Research in International Business and Finance 60 (2022) 101603

Contents lists available at ScienceDirect

Research in International Business and Finance

journal homepage: www.elsevier.com/locate/ribaf

Identifying bull and bear market regimes with a robust rule-based method

Piotr Zegadło

Kozminski University, Jagiellonska 57/59, 03-301 Warsaw, Poland

A R T I C L E I N F O

JEL classification:
G10
C18

Keywords:
Financial market
Market regimes
Financial cycle
Bull market
Bear market
Trading rules

A B S T R A C T

A new method for identifying bull and bear financial market regimes is proposed, related to a classic algorithm for picking turning points in the business cycle. Our approach uses only a single discrete parameter, adjusted to the periodicity of the data, which largely removes subjectivity from the regime identification process. Applying it to the benchmark Dow Jones Industrial Average index data, we show the method's capability of obtaining a classification similar to competing multi-parameter methods, without imposing any conditions on regime duration or amplitude. Despite its relative simplicity, our algorithm can be easily applied across different asset classes, where its direct competitors may fail, as we show in an out-of-sample identification example involving other stock indices, exchange rates and commodities. Our new market regime classification rule constitutes a relatively straightforward, but important methodological development that can be used in a broad palette of financial market research problems, where discerning different regimes is beneficial.

1. Introduction

"Bull market" and "bear market" are traditional terms used for periods of rising and falling asset prices, respectively. They feature extensively in academic literature, as classifying a series of asset prices into bull and bear periods has numerous theoretical and practical applications. The most basic advantage of accounting for multiple regimes in financial market returns is that it often improves the quality of models for returns and volatility. This helps with the examination of key financial variables, such as excess returns or systematic risk, and with asset pricing in general, as shown in the classic papers of Fabozzi and Francis (1977) and Turner et al. (1989), as well as later works by Ryden et al. (1998), Veronesi (1999), Gordon and St-Amour (2000), Diebold and Inoue (2001), and Maheu et al. (2012), among others. Estimation of market regime properties may also assist policymakers in various areas of economic policy, including financial stability, systemic risk or monetary policy, which was discussed in papers by Cashin et al. (2002), Rigobon and Sack (2003) or Bohl et al. (2007). Early identification of a regime change can also be used in the prediction of recessions and financial crises, as financial variables have been used as leading economic indicators (e.g. Estrella and Mishkin, 1998 or Nyberg, 2010). Determining the differences between price dynamics in bull and bear regimes may aid investors in the timing of investment decisions and in risk management – especially if the current market regime can be swiftly and correctly recognised. Important contributions within this research direction were made by Pagan and Sossounov (2003), Guidolin and Timmermann (2005), Cunado et al. (2010) or Maheu et al. (2012).

Still, no strict definition of the terms "bull market" and "bear market" exists, which leads to widespread subjectivity in market

E-mail address: pzegadlo@kozminski.edu.pl.

https://doi.org/10.1016/j.ribaf.2021.101603
Received 25 November 2020; Received in revised form 29 September 2021; Accepted 19 December 2021
Available online 24 December 2021
0275-5319/

regime identification. Some methods concentrate on the duration of the asset price's rises and falls, others consider the amplitude of the price changes. Dynamic properties of returns may also be investigated. Our paper proposes a way of reducing that widespread subjectivity by introducing a new market regime classification method. It uses only a single discrete parameter. In this paper we show its robustness to differing statistical properties of returns in the sense of the method being applicable to a variety of asset classes (stock indices, exchange rates, commodities) without any parameter adjustments.

Assorted methods of discerning financial market regimes have been employed in the literature. Contrary to the business cycle phase dating methodology which attracts much academic attention (see Hamilton, 2011 for a survey), the development of financial market regime dating methodology surprisingly seems to have been less dynamic. Markov regime-switching processes are often used to identify regimes, both in the business cycle and in the financial markets. The earliest financial applications include Turner et al. (1989) and Pagan and Schwert (1990). However, Harding and Pagan (2002) pointed out the subjectivity involved in specifying Markov-switching models within the context of business cycle dating. They suggested using possibly non-parametric techniques instead. Such techniques are often based on classification rules rather than fully specified models, like the algorithm described in the seminal work of Bry and Boschan (1971). It constitutes a rule-based way of picking local maxima and minima in economic activity data, based on a chosen time window. An important adaptation of this algorithm to financial data was a method presented by Pagan and Sossounov (2003), further abbreviated as PS. Lunde and Timmermann (2004) proposed a different rule-based method (abbreviated further as LT), concentrated not on the time aspect, but on the magnitude of asset returns within market regimes.

Harding and Pagan (2002) as well as Kole and van Dijk (2017) argued that rule-based methods were better suited to ex-post regime dating than Markov-switching models. Still, the approaches of Pagan and Sossounov (2003) as well as Lunde and Timmermann (2004) retain a large degree of subjectivity. Altering their parameters – which control assumed regime duration and/or amplitude characteristics – may lead to significantly different classification results.

Recently there has been renewed academic interest in market regime identification methods. Kole and van Dijk (2017) compared the performance of various methods both in sample and out of sample. Hanna (2018) closely examined PS and LT algorithms, and suggested a new approach based on a slight modification of the LT rule. This paper furthers the recent developments in financial market regime literature, proposing a new rule-based method, which largely removes subjectivity from the process of market regime classification. Similarly to its direct competitors, the algorithm we present is a heuristic rather than a formal model. We impose no conditions on regime duration or amplitude. The single discrete parameter used in the identification process is adjusted to the periodicity of the data and does not require optimisation. We illustrate our approach with an application to 3 stock market index time series, as well as other series exhibiting different dynamics – exchange rates and commodity prices.

Despite using just a single parameter, our algorithm is capable of obtaining an in-sample regime classification of the stock index which is fairly similar to those of the multi-parameter rule-based methods using typical parameter values. Out-of-sample, our method performs better or on par with the competitors, considering Sharpe ratio as a performance measure. Its advantages are conspicuous in case of time series whose characteristics differ from stock market data. No changes to the parameter value of our algorithm are necessary when applying it to different datasets, which makes it a robust alternative to methods with parameters chosen specifically to work in the equity market. This feature of the new method enables coherent cross-asset analysis of bull and bear markets. Such an analysis is not only useful for financial economists researching market interdependencies between geographies and asset classes, but also for portfolio managers studying long-term asset correlations and their changes in different market conditions. While using previously existing algorithms, comparing market regimes between asset classes or even assets in the same class required manual parameter adjustment. This interfered with the process, making it both more subjective and less immune to researcher's preconceived notions, for example about the duration and amplitude of market regimes in a specific market.

This paper is structured in the following way. Section 2 introduces our new regime classification rule, while Section 3 compares it in practice to two widely cited rules known in the literature. We consider 12 different datasets and compare the regime identification algorithms, taking into account the resulting bull and bear regime properties. Section 4 deals with out-of-sample performance of our rule, its direct competitors and two benchmarks. We construct simple trading rules based on regime classifications and use the resulting Sharpe ratio as a goodness-of-fit measure. Again datasets with different return characteristics are employed to highlight the differences between competing methods. Section 5 concludes. Dataset description is presented in Appendix A, while additional tables and figures are included in Appendix B.

2. New rule-based method for regime identification

2.1. Algorithm description

Two most widely cited rule-based methods for financial market regime identification are the PS and the LT algorithms. The former is based on the Bry and Boschan (1971) algorithm, which entails picking local maxima and minima in smoothed data series within a pre-specified time window. Moving averages are used as smoothing filters for the input data. The alternation of peaks and troughs is enforced by selecting the highest of multiple consecutive peaks not separated by troughs (or the lowest of multiple consecutive troughs), with assumptions regarding minimum duration of each phase and of the full cycle (consisting of two phases). The PS approach keeps the input data unsmoothed and allows for phases shorter than the minimum specified duration – if their amplitude is sufficiently large. Lunde and Timmermann (2004) proposed an algorithm concentrated on the magnitude of the returns in each market phase. It assumes that the regime had changed at turning point P if the rate of return λ since point P exceeded a pre-specified amount. The trigger return rate is treated as a parameter and allowed to differ between bull (λ_1) and bear (λ_2) regimes. Hanna (2018) suggested setting the trigger rates λ_1 and λ_2 in such a way that their equivalent logarithmic return rates were equal.

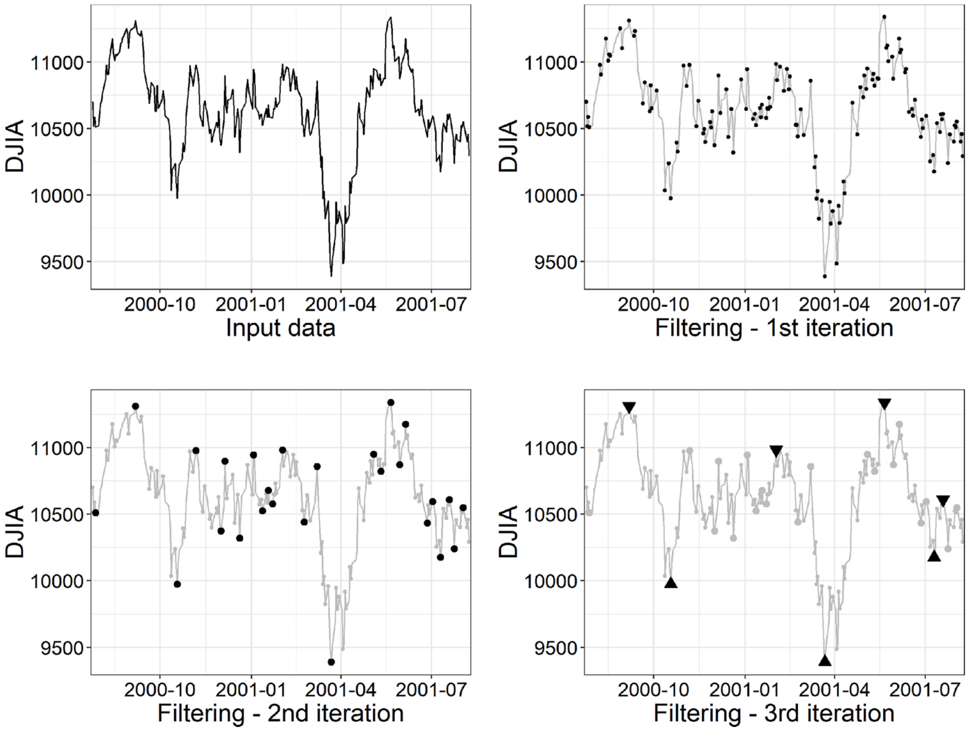

Fig. 1. Filtering algorithm application example. Note: Black markers indicate the detected turning points.

