Trends, Reversion, and Critical Phenomena in Financial Markets

Christof Schmidhuber

Zurich University of Applied Sciences
School of Engineering, Technikumstrasse 9
CH-8401 Winterthur, Switzerland
christof@schmidhuber.ch

June 16, 2020

Abstract

Financial markets across all asset classes are known to exhibit trends, which have been exploited by traders for decades. However, a closer look at the data reveals that those trends tend to revert when they become too strong. Here, we empirically measure the interplay between trends and reversion in detail, based on 30 years of daily futures prices for equity indices, interest rates, currencies and commodities.

We find that trends tend to revert before they become statistically significant. Our key observation is that tomorrow's expected return follows a cubic polynomial of today's trend strength. The positive linear term of this polynomial represents trend persistence, while its negative cubic term represents trend reversal. Their precise coefficients determine the critical trend strength, beyond which trends tend to revert.

These coefficients are small but statistically highly significant, if decades of data for many different markets are combined. We confirm this by bootstrapping and out-of-sample testing. Moreover, we find that these coefficients are universal across asset classes and have a universal scaling behavior, as the trends time horizon runs from a few days to several years. We also measure the rate, at which trends have become less persistent, as markets have become more efficient over the decades.

Our empirical results point towards a potential deep analogy between financial markets and critical phenomena. In this analogy, the trend strength plays the role of an order parameter, whose dynamics is described by a Langevin equation. The cubic polynomial is the derivative of a quartic potential, which plays the role of the energy. This supports the conjecture that financial markets can be modeled as statistical mechanical systems near criticality, whose microscopic constituents are Buy/Sell orders.

Keywords: trend following, mean reversion, futures markets, market efficiency, critical phenomena, social networks

1 Introduction

It is well-known that financial markets across all asset classes exhibit trends. These trends have been exploited very successfully by the tactical trading industry over the past decades, including the former "turtle traders" [1] and today's CTA industry.

A close look at the available data reveals that those trends tend to revert as soon as they become too strong. In this paper, we demonstrate this based on 30 years of daily futures returns for equity indices, interest rates, currencies and commodities. We analyze trends with 10 different time horizons, ranging from 2 days to 4 years, and empirically measure the critical strength, beyond which trends tend to revert. Here, the "strength" of a trend is defined in terms of its statistical significance, namely as the t-statistics of the trend.

In a first step, we measure the daily average return of a market as a function of the values of the 10 trend strengths on the previous day. In order to increase the statistical significance of the results, we aggregate across different markets and time scales. Our key observation is that tomorrow's average return can be quite accurately modeled by a polynomial of today's trend strength. It consists of a positive linear term that is responsible for the persistence of trends, and a negative cubic term that is responsible for the reversion of trends. Trends tend to revert beyond a critical trend strength, where the two terms balance each other. The corresponding regression coefficients are small, but statistically highly significant.

In a second step, we refine this quantitative analysis. Using multiple nonlinear regression, we empirically measure how the observed cubic function varies

- with the time scale of the trends: we find that trends of medium strength persist at scales of several days to several years, while reversion dominates at shorter or longer time scales. We model this scale-dependence by polynomial regression as well.
- with the asset class: we find that the available data do not allow us to fit different model parameters to different asset classes. Within the limits of statistical significance, the model parameters are thus universal, i.e., independent of the asset.

- over time: we find that the patterns have gradually changed over the decades. In particular, trends have become less persistent, and there is little evidence that classical trend-following can perform as well in the future as it did in the past.

Since financial market returns are only in a rough approximation independent, normally distributed random variables, we cannot trust the standard significance analyses for regression results. Instead, we use bootstrapping and cross validation to confirm that our results are statistically highly significant and robust out-of-sample. Throughout this paper, we try hard not to introduce a single parameter more than is absolutely necessary to capture the essence of the empirically observed patterns. We find that we may fit at most 6 parameters to our 30-year data set, and identify what we believe are the 4-6 most relevant parameters.

While trends have been exploited in the systematic trading industry for decades, they arrived relatively late in academia. Early observations on market trends appear, e.g., in [2, 3]. Early literature on the interplay of trends and reversion has focused on their cross-sectional counterparts (momentum and value) for single stocks [4]. With the advent of alternative beta strategies [5, 6], trend-following has become an active academic research area [7, 8, 9, 10]. There is now an extensive literature on trend-following, including backtests of its performance more than a century into the past [11, 12]. Our paper not only extends these previous analyses by precisely modeling when trends become unstable. It also deepens them by relating trends to critical phenomena in statistical physics.

As one application, our results signal to investors when to exit trends. Our observations also support the idea that market inefficiencies do exist, but disappear before they become strongly statistically significant, and they indicate how markets have become more efficient over the decades. Last but not least, our results imply that trends follow stochastic processes with cubic terms, as will be explained in section 5. Together with the observed universality and scaling behavior, this points towards a potential deep analogy between financial markets and statistical mechanical systems near second-order phase transitions.

2 Data and Definitions

2.1 Data

Our analysis is based on historical daily log-returns for the set of 24 futures contracts shown in table 1. This set is diversified across four asset classes (equity indices, interest rates, currencies, commodities), three regions (Americas, Europe, Asia) and three commodity sectors (energy, metals, agriculture). We use futures returns, instead of the underlying market returns, because futures returns are guaranteed to be marked-to-market daily. Moreover, they are readily available for all asset classes and net of the risk-free rate, which also makes returns in different currencies and interest rate regimes comparable with each other.

Table 1: Markets	America	Europe	Asia
Equities	S&P 500 TSE 60	DAX 30 FTSE 100	Nikkei 225 Hang Seng
Interest rates	US 10-year Canada 10-year	Germany 10-year UK 10-year	Japan 10-year Australia 3-year
Currencies	CAD/USD	EUR/USD GBP/USD	JPY/USD AUD/USD NZD/USD
Commodities	Crude Oil Natural Gas	Gold Copper	Soybeans Live Cattle
Com.-Sectors:	Energy	Metals	Agriculture

For all contracts, we consider 30 years of daily price data, covering the period from Jan 1, 1990, to Dec 31, 2019. The first two years 1990 and 1991 are merely used to compute the trend strengths at the beginning of 1992 (see below), so the actual regression analysis covers only 28 years. Daily prices $P_i(t)$ were taken from Bloomberg, where i labels the asset and futures are rolled 5 days prior to first notice. We define normalized daily log-returns $R_i(t)$:

$$R_i(t) = \frac{r_i(t)}{\sigma_i} \ , \quad r_i(t) = \ln \frac{P_i(t)}{P_i(t-1)} \ , \quad \sigma_i^2 = \mathrm{var}(r_i) \ , \quad \mu_i = \mathrm{mean}(r_i) \ , \qquad (1)$$

where the long-term daily risk premium μ_i and the long-term daily standard deviation σ_i of a market i are measured over the whole 30-year period. For some futures markets, the log-returns $r_i(t)$ had to be backtracked or proxied as follows:

1. TSE 60 futures: their history begins on Sep 9, 1999. Before, the TSE 60 futures returns are proxied by the S&P 500 futures returns, which, in our analysis, thus have double weight during that period.

