Testing the Weak-Form Efficient Market Hypothesis in Selected Central and Eastern European Markets

Timotej Bedenik, Žan Jan Oplotnik
University of Maribor, Faculty of Economics and Business, Maribor, Slovenia
timotej.bedenik@um.si, zan.oplotnik@um.si

This paper presents the results of testing the weak form of the Efficient Market Hypothesis (EMH) for the stock indices of Poland, Hungary, the Czech Republic, Slovakia, Slovenia, and Croatia using daily data for 2015–2024. The sample is split into two subperiods- 2015–2019 and 2020–2024 - with the cut at 1 January 2020 to capture the impact of the COVID-19 crisis on observed efficiency. The tests employed are the Lo–MacKinlay variance-ratio test, unit-root tests (ADF, Phillips–Perron), and serial-dependence tests (Ljung–Box Q and the runs test). For the full period, results are heterogeneous: the WIG (Poland) and BUX (Hungary) indices, on average, exhibit properties consistent with weak-form EMH, whereas the SAX (Slovakia) generally does not; the remaining indices display partial violations. After 2020, evidence in favour of weak-form efficiency declines noticeably for most indices, with outcomes varying in part by the test used.

DOI
https://doi.org/10.18690/um.epf.7.2026.23

ISBN
978-961-299-166-1

Keywords:
efficient markets, weak-form efficient market hypothesis, capital markets, statistical tests, Central and Eastern Europe, transition economies

University of Maribor Press

1 Introduction

The Efficient Market Hypothesis constitutes one of the fundamental concepts of modern financial theory, positing that the prices of financial instruments at any moment fully reflect all available information (Fama, 1970). In accordance with this definition, an efficient capital market enables the optimal allocation of resources to companies and thereby contributes to the effective functioning of the financial system and to a long-term economic growth. Fama (1970) distinguishes three forms of efficiency: weak, semi-strong, and strong. The weak form assumes that past prices do not contain information that would allow investors to systematically achieve above-average returns, whereas the semi-strong form encompasses all publicly available information, and the strong form also includes private or inside information.

The discussion of capital-market efficiency is of particular importance in transition economies which, after the collapse of central planning regimes (e.g., the Soviet Union, SFR Yugoslavia, Czechoslovakia), faced the challenge of constructing a modern financial system. The transition process included the establishment and strengthening of national capital markets as central mechanisms for the efficient allocation of financial resources. The initial period of transition was marked by low market capitalization, a limited number of financial instruments, a low degree of transparency, and a lack of investor confidence. Only after institutional reforms, accession to the European Union, and the opening of markets to foreign investors contributed to gradual stabilization, higher liquidity, and greater alignment with market standards (Claessens, Djankov, and Klingebiel, 2000).

Nevertheless, empirical studies (Gilmore and McManus, 2001; Hassan et al., 2006; Smith, 2012) indicate that, in their initial phases, the capital markets of transition economies mostly did not follow the weak form of efficiency, although over time improvements toward greater alignment with this concept were observed. This opens the research question of the further progress of selected capital markets in their capacity for efficient price formation.

The purpose of the study is to provide a comparative analysis of six selected capital markets in the Central and Eastern European (CEE) region over the recent period 2015–2024, which has not been previously examined in a unified framework for this group of transition economies. It empirically evaluates whether these markets follow the weak form of the Efficient Market Hypothesis using multiple statistical tests and introduces a novel ranking of the six capital markets based on the consistency with which they satisfy the conditions of weak-form efficiency. Furthermore, the analysis incorporates a post-COVID-19 perspective by dividing the sample into two sub-periods (2015–2019 and 2020–2024), allowing for a structured comparison of the observed level of market efficiency before and after the pandemic, while also providing additional insight into market behaviour under systemic shocks.

The paper is structured as follows: Section 2 presents the literature review. Section 3 describes the data and outlines the methodological framework, including the selected statistical tests. Section 4 presents the empirical results, while Section 5 provides a summary of the key findings and concluding remarks.

2 Literature review

Research on these markets began to take shape in the early to mid-1990s, when post-transition markets were in an early stage of development. Among the first contributions for Poland, Gordon and Rittenberg (1995) analysed 1991–1994 and–based on the predictability of returns following trading halts during periods of elevated volatility–indicated violations of weak-form efficiency. Although they did not employ the full set of classical statistical tests, the signal was clear: in the early transition period, return predictability was present. For Hungary, Macskási and Molnár (1996) detected statistically significant serial correlation using the Ljung–Box test over 1991–1996, inconsistent with a random walk. In Czechia, an early study often cited is Vošvrda et al. (1998), which, using the variance-ratio test and autocorrelation tests for 1995–1997, found departures from random price movements; here too, early inefficiency aligns with the state of the market in the initial stages of institutional development.

The turn of the millennium brought broader, comparatively oriented analyses and greater methodological variety. Gilmore and McManus (2001) employed a combination of the Ljung–Box test, the variance-ratio test, and unit-root tests on daily data for Czechia, Poland, and Hungary. The results were mixed: in some instances, the tests failed to reject randomness, while elsewhere the diagnostics indicated predictable patterns–consistent with the view that efficiency can be time-varying in transition markets, as year-by-year aggregations or rolling sub windows often reveal. Similarly, Smith and Ryoo (2003), using a multiple variance-ratio framework for Poland (1991–1998) and Hungary, found statistically significant departures from random price movements, consistent with observed inefficiency in the early stage of market development.

