here). As indicated in the previous paragraph in the case of regression models the data is smoothened, and the relations are linearized using logarithms. Drawing inference by comparing the original data remains an important matter for the profit-based approach. The non-parametric techniques applied here is a step forward for the analysis.

I argue that the non-parametric statistic 'transfer entropy' is appropriate. The reasons have to do with the properties and the insights that underlie the statistic. Specifically, the application of the 'transfer entropy' theory does not require that the investigated time series must follow any specific probability distribution. Every probability distribution can apply. Moreover, certain statistics measure the (asymmetric) transfer of information between two sets of time-series data. In other words, 'transfer entropy' is appropriate for non-linear processes. It is indicative that when estimating causality, between stationary time series the applied statistic reduces to the Granger causality as we will elaborate below. However, when the time series are non-stationary Granger and TE measure different things. In short, this technique is appropriate to measure the extent the Incremental Rate of Profit affects the returns on the S&P 500 without requiring any manipulation of the data. For the calculations, we will use again the logarithmic rate of growth of the real EPS as the proxy of the incremental rate of profit. This way we will estimate the full Shiller database starting from the 19th century, and not a calculation starting from 1947.[6]

4c.1. The Transfer Entropy (TE)

Before we move to this a brief outline of the notions of 'entropy' and 'transfer entropy' is appropriate. The exposition will be mainly conceptual before considering computational issues. 'Entropy' as a term has been coined by Rudolf Clausius from the Greek word for transformation (*τροπή*).[7] In modern science, it was associated with the second law of thermodynamics which states that the entropy of an isolated system does not diminish in time.[8] On the contrary, entropy maximizes when the system reaches thermodynamic equilibrium. This is a notion of equilibrium close to the perception of classical political economy and the profit-based approach. A turbulent process where the system (in our case the stock exchange) persistently transforms to new states through the sequel of positions of risky arbitrage.

[6] Shaikh's estimations of the incremental rate of profit begin in 1947 due to data availability.

[7] "I propose to call the magnitude *S* the *entropy* of the body, from the Greek word τροπή, *transformation*. I have intentionally formed the word *entropy* so as to be as similar as possible to the word *energy*; for the two magnitudes to be denoted by these words are so nearly allied in their physical meanings, that a certain similarity in designation appears to be desirable." (Clausius 1867: 357). This is a translation from the original German in Clausius 1865: 46, where it appears as *Entropie.*

[8] The second law of thermodynamics establishes the concept of *entropy* as a physical property of a thermodynamic system. Entropy predicts the direction of spontaneous processes and determines whether they are irreversible or impossible. The second law may be formulated by the observation that the entropy of isolated systems left to spontaneous evolution cannot decrease, as they always arrive at a state of thermodynamic equilibrium, where the entropy is highest. If all processes in the system are reversible, the entropy is constant.

The breakthrough in the calculation of entropy came from Claude Shannon (1948) a Bell Labs scientist, who developed concepts and formulas that can measure the microscopic disorder to the problem of random losses of information in telecommunication signals. This is the reason that the measurement of entropy took the name 'information theory'. In practice, the whole exercise is an effort to connect microscopic interactions to macroscopically observable behavior.
When it comes to random time series processes, in our case the rate of return of the stock exchange, and the incremental rate of profit, the concept of 'mutual information' is applied and extended. This means that the various algorithms attempt to calculate 1) How much uncertainty about the state of the stock exchange (S&P 500) returns is resolved by knowing the state of the incremental rate of profit (and *vice versa*)? 2) How much information is shared between the incremental rate of profit and S&P 500 returns? 3)How may we quantify the degree of statistical dependence between the two variables? In short, we attempt to calculate a non-linear correlation coefficient which, under additional assumptions, also specifies a causal relationship between the variables. The difference with the traditional measures is that the information is asymmetric it involves the impact of past values of the incremental rate of profit on the current rate of return of the stock exchange but also the impact of past values of the S&P 500 on its current price. This is a statistical notion remarkably close to the ideas of the reflexivity theory of Gorge Soros presented in Chapter 3 section 3f. The idea is that past values of the incremental rate of profit will eventually take over stock market returns, or that actual prices will gravitate around warranted prices like in the simulation presented in Figure 4.3 above.