As evident from the method's mechanics, the PS algorithm defines the bull (bear) market as a period when the price rises (falls) for at least a specific number of days (with exceptions possible for large enough price movements). The LT algorithm defines the bull (bear) market as a period in which the price rises (falls) by at least a specific amount. We propose a new rule-based approach to market regime identification. Its implicit definition of the bull market is that of a period when the price reaches new maxima (defined in a specific way). No minimum duration or amplitude of the price movement is required. The mechanics of the algorithm assure that each bull period starts from a local minimum that is the lowest price point during that period and ends with a local maximum which constitutes the highest price point within that period. Whenever the price exceeds the previous local maximum, the bull period is deemed to continue. Conversely, if the next local maximum is lower than the previous one, the bull period is recognised to have ended with the previous, higher maximum. The opposite rules are true for bear periods.

In practice, our algorithm takes a simplified version of the Bry and Boschan (1971) procedure and employs it to filter the input time series multiple times. For the first time, the unprocessed daily closing prices are used as the procedure input. Peaks and troughs are obtained from local maxima and minima of daily data. For the day t, the daily closing price C_t is defined as a local maximum (called peak P_t), when the preceding and the succeeding daily closing prices are lower ($C_{t-1} < C_t > C_{t+1}$). The local minima (local troughs T_t) are obtained accordingly, with the condition $C_{t-1} > C_t < C_{t+1}$. When two or more daily closing prices in a row are equal, the last of them is compared with the prices directly preceding and succeeding the string of equal prices. This step of the algorithm is equivalent to collapsing any runs of subsequent positive or negative price changes. It can also be viewed as running the PS algorithm with a 1-day identification window and no other conditions.

The two resulting series (peaks and troughs) are then used as the inputs for the next iteration of the filtering operation, in which such peaks P_i are identified, which are higher than the preceding and the succeeding peaks, that is $P_{i-1} < P_i > P_{i+1}$. Analogously, such troughs T_j are identified, which are lower than their neighbouring local troughs: $T_{j-1} > T_j < T_{j+1}$. The peaks and troughs thus obtained are a subset of extrema identified in the previous iteration. The procedure can be iterated as many times as desired, subsequently using a smaller and smaller subset of the peaks and troughs revealed initially in the daily closing prices. Starting from the second filtering iteration, peaks and troughs are processed separately. Therefore on rare occasions it may happen that two subsequent peaks (troughs) are identified, without a trough (peak) identified between them. This situation occurs also in the PS algorithm, where it is dealt with by selecting the highest (lowest) of multiple consecutive peaks (troughs). We solve this issue in the same manner.

For the sake of clarity, we reiterate the newly proposed algorithm in a series of steps:

1 Obtain the initial set of peaks and troughs from the local maxima ($C_{t-1} < C_t > C_{t+1}$) and local minima ($C_{t-1} > C_t < C_{t+1}$) of the input price series.
2 From the previous set of peaks and troughs, select a subset of peaks which constitute local maxima ($P_{i-1} < P_i > P_{i+1}$) and troughs which constitute local minima ($T_{j-1} > T_j < T_{j+1}$).

3 In case of identifying two subsequent peaks without a trough between them, remove the lower one. Do the opposite for troughs.
4 Repeat steps 2 and 3 as many times as desired.

Fig. 1 shows an example of filtering stock market index data using 3 iterations of the algorithm. The data used is the close values of the Dow Jones Industrial Average index time series for years 2001–2002, available in Refinitiv Datastream (mnemonic DJINDUS). Generally, we find 4 iterations to be well suited for daily data. The fourth iteration reveals primary peaks and troughs – thus classifying the data series into primary bull and bear periods. All observations between the date of a primary peak and the following primary trough are classified as belonging to a bear period and vice versa. The first observation of such a period is assumed to belong to the previous regime – and the last one is assumed to belong to the current period.

2.2. Comments on practical use

The new regime classification approach uses no minimum conditions for period duration or amplitude in the identification process, contrary to Pagan and Sossounov (2003) or Lunde and Timmermann (2004). The only parameter used in the whole process is the number of filtering operations to be performed. Following Pagan and Sossounov (2003), alternation of peaks and troughs is enforced to preserve logical coherence of the resulting regime classification. This algorithm is suitable for comparing characteristics of different financial markets. In contrast to many other methods, no adjustments to the parameter value are necessary when identifying regimes in different datasets and asset classes, provided that the data uses the same periodicity. The algorithm limits data snooping, because the parameter value (set to 4 for daily data) is not determined through optimisation and is scaled to the periodicity of the data. For different periodicities, we recommend setting the parameter by starting from the benchmark of 4 assigned to daily data and adjusting it lower for less granular data (for example lowering to 3 for weekly data) or higher for more granular data (raising it to 5 for hourly data). However, all depends on the required horizon of the analysis. Typically, using 4 algorithm iterations for daily financial time series, identifies regimes with duration between 50 and 500 days. In case of detailed intraday data (for example minute data), when one might be interested in examining microtrends happening within a single trading session, the value of the parameter set at 4 should result in uncovering bull and bear periods lasting probably between 50 and 500 minutes.

The regime classification can be made more granular if desired: within a single primary bull market regime, minor bull and bear subregimes may exist. Using this fact enables splitting of primary bull regimes into normal bull market regimes (during a bullish minor subregime) and bull market correction regimes (during bearish minor subregime). This "regime-within-regime" approach not only opens up new modelling opportunities, but is also attractive from a theoretical standpoint – it underscores different investment preferences of market participants, especially the diversity of investment horizons. More than two market regimes are sometimes found within the Markov-switching framework (e.g. Maheu et al., 2012), but are not commonly identified with rule-based methods. Hanna (2018) proposed a top-down approach enabling the identification of secondary trends, which was, however, based on a subjective decision regarding the return magnitude in a bull correction or a bear rally.

In the context of trend-cycle estimation, Nguyen et al. (2020) suggested adding a third, neutral market state to the regime identification procedure. This could be a beneficial expansion of our newly proposed algorithm in the cases where the price oscillates quickly with a narrow amplitude. Such price dynamics may cause the algorithm to identify bull and bear market regimes with insignificantly small amplitudes. This can be easily prevented by introducing a minimum condition for the amplitude of a market regime. If a regime is identified, but the minimum amplitude threshold is not exceeded, such a period is classified as neutral. However, such a modification adds a new subjective parameter to the method and should be applied at the discretion of a researcher encountering a relevant problem with the input time series.

Another important practical consideration is the delay between the moment of regime switch and its detection, inherent in rule-based methods. By definition, the PS algorithm always identifies turning points retrospectively – only after the minimum time window passed or there was a large enough move in the price. Similarly, the LT method identifies a turning point in the past only after a certain move in price occurred since. Therefore a built-in, pre-specified identification delay exists in these algorithms, either with regards to time or with regards to the change in price. The newly proposed algorithm is more flexible in this aspect, with no parameter creating a pre-specified delay in regime identification – although this delay clearly exists. It appears because for each of the peaks and troughs to be determined, the next peak or trough needs to be identified. In other words, in the notation used when defining the algorithm, to find a local maximum of daily close price C_t, the value of C_{t+1} is necessary. Analogously, to select a peak P_i constituting a local maximum, P_{i+1} needs to be known. This identification delay is largely irrelevant for *ex post* regime analysis, however bears large importance for regime nowcasting. Its practical consequences can be assessed by analysing the out-of-sample performance of the classification rules, which is done in Section 4.

3. Comparison of regime classification algorithms

3.1. Data description

For illustrative purposes, the new classification algorithm (abbreviated PZ), alongside the Pagan and Sossounov (2003) algorithm (PS) and the Lunde and Timmermann (2004) algorithm (LT), is applied to 12 different time series. The data was downloaded from Refinitiv Datastream. The time span for each asset was chosen to capture the whole history available in the Datastream database up to the cut-off point in the end of 2020. The times series were chosen to represent a possibly broad set of varied assets:

Table 1
Descriptive statistics for logarithmic returns of the stock indices and the exchange rates.

	DAX	DJIA	NIFTY	CHFJPY	EURPLN	GBPUSD
Number of observations	14,610	17,772	6442	7981	5738	12,682
Mean (%)	0.023	0.028	0.040	0.001	0.002	− 0.005
Median (%)	0.009	0.045	0.016	0.011	− 0.014	0.000
Variance (× 100)	0.015	0.010	0.022	0.005	0.004	0.004
Skewness	− 0.3	− 1.2	− 0.3	1.0	0.5	− 0.2
Kurtosis	11.4	39.6	12.2	37.9	9.1	9.9

Note: The values for variance were multiplied by 100 (× 100) for presentation reasons.

Table 2
Descriptive statistics for logarithmic returns of the commodities.

	Copper	Corn	Gold	Palladium	Sugar	Oil
Number of observations	11,199	12,863	13,327	8868	9912	9818
Mean (%)	0.020	0.011	0.034	0.034	0.009	0.005
Median (%)	0.003	0.000	0.000	0.000	0.000	0.000
Variance (× 100)	0.026	0.028	0.015	0.039	0.046	0.062
Skewness	− 0.2	− 1.0	0.2	− 0.3	− 0.2	− 1.3
Kurtosis	6.7	52.0	15.8	11.1	8.4	31.9

Note: The values for variance were multiplied by 100 (× 100) for presentation reasons.

- stock market indices: 2 for developed markets (DJIA and DAX) and 1 for an emerging market (NIFTY),
- exchange rates: 1 major currency pair (GBPUSD), 1 cross currency pair (CHFJPY) and 1 pair with an emerging market currency (EURPLN),
- commodities: metals (Copper, Gold, Palladium), Oil and agriculture (Corn and Sugar).

Detailed definitions of all the assets are presented in Appendix A. The basic descriptive statistics for the time series included in the dataset are presented in Table 1 (stocks and exchange rates) and Table 2 (commodities).

PS and LT algorithms were designed with the stock market in mind. Therefore their performance for series exhibiting dynamics different from stock market indices is of particular interest. Based on the descriptive statistics, there are 3 series which may constitute a particular challenge for the dating algorithms. Firstly – Oil, for which the variance of log-returns is the highest in the sample and skewness has the largest negative value. Secondly – GBPUSD, which is the only series with a negative mean return and the lowest return variance in the sample. Thirdly – Copper, which exhibits the smallest kurtosis and return skewness closest to 0. Fig. 2 shows the diversity of the 3 aforementioned time series, with the stock index DJIA serving as a benchmark. Plots of all 12 assets grouped by class are located in Appendix B.

3.2. Degree of classification similarity between the algorithms

For the practical comparison of regime dating output, each algorithm was run with its standard parameter values. Table 3 presents the parameter sets for each of the algorithms.