2. Hang Seng index futures: their history begins on Apr 2, 1992. Before, their returns are proxied by Nikkei 225 futures returns. As the regression analysis begins only on Jan 1, 1992, this is a minor data correction.

3. DAX futures: their history begins on Nov 26, 1990. Before, FTSE 100 futures returns are used as a proxy. This data correction is also minor: it merely slightly affects the initial trend strengths at the beginning of 1992, when the analysis begins.

4. EUR futures: their history begins on May 21, 1998. Before, the Deutsche Mark is used as a substitute for the Euro. We have reconstructed Deutsche Mark futures returns from the spot exchange rate and German/U.S. Libor differentials.

5. NZD futures: their history begins on May 9, 1997. Before, we have reconstructed the futures returns from the spot exchange rate and the Libor differentials as in the case of the Deutsche Mark.

6. German 10-year Bund futures: their history begins on Nov 27, 1990. Before, we have reconstructed futures returns from daily German 10-year and short-term interest rates, assuming a duration of 8. This data correction is also minor: it merely slightly affects the initial trend strengths at the beginning of 2002, when the analysis begins.

7. Natural gas futures: their history begins on Apr 5, 1990. Before, 1.5-fold levered crude oil futures returns are used as proxies for the natural gas futures returns, using the U.S. Libor rate as the cost of leverage. The 1.5-fold leverage reflects the higher volatility of natural gas compared with crude oil. Again, this data correction is minor, as it merely affects the initial trend strengths at the beginning of 1992, when the analysis starts.

2.2 Trend Strengths

We will examine the interplay between trends and reversion at 10 different time scales:

$$T_k = 2^k \text{ business days with } k \in \{1, 2, 3, ..., 10\}$$

This represents periods of approximately 2 days, 4 days, 8 days, 3 weeks, 6 weeks, 3 months, 6 months, 1 year, 2 years, and 4 years. Thus, there are 10 different trend strengths at each point in time. A given asset may well be, e.g., in a long-term up-trend at the 1-year time scale, and at the same time in a short-term down-trend at the 3-week time scale.

For a given time scale T, we define the trend strength $\phi_{i,T}(t)$ of a market i on day $t \in Z$ as a weighted average of past daily returns of that market - more precisely, of the normalized past daily log-returns (1) in excess of the long-term risk premium:

$$\phi_{i,T}(t) = \sum_{n=0}^{\infty} w_T(n) \cdot (R_i(t-n) - \frac{\mu_i}{\sigma_i}), \tag{2}$$

where $w_T(n)$ is a weight function for the time scale T. T is defined as twice the average lookback period:

$$T = 2 \cdot E(n) = 2 \sum_{n=0}^{\infty} n \cdot w_T(n) \cdot \big[\sum_{n=0}^{\infty} w_T(n)\big]^{-1}. \tag{3}$$

Removing the long-term risk premium μ_i in (2) ensures that the long-term expectation value of the trend $\phi_{i,T}$ is zero for each market and time scale. We also normalize the weight function such that $\phi_{i,T}$ has standard deviation 1. Assuming that market returns on different days are independent from each other (which is true to high accuracy), this implies:

$$\sum_{n=0}^{\infty} w_T^2(n) = 1.$$

With this normalization, $\phi_{i,T}$ can be regarded as the statistical significance of the trend. E.g., $\phi_{i,T} = 2$ represents a highly significant up-trend, while $\phi_{i,T} = -0.5$ represents a weakly significant down-trend. This normalization makes all trend strengths comparable with each other, and will thus allow us to aggregate across different markets and time scales below.

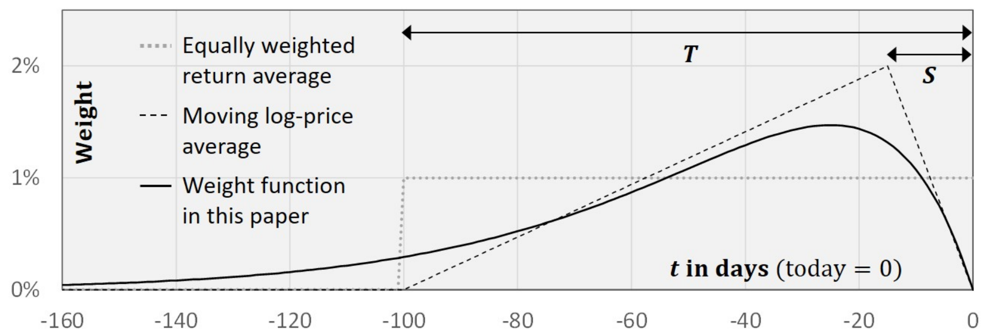

Figure 1: Our trend strength is defined as a weighted sum of past log-returns. The solid line shows the weight function used in this paper, compared with two standard alternatives.

The simplest weight function would be a step function (fig. 1, dotted line). Then the trend strength $\phi_{i,T}$ would just be proportional to the average log-return over the past T days. However, $\phi_{i,T}$ would then be quite volatile and jump on days when an outlier return enters or leaves the rolling time window. A more common definition of $\phi_{i,T}$ is based on moving price-averages: one subtracts the average log-price over the time period T from the average log price over a shorter time period S. This corresponds to a wedge-like weight function (fig. 1, dashed line). It makes the trend strength less volatile, as outlier returns affect it only gradually over time periods S and T. It also filters out short-term trends of time scales smaller than S. This is necessary to seperate trends at different time scales from each other.

However, the moving price average introduces the additional parameter S that must be fitted to the data in any analysis, which tends to reduce the statistical significance of the results. In this note, we will therefore use another natural weight function that involves only the single parameter T (fig. 1, solid line):

$$w_T(n) = N_T \cdot (n+1) \cdot \exp(-\frac{2n}{T}) \quad \text{with} \quad N_T = \frac{(1-e^{-4/T})^2}{\sqrt{1-e^{-8/T}}}.$$

The normalization factor N_T ensures that the squares of the weights sum up to 1. The average lookback period is $T/2$, as required by (3). We have verified that the results of this paper are not very sensitive to the precise definition of the trend strength. In fact, moving price averages yield very similar patterns if the ratio S/T is anywhere between 10-50%.

To limit the effect of outlier values of $\phi_{i,T}$, we will cut it of at ± 2.5 in the actual regression analyses of section 4, i.e., we will use the capped and floored version

$$\phi_{i,T}^{cap} = \min(2.5, \max(-2.5, \phi_{i,T}).$$

Without this cap and floor, the regression analyses of section 4 would yield higher R-squared's, and the results would be similar, but less robust. For computational efficiency, we also define an auxiliary variable $\psi_{i,T}(t)$ proportional to the standard EWMA return:

$$\psi_{i,T}(t) = M_T \sum_{n=0}^{\infty} e^{-2n/T} (R_i(t-n) - \frac{\mu_i}{\sigma_i}) \quad \text{with} \quad M_T = \sqrt{1 - e^{-4/T}}, \tag{4}$$

which also has expectation value 0 and variance 1, to obtain the recursion relations

$$\phi_{i,T}(t) - e^{-2/T} \phi_{i,T}(t-1) = \frac{N_T}{M_T} \cdot \psi_{i,T}(t), \quad \psi_{i,T}(t) - e^{-2/T} \psi_{i,T}(t-1) = M_T \cdot (R_i(t) - \frac{\mu_i}{\sigma_i}), \tag{5}$$

which allow us to compute our trend strength recursively.