Having outlined certain important aspects underlying 'transfer entropy' we can move to a more formal definition of the concept. In this definition, I will use at some point the Granger causality test as the benchmark. For now, we need to keep in mind that in this perception of entropy it is not only the past prices of the incremental rate of profit that must be considered but also the 'shared information' between past and present stock prices and returns. For our investigation, this indicates that for 'transfer entropy' prices and returns can be path-dependent as assumed by 'reflexivity theory' (Chapter 3 section 3f).

In light of the above, we can define (the one-period lag) transfer entropy keeping our investigation as the example and emphasizing non-stationary time series processes.

$$4.6\ T_{Irop \to smr(t)} \equiv I(smr_t : Irop_{t-1} \vdots smr_{t-1})\ and$$

$$I(smr_t : Irop_{t-1} \vdots smr_{t-1}) = H(smr_t : smr_{t-1}) - H(smr_t : smr_{t-1}, Irop_{t-1})$$

Equation 4.6 is a conditional 'mutual information function' it tells us if $T_{Irop \to smr(t)} = 0$ (where T stands for 'transfer entropy' and $I =$ 'information transfer') then the incremental rate of profit $Irop$ plays no part in the knowledge on the return of the

S&P500 in the next period denoted by smr_t. The reason is that the measure is always nonnegative (see footnote 4). The operators H are measures of uncertainty.[9] In the linearized model summarized in equation 4.5 above (for detrended prices) the statistical significance of the parameters of the (detrended) earnings per share $\log EPS_t$, $\log EPS_{t-1}$) prove that linearized prices do not depend only on their past values. Here we will examine whether the same result holds for the actual data comparing returns and fundamentals. The time subscript appearing in equation 4.6 indicates the non-stationary process. In the remaining of the chapter, the incremental rate of profit will be referred to also as the 'source' variable and the S&P 500 return as the 'target' variable.

Of course, the one-period lag of the 'source' variable history in equation 4.6 is by no means the only time lag considered. To see how the concept works we need to consider the appropriate variable history. It has been suggested (Lizier *et al.* 2012) that for non-stationary 'target' variables the history length should tend to infinity. This means that the information the past target prices provide about state transitions in the target goes back to the distant past. There is no such rule for the history of the source variable although everyone suggests that it is no harm to go back as possible for the source variable as well (Bossomaier *et al.* 2016: 71)

Given the points on the optimum history of both source and target variables transfer entropy definitions can be generalized in various directions. The lag between the source and target variable is based on the idea that information is stored in the past values of the target variable (in our case the past rate of return of the S&P 500, or the last period price) whereas the impact of the last period source variable (in our case the incremental rate of profit, or the past value of the earnings per share) reflects how much information the source variable provides about state transitions in the target variable. The first matter that needs to be defined in this framework is how much information is transferred from the past value of the target variable in its next period price or the Active Information Storage (AIS) as it is called TE terminology. It is a rationale that reflects a good part of empirical discussions on the Efficient Market Hypothesis, the assumptions of behavioral finance, and the profit-based approach. Let

---

[9] For example, we can measure the average conditional uncertainty of the stock market returns on their last period price using the fundamental Shannon conditional entropy formula:

$$H(smr_t\colon smr_{t-1}) = -\sum_{smr} p(smr\colon smr_{-1}) \cdot \log_2 p(smr\colon smr_{-1})$$

Where p is the average conditional probability of smr. The equation calculates the probability of a certain set of stock returns to appear from a particular value in the previous period. Returning to formula 4.6 this presentation indicates that if the incremental has no impact on stock returns then:

$$H(smr_t\colon smr_{t-1}, Irop_{t-1}) = H(smr_t\colon smr_{t-1})$$

Keep in mind that mutual information (MI) is non-negative only for the averaged forms of the Shannon formula, not for local forms as we will see below. In this case the minimum Shannon entropy is 0 and the maximum equal to the entropy of the target variable.

us assume that we wish to evaluate stock prices if the efficient market hypothesis holds all information is passively stored (AIS=0) in the past price and any variations are due to random shocks. If behavioral finance assumptions are valid, then all information comes from the past price (active storage different from zero) and variations in the fundamentals play no part. Finally, for the profit-based approach, both changes in the fundamentals and the past prices play a decisive part. In the latter case, average Transfer Entropy is greater than zero.