Table 4 shows the number of days having identical classification between each pair of the algorithms, expressed as a percentage of the analysed sample. The first and the last turning points for each asset, as detected by each of the algorithms, occur at different times. Therefore a comparison needs to take into account only those observations that have a regime assigned by both algorithms in the analyzed pair.

In terms of number of days with identical classification, the PZ and PS algorithms are the closest, with 91.3% matching days on average. However, for assets like CHFJPY and Sugar, their degree of similarity is as low as 81.9% and 77.0%, respectively. The LT algorithm is less similar to the PZ algorithm (86.2% matching days). The largest differences occur for Sugar again (74.7% matches) and Gold (79.0%). The PS and LT algorithms differ more from each other than from the PZ classification, with Sugar again being the most controversial asset (only 70.2% matching days). Classification similarity for all 3 examined stock indices amounts to at least 90% for all algorithm pairs, reminding the fact that the PS and LT methods were originally used in the context of the stock market.

The percentage of matching classifications for different algorithms may at first seem high, questioning the need for such a detailed investigation of competing methodologies. However, classification differences reach 10–20% of the sample for multiple combinations of assets and algorithms. This may translate to significant practical consequences, as shown in Section 4, where out-of-sample performance is examined.

We summarize this part of algorithm comparison with an illustration of regime identification for the DJIA between July 10, 1966 and April 28, 1971 in Fig. 3. This period proved to be complicated for the classification methods to tackle. The PS algorithm detected a single long bear market, because of the condition establishing the shortest possible cycle length at 16 months (320 working days). The LT algorithm found only a single bear period, as the amplitude of the earlier period of falling prices did not exceed the trigger rate

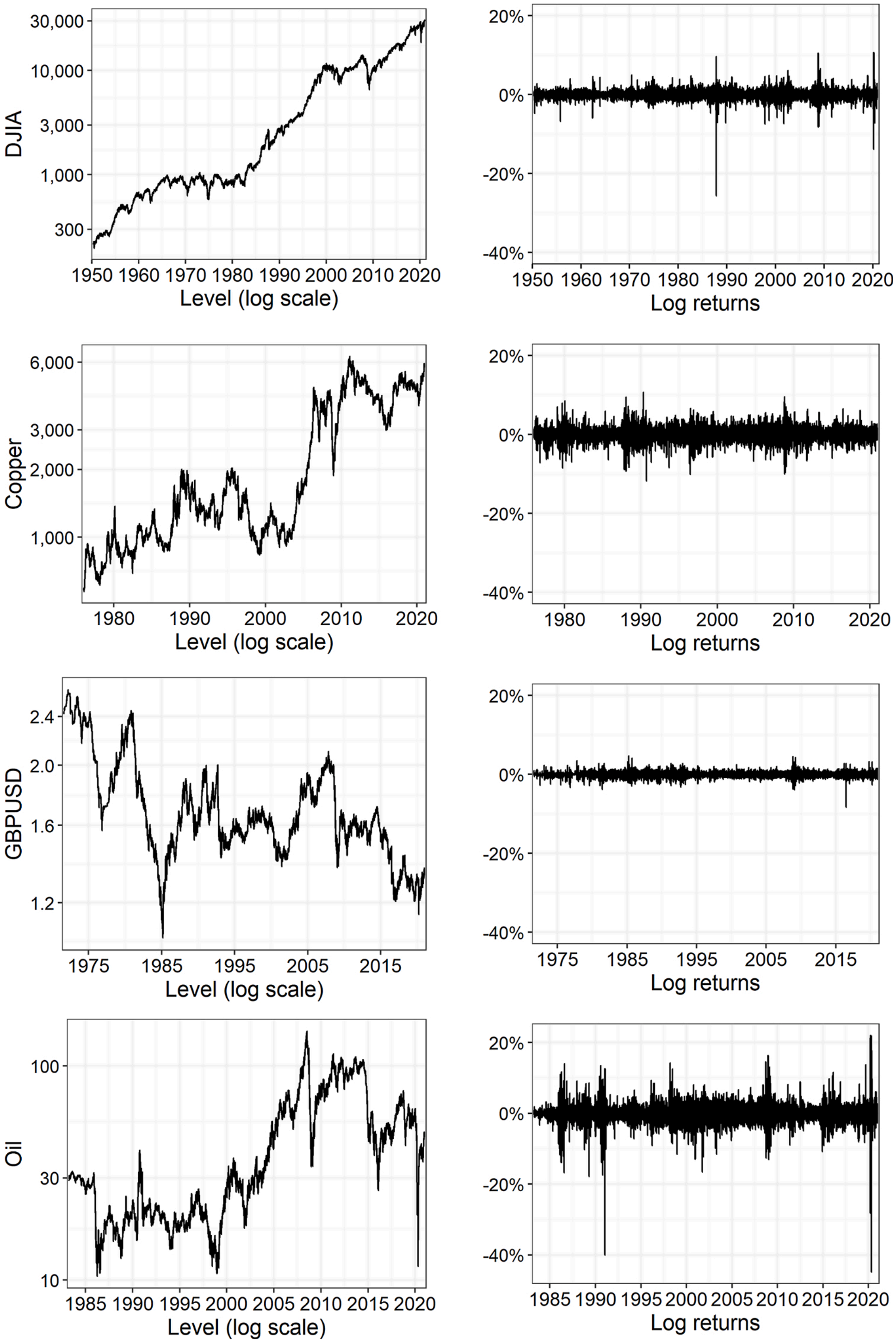

Fig. 2. Chosen time series used in the analysis.

necessary for a regime change. The PZ algorithm, operating without additional constraints stemming from multiple parameters, identified two bear periods, which qualitatively constituted two major periods of falling prices. It coped well in this ambiguous subsample, despite using only a single calibration parameter.

Table 3
Algorithm parameters.

PZ algorithm	
Number of filtering operations	4
PS algorithm	
Size of the peak and trough detection window	160 days
Censoring period at the start and at the end of the sample	120 days
Minimum cycle duration	320 days
Minimum phase duration	80 days
Rate of return overriding minimum phase duration	20%
LT algorithm	
Rate of return triggering the switch to a bull market	20%
Rate of return triggering the switch to a bear market	− 15%

PS and LT algorithms use parameter values stemming from Pagan and Sossounov (2003) as well as Lunde and Timmermann (2004), respectively. Monthly values were converted to days, assuming 20 workdays in a month.

Table 4
Algorithm comparison – days with identical classification (% of sample).

	PZ vs. PS	PZ vs. LT	PS vs. LT
CHFJPY	81.9	85.6	71.9
Copper	92.2	87.8	89.2
Corn	91.0	85.6	81.9
DAX	92.5	91.2	90.0
DJIA	96.6	92.7	92.1
EURPLN	95.8	94.6	96.7
GBPUSD	95.4	82.4	84.4
Gold	90.7	79.0	79.8
NIFTY	100.0	92.0	92.2
Oil	92.3	83.8	80.4
Palladium	90.2	85.6	85.4
Sugar	77.0	74.7	70.2
Average	91.3	86.2	84.5

3.3. Comparative analysis of the regime metrics

A more detailed comparison of the examined classification algorithms involves the use of regime metrics revealed for all the assets in the sample. Comparison of the algorithm output with regard to median regime duration and amplitude is shown in Table 5.

The PS algorithm exhibits most stable regime duration across assets, because by definition it bases its regime recognition on the time aspect. Regimes found by the PZ algorithm are usually shorter, with the exception of the NIFTY stock index. There is no clear correlation between neither PS nor PZ regime duration and the variance, skewness or kurtosis of asset returns.

The LT algorithm identifies regimes of varied duration - from extremely short (46 days for high-variance Oil series) up to very long (556 days for low-variance GPUSD series). This stems from the fact that its identification topology is based on asset returns. Therefore series exhibiting higher variance should lead the LT algorithm to detect more shorter regime periods and vice versa. This conclusion is supported by the statistically significant negative correlation (coefficient equal − 0.77) of the LT median regime duration with the asset return variance. The LT algorithm also has the most stable amplitude values, due to its return-based construction. In its case, positive skewness is connected with higher returns per regime (correlation coefficient of 0.56). Interestingly, both the PS and the PZ algorithms operate in an opposite manner to LT, with positive skewness being associated with lower regime returns. Regime amplitude for these two algorithms is driven by return variance (correlation coefficients of 0.87 and 0.78 for PS and PZ, respectively).

Based on these simple regime statistics, we see the strong similarity of regime properties identified by the PS and LT algorithms within the benchmark DJIA dataset. The PS and LT algorithm parameters were optimised for the US stock market, as they were designed and initially applied within that context. However, other results show considerable differences between these two algorithms. Regime amplitude for the PZ and PS algorithms is highly correlated (correlation coefficient of 0.90), as they share the Bry and Boschan (1971) lineage. Still, our new algorithm identifies more regime changes, which may stem from the lack of assumed minimum conditions. As a consequence, median duration and amplitude of regimes is lower for the PZ algorithm than for the PS one.

Investigating differences in various regime metrics between bull and bear periods gives more insight about the examined classification methods. The way the differences in regime duration and amplitude vary between the algorithms and the time series is presented in Table 6.