Table 2 displays a small extract of the resulting database for our analysis. Only two of the 7305 business days and only three of the 24 markets are shown. The third column shows the normalized daily log-returns (1), which have standard deviation 1. The 7305 business days cover only the 28-year period from Jan 1992 - Dec 2019, because the first two years 1990-1992 were only used to compute the initial trend strengths at the beginning of 1992. The full table thus has $7305 \times 24 = 175'320$ lines.

Table 2: Database			Trend strengths on previous day for 10 time scales									
Day	Market	$R_i(t)$	2d	4d	8d	3w	6w	3m	6m	1y	2y	4y
1	S&P 500	-0.2	0.3	0.7	1.0	0.6	0.2	0.3	0.6	0.6	1.0	1.6
1	EUR/$	-0.1	0.2	0.2	0.0	-0.4	-0.6	-0.6	-0.8	-0.9	-0.7	-0.8
1	Gold	-0.5	-0.3	-0.7	-0.7	0.1	1.1	1.5	1.1	0.4	0.1	-0.4
2	S&P 500	-0.3	-0.1	0.4	0.9	0.7	0.2	0.3	0.6	0.6	1.0	1.6
2	EUR/$	-1.0	-0.7	-0.2	-0.2	-0.4	-0.6	-0.6	-0.8	-0.9	-0.7	-0.8
2	Gold	0.8	0.3	-0.2	-0.6	0.1	1.1	1.5	1.1	0.4	0.1	-0.4

3 Qualitative Observations

This section begins with an exploratory analysis of our data. This will motivate the specific nonlinear regression analysis of the following section. Our aim is *not* to develop a futures trading strategy, which would have to include risk limits, trading cost minimization, and other features. Rather, we simply want to empirically measure and model the small autocorrelations of market returns as accurately as possible as a basis for future work.

3.1 Next-day Return vs. Trend Strength

We use the data of table 2 to measure the expected daily return of a futures market as a function of the trend strengths in that market on the previous day. To this end, we first construct $7305 \cdot 24 \cdot 10 = 1'753'200$ pairs of data. Each pair consists of the normalized log-return $R_i(t)$ in that market on day t, and one of the 10 trend strengths $\phi_{i,T}(t-1)$ on the previous day. So each return appears in 10 data pairs, each time paired with a different trend strength.

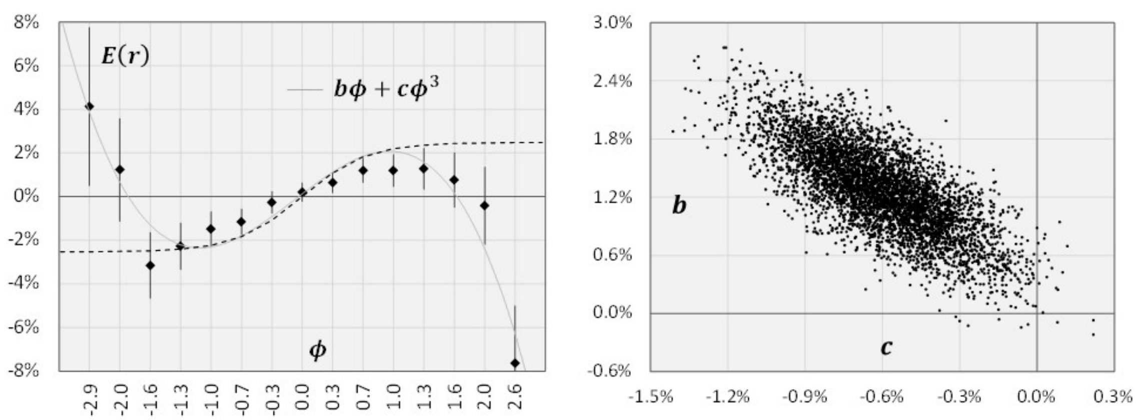

Figure 2: *Left*: The expectation value $E(r)$ of the next days return of a futures market can be modeled by a cubic polynomial of the current trend strength ϕ, where the linear term in ϕ represents trend-persistence and the cubic term in ϕ represents trend-reversion. *Right*: As confirmed by bootstrapping in section 4, the regression coefficients b and c corresponding to the linear and cubic terms are statistically highly significant.

We then group those pairs into 15 bins of increasing trend strength from $-\infty$ to -13/6, -13/6 to -11/6, ... ,-1/6 to+1/6, ... ,11/6 to 13/6, 13/6 to ∞. The mean trend strength within each bin is shown on the x-axis of fig. 2 (left). Within each bin, we average over the normalized return on the day after the trend has been measured. To obtain statistically significant results, we aggregate over the 28 years of daily returns for each market, across all 24 markets, and across different time scales. Fig. 2 (left) shows the results for the 4 monthly trend strengths (i.e., aggregated over $T = 6$ weeks, 3 months, 6 months, and 1 year).

We observe that the average next-day return is close to zero at zero trend strength, and grows linearly with the trend strength for small trend strengths. As the trend strength increases further, the average next-day return reaches a maximum. It then decreases again until it becomes zero somewhere below trend strength 2. For even stronger trends, the average return decreases dramatically. The same picture is mirrored on the left-hand side of the graph for down-trends. We have verified that this pattern remains almost the same if another day of delay is added, i.e., if the next-day return in our data pairs is replaced by the return 2 days later.

Thus, trends tend to revert when they become too strong. This makes sense intuitively: after strong trends, markets tend to be overbought or oversold, so one expects a reversion to "value". Our analysis quantify where exactly this happens: below a critical trend strength of 2, before trends become strongly statistically significant. Note that this is not in line with classical trend-following, which would follow the trend no matter how strong it becomes. The dashed line in fig. 2 (left) indicates the trading position that a classical trend-follower would take, depending on the trend strength.

3.2 Dependence on the Time Scale

In a next step, we analyze how the pattern observed in the previous sub-section depends on the time scale. To this end, we refine the bins used above: we split each bin into 10 bins, one for each of the 10 time scales. The resulting 15×10 refined bins are now too small and the results too noisy. To reduce the noise, we average the next-day returns over blocks of 3x3 neigboring bins (resp. 3x2 or 2x3 bins at the borders, 2x2 bins at the corners of the matrix of 15×10 bins), weighted by the number of returns in each bin. This yields the heat map of fig. 3 (left). Fig. 2 (left) can be thought of as a horizontal cross-section through this heat map along the dashed line.