In the early 2000s, both methodological practices and observation frequencies broadened. Worthington and Higgs (2004) situated Czechia, Poland, and Hungary within a wider comparative framework of developed and emerging markets. For Czechia, the runs and unit-root tests indicated departures from a random walk, leading to an assessment of inefficiency over the period examined. Subsequent research, however, points to improvement in Czechia: focusing on volatility modelling and martingale properties, Hančlová and Rublíková (2006) find that the PX gradually approaches a random process. This shift from "clear" inefficiencies toward at least partial random-walk characteristics is a recurring theme across the region.

For Croatia, Heininen and Puttonen (2008) provide a different yet complementary perspective: instead of standard efficiency tests, they examine calendar anomalies (day-of-the-week, month-of-the-year, turn-of-the-month, and Halloween effects). In the period 1997–2000, an above-average turn-of-the-month return is observed, and the Halloween effect is statistically significant; taken together, this points to partial inefficiency, consistent with findings from related markets in early development phases. The Slovenian market was first studied at this time by Deželan (2000), who, using autocorrelation tests, detected serial dependence and thus rejected the weak form of the efficiency hypothesis for the Ljubljana Stock Exchange.

The years following the global financial crisis reduced the perception of these markets as weak-form efficient in many cases. For Croatia, Šonje, Alajbeg, and Bubas (2011), using a rolling autoregression for 1997–2010, show that CROBEX was efficient on approximately 80% of trading days (at the 5% level) prior to the crisis, compared with 98.9% for the S&P 500 up to 2007; after including post-2007 data, the share of efficient days falls to around 55%. This underscores the sensitivity of smaller, less liquid markets to systemic shocks. In a similar time window, Čelebič (2009), using daily data–autocorrelation, the runs test, and the variance-ratio test–likewise rejects the weak form of efficiency for Western Balkan markets (including

Croatian and Slovenian). Omay and Karadagli (2010), by contrast, find weak-form efficiency using unit-root tests; however, their use of monthly data and a longer observation period (2002–2010) partly smooths the impact of the financial crisis. Meanwhile, Dritsaki (2011) reports mixed results for 1997–2010–both in terms of which markets exhibit efficiency and which tests indicate it.

Smith (2012) tested 15 emerging European markets over 2000–2009 using rolling variants of the variance-ratio test (with focal windows 2002–2005 and 2006–2009) and computed the share of windows in which the random-walk hypothesis was not rejected. The rankings place Hungary 3rd, Poland 4th, and Czechia 6th, followed by Croatia (9th) and Slovakia (11th), while Slovenia ranks 14th out of 18 countries. A notable temporal cut is that, after 2006, efficiency deteriorates on most markets.

After 2010, the literature increasingly turns to more advanced – often nonlinear – unit-root tests intended to better capture potential structural breaks and asymmetries. Erdas (2019) examines eleven Central and Eastern European markets for 2010–2018 and generally confirms the presence of a unit root in return series – i.e., results consistent with weak-form market efficiency – for Czechia, Croatia, Hungary, and Slovakia, thereby broadly supporting the view that these markets are adaptive.

3 Data and methods

3.1 Data

The dataset comprises six equity indices: WIG (Poland), SBITOP (Slovenia), CROBEX (Croatia), PX (Czechia), SAX (Slovakia), and BUX (Hungary). The sample covers daily observations from the first trading day of 2015 to the last trading day of 2024, totalling approximately 2,486 trading days; the actual number of observations may differ slightly across indices due to local holidays and trading-calendar specifics. For each index, we use closing prices and compute logarithmic daily returns as $rt = \ln(Pt) - \ln(Pt-1)$. The data were obtained from the Investing.com portal, and rows with missing or invalid values were removed without imputation.

To identify potential structural differences, we split the full sample into two subperiods: 2015–2019 and 2020–2024. All calculations and statistical tests (based on log-returns) were conducted in the Python programming language.

3.2 Methods

The empirical testing of the weak form of the Efficient Market Hypothesis rests on two complementary assumptions: (i) there is no serial dependence in returns, implying randomness of changes, and (ii) the logarithmic level of prices behaves as a unit-root process (I (1)), while its first differences – i.e., returns – are stationary. To test for serial dependence, tests of autocorrelation are used; in the present study the Ljung–Box test and the runs test will be included. The presence of a unit root will be examined using two complementary tests, namely the augmented Dickey–Fuller (ADF) unit-root test and the Phillips–Perron unit-root test. An intermediate bridge between the two groups of tests is provided by the Lo–MacKinlay test of the variance coefficient [variance-ratio test], which, on the basis of the properties of a random walk, derives a testable restriction concerning the linear growth of variance with the time lag. The joint use of these methods enables an expanded basis for inference, as they focus on different aspects of the same hypothesis: the correlation structure of returns, the randomness of the sequence of positive and negative returns, the temporal dynamics of variance, and the stationarity of the level of the price movement.

3.2.1. Ljung-Box Test

Ljung–Box is a test of joint autocorrelation designed to assess whether all autocorrelations up to a chosen lag h are simultaneously equal to zero. In the context of the weak form of the Efficient Market Hypothesis, its role is to examine whether the return series exhibits statistically significant serial (autocorrelative) structure that would permit forecasting future changes solely on the basis of past price movements. The method is standard in empirical analyses of market efficiency and, owing to its built-in small-sample correction, remains suitable for short and medium-sized samples–relevant for financial time series, where observations are often limited by data availability and regime shifts. The procedure is derived from the sum of squared sample autocorrelations up to h, with the finite-sample correction factor $n(n+2)/(n-k)$.