The degree of regime asymmetry varies between classification algorithms, but consistently bull regimes tend to last longer than bear regimes and have larger amplitudes. The only exceptions in the examined dataset are Gold, Sugar and the EURPLN exchange rate. Exchange rates constitute a possible natural exception here, as both mathematically and financially, the base currency and the quoted currency can always be swapped with each other. Interestingly, Gold and Sugar still exhibit bull periods with larger amplitude despite

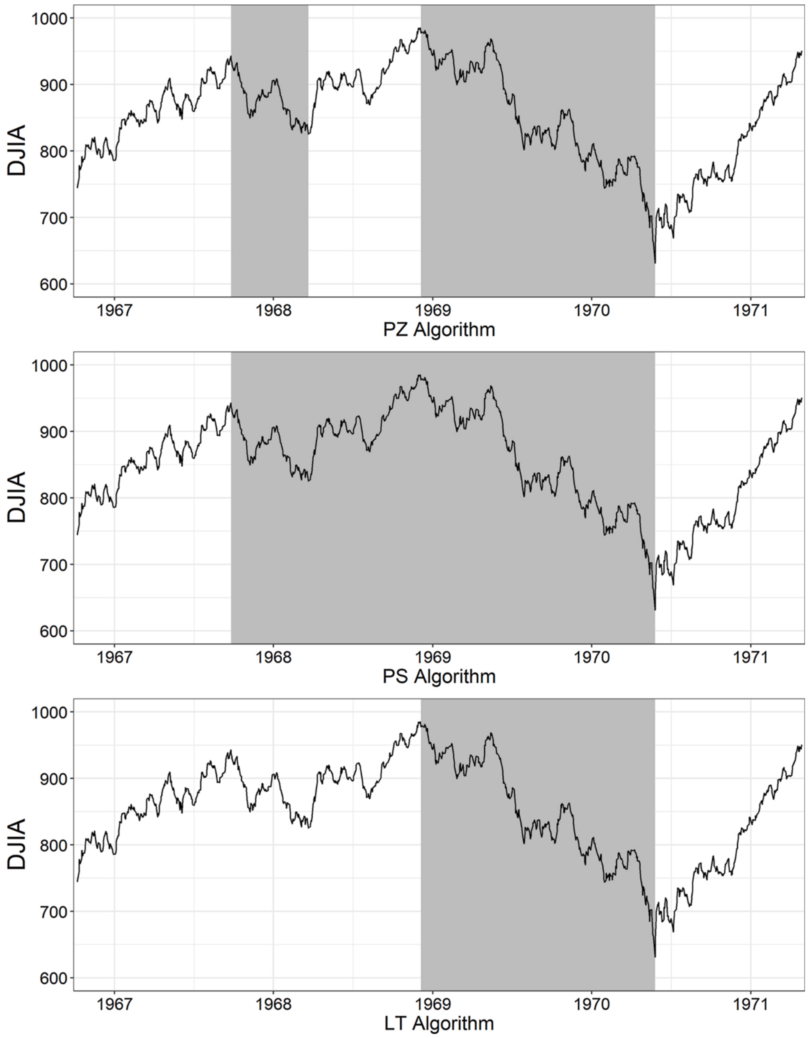

Fig. 3. Comparison of regime classification algorithms in a sample subperiod. Note: Grey background denotes detected bear periods.

Table 5
Algorithm comparison – median regime duration (days) and amplitude (%).

	Duration			Amplitude		
	PZ	PS	LT	PZ	PS	LT
CHFJPY	130	308	542	12.0	18.2	38.2
Copper	182	324	109	32.8	41.4	32.6
Corn	197	303	100	35.7	40.8	29.6
DAX	208	362	155	29.7	37.3	33.1
DJIA	195	329	285	26.9	32.4	31.6
EURPLN	145	306	417	9.0	12.7	34.6
GBPUSD	243	294	556	15.7	18.8	27.2
Gold	191	262	123	24.0	30.7	28.5
NIFTY	342	329	119	59.9	62.0	30.3
Oil	156	321	46	54.4	71.7	29.9
Palladium	169	373	67	37.0	71.8	33.2
Sugar	215	357	66	45.1	54.2	32.4

the bear periods having longer duration.

As presented in Table 7, bear period variance is greater for most assets, regardless of the classification method. At the same time, bear periods are characterized by larger absolute daily returns. Generally, the price movements in the bear regime tend to be more "abrupt" than in the bull regime – the periods are shorter, but the rate of change is greater. The main exceptions here are once again

Table 6
Algorithm comparison – difference in median regime duration (days) and amplitude (pp) between bull and bear periods.

	Duration			Amplitude		
	PZ	PS	LT	PZ	PS	LT
CHFJPY	42	60	188	5.3	4.7	12.9
Copper	21	105	16	6.6	33.3	13.9
Corn	127	153	35	20.7	34.2	9.1
DAX	86	193	233	18.1	59.2	20.6
DJIA	186	275	245	25.7	43.9	36.7
EURPLN	− 93	− 172	7	− 1.2	− 4.9	24.1
GBPUSD	58	80	605	0.1	0.9	3.7
Gold	− 4	− 48	31	8.2	16.1	24.9
NIFTY	462	346	90	87.1	79.7	31.5
Oil	147	40	26	47.6	53.3	12.2
Palladium	94	127	48	22.1	83.1	20.7
Sugar	− 6	− 53	− 5	34.6	43.8	6.6

Table shows differences in regime duration and amplitude between bull and bear periods. Positive values indicate a larger value for the bull periods, while negative values indicate a larger value for the bear periods.

Table 7
Algorithm comparison – difference in average variance and absolute daily returns (pp) between bull and bear periods.

	Variance (× 100)			Daily return		
	PZ	PS	LT	PZ	PS	LT
CHFJPY	− 0.001	− 0.001	− 0.003	− 0.04	− 0.01	− 0.05
Copper	− 0.002	− 0.002	− 0.009	− 0.03	− 0.04	− 0.06
Corn	− 0.007	− 0.006	0.000	− 0.17	− 0.06	− 0.06
DAX	− 0.011	− 0.009	− 0.009	0.00	− 0.01	− 0.07
DJIA	− 0.016	− 0.010	− 0.022	− 0.02	0.01	− 0.06
EURPLN	0.002	0.002	0.006	0.01	0.01	0.10
GBPUSD	− 0.001	− 0.001	− 0.001	0.00	− 0.02	− 0.03
Gold	0.007	0.003	− 0.003	0.03	0.04	0.01
NIFTY	− 0.023	− 0.022	− 0.012	− 0.02	− 0.01	− 0.14
Oil	− 0.070	− 0.022	− 0.024	− 0.16	− 0.05	− 0.28
Palladium	0.002	− 0.002	− 0.010	− 0.06	0.06	− 0.18
Sugar	0.001	− 0.010	− 0.012	0.03	− 0.01	0.05

Table shows differences in regime variance and daily returns between bull and bear periods. Positive values indicate a larger value for the bull periods, while negative values indicate a larger value for the bear periods. Note: The values for variance were multiplied by 100 (× 100) for presentation reasons.

Gold, Sugar and EURPLN. This observation is in line with the leverage effect – negative correlation of volatility and returns – which is a common finding, shared for example by Turner et al. (1989), Maheu and McCurdy (2000), Woodward and Marisetty (2005) and Ntantamis (2010).

4. Out-of-sample performance

4.1. A simple trading rule as an out-of-sample performance measure

Rule-based methods base their classification on various types of peak and trough detection in price series. Determining these peaks and troughs involves finding local extrema in a pre-specified time window (PS algorithm), picking extrema followed by a pre-specified return amplitude (LT algorithm) or determining a specific sequence of higher and lower extrema (PZ algorithm). These algorithms excel in in-sample applications, where they can capture alternating periods of positive and negative mean returns with high precision. However, forecasting or nowcasting current market state is also an important issue, especially in the area of broadly defined investment management – like trading or risk management. Early detection of regime change in financial markets may also be used in business cycle modelling and assist in predicting financial crises.

Gauging the regime identification precision of rule-based methods requires a goodness-of-fit measure which is based mostly on the direction of price changes, but also takes mean returns and their volatility into account. This is especially important if their performance is to be compared with the performance of other identification methods. As pointed out by Satchell and Timmermann (1995), when predicting financial time series with non-linear procedures, resulting mean squared error or mean absolute error of return forecasts may often prove to be higher than for a random walk. However, at the same time it may be possible to develop a simple

Table 8
Trading performance – significant differences in Sharpe ratios.

	Significantly outperforming at least 1 trading rule	Significantly underperforming at least 1 trading rule
CHFJPY	–	–
Copper	PZ	PS
Corn	PZ, BH, MS	LT
DAX	PZ, PS	BH, MS
DJIA	PS	MS
EURPLN	–	–
GBPUSD	PZ	PS, LT, BH, MS
Gold	PZ, PS, BH	LT, MS
NIFTY	PZ	PS, LT, BH
Oil	PZ, BH	LT
Palladium	PZ, PS, LT, BH	MS
Sugar	–	–

Statistical significance based on the Bootstrap Reality Check of White (2000) at 10% significance. Full p-value tables are available in Appendix B.

Table 9
Trading performance ranking for all 12 assets.

	Outperformance cases	Underperformance cases
1st place – PZ	8	0
2nd place – PS, BH	4	3
3rd place – LT, MS	1	5

trading strategy based on the non-linear forecasts, which yields higher returns than the buy-and-hold strategy. Therefore, trading performance of simple trading rules based on the obtained regime classification can serve as a goodness-of-fit measure, as shown by Kole and van Dijk (2017). They have found Markov switching models to perform better in out-of-sample applications than the popular regime classification algorithms of Pagan and Sossounov (2003) or Lunde and Timmermann (2004). Kole and van Dijk (2017) employed a trading strategy based on regime classification, using metrics of investor utility as forecast performance measures. They showed that strategies based on Markov switching models obtained higher investor utility in out-of-sample trading than rule-based methods. Still, risk-adjusted returns of Markov-based strategies were comparable to those of the market portfolio, thus providing no clear advantage over the buy-and-hold approach. To our knowledge, the paper of Kole and van Dijk (2017) is the only one dealing with the out-of-sample performance of rule-based market regime identification methods in such detail.

Our out-of-sample performance examination involves the comparison of trading rules based on the PS, LT and PZ algorithms against the backdrop of the buy-and-hold (BH) strategy and a trading rule based on a 2-regime Markov switching (MS) model, which serve as benchmarks. Buying the asset (long position) when a bull period is detected and selling the asset (closing the position) when a bear period is detected constitutes a natural trading rule. No short positions are allowed. Round trip transaction costs of 0.1% are taken into account in order to penalize algorithms with a large number of open positions (frequent changes of regime classification), which effectively translates into regime overfitting. Investing at the risk-free rate when the investor has no open position is not considered. The opening and closing of positions takes place one day after a new regime is identified. Identification delay of rule-based methods is fully accounted for. The 2-regime Markov switching model is re-estimated daily, simulating an investor's actual behaviour.

The main metrics used are the mean daily return $\widetilde{m}$ achieved by the investor, its volatility (standard deviation) σ and the annualised Sharpe ratio S. The number of opened positions n and the total share of days with an active open position A are also relevant, as they help in gauging the strategy's sensitivity to trading costs and changes in risk-free rate, respectively.

In order to compare the performance of trading rules, we propose the following approach as the test of the statistical significance of differences in risk-adjusted returns. Multiplying the vector of daily returns from a trading rule by a specific constant is equivalent to zero-cost financial leverage and preserves the Sharpe ratio, while changing mean return and return volatility. It is therefore possible to equalise the volatility levels for all trading rules by appropriate multiplication of return vectors. Then testing for significant differences in mean returns between trading rules is equivalent to a statistical test for difference in their Sharpe ratios. Similarly to Kole and van Dijk (2017), we decide to perform such a test through the Bootstrap Reality Check of White (2000). The stationary bootstrap proposed by Politis and Romano (1994) is employed, which is a moving blocks bootstrap based on blocks of random length from geometric distribution, with mean block length b. We use significance level of 10% as the test threshold in order to minimize the risk of missing existing differences (type II error).