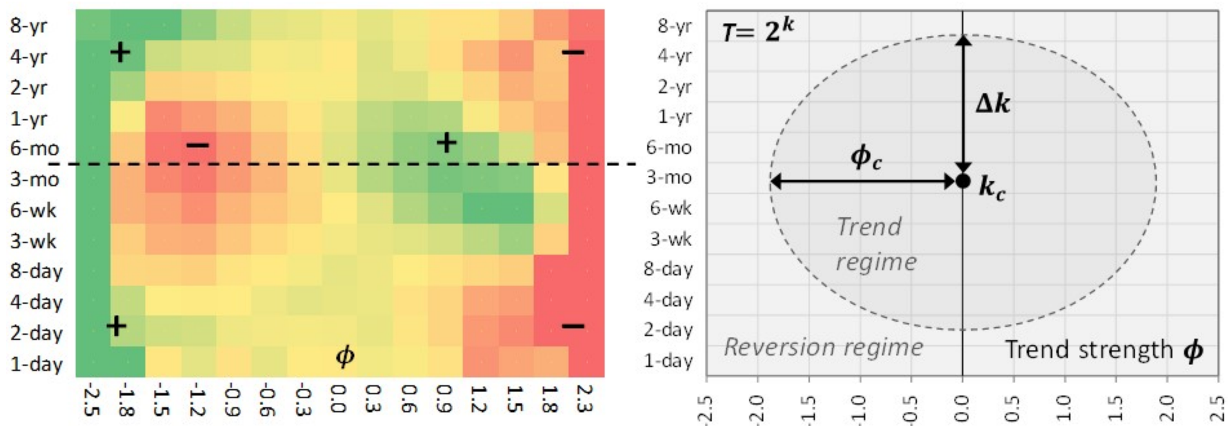

Figure 3: *Left*: A heat map shows how the expectation value of tomorrows return depends both on todays trend strength ϕ and its time horizon. Fig. 2 (left) can be thought of as a cross-section of Fig. 3 (left) along the dashed line. *Right*: The polynomial regression analysis of section 4 models the pattern of Fig. 3 (left) by an elliptic regime within which trends are persistent, and outside of which they revert. The values of the center and of the semi-axes of the ellipse are statistically highly significant, as confirmed by bootstrapping.

We observe that the trend- and reversion pattern of fig. 2 (left) is strongest on time scales from 1 month to 1 year. This is in line with the fact that typical trend-followers operate on those time scales. As the time scale increases or decreases, the pattern becomes weaker. The region where markets trend seems to disappear both for time scales of the order of economic cycles (several years) and for intra-week time scales. The phenomenon of reversion, on the other hand, appears to remain strong at all time scales.

3.3 Counting Degrees of Freedom

Before modeling these patterns in detail, let us do a back-of-the-envelope calculation of how many parameters we can hope to fit in our model without over-fitting our daily return data, and what fraction of the variance of these returns we can hope to explain by trend factors.

Our $7'305 \cdot 24 = 175'320$ daily log-returns are not independent, because the 24 markets are correlated with each other. How many independent markets are there? The daily returns are normalized to have variance 1. For a portfolio that invests 1/24 in each market, we find a variance of $\sigma^2 \sim 1/8$, just as if it contained $n_m = 8$ independent assets. A principal component analysis confirms that the first 8 (resp. 12) principal components explain 65% (resp. 80%) of the variance of the returns of our 24 markets. In this sense, these returns effectively live in a space of dimension $n_m \sim 8$. Adding more markets to our 24 time series does not significantly increase n_m.

What is the highest annualized Sharpe ratio S that one can hope to achieve by systematically trading a broadly diversified set of highly liquid futures markets based on trends and reversion? Experience with the Managed Futures ("CTA") industry suggests that S can be at best 1. The small number of CTA's that have achieved a higher Sharpe ratio for several years in a row presumably also pursue other strategies that are not purely based on trends, or they are not market-neutral (zero net exposure to each market over time).

An annualized Sharpe ratio of $S = 1$ implies a daily Sharpe ratio ρ for each market of

$$\rho = \frac{S}{\sqrt{260 \cdot n_m}} \sim 0.02.$$

So the predicted next-day return of a market has a correlation of $\rho = 0.02$ with the actual next-day return. E.g., if we only try to predict the sign of the next return, we can at best hope to be right on 51 and wrong on 49 out of 100 days. The adjusted R-squared (achieved out-of-sample in real trading) is then $R^2_{adj} \sim \rho^2 = 4$ basis points ($1bp = 4 \cdot 10^{-4}$). Clearly, the variance of financial market returns is overwhelmingly due to random noise.

If we fit k parameters to our data, and if our returns were independent and identically distributed ("iid"), then, for small R^2, the adjusted R-squared would be approximately

$$R^2_{adj} \sim R^2 - \frac{k}{N} \sim 4\ bp \quad \text{with} \quad N = 260 \cdot n_m \cdot Y \text{ data points} \tag{6}$$

for Y years of daily data. If we require that the correction for the in-sample bias does not erode more than 20% of our R^2, then we conclude that we cannot fit more than $k \sim N \cdot 1\ bp \sim 6$ parameters to our 28 years of data. This 20%-requirement is not too conservative, as our returns are only approximately "iid", and therefore the actual correction per fitted parameter will be higher than 20% (below, cross-validation will show that it is indeed about twice as big). We conclude that we must use parameters wisely, and not "waste" them on features that may be artifacts of our limited data set.

4 Regression Analysis

In this section, we confirm and quantify the observations of the previous section by nonlinear regression based on ordinary least squares. To this end, we model the next-day return of a market as a polynomial function of *both* the current trend strength in that market *and* its time scale. This regression is performed directly on the underlying 1'753'200 pairs of data, not on the bins we have defined in the previous section. Thus, our results are independent of any choice of how to split the data into bins.

4.1 Dependence on the Trend Strength

The graph in fig. 2 (left) suggests to model the next-day normalized log-return $R(t+1)$ (1) as a polynomial of the current trend strength $\phi(t)$ (2) across all markets and time scales:

$$R(t+1) = a + b \cdot \phi(t) + d \cdot \phi^2(t) + c \cdot \phi^3(t) + ... + \epsilon(t+1), \tag{7}$$

where ϵ represents random noise, and a measures the average risk premium μ_i/σ_i across all assets. We have performed a corresponding regression analysis on the 1'753'200 data pairs $\{r_{t+1}, \phi_t\}$. Using only the linear and cubic terms of (7) yields the results of table 3.

Table 3	Regression 1 with linear and cubic terms		
Coefficient	Value	Error	t-statistics
a	1.33%	±0.41%	3.3
b	1.29%	±0.44%	2.9
c	−0.62%	±0.24%	2.6
R-squared	Single time scales	Aggregated across time scales	
R^2	1.31 bp	4.91 bp	
R^2_{adj}	1.07 bp	4.01 bp	

Since market returns cannot be assumed to be independent, identically distributed normal variables, we cannot trust the usual estimates of the t-statistics, adjusted R-squared, and F-statistic. Instead, the test statistics shown in table 3 are measured empirically as follows:

- The standard errors of the coefficients and their t-statistics are computed by bootstrapping: from the 7305 days, we randomly create 5000 new samples of 7305 days each, with replacement. I.e., some days occur several times in a new sample, while other days do not occur at all. Regression on each new sample of days yields the distribution of 5000 regression coefficients b and c shown in fig. 2 (right). The errors of the coefficients in table 3 represent half the difference between the 84th and 16th percentile, which equals the standard deviation in the case of a normal distribution.