Additionally, there are categories of conditional TE relating to a common driver effect that determines both the source and the target variables. In this regard, transfer entropy is redefined. The degree of uncertainty about the current target variable resolved by the past state of the source variable, the target variable, and the common driver together is subtracted from the degree of uncertainty of the current target variable already resolved by its past state and the past state of the common driver. This way the direct impact of the source variable is identified. Similarly, in concept, TE can be extended to various multivariate processes like 'global entropy' (Barnett *et al.* 2013).

Returning to the one-period lag question we can conclude that it is not binding for the source variable. TE can be calculated for any lagged value of the source variable. A fixed period lag is binding only for the storage target variable. In other words, the calculation of the impact of the lagged value of the target variable can only have a specific period lag. It has been shown that this way the Wiener principle of causality is preserved (Wibral *et al.* 2012). The Wiener 'principle of observational causality' argues that a time series X is called causal to a second time series Y if knowledge about the past of X and Y together allows one to predict the future of Y better than knowledge about the past of Y alone. In our case, if the incremental rate of profit or the earnings per share can predict the rate of return or the price of the S&P 500 better than their past value alone this constitutes a causal relation. We will elaborate on this matter further in relation to the traditional Granger causality test later in this section (4c.3).

For now, we need to consider an additional aspect of transfer entropy. So far, we have outlined how a calculation of the average TE can provide us with inference about the effect of a (source) variable on a target variable although their relationship is not (necessarily) linear. We have implied further that we have tools (not presented yet) to calculate this relation. Moreover, if such a relation exists it can imply a causal link between the source and the target variable. Nevertheless, averages hide the dynamical structure of the relation between the source and the target variable. On the contrary, the local perspective can reveal the dynamic structure. Applied to time-series data, local measures tell us about the dynamics of information in the system, since they vary with the specific observations in time. To be specific, a measured average of 'mutual information' and/or 'transfer entropy' does not tell us about how the symmetric or directed relationship between two variables fluctuates through time, how different specific source states may be more predictive of a target than other states, or how coupling strength may relate to changing underlying experimental conditions. On the contrary local measures can be revealing of these relations helping the resolutions of problems we have encountered already in the simulation originating from Shaikh (2016) and presented in section 4.1 (Figure 4.3). Local estimations is the part of TE theory we will apply in the present work.

4c.2. Transfer Entropy (TE) Estimators and calculations

I will present the transfer entropy estimators based on the assumption that both the incremental rate of profit and the rate of return of the stock exchange are discrete-time variables. As it will become evident shortly this means that we can estimate TE directly from the probability distribution functions of the source and the target variable. This is in accordance with the basic assumption of the profit-based approach (Shaikh 1997) where stock market returns must react to the underlying fundamentals. Moreover, the estimation does not rely on specific assumptions of the probability distribution of the variables.

To understand how the process works we can consider the Shannon mutual information (MI) formula. It reads as follows:

$$4.7\ I(smr_t, Irop_t) = \sum_{smr_t \in \boldsymbol{smr}} \sum_{Irop_t \in \boldsymbol{Irop}} prob_{smr,Irop}(smr_t, Irop_t) \cdot \log_2 \frac{prob_{smr,Irop}(smr_t, Irop_t)}{prob_{smr}(smr_t) \cdot prob_{Irop}(Irop_t)}$$

Where $prob_{smr,Irop}(smr_t, Irop_t)$ is the joint mass probability function of the stock market returns smr_t and the incremental rate of profit $Irop_t$ and $prob_{smr}$, $prob_{Irop}$ are the marginal probability density functions of the source and the target variable. MI tells us how much uncertainty about stock returns is reduced from knowing the incremental rate of profit.[10] For example, if the two variables are independent then $I(smr_t, Irop_t) = 0$ proven as follows:

$$If\ the\ variables\ are\ indipendent\ \rightarrow prob_{smr,Irop}(smr_t, Irop_t) = prob_{smr}(smr_t) \cdot prob_{Irop}(Irop_t)$$

$$\rightarrow \log_2 \frac{prob_{smr,Irop}(smr_t, Irop_t)}{prob_{smr}(smr_t) \cdot prob_{Irop}(Irop_t)} = \log_2 1 = 0$$

Equation 4.7 can be calculated straight forward provided we know the probability functions. However, the calculation is not always so easy since it involves sample and other biases that can increase MI. The Kozachenko-Leonenko entropy estimator (Delattre and Fournier 2016) and its development in the KSG algorithm (Kraskov, Stögbauer and Grassberger 2004) enables the approximation of equation 4.7 without knowing the probability function. More importantly, these algorithms can estimate

[10] There is no rule about using a particular unit of measurement for the log values. In discrete variables it involves logarithms of base 2 and 10. I use logarithms with base 2 that are the most common in bibliography.

also TE limiting any possible biases. However, in our estimation, we will be able to estimate probabilities directly (see section 4.3.6 below).