The goal of this trading exercise is not to develop a viable trading strategy. Sharpe ratios rather serve as goodness-of-fit measures for the examined regime classification methods. Still, if any of these simple trading rules delivers promising trading results, it may be incorporated into a trading strategy, for example as a filter allowing only the opening of positions in agreement with the current market regime.

Table 10
Out-of-sample trading performance – DJIA.

	PZ	PS	LT	BH	MS
Sharpe ratio	0.41	0.48	0.42	0.45	0.27
Mean return (%)	0.019	0.022	0.019	0.028	0.009
Volatility	0.74	0.72	0.71	0.99	0.54
% of active days	68.1	64.2	78.1	100.0	57.9
Number of positions	30	21	22	1	743

Sharpe ratio is annualized. Mean return and volatility (standard deviation) are shown as daily values. % of active days denotes the percentage of the sample with an open long position. PS significantly outperformed MS (p-value 0.072).

Table 11
Out-of-sample trading performance – NIFTY.

	PZ	PS	LT	BH	MS
Sharpe ratio	0.59	0.45	0.39	0.43	0.44
Mean return (%)	0.041	0.032	0.026	0.040	0.027
Volatility	1.11	1.14	1.07	1.47	0.96
% of active days	60.6	69.0	66.4	100.0	72.4
Number of positions	6	7	17	1	255

Sharpe ratio is annualized. Mean return and volatility (standard deviation) are shown as daily values. % of active days denotes the percentage of the sample with an open long position. PZ significantly outperformed PS (p-value 0.082), LT (p-value 0.098) and BH (p-value 0.100).

Table 12
Out-of-sample trading performance – Oil.

	PZ	PS	LT	BH	MS
Sharpe ratio	0.12	0.02	− 0.15	0.03	− 0.04
Mean return (%)	0.013	0.002	− 0.016	0.005	− 0.003
Volatility	1.71	1.65	1.67	2.50	1.50
% of active days	57.3	55.4	50.0	100.0	53.7
Number of positions	17	12	67	1	1077

Sharpe ratio is annualized. Mean return and volatility (standard deviation) are shown as daily values. % of active days denotes the percentage of the sample with an open long position. Both PZ and BH significantly outperformed LT (p-values of 0.034 and 0.078, respectively).

Table 13
Out-of-sample trading performance – Corn.

	PZ	PS	LT	BH	MS
Sharpe ratio	0.12	0.11	− 0.03	0.10	0.12
Mean return (%)	0.010	0.009	− 0.002	0.011	0.009
Volatility	1.32	1.35	1.17	1.67	1.24
% of active days	49.7	53.2	42.0	100.0	31.7
Number of positions	25	17	47	1	966

Sharpe ratio is annualized. Mean return and volatility (standard deviation) are shown as daily values. % of active days denotes the percentage of the sample with an open long position. PZ, MS and BH significantly outperformed LT (p-values of 0.076, 0.092 and 0.098, respectively).

Table 14
Out-of-sample trading performance – GBPUSD.

	PZ	PS	LT	BH	MS
Sharpe ratio	0.19	− 0.01	− 0.08	− 0.12	− 0.38
Mean return (%)	0.005	0.000	− 0.002	− 0.005	− 0.006
Volatility	0.38	0.40	0.37	0.59	0.26
% of active days	48.9	50.5	40.2	100.0	31.0
Number of positions	19	17	8	1	1037

Sharpe ratio is annualized. Mean return and volatility (standard deviation) are shown as daily values. % of active days denotes the percentage of the sample with an open long position. PZ significantly outperformed BH, LT, PS and MS (p-values of 0.004, 0.006, 0.008 and 0.050, respectively).

4.2. Performance analysis

Firstly, Sharpe ratios from the aforementioned trading rules are checked for statistically significant differences. This general view of trading results, as presented in Table 8, leads to a conclusion that the PZ algorithm is a clear winner of the exercise. A short performance summary for each of the trading rules is shown in Table 9.

These results require a closer investigation in order to assess the performance in more detail. We concentrate on discussing the results for these assets, which reveal the most striking differences between the trading rules and – in consequence – regime classification methods:

- Corn and Oil, where the LT-based trading rule achieves negative Sharpe ratios,
- GBPUSD, where all trading rules except for PZ achieve negative Sharpe ratios,
- NIFTY, a stock index for which the PZ-based rule outperforms most of the competition,
- DJIA, serving as a benchmark.

Tables containing detailed trading performance for all the other assets are available in Appendix B. Another set of tables located in Appendix B shows the *p*-values for the test of differences between Sharpe ratios for each asset and trading rule pair.

Starting with the benchmark DJIA series, Table 10 presents the relevant trading results.

In out-of-sample trading for the DJIA sample, the PZ algorithm fares similarly to its peer group. This occurs despite the parameters of the PS and LT approaches being tuned to equity index data. The extremely high number of positions opened by the MS rule makes its Sharpe ratio susceptible to transaction costs, which makes it underperform significantly with regard to the best performer – PS (*p*-value 0.072). Large percentage of active days for the LT rule limits the possibility of improving its Sharpe ratio by investing capital at the risk-free rate when not used otherwise.

Moving on to the next stock index, NIFTY (Table 11), PZ outperforms 3 other contenders – PS, LT and BH. Still, LT's and BH's underperformance is on the verge of statistical significance. MS fares relatively well, despite the high number of open positions, when compared with the peer group. None of the trading rules strongly falls behind the others, as is the case for the Oil dataset (Table 12). Here the high volatility of the asset seems to have proven particularly troublesome for the LT algorithm, resulting in a negative Sharpe ratio. LT performs significantly worse than PZ, but also worse than a buy-and-hold strategy. The MS algorithm exhibits a very large number of open positions, making its performance vulnerable to transaction costs.

Corn is another asset with relatively high variance (and extreme kurtosis, too), where the LT algorithm does not cope well, as evident from Table 13. It is significantly outperformed by PZ (*p*-value of 0.076) as well as MS and BH, admittedly with *p*-values on the verge of significance.

Finally, GBPUSD is the only series in the sample with a negative mean over the analyzed period (Table 14). It is also characterized by the lowest volatility among all examined assets. In this low-volatility environment, the PZ algorithm significantly beats all other contestants in terms of the Sharpe ratio. What is more, ratios for all trading rules except PZ are negative. Performance of the MS method is hampered by transaction costs caused by opening over a thousand positions.

Our findings show that the PZ algorithm is capable of out-of-sample regime identification which is better or on par with other common identification methods – especially for time series with characteristics differing from equity index data. Its parsimonious construction makes it avoid the potential pitfalls of parameter optimisation and allows application to a diverse set of financial series with no parameter changes. We find this robustness of the newly proposed method to constitute its greatest advantage over the analysed alternatives. Especially the LT algorithm is prone to fail for such assets, where return variance is either much higher or lower than expected for a stock market index. The PS algorithm outperformed the buy-and-hold strategy only once. This effectively suggests that from an out-of-sample standpoint, the PS classification is often not better than simply assuming that the whole time series encompasses just a single bull period. Markov-switching employed in an out-of-sample manner always had numerous open positions. Not only the MS approach suffers heavily from the impact of transaction costs, but also its usefulness in the context of regime dating is low, as it frequently switches between quickly changing bull and bear periods, rarely identifying more stable long-term price tendencies.

The above results differ from the conclusion of Kole and van Dijk (2017), who found in their research that rule-based methods of regime identification did not fare well in out-of-sample trading. However, none of the trading rules provides risk-adjusted returns consistently and significantly dominating the market portfolio. Still, we use these simple trading rules as a way of obtaining goodness-of-fit measures. They are not supposed to be full-fledged trading strategies aimed at maximising risk-adjusted returns.

5. Conclusion

Our research adds to the important, but still relatively underdeveloped area of financial market regime identification methodology. The newly proposed rule-based method of market regime classification uses only a single discrete parameter, adjusted to the periodicity of the data. Despite its parsimonious construction, it leads to in-sample regime classification which is not drastically different from the established methods of Pagan and Sossounov (2003) as well as Lunde and Timmermann (2004) when applied to an equity index. PZ rule's single-parameter specification has the important advantage of removing subjectivity and the risk of over-optimising algorithm parameters to a specific dataset. No parameter changes are necessary when applying the rule to different financial time series of the same periodicity, which enables comparing market characteristics without distortions generated by algorithm adjustments. This robustness to input data characteristics is accentuated by the comparison of out-of-sample performance for a variety of different assets. In all cases, our new algorithm either performs the best or is not significantly different than the best method, taking into account the Sharpe ratio of a simple trading rule used as a goodness-of-fit measure.

Due to the broad palette of applications for market regime classification methods, we find our regime identification rule to be a relatively simple, yet methodologically important tool, relevant for a number of research problems. A research direction naturally related to our findings is a study of synchronicity in price cycles between different markets and its relationship with risk-on and risk-off periods, as the robustness of our rule allows its application to different datasets without additional adjustments. Examining regime persistence could reveal more about the factors influencing asset prices in the long term. Further research directions may also include investigating the properties of market regimes in detail as well as the potential significance of multiple subregimes (like bull market corrections and bear market rallies). Moreover, regime classification resulting from the PZ algorithm can be used as an exogenous threshold variable in an econometric model allowing for regime-switching. Then the properties of bull and bear regimes can be investigated within a formal econometric framework.

Another important research problem is the issue of out-of-sample regime identification delay of all the rule-based methods. Developing approaches to limit such delay would constitute an important enhancement to this strand of financial literature. Also within this context, we consider more detailed analysis of trading applications of the PZ algorithm to be a rewarding path of further research. The simple trading rule of opening long positions when a bull period is identified and closing them when a bear period starts serves well as a tool to compare algorithm performance, but does not constitute a full-fledged trading strategy. More work would be needed to make the algorithm more useful in the context of portfolio management, although already in its current state it may serve as a filter for trading signals.

Declaration of competing interest

None.

Acknowledgements

The author would like to thank the anonymous reviewers for their insightful suggestions and comments. Many thanks also to Stanisław Kirejczyk, Stanisław Zegadło and Roald Versteeg for fruitful discussions that resulted in this paper.

Appendix A

Stock indices included in the dataset:

- DAX – the daily closing values of the German DAX index for the Frankfurt Stock Exchange from December 31, 1964 to December 31, 2020 (Datastream mnemonic DAXINDX),
- DJIA – the daily closing values of the Dow Jones Industrial Average index from May 4, 1950 to December 31, 2020 (Datastream mnemonic DJINDUS),
- NIFTY – the daily closing values of the NIFTY 50 index for the Indian National Stock Exchange from April 23, 1996 to December 31, 2020 (Datastream mnemonic INNSE50).