- The adjusted R-squared is computed by 15-fold cross validation: we split our 7305-day time window into 15 sub-windows of 487 consecutive days each. For each sub-window, we predict the next-day returns based on the betas obtained by regression on the other 14 sub-windows. The square of the correlation between the predicted and the actual returns is the out-of-sample R-squared R^2_{adj} reported in table 3.

- Table 3 also reports R^2 and R^2_{adj} "aggregated across time scales". Those are based on using the equally-weighted mean of the 10 trend strengths on each day to predict the next-day return for each market. I.e., we combine the 10 different trend factors into a single one, which naturally has a higher predictive power than each single factor by itself. This regression is thus performed on only $7305 \cdot 24 = 175'320$ pairs of data.

- The F-statistics can be computed numerically to be $F = 4.6$ with a p-Value of 0.7% by modelling the distribution of regression coefficients in fig. 2 (right) by an elliptical distribution. However, the distributions in subsequent sections are not even approximately elliptical. We will therefore use R^2_{adj} and not F to compare the out-of-sample explanatory power of our models with each other.

The regression results of table 3 confirm and quantify our conclusions from the previous section. We see that the values of b and c - although very small - are statistically highly significant, despite the fact that market returns are neither normally distributed, nor independent, nor identically distributed. So is the average long-term risk premium a. The overall result is significant at the 99% level. The aggregated out-of-sample R^2_{adj} that combines the predictions from all 10 time scales matches our initial expectation of 4 *bp* (6). Note that the correction for the aggregated in-sample bias, $R^2 - R^2_{adj} = 0.90$ *bp*, is much bigger than what would have been expected if returns were "iid", namely $2/(260 \cdot 8 \cdot 28) = 0.34$ *bp*.

We have also tested the quadratic, quartic and quintic terms in ϕ_T in (7). None of them turned out to be statistically significant at the 95% level. We therefore drop them from our analysis to avoid over-fitting the historical data (i.p., the t-statistics for d is below 1).

4.2 Dependence on the Time Scale

Next, we try to refine our model by measuring the dependence of the coefficients b and c on the time scale, the asset class, and the time period. We begin with the time scale T: we model the expected return as a function of both the trend strength *and* T, trying to replicate fig. 3 (left). We first repeat the linear and cubic regression (7) of the previous sub-section for each of the 10 time scales seperately. The resulting coefficients $b(T)$ and $c(T)$ are plotted in fig. 4 (we neglect the overall risk premium a, which is not the focus of this paper).

From the coefficient b of the linear term, which models trends, we observe that trend-following works best at time scales from 3 months to 1 year, where b peaks. This appears to be in line with the time scales on which typical CTAs follow trends. Even at those scales, the critical trend strength

$$\phi_c = \sqrt{-b/c} \ \leq 1.91 \ ,$$

beyond which trends tend to revert, is below 2. So trends never become strongly significant. For scales below a few days and above several years, b seems to go to zero, which means that trends are not persistent there. This is consistent with the heat map in fig. 2 (right).

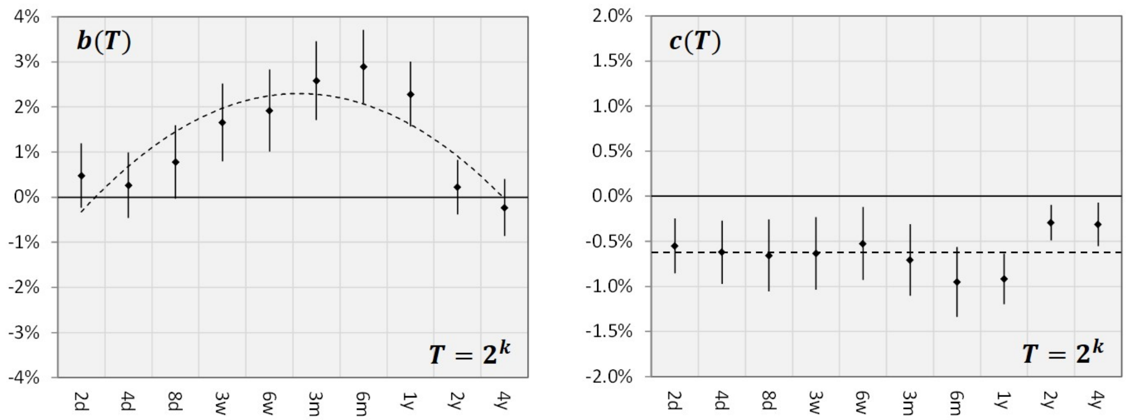

Figure 4: *Left:* The coefficient $b(T)$ of the linear term (corresponding to the trending of markets) peaks at time scales T of 3 months to 1 year. Its scale dependence is modeled by a parabola. *Right:* The coefficient c of the cubic term (corresponding to the reversion of markets) does not show a clear dependence on the time scale.

On the other hand, the coefficient c of the cubic term, which ensures that trends revert, is quite stable, except that its magnitude appears to be somewhat lower for the 2- and 4-year scales. The 2- and 4-year results must be taken with a grain of salt, though, as there are only 14 independent 2-year trends and 7 independent 4-year trends in our 28-year time window. Indeed, a preliminary check based on 60 years of monthly returns resulted in $c \sim -0.6\%$ at the 8-year scale. The available data thus indicate that, unlike trend-following, mean reversion works at all time scales. This is also consistent with our earlier observation from the heat map in fig. 3 (left).

To quantify these observations, we refine our regression ansatz (7). We continue to model the cubic coefficient c by a constant, but we model the dependence of the linear coefficient $b(k)$ on the logarithm k of the time scale $T = 2^k$ by a parabola:

$$\begin{aligned} b(k) &= b - e \cdot (k - k_0)^2 \\ \Rightarrow R(t+1) &= b \cdot \left\{ 1 - \frac{(k-k_0)^2}{(\Delta k)^2} \right\} \cdot \phi(t) + c \cdot \phi^3(t) + \epsilon(t+1) \end{aligned} \tag{8}$$

with $(\Delta k)^2 = b/e$. The critical trend strength $\phi_c(k) = (-b(k)/c)^{1/2}$, at which the expected return $E(R_{t+1})$ is zero (without the noise ϵ), and beyond which trends revert, is then an ellipse with semi-axes Δk and $\phi_c(k_0)$. Altogether, we now fit 4 parameters to our data:

- The "persistence of trends" b, i.e. the value of $b(k)$ at its peak
- The "strength of reversion" c
- The range $k_0 \pm \Delta k$ of the log of the time scales $T = 2^k$ at which markets may trend.

A nonlinear regression on the full underlying data set yields the results of table 4. Fig. 3 (right) plots the elliptic region, which seperates the trend regime (inside) from the reversion regime (outside). For its second semi-axis, we find $\phi_c(k_0) = 1.78 \pm 0.32$. This quantifies the empirical heat map in fig. 3 (left) and confirms that highly significant trends of strength (i.e., t-statistics) $\phi_c \geq 2$ always tend to revert.