As stated already while commenting equation 4.6 TE is a case of conditional MI presented in the following Shannon equation:

$$4.8\ T_{Irop\rightarrow smr(t)} \equiv I(smr_t : Irop_{t-1} \vdots smr_{t-1})$$

$$T_{Irop\rightarrow smr(t)} = \sum_{smr_{t-u}\in \boldsymbol{smr_{-1}}} \sum_{Irop_{t-i}\in \boldsymbol{Irop}} \sum_{smr_t\in \boldsymbol{smr}} prob_{smr_{-1},Irop,smr}(smr_{t-1}, Irop_{t-1}, smr) \cdot \log_2 Z$$

$$Z = \frac{prob_{smr_{-1},Irop,smr}(smr_{t-1}, Irop_{t-1}, smr) \cdot prob_{smr_{-1}}(smr_{t-1})}{prob_{smr,smr_1}(smr_t, smr_{t-1}) \cdot prob_{Irop,smr_1}(Irop_{t-1}, smr_{t-1})}$$

Like equation 4.7 equation 4.8 is difficult to estimate directly. Nevertheless, the algorithms mentioned in the previous paragraph, especially the KSG algorithm can provide reliable estimations. In fact, the algorithms are incorporated in open-source software applicable in MATLAB and R.

Before we move to the actual calculation, we need to address the issues of causality, inference, and applications of transfer entropy in financial time series. This way we will have a more complete understanding of TE theory before applying the appropriate statistic in testing the profit-based approach.

4c.3. Transfer Entropy and Causality a Comparison with Granger Causality

Here I will not enter the discussion of whether the Granger causality test is a true statistical estimation of causal relations between two or more time series. I will take Granger causality for granted. The reason is that at the conceptual level Granger causality and Transfer Entropy appear to be almost identical. In short, Granger suggested that if the past values of a certain explanatory (source) variable explained the fluctuations of the dependent (target) variable above its past values then this means that a causal relationship exists between the explanatory variable and the dependent variable.

Up to this point, it is hard to find a difference between the Granger test and the concept of TE. However, when moving to the calculation of causality under this concept Granger applied linear models and more importantly focused on predictability and not information transfer. Specifically, he suggested that causality can be estimated from the comparison of what he called the *full* and the *reduced* model. If we wanted to test our assumption of the relation between the incremental rate of profit and the stock market returns our Granger would look as follows:

$$4.9\ smr_t = a_1 \cdot smr_{t-1} + a_2 \cdot smr_{t-2} + \cdots a_k \cdot smr_{t-k} + \beta_1 \cdot Irop_{t-1} + \beta_2 \cdot Irop_{t-2} + \cdots + \beta_l \cdot Irop_{t-l} + \epsilon_t$$

$$4.10\ smr_t = a'_1 \cdot smr_{t-1} + a'_2 \cdot smr_{t-2} + \cdots + a'_k \cdot smr_{t-k} + \epsilon'_t$$

Equation 4.9 is the full model and 4.10 the reduced model. Both are VAR models. The model parameters are the coefficient matrices a_ι, β_j, a_i' and the covariance matrices $\Sigma \equiv c(\epsilon_t),\ \Sigma' \equiv c'(\epsilon_t')$. The elements ϵ_t, ϵ_t' are the serially uncorrelated residuals of the estimation. Of course, the estimation of the relation between $smr_t\ and\ Irop_{t-i}$ with such a model is not adequate. First, the relation is assumed linear, and second, the probability distribution of both variables is assumed stationary. This implies that both variables follow a Markov process something that has proven inadequate for the calculation of risk and volatility in financial asset time series as discussed in chapters 1 and 2.