Exchange rates included in the dataset:

- CHFJPY – the daily closing spot exchange rate of the Swiss franc (CHF) to the Japanese yen (JPY) from May 30, 1990 to December 31, 2020 (Datastream mnemonic TCHJPSR),
- EURPLN – the daily closing spot exchange rate of the euro (EUR) to the Polish zloty (PLN) from January 4, 1999 to December 31, 2020 (Datastream mnemonic TEPLNSP),
- GBPUSD – the daily closing spot exchange rate of the British pound (GBP) to the US dollar (USD) from August 20, 1971 to December 31, 2020 (Datastream mnemonic USDOLLR).

Commodities included in the dataset:

- Copper – the daily closing GBP-denominated price (pounds per metric tonne) of copper grade A at the London Metal Exchange from January 2, 1976 to December 31, 2020 (Datastream mnemonic COPHGRD),
- Corn – the daily closing USD-denominated price (cents per pound) of yellow corn grade 2 at the Chicago Board of Trade, first positional futures from December 31, 1969 to December 31, 2020 (Datastream mnemonic CORNYF1),
- Gold – the daily closing GBP-denominated price (pounds per troy ounce) of gold bullion quoted by the London Bullion Market Association from April 1, 1968 to December 31, 2020 (Datastream mnemonic GOLDBN€),
- Oil – the daily closing USD-denominated price of the New York Mercantile Exchange West Texas Intermediate crude oil barrel first month futures contract from March 30, 1983 to December 31, 2020 (Datastream mnemonic NYMWTI1),
- Palladium – the daily closing USD-denominated price (dollars per troy ounce) of palladium at the London Metal Exchange from January 5, 1987 to December 31, 2020 (Datastream mnemonic PALLADM),
- Sugar – the daily closing USD-denominated price (cents per pound) of raw sugar quoted by the International Sugar Organisation from January 4, 1983 to December 31, 2020 (Datastream mnemonic WSUGDLY).

Appendix B

Figs. B.1–B.4 and Tables B.1–B.19.

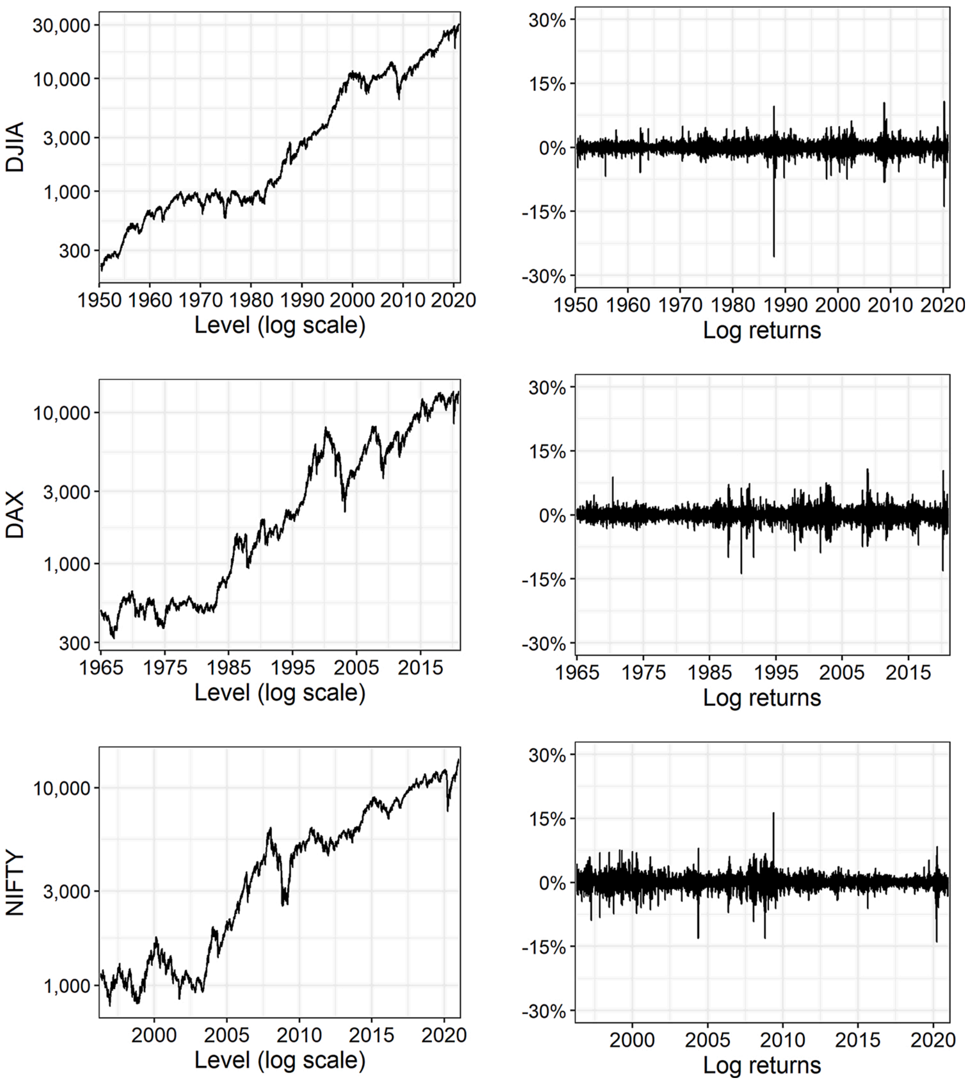

Fig. B.1. Stock indices used in the analysis.

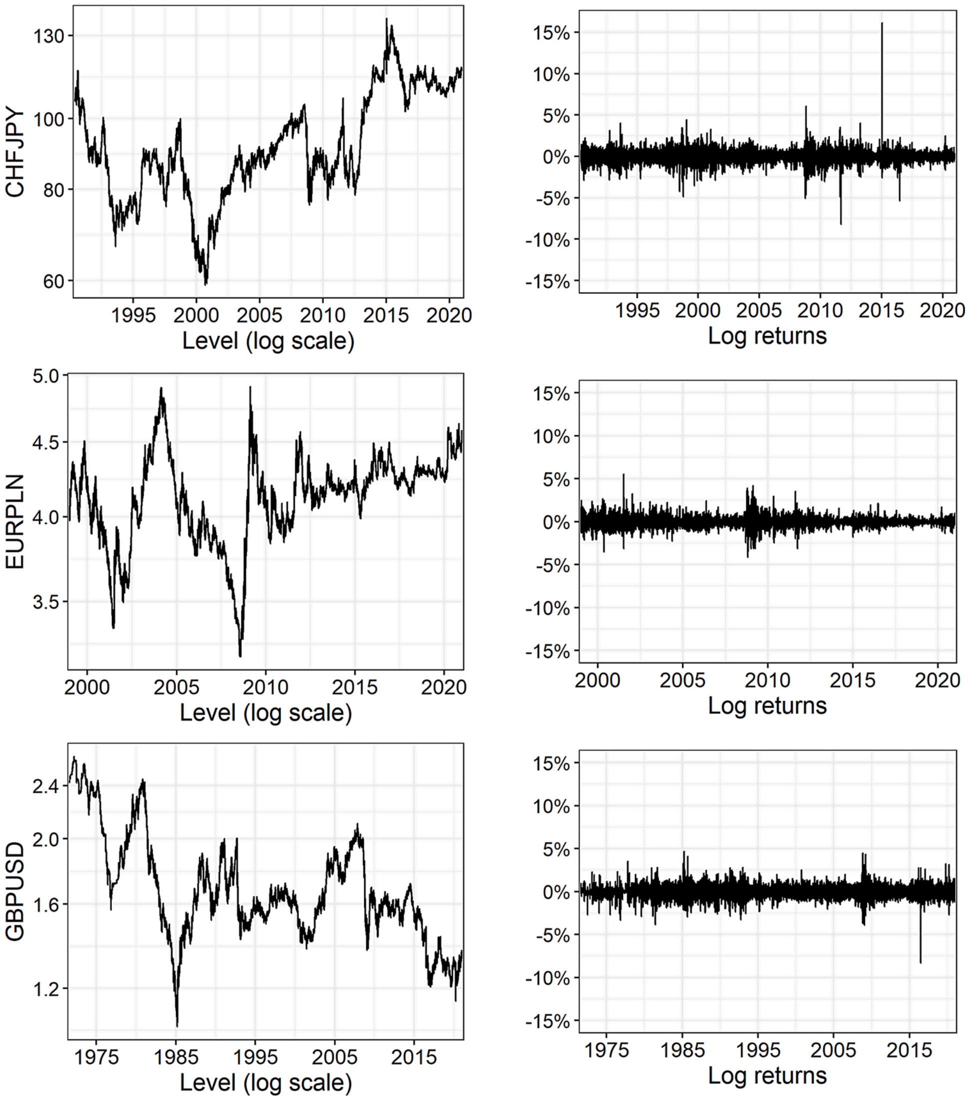

Fig. B.2. Exchange rates used in the analysis.