The Q statistic is given by:

$$Q = n(n+2)\sum_{k=1}^{h}\frac{\hat{\rho}k^2}{n-k}$$

The correction factor can equivalently be written as:

$$Q = n\sum_{k=1}^{h} w_k\ \hat{\rho}k^2, \quad w_k = \frac{n+2}{n-k}$$

The decision rule employs the χ^2 distribution with h degrees of freedom:

$$\text{Reject } H_0 \text{ if } Q > \chi^2_{h,1-\alpha}, p\text{-value} = 1 - F_{\chi^2_h}(Q)$$

In practical implementation, a transparent choice of h is crucial: for small samples, a few lags suffice, whereas for longer series – especially when the aim is to detect weak and diffuse correlation patterns – the literature recommends a wider range (e.g., up to 20 or more), while recognizing the trade-off between test power and the increased risk of false rejections. Results are reported as pairs (Q, p) for each selected h (or for a set of h values), together with an explicit decision at a pre-specified significance level α and a statement on robustness across similar h-specifications and subsamples. The test's role within the suite of complementary procedures is diagnostic: it does not measure the magnitude of departures from a random walk in terms of the dynamic structure of variance or the stationarity of levels, but rather the aggregate presence of linear dependencies of the first (and higher) orders; consequently, its findings should be interpreted in combination with other tests.

3.2.2 Runs Test

Runs test is a nonparametric test of the randomness of the direction of changes in a time series. A run is defined as the maximal sequence of consecutive observations with the same sign (e.g.$++--+$ contains the runs $++$, $--$, and $+$). In the methodological exposition, we first determine the numbers of positive and negative changes, N^+ and N^- , and the total number of observations $N = N^+ + N^-$. The expected number of runs under the null hypothesis of a random arrangement of signs is given by

$$m = 1 + \frac{2\,N^{+}N^{-}}{N}$$

The result of the runs test is expressed by the value of the Z-statistic, which is computed according to the following equation:

$$Z = \frac{R - m}{\sigma_m}$$

where the standard deviation of the expected number of runs is computed using the equation:

$$\sigma_m = \sqrt{\frac{2\,N^{+}N^{-}\,(2\,N^{+}N^{-} - N)}{N^2(N-1)}}$$

Results are interpreted based on the absolute value of the Z-statistic: values exceeding 1.96 imply rejection of the null of randomness at the 5% significance level, while values exceeding 2.58 imply rejection at the 1% level. Rejection of the null indicates the presence of non-random patterns in return changes–i.e., serial dependence–which constitutes a departure from the assumptions of weak-form market efficiency.

3.2.3 Variance-Ratio Test

The Lo–MacKinlay variance-ratio test is a statistical procedure for testing the weak form of market efficiency. It rests on the premise that, under a random walk, the variance of cumulative (k-period) returns is linearly proportional to the length of the observation interval.

The basic variance-ratio can be written as:

$$\mathrm{VR}(q) = \frac{Var\big(R_t(q)\big)}{q\,Var\big(R_t(1)\big)}$$

Under a random walk, the variance ratio, $VR(q)$, should be close to 1. Deviations from this value indicate two cases: if $VR(q)<1$, this points to mean reversion (negative autocorrelation), whereas $VR(q)>1$ indicates momentum (positive autocorrelation).

The computation of the variance-ratio test statistic depends on assumptions about the distribution of returns. Lo and MacKinlay (1988) define two versions of the statistic: $M1(q)$, which assumes independent and identically distributed (i.i.d.) returns (the RW1 framework) with the corresponding asymptotic variance $\phi(q)$; and $M2(q)$, which is robust to heteroskedasticity (the RW3 framework).

The $M1$ statistic can be written as:

$$M_1(q) = \frac{\mathrm{VR}(q) - 1}{\sqrt{\phi(q)}}$$

where the asymptotic variance can be computed as:

$$\phi(q) = \frac{2(2q-1)(q-1)}{3qT}$$

In the case where heteroskedasticity of return distributions is assumed–which is characteristic of financial time series–we use the $M2(q)$ statistic, defined as:

$$M_2(q) = \frac{\mathrm{VR}(q)-1}{\sqrt{\phi^*(q)}}, \text{ where } \phi^*(q) = \sum_{j=1}^{q-1} \left[\frac{2\,(q-j)}{q}\right]^2 \delta(j)$$

$$\delta(j) = \sum_{t=j+1}^{T} (x_t - \bar{x})^2 \left(x_{t-j} - \bar{x}\right)^2 \left[\sum_{t=1}^{T} (x_t - \bar{x})^2\right]^2$$

In the equations, xt denotes the one-period logarithmic return, $\bar{x}$ the arithmetic mean of the time series, q the selected lag, and T the number of observations in the sample.

The statistics $M1(q)$ and $M2(q)$ are Z-statistics that are compared with the critical values of the standard normal distribution. If the absolute value of the Z-statistic exceeds the critical threshold (e.g., ±1.96 for the 5% significance level), we reject the null hypothesis of a random walk. This means that the return series exhibits a statistically significant autocorrelation structure, which is contrary to the weak form of market efficiency.

Since financial time series often display heteroskedasticity, the empirical part of the study will employ the more robust $M2(q)$ statistic.

3.2.4 Unit Root Tests

Unit root tests are statistical tools for assessing the stationarity of time series. In the context of the weak form of the EMH, it is crucial to determine whether the levels of log prices follow a random walk, implying non-stationarity and the presence of a unit root. If the series is stationary, this points to mean-reverting mechanisms, which allow a degree of predictability and run counter to the weak form of the EMH.

Two complementary tests are employed: the Augmented Dickey–Fuller (ADF) test and the Phillips–Perron (PP) test. Both test the null hypothesis $H0$: the time series contains a unit root (is non-stationary), against the alternative $H1$: the time series is stationary.

