

ISSN: 2543-6821 (online)

Journal homepage: <http://ceej.wne.uw.edu.pl>

Iskandar Sheraliev,
Robert Ślepaczuk

Cross-Country Differences in Return and Volatility Metrics of World Equity Indices

To cite this article

Sheraliev, I., Ślepaczuk, R. (2023). Cross-Country Differences in Return and Volatility Metrics of World Equity Indices. *Central European Economic Journal*, 10(57), 90-115.

DOI: 10.2478/ceej-2023-0006

 To link to this article: <https://doi.org/10.2478/ceej-2023-0006>



Iskandar Sheraliev

University of Warsaw, Quantitative Finance Research Group, Długa 44/50, 00-241 Warsaw, Poland

Robert Ślepaczuk 

University of Warsaw, Faculty of Economic Sciences, Quantitative Finance Research Group,

Długa 44/50, 00-241 Warsaw, Poland

corresponding author: rslepaczuk@wne.uw.edu.pl

Cross-Country Differences in Return and Volatility Metrics of World Equity Indices

Abstract

This research seeks to determine whether the cross-country differences in return and volatility metrics in various country equity indices can be explained by differences in economic development. We base the study on the MSCI IMI net income indices on two samples: a 51-country sample from the period 31 May 2002 to 28 February 2022, and a 75-country sample from the period 30 November 2010 to 28 February 2022. In this study, countries are grouped into four categories: frontier, emerging, early-developed, and developed, based on gross domestic product (GDP) per capita. The Kruskal–Wallis rank sum test is used to find cross-group differences, and the results are further analyzed with the pairwise Wilcoxon rank sum test with the Holm–Bonferroni p value adjustment method. The results are relatively unintuitive and show that there is no significant cross-group difference in daily and monthly returns. There is evidence of a considerable difference in volatility metrics, especially in the case of the emerging market group, which is significantly different from the three other groups. The results are slightly sensitive to time period change and very sensitive to changes in income categories of some countries.

Keywords

Frontier markets | Emerging markets | Developed markets | Cross-country differences | Volatility | GDP per capita

JEL Codes

1. Introduction

The aim of this paper is to determine whether economic development can be used to explain cross-country differences in return and volatility metrics. In previous studies, cross-country differences in risk-return characteristics were explained using various other metrics, including market capitalization, degree of financial liberalization, and financial development metrics. There have been only a few types of research done using gross domestic product (GDP) per capita to explain the cross-country differences. Therefore, we decided to develop this topic further.

Additionally, after preliminary research on this topic, we noticed that the relationship of volatility metrics between tested groups of countries is significant and inconsistent with the existing literature. It was one of the reasons that we decided to

focus on this research. Our main contributions show results contrary to the established literature and add to the discussion of this phenomenon. We have found numerous results in the quantitative finance literature focused on asset allocation and portfolio management showing that the established relationships (i.e., that higher volatility is closely connected with higher returns) do not necessarily work. Examples of such are the results of high-volatility portfolios versus low-volatility portfolios, and low-beta stocks versus high-beta stocks, which contradicts the information from Markowitz theory or the single-index Sharpe model and can be based on premises similar to our results.

In this context, we state the following research hypotheses:

- RH1: Do daily and monthly return distributions of country equity indices differ with regard to the level of economic development?

- RH2: Do distributions of volatility metrics of country equity indices differ with regard to the level of economic development?

Additionally, in order to test the robustness of results of the main hypotheses, the following research questions were developed:

- RQ1: Is the result obtained robust to the change in time period used?
- RQ2: Is the result obtained robust to the change in the income categories of countries?

In the process of verification of the above-mentioned research hypotheses and questions, we used gross domestic product (GDP) per capita in current USD in 2020 collected by the World Bank and, in the case of Taiwan, initial calculations of 2020 GDP per capita made by the International Monetary Fund (IMF) and published in the IMF *World Economic Outlook, October 2021* (International Monetary Fund, 2021). Furthermore, we used Morgan Stanley Capital International Investable Market Indexes (MSCI IMI) net income country indices that include big, medium, and small capitalization equity in two samples based on the availability of data: 51 country samples from the period 31 May 2002 to 28 February 2022 and 75 countries from the period 30 November 2010 to 29 February 2022. The frequency of the index values is daily.