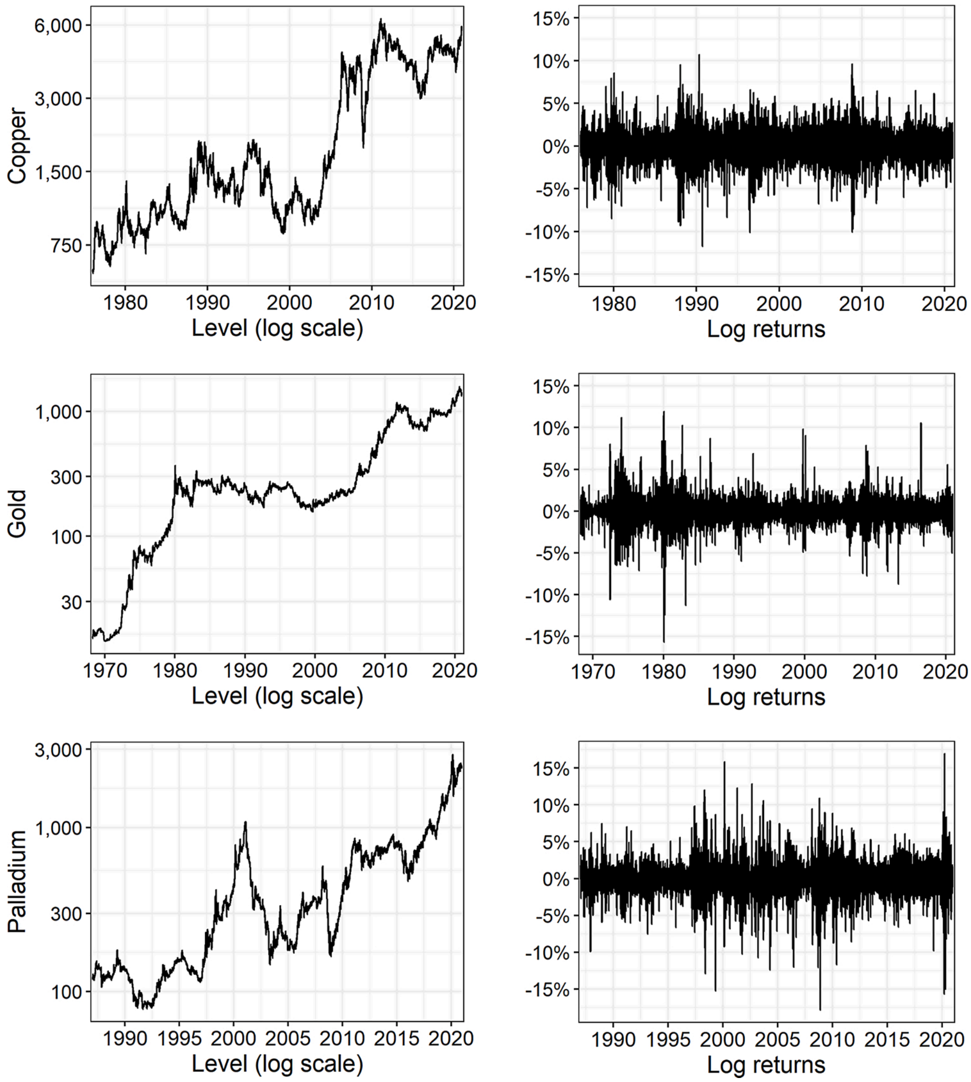

Fig. B.3. Metals used in the analysis.

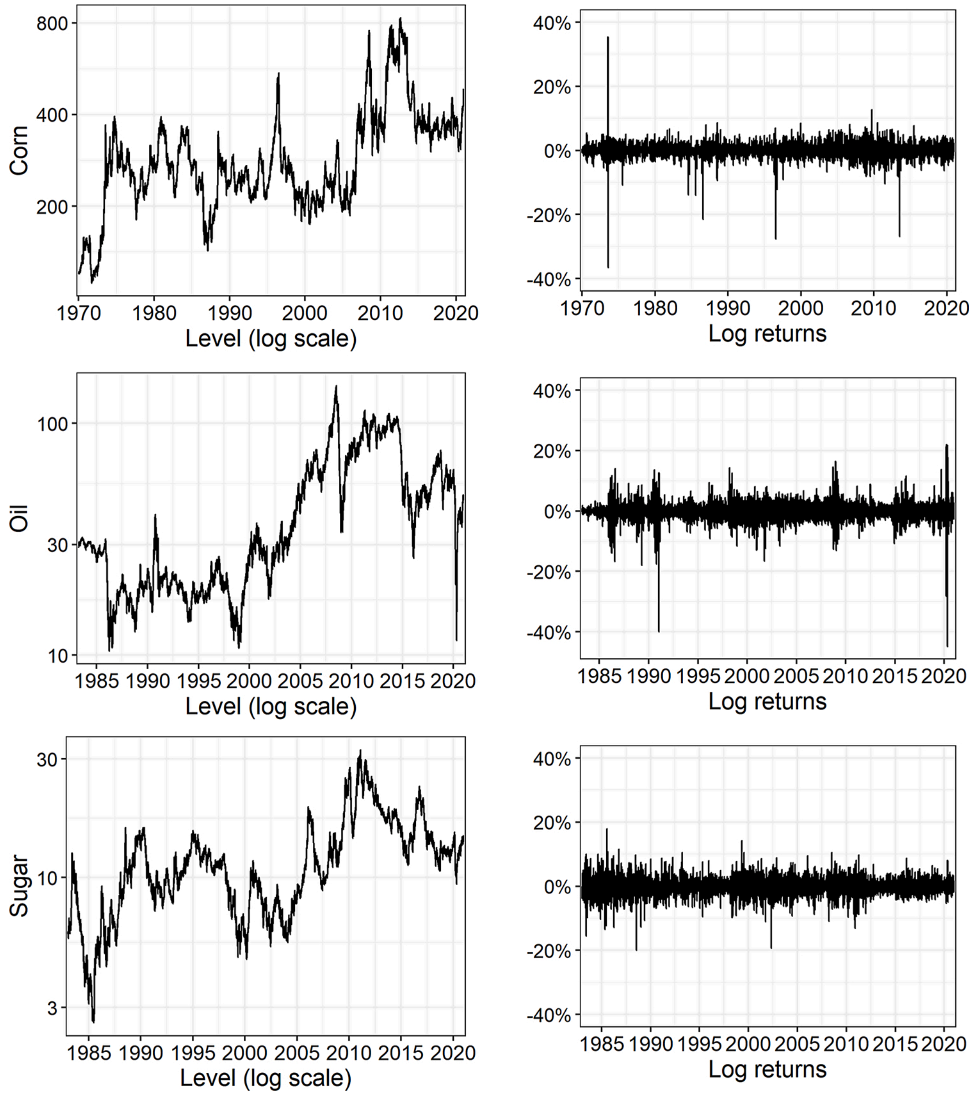

Fig. B.4. Other commodities used in the analysis.

Table B.1
Comparison of Sharpe ratios – CHFJPY *p*-values.

	BH	PS	LT	PZ	MS
BH	–	0.450	0.746	0.678	0.168
PS	0.552	–	0.752	0.740	0.282
LT	0.248	0.268	–	0.420	0.108
PZ	0.324	0.282	0.584	–	0.148
MS	0.820	0.728	0.890	0.882	–

Test *p*-values showing whether the trading rule in a specific row achieved a Sharpe ratio equal (H_0) or greater (H_1) compared to the trading rule in a specific column.

Table B.2
Comparison of Sharpe ratios – Copper *p*-values.

	BH	PS	LT	PZ	MS
BH	–	0.166	0.376	0.608	0.226
PS	0.826	–	0.688	0.904	0.482
LT	0.660	0.326	–	0.644	0.422
PZ	0.496	0.094	0.378	–	0.250
MS	0.754	0.472	0.612	0.720	–

Test *p*-values showing whether the trading rule in a specific row achieved a Sharpe ratio equal (H_0) or greater (H_1) compared to the trading rule in a specific column.

Table B.3
Comparison of Sharpe ratios – Corn *p*-values.

	BH	PS	LT	PZ	MS
BH	–	0.560	0.098	0.540	0.656
PS	0.456	–	0.104	0.524	0.654
LT	0.884	0.872	–	0.912	0.942
PZ	0.444	0.484	0.076	–	0.594
MS	0.352	0.388	0.092	0.410	–

Test *p*-values showing whether the trading rule in a specific row achieved a Sharpe ratio equal (H_0) or greater (H_1) compared to the trading rule in a specific column.

Table B.4
Comparison of Sharpe ratios – DAX *p*-values.

	BH	PS	LT	PZ	MS
BH	–	0.936	0.720	0.894	0.200
PS	0.056	–	0.136	0.318	0.038
LT	0.296	0.866	–	0.756	0.114
PZ	0.120	0.684	0.260	–	0.052
MS	0.824	0.962	0.882	0.964	–

Test *p*-values showing whether the trading rule in a specific row achieved a Sharpe ratio equal (H_0) or greater (H_1) compared to the trading rule in a specific column.

Table B.5
Comparison of Sharpe ratios – DJIA *p*-values.

	BH	PS	LT	PZ	MS
BH	–	0.678	0.370	0.296	0.120
PS	0.340	–	0.196	0.148	0.072
LT	0.616	0.782	–	0.420	0.178
PZ	0.698	0.864	0.568	–	0.212
MS	0.866	0.916	0.812	0.814	–

Test *p*-values showing whether the trading rule in a specific row achieved a Sharpe ratio equal (H_0) or greater (H_1) compared to the trading rule in a specific column.

Table B.6
Comparison of Sharpe ratios – EURPLN *p*-values.

	BH	PS	LT	PZ	MS
BH	–	0.650	0.568	0.730	0.342
PS	0.358	–	0.414	0.608	0.254
LT	0.444	0.588	–	0.676	0.284
PZ	0.182	0.354	0.330	–	0.192
MS	0.716	0.768	0.778	0.806	–

Test *p*-values showing whether the trading rule in a specific row achieved a Sharpe ratio equal (H_0) or greater (H_1) compared to the trading rule in a specific column.

Table B.7
Comparison of Sharpe ratios – GBPUSD *p*-values.

	BH	PS	LT	PZ	MS
BH	–	0.838	0.652	0.990	0.576
PS	0.126	–	0.254	0.984	0.300
LT	0.444	0.712	–	0.988	0.460
PZ	0.004	0.008	0.006	–	0.050
MS	0.446	0.720	0.512	0.962	–

Test *p*-values showing whether the trading rule in a specific row achieved a Sharpe ratio equal (H_0) or greater (H_1) compared to the trading rule in a specific column.

Table B.8
Comparison of Sharpe ratios – Gold *p*-values.

	BH	PS	LT	PZ	MS
BH	–	0.828	0.242	0.700	0.040
PS	0.176	–	0.084	0.352	0.020
LT	0.790	0.932	–	0.920	0.184
PZ	0.272	0.662	0.090	–	0.014
MS	0.970	0.988	0.812	0.990	–

Test *p*-values showing whether the trading rule in a specific row achieved a Sharpe ratio equal (H_0) or greater (H_1) compared to the trading rule in a specific column.

Table B.9
Comparison of Sharpe ratios – NIFTY *p*-values.

	BH	PS	LT	PZ	MS
BH	–	0.546	0.402	0.872	0.630
PS	0.176	–	0.330	0.932	0.548
LT	0.790	0.618	–	0.918	0.648
PZ	0.272	0.082	0.098	–	0.210
MS	0.970	0.440	0.314	0.742	–

Test *p*-values showing whether the trading rule in a specific row achieved a Sharpe ratio equal (H_0) or greater (H_1) compared to the trading rule in a specific column.

Table B.10
Comparison of Sharpe ratios – Oil *p*-values.

	BH	PS	LT	PZ	MS
BH	–	0.426	0.078	0.786	0.434
PS	0.552	–	0.132	0.814	0.446
LT	0.956	0.842	–	0.966	0.846
PZ	0.284	0.154	0.034	–	0.214
MS	0.558	0.532	0.146	0.780	–

Test *p*-values showing whether the trading rule in a specific row achieved a Sharpe ratio equal (H_0) or greater (H_1) compared to the trading rule in a specific column.

Table B.11
Comparison of Sharpe ratios – Palladium *p*-values.