The Augmented Dickey–Fuller (ADF) test extends the basic DF test by including lags of the first differences of the dependent variable to explicitly capture autocorrelation in the errors. The test is based on estimating the following regression equation:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{i=1}^{p} \delta_i \, \Delta y_{t-i} + \varepsilon_t$$

Where Δyt is the first difference; α is a constant (drift); βt represents a deterministic linear trend; γy_{t-1} is the key term for the unit-root test $\sum_{i=1}^{p} \delta_i \, \Delta y_{t-i}$; includes lagged differences to capture short-run dynamics and remove serial correlation; and εt is a random disturbance (white noise).

The test statistic is compared with the critical values of the Dickey–Fuller distribution. If the statistic falls below the negative critical value (e.g., −2.86 at the 5% significance level), H_0 is rejected and the series is concluded to be stationary.

The Phillips–Perron (PP) test takes a different approach. Instead of including lags of first differences of the dependent variable, it applies a nonparametric correction to adjust the standard errors of the test statistic so as to account for heteroskedasticity and higher-order autocorrelation. It is based on the basic DF regression without additional dynamic terms:

$$\Delta y_t = \alpha + \beta t + \phi\, y_{t-1} + \varepsilon_t$$

The principal advantage of the PP test is its robustness to a wide range of heteroskedasticity and autocorrelation patterns without requiring the specification of the lag order p. The test statistic is computed as a modified version of the DF statistic with a correction factor for the long-run variance $\hat{\sigma}^2$.

It can be written as follows:

$$Z_t = \frac{\hat{\gamma}}{\mathrm{SE}(\hat{\gamma})} - \frac{1}{2}\left(\frac{\widehat{\lambda^2} - \widehat{\sigma_u^2}}{\widehat{\sigma^2}}\right) \cdot \frac{T \cdot \mathrm{SE}(\hat{\gamma})}{\widehat{\sigma_u}}$$

where $\hat{\lambda}^2$ is the estimate of the long-run variance of the residuals, while $\hat{\sigma}_u^2$ is the estimate of the short-run (contemporaneous) variance of the residuals. The decision rules are identical to those for the ADF test: reject H_0(unit root) if the PP test statistic Zt is more negative than the corresponding Dickey–Fuller critical value for the chosen deterministic specification (no constant / with constant / with trend) at the selected significance level.

4 Results and discussion

4.1 Ljung-Box Test

The Ljung–Box autocorrelation test results are reported in Appendix A (full period: Table A1; subperiods: Table A2). The full-period analysis covers ten years between 2015 and 2024 (n=2,486 observations). Testing the null hypothesis H_0:

$\rho 1 = \rho 2 = \cdots = \rho m = 0$ (no serial dependence up to the chosen lag) shows that H_0 is rejected for all indices except Poland's WIG, which alone does not exhibit departures from a random walk. In terms of rejection strength, Croatia's CROBEX stands out: its Q-statistics substantially exceed those of the other indices at all inspected lags, indicating a stronger joint departure from H_0 either due to a few larger autocorrelation coefficients or to numerous smaller ones that cumulate.

We split the sample into two subperiods with a breakpoint in 2020 to capture potential structural changes following the COVID-19 shock. Before 2020, in addition to WIG, we do not detect statistically significant serial dependence for SBITOP or BUX (i.e., H_0 is not rejected), whereas the PX results do not materially differ from the full-period findings and are not consistent across lags. After 2020, however, Hungary's BUX and Slovenia's SBITOP begin to exhibit statistically significant serial dependence, suggesting a deterioration in weak-form efficiency on these markets in the post-pandemic period. For Croatia's CROBEX and Slovakia's SAX, serial dependence is present across all periods considered.

4.2 Runs Test

The runs-test results are provided in Appendix B–Table B1 for the full period and Table B2 for the subperiods. Over the full sample there are approximately 2,486 observations; for Slovakia's SAX, the number is lower–about 1,634–because days with no change in the index price are excluded from the runs test. Approximately 1,200 runs are observed.

For four indices, p-values exceed 0.05; thus, the null hypothesis H_0 (runs are random) is not rejected, meaning the sign sequence of returns is consistent with a random walk and with the weak form of efficiency over the tested horizon. By contrast, for PX and SAX the p-values fall below 0.05, so H_0 is rejected, implying that the sign sequence is not random. The direction of the departure accords with the sign of the Z-statistic: too many runs ($Z > 0$) indicate mean reversion, whereas too few runs ($Z < 0$) indicate momentum (trend persistence). In our case, this suggests mean reversion for PX and momentum for SAX. For the remaining indices, any tendencies (e.g., a mild tilt toward mean reversion) are not statistically significant ($p > 0.05$).

In the subperiod split, the key change concerns PX: in 2015–2019, H_0 is rejected ($\alpha = 0.05$), indicating a non-random sign sequence. For the other indices, results are broadly similar to the baseline in terms of strength (rejections vs. non-rejections of H_0) and the direction of departures, with the exception of SBITOP, which shows slight momentum (too few runs; $Z < 0$). In 2020–2024, $H_0$is not rejected only for WIG, CROBEX, and BUX ($p > 0.05$; results for WIG and CROBEX are strongly consistent with H_0). PX is at the threshold of statistical significance ($p \approx 0.05$), while SAX and SBITOP reject H_0. Based on the Z-values, the direction of departures is clear: SAX exhibits strong momentum (too few runs; $Z < 0$), whereas SBITOP exhibits mean reversion (too many runs; $Z > 0$).