Table 4	Refined regression 2 with 4 parameters		
Coefficient	Value	Error	t-Stat.
b	2.00%	±0.51%	3.9
c	−0.63%	±0.24%	2.6
k_0	5.78	±0.72	8.0
Δk	4.87	±1.16	4.2
R-squared	Single time scales	Aggregated across time scales	
R^2	1.64 bp	7.51 bp	
R^2_{adj}	1.22 bp	5.81 bp	

The errors of the regression parameters in table 4 are again computed by bootstrapping. The distribution of b and c looks the same as in the univariate case (Fig. 2, right). Fig 5 (left) plots the distribution of the values of the center k_0 and the semi-axis Δk of the ellipse that separates the trending regime from the reversion regime. The "aggregate" R^2 and R^2_{adj} in table 4 now refer to a single factor that is a linear combination of the 10 trend strengths for the 10 time scales, weighted by a parabolic weight function proportional to $b(k)$. Note that the aggregated adjusted R-squared now exceeds our original expectation (6) of 4 *bp*.

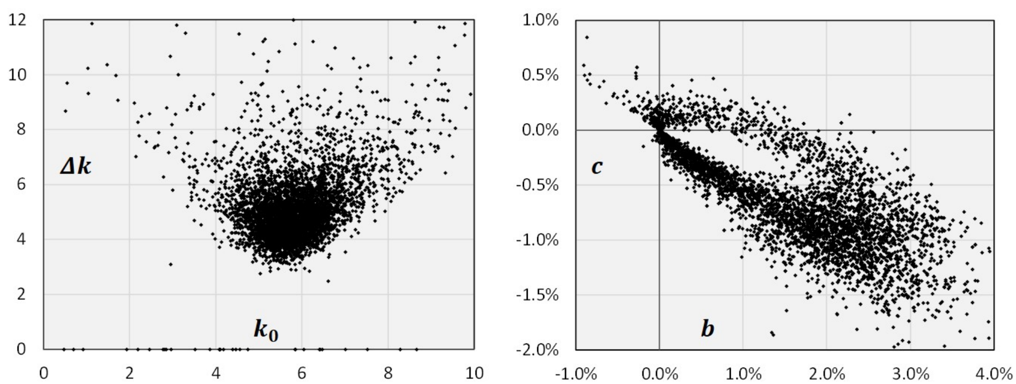

Figure 5: *Left:* Distribution of the regression coefficients for the center k_0 and width Δk of the elliptic region within which trends are persistent, as obtained by bootstrapping. *Right*: Distribution of the linear and cubic regression coefficients b, c for a rejected alternative model.

We have tried to further refine ansatz (8). First, $b(k)$ in fig. 4 (left) seems to be tilted to the right, which could be accounted for by models such as $b(k) \sim b - e\cdot(k-k_0)^2 + f\cdot(k-k_0)^3$, or $b(k) \sim \exp(f\cdot k)\cos((k-k_0)/\Delta k)$. We find that such models increase the adjusted R-squared at best marginally. So we use the simplest model (8) in this paper, to avoid over-fitting the historical data. Second, we also tested for a polynomial dependence of c on k. The most significant ansatz was that $c(k)$ is also a parabola proportional to $-b(k)$. In this case, the critical trend strength is constant across all time scales, and the region within which markets trend is rectangular instead of elliptic. The distribution of the parameters $b(k_0), c(k_0)$ then turns out to have the shape of the stretched annulus shown in fig. 5 (right). However, this scenario seems less likely, as it yields a much lower adjusted R-squared (0.77 bp).

4.3 Dependence on the Asset Class

Can we refine our 4-parameter-model further by distingushing between asset classes, i.e., by fitting seperate regression parameters for equities, bonds, currencies and commodities?

To test this, we have repeated the regression analysis of the previous section for these 4 sub-sets of our data. Fig. 6 (left) shows the 16th, 50th and 84th percentile of the values of the 4 regression parameters for each asset class, divided by the values of the regression parameters for the overall sample. E.g., for equities, the quantiles for b are $(1.80\%, 2.82\%, 4.01\%)$, which are multiples of $(0.90, 1.41, 2.01)$ of the overall regression coefficient 2.00% (see table 3). Those multiples are what is shown in the first bar of fig. 6 (left).

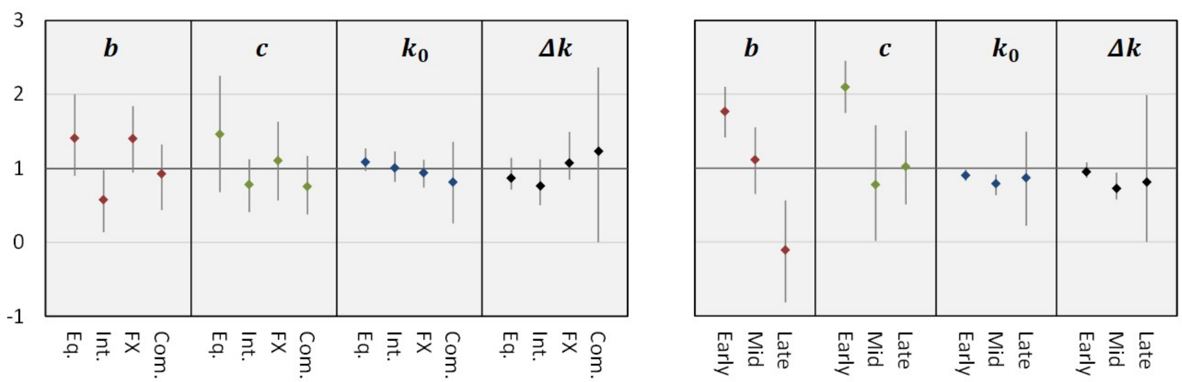

Figure 6: *Left:* Ratios of the values of our 4 regression parameters for equities, interest rates, FX rates and commodities, divided by their overall values across all asset classes. The ratios do not differ significantly from 1. *Right:* The analoguous ratios for the early, middle and late third of the time period. At least b has decreased significantly over time.

For each asset class, we observe that the values of all four parameters are within one standard error of the overall parameter values. Thus, based on our data set, we cannot justify fitting different parameters of our trend-reversion model to individual asset classes, let alone to individual assets. This is consistent with our back-of-the-envelope caculation of subsection 3.4, which suggests that we cannot fit as many as $4 \times 4 = 16$ parameters for the four different asset classes to our data. We thus use a single, universal model for all assets.

4.4 Dependence on the Time Period

Lastly, we investigate how our patterns have been evolving in time. First, we split up the 28-year (7305-day) time window into three non-overlapping sub-periods of 2435 days each: An early period from Jan 1992 to Apr 2001, a middle period from May 2001 to Aug 2010, and a late period from Sep 2010 to Dec 2019.

Fig. 6 (right) shows the 16th, 50th and 84th percentile of the values of the 4 regression parameters for each of these time-windows, again divided by the values of the regression parameters for the overall sample. The persistence of trends b has consistently and significantly decreased over time. With less consistency, this can also be observed for strength of reversion c, while no clear trend is visible for the range $k_0 \pm \Delta k$, within which trends persist. The decrease of b is in line with the industry observation that trend-following no longer works as well as it used to: markets seem to have become more efficient in this respect. Given the decrease in trading costs, an increase in algorithmic trading, and an increase in assets under management invested in trend-following, this is not surprising.