	BH	PS	LT	PZ	MS
BH	–	0.734	0.488	0.704	0.002
PS	0.272	–	0.300	0.412	0.002
LT	0.522	0.692	–	0.642	0.030
PZ	0.318	0.604	0.336	–	0.004
MS	0.998	0.994	0.970	0.996	–

Test *p*-values showing whether the trading rule in a specific row achieved a Sharpe ratio equal (H_0) or greater (H_1) compared to the trading rule in a specific column.

Table B.12
Comparison of Sharpe ratios – Sugar *p*-values.

	BH	PS	LT	PZ	MS
BH	–	0.478	0.692	0.802	0.702
PS	0.514	–	0.682	0.784	0.664
LT	0.286	0.306	–	0.612	0.452
PZ	0.246	0.226	0.434	–	0.394
MS	0.326	0.376	0.484	0.638	–

Test *p*-values showing whether the trading rule in a specific row achieved a Sharpe ratio equal (H_0) or greater (H_1) compared to the trading rule in a specific column.

Table B.13
Out-of-sample trading performance – CHFJPY dataset.

	PZ	PS	LT	BH	MS
Sharpe ratio	0.02	0.00	0.12	0.09	− 0.18
Mean return (%)	0.001	0.000	0.004	0.003	− 0.005
Volatility	0.73	0.46	0.49	0.53	0.42
% of active days	100	47.4	45.5	55.3	54.6
Number of positions	1	11	5	20	339

Sharpe ratio is annualized. Mean return and volatility (standard deviation) are shown as daily values. % of active days denotes the percentage of the sample with an open long position.

Table B.14
Out-of-sample trading performance – Copper dataset.

	PZ	PS	LT	BH	MS
Sharpe ratio	0.20	0.11	0.17	0.21	0.08
Mean return (%)	0.020	0.009	0.012	0.016	0.005
Volatility	1.60	1.26	1.12	1.22	1.01
% of active days	100	55.9	45.3	58.1	61.4
Number of positions	1	15	34	22	642

Sharpe ratio is annualized. Mean return and volatility (standard deviation) are shown as daily values. % of active days denotes the percentage of the sample with an open long position.

Table B.15
Out-of-sample trading performance – DAX dataset.

	PZ	PS	LT	BH	MS
Sharpe ratio	0.30	0.46	0.35	0.42	0.18
Mean return (%)	0.023	0.024	0.019	0.023	0.010
Volatility	1.22	0.83	0.87	0.86	0.86
% of active days	100	56.8	63.5	60.7	51.2
Number of positions	1	17	27	25	728

Sharpe ratio is annualized. Mean return and volatility (standard deviation) are shown as daily values. % of active days denotes the percentage of the sample with an open long position.

Table B.16
Out-of-sample trading performance – EURPLN dataset.

	PZ	PS	LT	BH	MS
Sharpe ratio	0.05	0.12	0.09	0.16	− 0.10
Mean return (%)	0.002	0.003	0.002	0.004	− 0.003
Volatility	0.61	0.39	0.32	0.39	0.40
% of active days	100	42.3	14.6	39.4	18.3
Number of positions	1	8	2	12	287

Sharpe ratio is annualized. Mean return and volatility (standard deviation) are shown as daily values. % of active days denotes the percentage of the sample with an open long position.

Table B.17
Out-of-sample trading performance – Gold dataset.

	PZ	PS	LT	BH	MS
Sharpe ratio	0.43	0.51	0.36	0.48	0.26
Mean return (%)	0.034	0.032	0.021	0.031	0.016
Volatility	1.23	0.99	0.93	1.02	0.99
% of active days	100	49.7	50.8	54.4	33.6
Number of positions	1	15	24	26	801

Sharpe ratio is annualized. Mean return and volatility (standard deviation) are shown as daily values. % of active days denotes the percentage of the sample with an open long position.

Table B.18
Out-of-sample trading performance – Palladium dataset.

	PZ	PS	LT	BH	MS
Sharpe ratio	0.27	0.34	0.26	0.32	− 0.05
Mean return (%)	0.034	0.033	0.023	0.032	− 0.005
Volatility	1.97	1.52	1.39	1.59	1.54
% of active days	100	59.5	45.7	65.2	55.0
Number of positions	1	10	41	15	740

Sharpe ratio is annualized. Mean return and volatility (standard deviation) are shown as daily values. % of active days denotes the percentage of the sample with an open long position.

Table B.19
Out-of-sample trading performance – Sugar dataset.

	PZ	PS	LT	BH	MS
Sharpe ratio	0.07	0.05	0.13	0.16	0.07
Mean return (%)	0.009	0.004	0.012	0.015	0.007
Volatility	2.14	1.34	1.43	1.44	1.45
% of active days	100	49.0	39.5	47.6	52.8
Number of positions	1	10	49	13	1370

Sharpe ratio is annualized. Mean return and volatility (standard deviation) are shown as daily values. % of active days denotes the percentage of the sample with an open long position.

References

Bohl, M.T., Siklos, P.L., Werner, T., 2007. Do central banks react to the stock market? The case of the Bundesbank. J. Bank. Finance 31, 719–733. https://doi.org/10.1016/j.jbankfin.2006.04.003.

Bry, G., Boschan, C., 1971. Cyclical Analysis of Time Series: Selected Procedures and Computer Programs. National Bureau of Economic Research, New York.

Cashin, P., McDermott, C.J., Scott, A., 2002. Booms and slumps in world commodity prices. J. Dev. Econ. 69, 277–296. https://doi.org/10.1016/S0304-3878(02)00062-7.

Cunado, J., Gil-Alana, L.A., de Gracia, F.P., 2010. Mean reversion in stock market prices: new evidence based on bull and bear markets. Res. Int. Bus. Finance 24, 113–122. https://doi.org/10.1016/j.ribaf.2009.10.001.

Diebold, F.X., Inoue, A., 2001. Long memory and regime switching. J. Econom. 105, 131–159. https://doi.org/10.1016/S0304-4076(01)00073-2.

Estrella, A., Mishkin, F.S., 1998. Predicting U.S. recessions: financial variables as leading indicators. Rev. Econ. Stat. 80, 45–61 doi:10.1162/003465398557320.

Fabozzi, F.J., Francis, J.C., 1977. Stability tests for alphas and betas over bull and bear market conditions. J. Finance 32, 1093–1099 doi:10.1111/j.1540-6261.1977.tb03312.x.

Gordon, S., St-Amour, P., 2000. A preference regime model of bull and bear markets. Am. Econ. Rev. 90, 1019–1033 doi:10.1257/aer.90.4.1019.

Guidolin, M., Timmermann, A., 2005. Economic implications of bull and bear regimes in UK stock and bond returns. Econ. J. 115, 111–143 doi:10.1111/j.1468-0297.2004.00962.x.

Hamilton, J.D., 2011. Calling recessions in real time. Int. J. Forecast. 27, 1006–1026. https://doi.org/10.1016/j.ijforecast.2010.09.001.

Hanna, A.J., 2018. A top-down approach to identifying bull and bear market states. Int. Rev. Financ. Anal. 55, 93–110. https://doi.org/10.1016/j.irfa.2017.11.001.

Harding, D., Pagan, A., 2002. A comparison of two business cycle dating methods. J. Econ. Dyn. Control 27, 1681–1690. https://doi.org/10.1016/S0165-1889(02)00076-3.

Kole, E., van Dijk, D., 2017. How to identify and forecast bull and bear markets. J. Appl. Econom. 32, 120–139 doi:10.1002/jae.2511.

Lunde, A., Timmermann, A., 2004. Duration dependence in stock prices: an analysis of bull and bear markets. J. Bus. Econ. Stat. 22, 253–273 doi:10.1198/073500104000000136.

Maheu, J.M., McCurdy, T.H., 2000. Identifying bull and bear markets in stock returns. J. Bus. Econ. Stat. 18, 100–112 doi:10.1080/07350015.2000.10524851.

Maheu, J.M., McCurdy, T.H., Song, Y., 2012. Components of bull and bear markets: bull corrections and bear rallies. J. Bus. Econ. Stat. 30, 391–403 doi:10.1080/07350015.2012.680412.

Nguyen, L., Novák, V., Mirshahi, S., 2020. Trend-cycle estimation using fuzzy transform and its application for identifying bull and bear phases in markets. Intell. Syst. Account., Finance Manag. 27, 111–124 doi:10.1002/isaf.1473.

Ntantamis, C., 2010. A Duration Hidden Markov Model for the Identification of Regimes in Stock Market Returns. CREATES Research Papers Department of Economics and Business Economics. Aarhus University.

Nyberg, H., 2010. Dynamic probit models and financial variables in recession forecasting. J. Forecast. 29, 215–230 doi:10.1002/for.1161.

Pagan, A.R., Schwert, G.W., 1990. Alternative models for conditional stock volatility. J. Econom. 43, 267–290. https://doi.org/10.1016/0304-4076(90)90101-X.

Pagan, A.R., Sossounov, K.A., 2003. A simple framework for analysing bull and bear markets. J. Appl. Econom. 18, 23–46 doi:10.1002/jae.664.

Politis, D.N., Romano, J.P., 1994. The stationary bootstrap. J. Am. Stat. Assoc. 89, 1303–1313 doi:10.2307/2290993.
Rigobon, R., Sack, B., 2003. Measuring the reaction of monetary policy to the stock market. Q. J. Econ. 118, 639–669 doi:10.1162/003355303321675473.
Ryden, T., Terasvirta, T., Asbrink, S., 1998. Stylized facts of daily return series and the hidden Markov model. J. Appl. Econom. 13, 217–244 doi:10.1002/(SICI)1099-1255(199805/06)13:3<217::AID-JAE476>3.0.CO;2-V.
Satchell, S., Timmermann, A., 1995. An assessment of the economic value of non-linear foreign exchange rate forecasts. J. Forecast. 14, 477–497 doi:10.1002/for.3980140602.
Turner, C.M., Startz, R., Nelson, C.R., 1989. A Markov model of heteroskedasticity, risk, and learning in the stock market. J. Financ. Econ. 25, 3–22. https://doi.org/10.1016/0304-405X(89)90094-9.
Veronesi, P., 1999. Stock market overreaction to bad news in good times: a rational expectations equilibrium model. Rev. Financ. Stud. 12, 975–1007 doi:10.1093/rfs/12.5.975.
White, H., 2000. A reality check for data snooping. Econometrica 68, 1097–1126 doi:10.1111/1468-0262.00152.
Woodward, G., Marisetty, V.B., 2005. Introducing non-linear dynamics to the two-regime market model: evidence. Q. Rev. Econ. Finance 45, 559–581. https://doi.org/10.1016/j.qref.2005.04.001.