4.3 Variance-Ratio Test

The full-period results are presented in Appendix C, Table C1. The subperiod results are reported in Appendix C, Table C2. For the full sample, the Lo–MacKinlay variance-ratio test indicates that the null hypothesis H_0(prices follow a random walk) is not rejected only for two indices–Poland's WIG and Hungary's BUX–and this holds consistently across all inspected aggregation intervals q. For PX and SBITOP, $H_0$is rejected at all aggregation intervals except one, whereas for CROBEX and SAX it is rejected at all intervals. In terms of direction, nearly all indices (except SAX) exhibit momentum for most q(i.e., $VR(q) > 1$), while SAX points to mean reversion (i.e., $VR(q) < 1$).

In 2015–2019, only SAX rejects H_0; CROBEX shows mixed outcomes depending on the chosen aggregation interval q. For all other indices $H_0$is not rejected, with the caveat that WIG displays a departure at the first interval. By direction, most indices in this subperiod indicate mean reversion, whereas CROBEX exhibits more marked momentum.

After 2020, the findings align more closely with the full-period results: $H_0$is not rejected only for BUX and WIG. The direction of departures mostly indicates momentum, with SAX again standing out for marked mean reversion. Results are consistent across aggregation intervals, with a minor exception for WIG, where estimates at some intervals lie very close to the rejection thresholds

4.4 Unit Root Tests

The unit-root test results are provided in Appendix D–Table D1 for the 2015–2024 period and Table D2 for the subperiods. Over the full sample, for all indices the ADF and Phillips–Perron (PP) tests do not reject the null hypothesis H_0(presence of a unit root, i.e., $I(1)$); for CROBEX, the heteroskedasticity-robust PP result is at the threshold of statistical significance (p-value around 0.05), yet H_0still cannot be reliably rejected. Because we test log prices, such outcomes–i.e., non-rejection of $H_0$and thus $I(1)$–are compatible with the weak form of the EMH (the random-walk property of prices).

In the subperiod analysis, there are no material departures. Notably, for CROBEX the subperiod results provide clearer support for H_0(hence $I(1)$), in contrast to the borderline outcome in the full sample. Differences between the two tests (ADF vs. PP) appear only as minor variations in p-values, while across subperiods we observe some fluctuation in the "strength" of (non-)rejection for specific indices (especially BUX and PX). These do not alter the central conclusion: in all cases, the unit-root null $H_0$is not rejected.

4.5 Reasons for Divergence in Empirical Tests Results of Weak-Form Market Efficiency

Differences in assessing the ability of individual indices to follow the weak form of the Efficient Market Hypothesis arise primarily because the statistical tests used measure different aspects of efficiency. As already highlighted in the methodology section, testing the weak form of the EMH through the random walk model implies two verifiable properties of financial time series: (i) the absence of serial dependence in returns and (ii) the non-stationarity of price series (i.e., the presence of a unit root). Non-stationarity implies that the statistical properties of a time series, such as the mean, variance, and covariance, change over time. However, satisfying this condition represents only a necessary, but not sufficient, condition for the validity of the weak form of the EMH. As emphasized by Rahman and Saadi (2008), a time series can be non-stationary (in prices) or stationary (in returns) while still being predictable, thereby violating the core requirement of the Efficient Market Hypothesis. The reliability of unit root tests may be further limited by the presence of structural breaks.

Additional discrepancies in results also arise from tests that examine the absence of predictability or independence of returns. The random walk model, as a framework for testing the weak form of the EMH, was formalized into three hierarchical forms by Campbell, Lo, and MacKinlay (1997). The first-order random walk model (RW1) assumes independent and identically distributed (i.i.d.) returns with constant variance (homoskedasticity). The second-order model (RW2) relaxes the identical distribution assumption by allowing returns to be independent but not identically distributed – permitting unconditional heteroskedasticity – while still requiring full statistical independence between observations. The third-order model (RW3) further relaxes the independence assumption by allowing for conditionally dependent increments, requiring only that returns remain serially uncorrelated rather than fully independent. This formulation accommodates time series in which dependence exists in higher-order moments, such as the conditional variance, even in the absence of linear autocorrelation in returns.

The remaining statistical tests used in this study can be classified according to the aspect of dependence or randomness they detect. The variance ratio test directly examines consistency with the random walk model, and under its heteroskedasticity-robust specification (Lo and MacKinlay,1988), is robust to the conditional heteroskedasticity characteristic of RW3. Consequently, under less restrictive assumptions, it more often confirms weak-form market efficiency than tests requiring full independence. In contrast, the Ljung–Box test examines only linear uncorrelatedness of returns, which is a necessary but not sufficient condition for a random walk, as it does not capture nonlinear dependencies nor ensure full independence of returns. Moreover, the Ljung–Box test often leads to more frequent rejections of the null hypothesis because it requires the absence of linear autocorrelation across all observed lags; even a single statistically significant autocorrelation coefficient is sufficient to reject the null. Since financial time series often exhibit weak but persistent short-term dependencies and heteroskedasticity, this condition is rarely fully satisfied in practice. The runs test, on the other hand, evaluates the randomness of the sequence of returns and is therefore predominantly associated with testing the stricter form of the random walk (RW1), which assumes complete independence of returns. Its rejection thus indicates the presence of nonlinear patterns in the series, even in the absence of linear autocorrelation (Roldán-Casas and García-Moreno García, 2022).

5 Conclusions

The summary of the tests performed is reported in Appendix E (Table E1)**.** The results point to differences in compliance with the conditions for weak-form efficiency across the markets examined. WIG and BUX stand out as relatively weak-form efficient, meeting the conditions in almost all tests, with a slight edge for WIG (in terms of the strength of rejections in the statistical procedures employed). PX, SBITOP, and–conditionally–CROBEX can be classified as partially weak-form efficient: although the number of null-hypothesis rejections for CROBEX is comparable to PX and SBITOP, the departures in the tests where conditions were not met are more pronounced and, in their characteristics, closer to SAX, which we classify as weak-form inefficient. In a direct SBITOP–PX comparison we give preference to PX, since for it we observe stronger rejections of the null hypothesis in the selected tests. final ranking of indices is presented in Appendix E (Table E2).