Based on GDP per capita values, countries were divided into four categories: frontier, emerging, early-developed, and developed. There are five possible grouping scenarios of our categorization methodology. The first scenario is used as the baseline for verification of the main hypothesis, and four other outcomes are used to answer the second research question.

In order to verify the main hypotheses, the Kruskal–Wallis rank sum test is used. The results of the Kruskal–Wallis test are then further elaborated using the pairwise Wilcoxon rank sum test with adjusted p values based on the Holm–Bonferroni adjustment method. Additionally, time period is changed from 2002–2022 to 2007–2022 for the first sample, and from 2010–2022 to 2015–2022 for the second sample to answer the corresponding research question. We expect that the Kruskal–Wallis test will show that there exist cross-group differences in daily and monthly return as well as volatility metrics of country equity indices. However, we expect the result to depend on the time period and changes in the income categories of some countries.

The paper is structured as follows: In section 2, the literature review of GDP per capita and chosen volatility metrics and the results of previous research on the cross-country risk and return differences are provided; in section 3, we explain the methodology of categorization of countries based on GDP per capita, the calculation process of volatility metrics, the Kruskal–Wallis rank sum test, the pairwise Wilcoxon rank sum test, and the Holm–Bonferroni adjustment method; in section 4, the data samples are fully described; in section 5, empirical results are presented and discussed; in section 6, we conduct a sensitivity analysis to test the robustness of the empirical results; in section 7, we draw conclusions and suggest extensions and ideas for further research on this topic.

2. Literature

The relationship between economic growth measured in GDP (real/nominal, aggregate/per capita) and stock market performance is well researched. Amtiran et al. (2017) in their study of the Indonesian capital market in the period from 2007 to 2014 with a sample of 80 companies using ordinary least square regression (OLS) found that nominal GDP growth rate has a positive impact on stock returns but is insignificant. A similar study was done by Amaresh et al. (2020) using the Colombo Stock Exchange All-Share Price Index as a dependent variable and inflation, interest rate, and GDP as independent variables in the OLS model. They studied 120 observations in the period from January 2009 to December 2018, and a positive relationship between the GDP of Sri Lanka and the Index was found, but it was insignificant. Montes and Tiberto (2012), using OLS and the generalized method of moments (GMM), explored the relationship between macroeconomic variables, country risk, and Brazilian stock performance. They used Index Bovespa (IBOVESPA) values in the period from December 2001 to September 2010. They found that the GDP of Brazil and IBOVESPA performance were positively related and that this relationship was significant in both the OLS and GMM models. Giri and Joshi (2017) used the autoregressive distributed lag (ARDL) approach and the vector error correction model (VECM) to examine the relationship between Indian stock market performance and certain macroeconomic variables, using annual data from 1979 to 2014 of the Bombay Stock Exchange Sensitivity Index. They discovered that economic growth (real GDP growth)

had a significant positive short- and long-term effect on stock prices, and that the stock price growth was unidirectionally caused by real GDP growth. The evidence from the Taiwan Stock Exchange found by Singh et al. (2011) suggests that GDP positively affects prices of stock portfolios regardless of the size of the firm. The study of stock market performance of the United Arab Emirates in the period from 1990 to 2005 (Al-Tamimi et al., 2011) revealed a positive but insignificant relationship between GDP value and stock price. A similar study (Kalam, 2020) was done in Malaysia in the period from 2000 to 2019, and a positive relationship between GDP value and stock price was found. In Nigeria, using the sample period of 1975 to 2005, Osamwonyi and Evbayiro-Osagie (2012) came to a similar conclusion. Overall, a positive relationship between GDP and stock price was found.

There were also many studies on how macroeconomic factors affect stock market development. Cherif and Gazdar (2010), in their study of 14 the Middle East and North Africa (MENA) countries in a sample period from 1990 to 2007, examined the institutional and macroeconomic determinants of stock market development. They used market capitalization divided by GDP as a proxy for stock market development. One of the important factors of market capitalization was real income level (the real GDP in USD), which they found significant in nine out of ten of their regressions. A similar study was conducted by Yartey (2008) using panel data of 42 emerging economies over the period from 1990 to 2004. One of the findings of the study was that the GDP per capita significantly and positively affects stock market development measured by market capitalization as a percentage of GDP. Moreover, Ho and Iyke (2017), in their review of the literature, found that previous research indicates that real income level and its growth positively affect stock market development.