To put it differently, Granger tests were applied for linear models involving stationary variables. The calculation of causality was performed by applying two roughly equivalent approaches. The first had to do with the calculation of a statistic indicating causality on the grounds of better predictability of the full model compared to the reduced model. The second had to do with the calculation of the likelihood ratio. Let us begin with the first approach:

$$4.11\ F_{Irop \to smr} = log \frac{|\Sigma'|}{|\Sigma|}$$

Equation 4.11 is the log ratio of the determinants of the covariance matrices defined above. It is presented here by applying the approach of Geweke (1982) on Granger causality for linear models and not the traditional approach applied by Sims (for example) in his (1972) paper. The statistic is based on the *generalized* rather than the *total variance* of the residuals[11]. The meaning of 4.11 is simple if the generalized variance of the full model $|\Sigma|$ is equal to the generalized variance of the reduced model $|\Sigma'|$this means there is no causal relation between the variables. Moreover, the greater value of the statistic the stronger the causal relation. Geweke (1982: 306) gives a full account of the properties of the statistic pointing also that if the time series is Gaussian the maximum likelihood estimate of F is easy to construct.

The latter brings us to the second equivalent approach. If the time series is Gaussian, then the statistic 4.11 is the log-likelihood ratio with a null hypothesis:

$$4.12\ H_0: \beta_1 = \beta_2 = \cdots = \beta_l = 0$$

4.12 shows a null hypothesis where the parameters of β_i in equation 4.9 are simultaneously zero. The interesting part with the maximum likelihood approach, in this case, is that the F statistic described in equation 4.11 is associated with an (asymptotic) χ^2 distribution with degrees of freedom equal to the difference in the number of free parameters between the full and the reduced model. Therefore, causality can be formally estimated statistically in this context.

[11] The total variance is the sum of variances whereas generalized variance is, by definition, the determinants of the covariance matrices.

Granger causality has certain additional properties. It can be extended from the time to the spectral/ frequency domain. This means that causal interactions can be decomposed by frequency. Moreover, the 4.11 statistic is invariant in the time and frequency domain if stationarity is strengthened through filtering. However, filtering leads to poor modeling. These issues will prove intuitive in understanding the relation between Granger causality and Transfer Entropy.
To start Granger causality and transfer entropy are linearly related only in the case the underlying time series reflect a Gaussian joint process. The relation is the following:

$$4.13\ T_{Irop \to smr} = \frac{1}{2} \cdot F_{Irop \to smr}$$

The proof is provided in Bossomaier *et al.* (2016: 86).

In the same fashion if the underlying processes are Markov processes with a certain degree of ergodicity the Maximum likelihood transfer entropy estimator will converge towards the TE estimator defined in equation 4.8. The proof is provided in Bossomaier *et al.* (2016: 87).

Nevertheless, the association between a non-linear Granger causality test with a parametric (Maximum Likelihood) TE statistic can hold only under these restrictive assumptions. If we move from the Markov ergodic world any association between Granger and TE theories is lost. In short for non-Gaussian processes TE and Granger statistics (even if the latter incorporate certain non-linear relations) do not calculate the same thing. Transfer entropy calculates information flow whereas Ganger tests emphasize on predictability.

4c.4. Transfer Entropy and Statistical Significance/ Inference

Due to bias issues the statistical significance and confidence intervals of TE measures are calculated using sub-sample techniques. In practice, we set a *null hypothesis* and check whether it is true or false. Of course, this requires knowledge of what the probability distribution would look like if the null hypothesis H_0 is true. One way is to use surrogate variables with the same statistical properties as the tested variables generated under the null hypothesis. For example, if we assume that the null hypothesis is that $Irop$ and smr are not associated then if the null hypothesis holds this means that the distribution of the null hypothesis will be that of smr_t conditional on smr_{t-u}.

In the case of known distributions of the underlying variables, the task is easier. For discrete Gaussian processes, for example, we know that $T_{Irop \to smr}$ has an asymptotic distribution (discussed in the previous section $x^2/2 \cdot N \cdot \log 2$. The degrees of freedom are $(M_{smr} - 1) \cdot (M_{Irop} - 1) \cdot M_{smr}$where M stands for cardinalities and that the two variables (smr, $Irop$) have the same history. In general, even for skewed distributions as $N \to \infty$ the x^2 distribution described above holds asymptotically. These issues can prove complicating and this is one of the reasons we have applied the MI instead of the TE statistic.