We also note divergences in results across statistical procedures, as different statistical tests are designed to examine different assumptions or conditions of the random walk model, which in turn determine the acceptance or rejection of the weak-form Efficient Market Hypothesis. Unit root tests most often fail to reject the null hypothesis, which in this case implies the presence of a unit root in log price series, i.e. non-stationarity. This represents a necessary, but not sufficient, condition for the validity of the weak-form EMH. Following these, the variance ratio test examines the broadest form of the random walk model (RW3), while the runs test is consistent with the assumptions of RW1. The Ljung–Box test, similarly to unit root tests, assesses only the condition of no autocorrelation; however, due to the requirement that the null hypothesis must hold across all lags, it is frequently rejected.

In the subperiod split, a structural shift after 2020 is evident: rejections of the null hypothesis are markedly rarer after the COVID-19 period than before it. Özkan (2021) cites as possible reasons the strengthened role of monetary and fiscal policy, the increased participation of retail investors, and departures of macroeconomic indicators from typical levels, which complicate their interpretation; additional macroeconomic shocks and geopolitical events may further have reduced market efficiency. The variability of observed efficiency is also consistent with Lo's (2004) Adaptive Markets Hypothesis, according to which markets temporarily depart from weak-form efficiency during periods of major shocks and heightened uncertainty.

References

Borges, M. R. (2010). Efficient market hypothesis in European stock markets. *The European Journal of Finance, 16*(7), 711–726. https://doi.org/10.1080/1351847X.2010.495477

Campbell, J. Y., Lo, A. W., & MacKinlay, A. C. (1997). *The econometrics of financial markets*. Princeton University Press.

Claessens, S., Djankov, S., & Klingebiel, D. (2000). *Stock markets in transition economies* (Financial Sector Discussion Paper No. 5). World Bank. https://documents1.worldbank.org/curated/en/303211468760761590/pdf/multi0page.pdf

Čelebić, M. (2009). *Učinkovitost trgov kapitala držav Zahodnega Balkana: Empirična analiza* (Bachelor's thesis, University of Ljubljana). http://www.cek.ef.uni-lj.si/u_diplome/celebic4076.pdf

Deželan, S. (2000). Efficiency of the Slovenian equity market. *Economic and Business Review, 2*(1), 61–83.

Dritsaki, C. (2011). The random walk hypothesis and correlation in the Visegrad countries' emerging stock markets. *Romanian Economic Journal, 14*(40), 25–56. https://rejournal.eu/sites/rejournal.versatech.ro/files/articole/2011-06-01/2083/je40-dritsaki.pdf

Erdas, M. L. (2019). Validity of weak-form market efficiency in Central and Eastern European countries. *Review of Economic Perspectives, 19*(4), 399–428. https://doi.org/10.2478/revecp-2019-0020

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance, 25*(2), 383–417. https://doi.org/10.2307/2325486

Frank, J., & Öhrström, V. (2019). *Testing the weak form EMH: An empirical study of the Swedish stock market* (Bachelor's thesis). University of Gothenburg. https://gupea.ub.gu.se/bitstream/handle/2077/61212/gupea_2077_61212_1.pdf

Gilmore, C. G., & McManus, G. M. (2001). Random walk and efficiency tests of Central European equity markets. *SSRN Working Paper*. https://doi.org/10.2139/ssrn.269510

Gordon, B., & Rittenberg, L. (1995). The Warsaw stock exchange: A test of market efficiency. *Comparative Economic Studies, 37*(2), 1–27. https://doi.org/10.1057/ces.1995.7

Hančlová, J., & Rublíková, E. (2006). Testing the weak form of efficiency on Czech and Slovak stock market. *Badania Operacyjne i Decyzje*, (1), 21–38.

Heininen, P., & Puttonen, V. (2008). Stock market efficiency in transition economies. In *Proceedings of the EACES Conference.*

Ljung, G. M. (1986). Diagnostic testing of univariate time series models. *Biometrika, 73*(3), 725–730. https://doi.org/10.2307/2336539

Lo, A. W. (2004). The adaptive markets hypothesis. *Journal of Portfolio Management, 30*(5), 15–29.

Lo, A. W., & MacKinlay, A. C. (1988). Stock market prices do not follow random walks. *Review of Financial Studies, 1*(1), 41–66. https://doi.org/10.1093/rfs/1.1.41

Macskási, Z., & Molnár, J. (1996). *The predictability of the Hungarian stock exchange* (ACE Project Memoranda No. 96/13).