To verify and quantify these observations, let us introduce time t, measured in years, with its origin $t = 0$ on Dec 31, 2005, the center of our 28-year time window. We now further refine our model (8) by including linear trends in b and c while leaving k_0 and Δk constant:

$$b(t) = \bar{b} \cdot (1 - Q_b \cdot t) \quad , \quad c(t) = \bar{c} \cdot (1 - Q_c \cdot t)$$

The results of a regression analysis, including bootstrapping and cross-validation, are shown in table 5. The decrease of c, which measures the strength of reversion, is only weakly significant. However, the decrease of b, which measures the persistence of trends, is significant at the 97% confidence level. In principle, we could compute the year Y_0, in which $b(t) = 0$:

$$Y_0 = 2005 + \frac{1}{Q_b} \ \in \ \{2012, 2028\} \quad \text{with expectation value} \quad Y_0 \sim 2017.$$

Thus, if one were to take this linear down-trend of b literally, one would conclude that the phenomenon of persistent market trends may have already disappeared. However, there are other scenarios for the time decay of the persistence of trends that are consistent with our data. E.g., for an exponential decay scenario, in which trends never disappear, we find an only slightly lower adjusted R-squared of 1.39 *bp* instead of 1.52 *bp*, with

$$b \sim \bar{b} \cdot e^{-Qt} \quad \text{with decay rate} \quad Q \sim (24 \text{ years})^{-1}.$$

Table 5	Refined regression 4 with 6 parameters		
Coefficient	Value	Error	t-Stat.
$\bar{b}$	1.91%	±0.53%	3.6
$\bar{c}$	−0.62%	±0.22%	2.8
k_0	5.83	±0.55	10.6
Δk	4.97	±0.51	9.7
Q_b	0.088	± 0.047	1.9
Q_c	0.047	± 0.049	1.0
R-squared	Single time scales	Aggregated across time scales	
R^2	1.97 bp	8.90 bp	
R^2_{adj}	1.52 bp	6.97 bp	

We have also tested scenarios where all 4 parameters or other subsets of them change at different rates, but found that all of these scenarios significantly reduce the adjusted R-squared. It is left for future work to investigate the time evolution of the pattern of trends and reversion in more detail.

5 Analogies with Critical Phenomena

We conclude by pointing out some striking analogies between the empirical observations of sections 3 and 4 and critical phenomena in statistical mechanics.

Analogies with critical phenomena are plausible, if financial markets are regarded as statistical mechanical systems, whose microscopic constituents are the Buy/Sell orders of individual traders. It is conceivable that these orders can be modeled by degrees of freedom that sit on the vertices of a hypothetical "social network" of traders. These degrees of freedom may interact with each other in analogy with spins on a lattice, thereby creating the macroscopic phenomena of trends (herding behavior) and reversion (contrarian behavior).

Candidates for such networks include small-world networks [13], scale-free networks [14], or the Feynman diagrams of large-N field theory [15]. To our knowledge, no convincing specific model has emerged as a consensus so far. Our results provide an empirical basis for accepting or rejecting such candidates: any statistical-mechanical model of financial markets, if accurate, must replicate the interplay of trends and reversion observed in this paper.

To be specific, let us focus on the 6-month time horizon, i.e., $T = 2^7 = 128$ trading days (the results for other horizons are similar). In the continuum limit of infinitesimal time steps, equation (7), combined with the recursion relations (5), then implies the following second-order stochastic differential equation for the trend strength ϕ:

$$(\frac{d}{dt} + \frac{1}{64})^2 \phi(t) = -\frac{\partial}{\partial \phi} V(\phi) + \epsilon(t) \quad \text{with} \quad V(\phi) = -\frac{b}{512} \cdot \phi^2 + \frac{|c|}{1024} \cdot \phi^4. \tag{9}$$

A similar first-order equation can also be derived for its simpler cousin ψ in (4):

$$(\frac{d}{dt} + \frac{1}{64}) \psi(t) = -\frac{\partial}{\partial \phi} \tilde{V}(\psi) + \epsilon(t) \quad \text{with} \quad \tilde{V}(\psi) = -\frac{\tilde{b}}{4} \cdot \psi^2 + \frac{|\tilde{c}|}{8} \cdot \psi^4, \tag{10}$$

with the following empirical parameter values, as measured by regression:

$$b = 2.94\%, \ c = -0.95\%, \ \tilde{b} = 1.79\%, \ \tilde{c} = -0.66\%.$$

(10) is the purely dissipative Langevin equation, which describes the dynamics of the order parameter of certain statistical mechanical systems near second-order phase transitions [16, 17]. This is consistent with the conjecture that the trend strength (defined as either ϕ or ψ) plays the role of an order parameter, in analogy with the magnetization in spin models.

To take the analogy further, statistical mechanical systems near second-order phase transitions are characterized by a universal scaling behavior. E.g., a scalar field theory with a ϕ^4

potential similar to the potentials V in (9,10) describes the Ising model and other physical systems in the same universality class (such as water and steam, or CO_2) near their critical points. For all systems within this universality class, the parameters b and c show the same dependence on the scale of the system [16, 17].

In section 4, we have seen that - within the limits of statistical significance - the values of the coefficients b and c are the same for very different markets, such as equity indices, bonds, FX-rates, and commodities. The parameters k_0 and Δk in (8), which characterize how b behaves under a rescaling of the time horizon T, are also the same. This could be an expression of universality and scaling in financial markets. To confirm this, it will be key to examine how the scaling behavior in (8) extends to intra-day and multi-year time horizons $T = 2^k$ with $k > 10$ or $k < 1$. For example, it might reflect a complex critical exponent [18]. Together with the stochastic differential equations (9,10), the empirically observed scaling behavior may uniquely specify a particular social network that models financial markets.

6 Summary and Discussion

In this paper, we have empirically observed the interplay of trends and reversion in financial markets, based on 30 years of daily futures returns across equity indices, interest rates, currencies and commodities. We have considered trends over ten different time horizons of $T = 2^k$ days with $k \in \{1, 2, ..., 10\}$, ranging from 2 days to approximately 4 years. For a given market i on a given day t, we have defined the trend strength $\phi_{i,k}(t)$ as the statistical significance (t-statistics) of a smoothed version of its mean return over the past 2^k days.

Our key results, as illustrated in figs 2 and 3, are the following: for a given market i and each time horizon labeled by k, tomorrow's normalized log-return $R_i(t+1)$ can accurately be modeled by a cubic polynomial of today's trend strength in that market:

$$R_i(t+1) = \alpha_i + b \cdot f_k \cdot \phi_{i,k}(t) + c \cdot \phi_{i,k}^3(t) + \epsilon_i(t+1). \tag{11}$$

Here, ϵ_i represents random noise. α_i is the normalized long-term risk premium of market i, which has not been not the focus of this paper. Instead, we have concentrated on determining the coefficients b, c, and the function f_k, which measure how the expected return of an asset varies in time. As discussed, we interpret b as the persistence of trends, and c as the strength of trend reversion. Within the limits of statistical significance, we find that they are universal, i.e., the same for all assets. Over the past 30 years, we find from table 4:

$$b \sim +2.0\% \ , \quad c \sim -0.6\% \tag{12}$$

While the strength of reversion is approximately constant, we find that the persistence of trends depends on the time horizon of the trend. Within the range of time scales considered here, it can be approximated by a parabolic function of the log of the time scale:

$$f_k \sim 1 - \frac{(k-k_0)^2}{\Delta k^2} \quad \text{with} \quad k_0 \sim 6 \ , \quad \Delta k \sim 5. \tag{13}$$

This implies that trends may only be stable if the log of the time horizon is within the range $k_0 \pm \Delta k$, corresponding to time scales from a few days to several years. The parameters k_0 and Δk are also universal. By bootstrapping and cross-validation, we have found that all four parameters in (12) and (13) are statistically highly significant out-of-sample.