4c.5 Applications of Transfer Entropy in Financial Time Series Data

I will conclude this brief outline with some applications of TE for financial time series. As we have seen already both analytically and to a certain extend empirically the profit-based approach combines both the 'market sentiment and market fundamentals' (Bossomaier *et al.* 2016: 127). This is precisely the understanding of the calculation of Transfer Entropy for stock market returns in Bossomaier *et al.* as quoted. Nevertheless, the authors of the cited book on transfer entropy are not familiar with any part of the profit-based approach argument. This is a strong indication of how adequate is TE theory for evaluating statistically the argument of the profit-based approach on stock returns.

Despite this hint, most of the applications of TE in the financial market were interested in the direction of the causal arrow between different financial variables rather than the direct estimation of the information flows between fundamentals and stock prices and returns. Specifically, some investigations were focused on identifying whether information flows from stocks to indexes or the opposite. The idea was that if indexes were the crucial factor then the momentum overrides the fundamentals and if the causality is the other way around (form equities to indexes) fundamentals rule. This, besides other issues, indicates why it is important to estimate stock prices and returns directly from the fundamentals instead of trying to infer their influence through assumptions that are not directly tested. The latter will become evident when considering specific stock-index empirical models.

The most known study is that of Kwon and Oh (2012) who worked with indexes and stocks from the Us, Europe, UK, and the Asia Pacific finding in all nine indexes considered a univocal flow of information from indexes to stocks rather than the other way around. In fact, in all cases, there was almost no indication of information transfer from stocks to indexes. This finding was considered as a verification of an investment behavior where traders try to anticipate what the other buyers and sellers believe. In this regard, they referred to Keynes' quote from the *General Theory* (Keynes 1936, Ch. 12: 156) on the newspaper beauty competition to emphasize the effort of traders to understand the market beliefs. In other words, the model implies an 'inefficient market' where the next movement can be predicted through 'technical analysis' although everybody knows this has not proven to be the case. Moreover, our findings on the association of the S&P 500 prices and returns with the Incremental Rate of Profit (Sections 4.1, 4.2) do not just justify the conclusion. The theoretical reasoning is so arbitrary that someone could even argue that the stock index is a proxy of the market portfolio and the findings are supportive of the capital asset pricing model where the market portfolio is the only source of undiversifiable risk. For our purposes, it must be clear that nonparametric models can lead to the inference, regarding the underlying theory only when the actual assumption is the one tested. Otherwise, the findings can be associated with practically any theory.

A more interesting application of TE in financial markets can be found in the effort of looking into financial market fluctuations and crashes as a 'phase transition' like the ones studied in physical phenomena. In a (2006) paper Kiyono, Struzik, Yamamoto studied the Black Monday of the New York Stock Exchange. They found that the financial system behaves a lot like a physical system through a varying underlying parameter that proves the non-stationarity of financial data. To put it differently, an underlying factor parameter varies as the system approaches the critical point and this alters the probability distribution of prices. Subsequently, Wicks, Chapman, and Dendy (2007) have shown that 'mutual information' (MI) can be a tool for detecting order/ disorder transitions in various systems. Harré and Bossomaier (2009) applied this methodology for financial markets. However, again the argument was not one relating fundamentals with prices but different stocks of the S&P 100 from 1995-2008 and the MI between them.

This discussion closes the rather lengthy, but I hope useful, introduction to Transfer Entropy. From what follows it will become evident that our application to transfer entropy is an original contribution since it attempts to infer on a theory and not simply to the properties of the time series. In other words, we pick the estimation technique basis the anticipated properties of the time series and not the opposite as it frequently happens in the empirical analysis.

4c.6. Mutual Information Local Estimations for the S&P 500 and EPS Growth 1880-2020

I work on the Wicks, Chapman, and Dendy (2007) methodology. The difference is that I calculated the local MI measures comparing the log growth of the S&P 500 directly to the log growth of the Earnings per Share (EPS). To calculate probabilities, I defined four possible states. One where the EPS increases and so does the S&P 500, one where both decline, and two cases where they move in the opposite direction. The case where the S&P 500 increases whereas the EPS declines will lead to a bubble if it is persistent. The opposite 'state' (the index drops although EPS increases) if persistent it will lead to underpricing. The MI statistic was calculated by applying equation 4.7 for the log-returns going back ten years. The table that follows is an example of the first calculation covering the period 1872-1881.