There is some research that explains cross-country differences in stock market risk-return characteristics with the level of financial development. Dellas and Hess (2005), in their study of 49 countries in total emerging and developed markets over the period of 1980 to 1999 found that cross-country difference in stock returns is significantly explained by the degree of financial development measured by four indicators: liquid liabilities by GDP, commercial-central bank ratio, private credit divided by GDP, and total value of shares traded as a percentage of GDP. Countries with more developed banking systems also had less

volatile stock returns. Additionally, it was found by Pradhan et al. (2014) in their research on the relationship between economic growth, banking sector development, and other factors in ASEAN countries in the period 1961 through 2012, that banking sector advancement Granger-causes stock market development unidirectionally, and that the relationship between economic growth and stock market development can be both unidirectional and bidirectional. Similarly, a review of literature done by Ho and Iyke (2017) suggests that banking sector development can be both a substitute and a complement to the stock market, meaning it can both hinder and help in its development.

Umutlu et al. (2010) examined the relationship between the aggregate total volatility of 25 emerging economies and the degree of financial liberalization in the period from 1991 to 2005. They used several measures of financial liberalization: LMF, FEL, IC, and EW. LMF is “the sum of a country’s foreign equity assets and liabilities and the foreign direct investment assets and liabilities as a share of the GDP”. FEL is a ratio of the capitalization of foreign firms in the local stock exchange by the whole stock market capitalization in a given country. IC measures the openness in capital controls. EW measures the accessibility of stock exchange by foreign investors. In their study, they found that all these measures are significant and reduce the aggregated total volatility. They also found that when the countries were divided by GDP into small, medium, and large, the measures of financial liberalization were significant only for small countries while these measures lost significance at higher GDP levels. Their reasoning was that as the size of the economy increases, additional foreign investors are of lesser importance, whilst smaller GDP countries benefit from the bigger investor base the most. A conclusion similar to this research was reached by James and Karoglou (2010) in their study of the Indonesian market in the period from April 1983 to January 2006, when the opening of the market to foreign investors significantly reduced the volatility of the market index.

Downside risk is one of the measures that can explain the difference in risk-return characteristics among countries. Downside risk, specifically mean-semi-variance and downside beta, explains returns much better in comparison to mean-variance and beta, according to Estrada (2007). His study used data from 23 developed and 27 emerging markets in the period from January 1988 to December 2001. The

Table 1. Summary of the Literature Review

Subject	Summary	Authors
Economic growth	Increases stock market return	Al-Tamimi et al. (2011)
		Amaresh et al. (2020)
		Amtiran et al. (2017)
		Giri and Joshi (2017)
		Kalam (2020)
		Montes and Tiberto (2012)
		Osamwonyi and Evbayiro-Osagie (2012)
Economic growth	Causes stock market development	Singh et al. (2011)
		Cherif and Gazdar (2010)
		Ho and lyke (2017)
Banking sector development	Causes stock market development	Yartey (2008)
		Dellas and Hess (2005)
		Ho and lyke (2017)
Banking sector development	Reduces stock market volatility	Pradhan et al. (2014)
		Dellas and Hess (2005)
Financial liberalization	Reduces stock market volatility	James and Karoglou (2010)
		Umutlu et al. (2010)
Downside volatility	Increases stock market return	Ali (2019)
		Ang et al. (2006)
		Estrada (2007)
Risk-adjusted returns	Higher in emerging markets	Atilgan and Demirtas (2016)

importance of downside risk is further supported by Ali (2019), who used a sample of 3,658 companies listed on the Chinese stock market from 1998 to 2017. The study yielded similar results regarding downside risk, semi-deviation, and downside beta in particular: “the results show a positive reward for holding stocks with high downside risk, and this reward is not explained by other cross-sectional effects.” Ang et al. (2006), using the returns of companies listed on the New York Stock Exchange (NYSE) from July 1963 to December 2001 found that “cross section of stock returns reflects a downside risk premium of approximately 6% per annum.” Overall, the downside risk is associated with a positive premium for stock returns.