Let us now discuss these results. First, they imply that trends tend to revert above a critical trend strength, where the linear and cubic term in (11) balance each other. This critical trend strength lies below 2 in all cases. In other words, by the time a trend has become statistically significant, such that it is obvious in a price chart, it is already over. This supports a variant of the efficient market hypothesis [19, 20, 21]: inefficiencies in financial markets are eliminated before they become strongly statistically significant.

Despite being insignificant, small trends can add value for investors through tactical asset allocation strategies, if accompanied by appropriate risk management and broad diversification across assets. While this paper does not recommend investment strategies, we note that the inclusion of the cubic term in (11) appears to be a major improvement over classical trend-following, as it takes investors out of trends before they are likely to revert. We believe that publishing such strategies and subjecting them to an academic discussion and independent review will ensure a high level of professionality in asset management.

Trend-following has been very successful in the 80's and 90's, when it was the proprietary strategy of a limited number of traders. By now, large amounts of capital have flown into this strategy, so it can no longer be expected to provide a free lunch. Indeed, while we have not observed a consistent weakening of the strength of reversion c, we have seen that the persistence of market trends b has clearly decreased over the decades. This measures the rate, at which markets are becoming more efficient with respect to trends.

What will happen, when all investors try to exploit trends and reversion? Then both phenomena should weaken, until they earn a moderate equilibrium return that just compensates for the systematic risk of these strategies and their implementation costs. In this sense, trend-following and mean reversion may just become "alternative market factors" as part of the general market portfolio. In fact, the weakening of b that we have observed here indicates that this development is already well underway at least for traditional trend-following.

On a conceptual level, our precise measurement of trends and reversion reveals intriguing analogies with critical phenomena in physics. They support the conjecture that financial markets can be modeled by statistical mechanical systems near second-order phase transitions. In such a model, Buy/Sell orders would represent microscopic degrees of freedom that live on a "social network" of traders. The trend strength would play the role of an order parameter, whose dynamics is described by the stochastic differential equations (9,10). Together with an extension of the scaling behavior (13) to shorter and longer horizons, these equations provide an empirical starting point for developing such a model.

If such a statistical mechanical theory of financial markets can be established, it will introduce powerful concepts from field theory into finance, such as the renormalization group, critical exponents, and Feynman diagrams. This will lead to a new and deeper understanding of financial markets, and phenomena such as trends, reversion, and shocks will become more accessible to scientific analysis. Further research in this direction is underway.

Acknowledgements

I would like to thank W. Breymann for discussions and encouragement to publish these observations, and A. Ruckstuhl for advice on statistical matters. I would also like to thank J. Behrens for many interesting conversations and cooperation at Syndex Capital Management from 2008-2014, where some of these observations were initially made. This research is supported by grant no. CRSK-2 190659 from the Swiss National Science Foundation.

References

[1] For a popular review, see, e.g., Covel, M., 2007. The Complete TurtleTrader: The Legend, the Lessons, the Results. Collins.

[2] Cutler, D.M., Poterba, J.M. and Summers, L.H., 1991. Speculative dynamics. The Review of Economic Studies, 58(3), pp.529-546.

[3] Silber, W.L., 1994. Technical trading: when it works and when it doesn't. The Journal of Derivatives, 1(3), pp.39-44.

[4] Asness, C.S., 1997. The interaction of value and momentum strategies. Financial Analysts Journal, 53(2), pp.29-36

[5] Fung, W. and Hsieh, D.A., 1997. The risk in hedge fund strategies: Theory and evidence from trend followers. The review of financial studies 14, no. 2 (2001): 313-341.

[6] Jaeger, L., 2002. Managing risk in alternative investment strategies: Successful investing in hedge funds and managed futures. Pearson Education.

[7] Miffre, J. and Rallis, G., 2007. Momentum strategies in commodity futures markets. Journal of Banking & Finance, 31(6), pp.1863-1886.

[8] Shen, Q., Szakmary, A.C. and Sharma, S.C., 2007. An examination of momentum strategies in commodity futures markets. Journal of Futures Markets: Futures, Options, and Other Derivative Products, 27(3), pp.227-256; Trend-following trading strategies in commodity futures: A re-examination. Journal of Banking & Finance, 34(2), pp.409.

[9] Moskowitz, T.J., Ooi, Y.H. and Pedersen, L.H., 2012. Time series momentum. Journal of financial economics, 104(2), pp.228-250.

[10] Menkhoff, L., Sarno, L., Schmeling, M. and Schrimpf, A., 2012. Currency momentum strategies. Journal of Financial Economics, 106(3), pp.660-684.

[11] Lemprire, Y., Deremble, C., Seager, P., Potters, M. and Bouchaud, J.P., 2014. Two centuries of trend following. arXiv preprint arXiv:1404.3274.

[12] Hurst, B., Ooi, Y.H. and Pedersen, L.H., 2017. A century of evidence on trend-following investing. The Journal of Portfolio Management, 44(1), pp.15-29;

[13] D.J. Watts and S.H. Strogatz, "Collective dynamics of 'small-world' networks", Nature 393 (1998)

[14] Barabsi, A.L. and Albert, R., 1999. Emergence of scaling in random networks. Science, 286(5439), pp.509-512.

[15] Brzin, E., Itzykson, C., Parisi, G. and Zuber, J.B., 1993. Planar diagrams. In The Large N Expansion In Quantum Field Theory And Statistical Physics (pp. 567-583).

[16] Hohenberg, P.C. and Halperin, B.I., 1977. Theory of dynamic critical phenomena. Reviews of Modern Physics, 49(3), p.435.

[17] J. Zinn-Justin, Quantum Field Theory and Critical Phenomena, Oxford U. Press, 1989

[18] Sornette, D., 1998. Discrete-scale invariance and complex dimensions. Physics reports, 297(5), pp.239-270.

[19] Mandelbrot, Benoit (January 1963). "The Variation of Certain Speculative Prices". The Journal of Business. 36 (4): 394. doi:10.1086/294632. ISSN 0021-9398.

[20] Samuelson, Paul (1965). "Proof That Properly Anticipated Prices Fluctuate Randomly". Industrial Management Review. 6: 4149.

[21] Fama, Eugene (1970). "Efficient Capital Markets: A Review of Theory and Empirical Work". Journal of Finance. 25 (2): 383. doi:10.2307/2325486. JSTOR 2325486.