Table 4.3

Probability Table 1872-1881				
		S&P		
		Incr.	Dec.	10

EPS	Incr.	0.50	0.00	0.50
	Dec.	0.20	0.30	0.50
		0.70	0.30	
H S&P	**H EPS**	**I(S&P, EPS)**	**H(S&P, EPS)**	**MI (S&P, EPS)2**
-0.3602	-0.5	0.257287	-0.5	
-0.52109	-0.5	0	0	
0.881291	1	-0.16147	-0.46439	
		0.3	-0.52109	
MI		**0.395816**	1.485475	**0.395815602**

Table 4.3 shows that during the decade the S&P 500 increased together with the EPS five years, they simultaneously decreased for 3 years and in the remaining 2 years of the decade, the index increased although EPS fell. The mutual information (MI) is almost 40% and explains about 45% of the entropy of the S&P 500 which is 0.88. I repeated this same calculation until the present dropping the first year of the calculation and including the next (1873-1882 and so on). The findings are summarized in Figure 4.5 that follows.

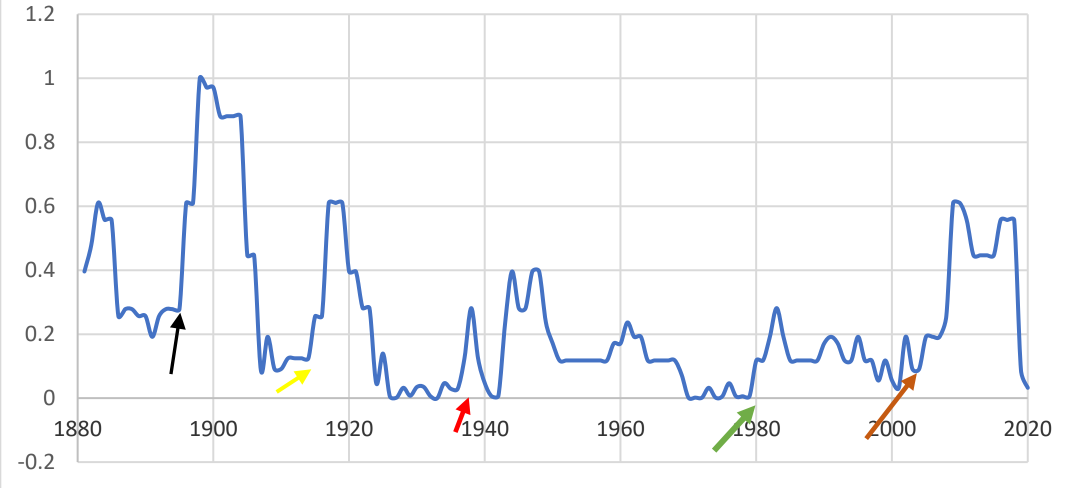


Figure 4.5 MI Estimation

It is evident that the Local MI statistic experiences severe fluctuations over the past 140 years.

However, these fluctuations are indicative of an interesting pattern. The value of the statistic experiences a strong decline in almost all phase transitions that took place during the past century and a half. The black arrow points to the year 1893 it is the year that marked the end of the 'long crisis' (1872-1893). Thereafter, the price of the statistic MI surges and tends to explain the total of the S&P 500 returns entropy for a short time. The time between 1882 (when MI begins to fall) and 1885 (when the statistic begins to recover) was a period of depression in the United States as recorded by the National Bureau of Economic Research (NBER). The period from 1879 to 1881 was a period of prosperity (railroad growth). The latter explains MI prices of around .0.6 as it will become clear shortly. A similar reason (a recession involving a decline in industrial activity of 16%) stands behind the decline of the (MI) statistic for the period (1906-1913) indicated by the yellow arrow. The statistic, as well as the market, recovered from 1913 to 1926 having a cumulative increase of 36% over the period. However, following 1926 it became clear that the S&P 500 was a bubble. The MI index becomes almost zero (red arrow) indicating an increasing market with deteriorating fundamentals. Following the crash of 1929-1932 the opposite happened, the market did not reflect the recovery of the corporate fundamentals. It was only after the end of the so-called 'Roosevelt depression' in 1938 that the (MI) recovered to drop in the first war years (1940-1041) and recover for good after 1943. However, the dependence of the stock market on the earnings per share calculated by the (MI) statistic did not reach the pre-war levels during the Golden fifties and sixties. It was stable for values between 10 and 20%. Nevertheless, when corporate fundamentals began to deteriorate in the late 1960's early 1970s' the pattern persisted as indicated by the green arrow. Again the (MI) value dropped to zero and remained weak throughout the great 'stagflation'. It was only after 1980 when the crisis ended that corporate fundamentals were reflected in the market. The neoliberal era did not change this strong and long-lasting pattern. Although corporate fundamentals did not deteriorate the (MI) statistic reacted to the burst dot–com bubble of 2000 as pointed by the brown arrow. However, because the bubble was not hiding a depression the market recovered quickly and remained in line with the fundamentals throughout the first decade of the new century. The latter is confirmed also by the simulation of Figure 4.3. We can see there that in 2008 the stock market collapsed shortly after the collapse of the corporate fundamentals. Actually, after the millennium (MI) took high prices that were not witnessed since the second decade of the previous century. The reason is that the bubble did not burst together with the depression but 10 years earlier in the dot-com crisis.