Omay, N. C., & Karadağli, E. C. (2010). Testing weak-form market efficiency. *MPRA Paper No. 27312.*

Özkan, O. (2021). Impact of COVID-19 on stock market efficiency. *Research in International Business and Finance, 58*, 101445. https://doi.org/10.1016/j.ribaf.2021.101445

Phillips, P. C. B., & Perron, P. (1988). Testing for a unit root. *Biometrika, 75*(2), 335–346. https://doi.org/10.1093/biomet/75.2.335

Rahman, A., & Saadi, S. (2008). Random walk and breaking trend in financial series. *Journal of Economic Studies, 35*(2), 91–103. https://doi.org/10.1016/j.rfe.2007.05.002

Roldán-Casas, J. A., & García-Moreno García, M. B. (2022). A procedure for testing the hypothesis of weak efficiency in financial markets. *Annals of Finance*. https://doi.org/10.1007/s10260-022-00627-4

Smith, G. (2012). The changing efficiency of European emerging stock markets. *The European Journal of Finance, 18*(8), 689–708. https://doi.org/10.1080/1351847X.2011.628682

Smith, G., & Ryoo, H. J. (2003). Variance ratio tests of the random walk hypothesis. *The European Journal of Finance, 9*(3), 290–300. https://doi.org/10.1080/1351847021000025777

Šonje, V., Alajbeg, D., & Bubas, Z. (2011). Efficient market hypothesis: Croatian market. *Financial Theory and Practice, 35*(3), 301–326. https://doi.org/10.3326/fintp.35.3.3

Worthington, A. C., & Higgs, H. (2004). Random walks and market efficiency. *Global Journal of Finance and Economics, 1*(1), 59–78.

Wright, J. H. (2000). Alternative variance-ratio tests using ranks and signs. *Journal of Business & Economic Statistics, 18*(1), 1–9. https://doi.org/10.1080/07350015.2000.10524842

Appendix a

Table A1. Ljung-Box test results for the full period (2015-2024).

Indices	Lag5 (Q value, p-value)	Lag10 (Q value, p-value)	Lag20 (Q value, p-value)
Period	2015-2024	2015-2024	2015-2024
BUX	26,92 (0,0001)	40,94 (0,0000)	63,80 (0,0000)
CROBEX	129,84 (0,0000)	157,24 (0,0000)	179,88 (0,0000)
PX	24,09 (0,0002)	33,06 (0,0003)	61,12 (0,0000)
SAX	33,07 (0,0000)	45,98 (0,0000)	52,14 (0,0001)
SBITOP	41,99 (0,0000)	55,22 (0,0000)	69,78 (0,0000)
WIG	7,43 (0,1906)	12,64 (0,2443)	20,54 (0,4241)

Source: Author's calculations.

Table A2. Ljung-Box test results for subperiods.

Indices	Lag 5 (Q value, p-value)		Lag 10 (Q value, p-value)		Lag 20 (Q value, p-value)	
Period	2015–2019	2020–2024	2015–2019	2020–2024	2015–2019	2020–2024
BUX	8,12 (0,1496)	27,68 (0,0000)	16,42 (0,0881)	40,55 (0,0000)	25,82 (0,1717)	63,76 (0,0000)
CROBEX	35,53 (0,0000)	122,23 (0,0000)	69,01 (0,0000)	148,20 (0,0000)	88,75 (0,0000)	174,60 (0,0000)
PX	12,64 (0,0269)	27,28 (0,0001)	15,09 (0,1287)	41,83 (0,0000)	27,11 (0,1321)	75,43 (0,0000)
SAX	21,59 (0,0006)	16,76 (0,0050)	28,71 (0,0014)	26,35 (0,0033)	38,26 (0,0082)	35,51 (0,0175)
SBITOP	6,07 (0,2991)	41,52 (0,0000)	9,52 (0,4835)	53,71 (0,0000)	23,71 (0,2553)	69,23 (0,0000)
WIG	13,96 (0,0158)	7,11 (0,2119)	16,08 (0,0973)	13,95 (0,1751)	24,74 (0,2114)	25,07 (0,1987)

Source: Author's calculations.

Appendix b

Table B1. Runs test results for full period (2015-2024).

Runs test (2015-2024)					
Indices	**Observations**	**Observed runs**	**Expected runs**	**Z-statistics**	**p-value**
BUX	2485	1204	1234	-1,20	0,2301
CROBEX	2483	1234	1236	-0,08	0,9341
PX	2485	1182	1236	-2,18	0,0292
SAX	1634	958	816	7,07	0,0000
SBITOP	2486	1215	1240	-1,01	0,3132
WIG	2486	1230	1242	-0,50	0,6172

Source: Author's calculations.

Table B2. Runs test results for subperiods.

Runs test for subperiods										
Indices	**Observations**		**Observed runs**		**Expected runs**		**Z-statistics**		**p-value**	
Period	2015-2019	2020-2024	2015-2019	2020-2024	2015-2019	2020-2024	2015-2019	2020-2024	2015-2019	2020-2024
BUX	1241	1244	612	593	616	619	-0,23	-1,47	0,8194	0,1428
CROBEX	1241	1242	614	620	621	615	-0,38	0,31	0,7009	0,7543
PX	1241	1244	600	583	619	617	-1,10	-1,96	0,2700	0,0501
SAX	965	669	539	419	480	335	3,86	6,46	0,0001	0,0000
SBITOP	1242	1244	641	575	622	616	1,08	-2,36	0,2797	0,0181
WIG	1242	1244	600	630	621	622	-1,22	0,46	0,2223	0,6458

Source: Author's calculations.

Appendix c

Table C1: Variance-ratio test results for full period (2015-2024).

Indices	Lag (q)	VR	Z-statistics	p-value
BUX	2	1,01	0,73	0,4641
	4	1,01	0,30	0,7664
	8	1,02	0,39	0,6999
	16	1,05	0,52	0,6056
	32	1,09	0,70	0,4848
CROBEX	2	0,97	-1,38	0,1668
	4	1,12	3,29	0,0010
	8	1,37	6,32	0,0000
	16	1,65	7,31	0,0000
	32	1,69	5,42	0,0000
PX	2	1,04	1,91	0,0562
	4	1,15	4,05	0,0001
	8	1,22	3,77	0,0002
	16	1,28	3,18	0,0014
	32	1,21	1,61	0,1079
SAX	2	0,89	-5,39	0,0000
	4	0,79	-5,49	0,0000
	8	0,73	-4,53	0,0000
	16	0,62	-4,26	0,0000
	32	0,48	-4,04	0,0000
SBITOP	2	1,04	1,83	0,0670
	4	1,16	4,15	0,0000
	8	1,30	5,05	0,0000
	16	1,51	5,77	0,0000
	32	1,56	4,35	0,0000
WIG	2	1,02	0,87	0,3820
	4	1,04	1,02	0,3067
	8	1,07	1,17	0,2427
	16	1,15	1,67	0,0959
	32	1,13	0,98	0,3279

Source: Author's calculations.