There is both an economic theory but also intuition coming from physics behind this pattern. In physical phenomena 'a few particles' are sufficient for 'system transformation' (Wicks, Chapman, and Dendy 2007, Kiyono, Struzik, and Yamamoto 2006). In our context, this means that a few years of one-sided motion between the market and the fundamentals is sufficient to destabilize the market provided that the fundamentals keep deteriorating. As we saw in chapter 3 this is not simply a property of the (MI) statistic it is also indicative of a pattern in investment behavior. This is no other than the 'reflexivity theory' of George Soros. Financial capital controls stock market returns making positive expectations a self-fulfilling prophesy for some time. However, at some point, everybody realizes that the market is a bubble, and this leads to a sharp correction. If the market remains in line with the fundamentals like in 2008 the correction is dramatic, but the market recovers soon.

Conclusion

The models of the profit-based approach are complementary to each other. Simulation models like Shaikh (2016) presented in Figure 4.3 indicates a strong and long-lasting correlation between warranted and actual prices. The regression presented in section

4.2 gives clear inference that stock market returns depend on corporate fundamentals. Finally, the 'Mutual Information' model presented in paragraph 4.3.6 proves that this is not the result of manipulations in the data (detrending) but reflects long-lasting patterns. This means that the stock market volatility is perfectly rational since it reflects turbulent fundamentals but also path-dependent investment expectations. Overall, this research can open investigations both on the profit-based approach for stocks, but also other assets priced by the theory. In case these models are applied for professional use they must be evaluated together. If treated separately they can prove misleading.

Appendix 4.1

Calculation of the modified incremental rate of profit in Shaikh (1997, 2016).

The table that follows is an extract of the calculation of the incremental rate of profit. The nominal data come directly from the BEA tables as indicated in the description. This data is deflated by the Implicit Price Deflator to give real gross Investment (IRG corp.). This concludes step1. Step 2 calculates Gross corporate Profits (including depreciation) from BEA tables and deflates the nominal amount arriving at Real Gross Profit (using the Implicit Price Deflator). The amended incremental rate of profit for 1948 is the Difference in Real Gross Profit divided by Real Gross Corporate investment in the previous period.

Step1 Implicit Price Deflator Gross Investment				
	1947	**1948**	**1949**	**1950**
FA T.6.7 line 2	17.30	19.50	17.80	19.50
FA T.6.8 line 2	4.96	5.14	4.58	4.91
(IGRcorp. index bea (t)/100)*IGCcorpbea(2005): Bills-2005$	64.28	66.62	59.25	63.54
Implicit Price Deflator, Gross Investment, pIG corp bea	26.92	29.27	30.04	30.69
Step2				
BEA Table 6.4 line 2 Current - Cost Corporate Depreciation	9.80	11.50	12.40	13.30
BEA Table 1.14 line 11	23.20	30.10	27.90	34.80
Gross Corporate NIPA Profit (sum of excel lines 8+9)	33.00	41.60	40.30	48.10
Real Gross Corporate Profit = line 10x100/ line 6	122.61	142.12	134.15	156.73
Modified Incremental Rate of Profit		**0.30**	-0.12	0.38
	formula	**(C11-B11)/B5**		