Table C2: Variance ratio test results for subperiods.

Indices	Lag (q)	VR	Z-statistics	p-value	VR	Z-statistics	p-value
Period		2015-2019			2020-2024		
BUX	2	1,01	0,29	0,7704	1,02	0,64	0,5232
	4	0,97	-0,61	0,5439	1,03	0,61	0,5408
	8	0,85	-1,74	0,0810	1,11	1,29	0,1975
	16	0,81	-1,51	0,1306	1,16	1,31	0,1904
	32	0,82	-0,97	0,3330	1,23	1,27	0,2056
CROBEX	2	1,07	2,51	0,0120	0,93	-2,43	0,0149
	4	1,03	0,58	0,5620	1,16	3,05	0,0023
	8	1,12	1,43	0,1539	1,48	5,76	0,0000
	16	1,30	2,44	0,0147	1,79	6,34	0,0000
	32	1,56	3,09	0,0020	1,76	4,21	0,0000
PX	2	1,01	0,23	0,8218	1,06	1,97	0,0489
	4	1,02	0,31	0,7577	1,23	4,26	0,0000
	8	0,95	-0,54	0,5861	1,37	4,42	0,0000
	16	0,86	-1,10	0,2715	1,51	4,10	0,0000
	32	0,71	-1,62	0,1049	1,48	2,68	0,0073
SAX	2	0,88	-4,09	0,0000	0,90	-3,45	0,0006
	4	0,77	-4,35	0,0000	0,83	-3,30	0,0010
	8	0,66	-4,06	0,0001	0,82	-2,13	0,0332
	16	0,56	-3,49	0,0005	0,69	-2,46	0,0138
	32	0,40	-3,34	0,0009	0,57	-2,37	0,0179
SBITOP	2	0,98	-0,57	0,5675	1,06	2,12	0,0336
	4	0,98	-0,31	0,7551	1,23	4,32	0,0000
	8	0,98	-0,21	0,8335	1,43	5,13	0,0000
	16	1,06	0,46	0,6473	1,70	5,62	0,0000
	32	1,18	0,99	0,3231	1,71	3,91	0,0001
WIG	2	1,07	2,49	0,0128	1,00	-0,11	0,9118
	4	1,05	0,95	0,3431	1,03	0,66	0,5112
	8	0,98	-0,27	0,7882	1,11	1,28	0,2018
	16	0,89	-0,89	0,3726	1,25	2,04	0,0417
	32	0,93	-0,37	0,7113	1,22	1,21	0,2281

Source: Author's calculations.

Appendix d

Table D1: Unit root test results for full period (2015-2024).

Unit root test				
2015-2024				
Indices	ADF Statistics	ADF p-value	PP Statistics	PP p-value
BUX	-0,34	0,9200	-0,42	0,9069
CROBEX	-2,92	0,0428	-2,86	0,0496
PX	-2,19	0,2095	-2,33	0,1619
SAX	0,11	0,9664	0,11	0,9668
SBITOP	-2,26	0,1850	-2,20	0,2074
WIG	-2,25	0,1872	-2,36	0,1533

Source: Author's calculations.

Table D2: Unit root test results for subperiods.

Unit root test								
Indices	ADF Statistics	ADF p-value	ADF Statistics	ADF p-value	PP Statistics	PP p-value	PP-statistics	PP p-value
Period	2015-2019		2020-2024		2015-2019		2020-2024	
BUX	0,85	0,9924	-2,15	0,2251	1,12	0,9954	-2,21	0,2044
CROBEX	-2,05	0,2659	-2,16	0,2217	-2,06	0,2599	-2,09	0,2493
PX	-1,76	0,4000	-1,63	0,4700	-1,77	0,3973	-1,77	0,3978
SAX	0,19	0,9715	-2,02	0,2800	0,64	0,9886	-2,00	0,2871
SBITOP	-2,03	0,2745	-1,72	0,4226	-2,07	0,2561	-1,69	0,4348
WIG	-1,39	0,5887	-1,69	0,4367	-1,33	0,6150	-1,81	0,3736

Source: Author's calculations.

Appendix e

Table E1: Tests results summary.

(2015-2024)	LJUNG-BOX TEST	RUNS TEST	VR	UNIT ROOT TEST
BUX	-	+	+	+
CROBEX	-	+	-	-/+
PX	-	-	-/+	+
SAX	-	-	-	+
SBITOP	-	+	-	+
WIG	+	+	+	+
2015-2019				
BUX	+	+	+	+
CROBEX	-	+	+	+
PX	-/+	+	+	+
SAX	-	-	-	+
SBITOP	+	+	+	+
WIG	-/+	+	+	+
2020-2024				
BUX	-	+	+	+
CROBEX	-	+	-	+
PX	-	-/+	-	+
SAX	-	-	-	+
SBITOP	-	-	-	+
WIG	+	+	+	+

Source: Author's calculations.

Table E2: Final rankings of selected indices.

Indices	Number of rejected H_0
WIG	11,5
BUX	10
PX	6,5
SBITOP	7
CROBEX	6,5
SAX	3

Source: Author's calculations.