There are a few studies similar to this. Atilgan and Demirtas (2016) compared the risk-adjusted performance of countries using the ordinary Sharpe ratio, a variation of it that uses value-at-risk as the denominator, and another variation that uses expected shortfall (ES). The data used in the study include the returns in the

period from January 1973–to December 2011 from 28 developed markets and 24 emerging markets. They found that emerging markets had a better risk-adjusted performance in the whole period, in the period from January 1973– to September 2008, and in the period from 2008– to December 2011. Furthermore, using Fama–Macbeth regression, they found that expected returns for horizons from one month to twelve months are significantly higher for indices that had a higher risk-adjusted ratio based on the risk calculated over 100 trading days and return for the previous month.

A summary of the literature review can be seen in Table 1.

Overall, there is evidence of a correlation between GDP and stock market development and a causal relationship between stock market development and economic growth of a country; thus it is possible to use one as a proxy for the other. Additionally, stock market development is partially determined by the economic growth of a country. There is also evidence

that the volatility of stock returns is lower in countries with more developed banking systems. Moreover, as stock markets become more open to foreign participation, they become less volatile; this effect is especially evident in small economies. There exists evidence that downside volatility is associated with a positive premium to stock returns, which may lead us to expect higher returns in frontier and emerging markets. Furthermore, emerging markets are known to have higher risk-adjusted returns compared to more developed economies.

3. Methodology

3.1. Income Categorization Method

The most important aspect of this study is the categorization of countries. Based on the GDP per capita measured in current USD in 2020, 75 countries were divided into four distinct groups, from smallest value to highest value: frontier, emerging, early-developed, and developed. To categorize countries into the four groups, incomes in 75 countries were ordered in ascending order. Then three metrics were calculated to find which country would be divided into which category: the percentage difference in income between countries with one and two positions' difference, and the sum of the percentage differences in income.

The percentage difference in income between countries with n position difference:

$$IM_n = \frac{GDP_i}{GDP_{i-n}} - 1, \quad (1)$$

where IM_n is the percentage difference in income between countries with a difference of n position below; GDP is ordered in ascending order; GDP_i or GDP_{i-n} is the GDP per capita of a country numbered i or $i-n$ in ascending order; and $n-a$ is the number of positions below the higher income country.

The sum of percentage difference in income:

$$SIM = IM_1 + IM_2, \quad (2)$$

where IM_1 is the percentage difference with a country one position below in income; GDP is ordered in ascending order; IM_2 is the percentage difference with a country two positions below; and SIM is the sum of the percentage differences.

Then the country that will start the next income category is determined based on four rules: $IM_1 > 5\%$, $IM_2 > 15\%$, $SIM > 25\%$; there should be 15 or more countries in each category. Based on this method, countries can be categorized in five ways, which are illustrated in Table 2. The first country is used as the baseline and the others as the robustness checks. This kind of division into five different versions of categorization partly refers to the very important issue of countries changing category due to the fact the GDP per capita for each country is different each year. We did not change the constituents of each group in each year for the baseline scenario, but by repeating the calculations for each version of the initial categorization, we referred to the issue of rebalancing (i.e. a possible change of the group when the new GDP data are released).

Additionally, in Tables 3 and 4 the comparison between modified baseline classification and MSCI classification of country development is shown.

Tables 3 and 4 show that in comparison to the baseline classification, the MSCI classification shows that there are some differences between the constituents of each group but at the same time they are not larger than the differences among our five versions presented in Table 2.

3.2. Volatility Metrics

Six volatility metrics are used to compare the four different market groups: annualized standard deviation (STD), annualized downside semi-deviation (DSTD), Ulcer index (UI), maximum drawdown (MDD), 97.5% value-at-risk (VaR.N, VaR.H), 97.5% expected shortfall (ES.N, ES.H). The justification for selecting six various volatility metrics can be explained by the need to quantify the risk in many various dimensions, thus enabling us to finally treat our results as robust ones.

STD is calculated using the following formula:

$$\sigma = \sqrt{\frac{1}{N} \sum_{t=1}^N (R_t - \bar{R})^2}, \quad (3)$$

where N is the number of daily returns; R_t given month; R_t is the return on a given day t ; \bar{R} is the average daily return in a month; σ is the STD of a given country equity index in a given month.

Table 2. Income Categorization Method Outcomes

	Baseline	Version 2	Version 3	Version 4	Version 5
Frontier	Pakistan, Zimbabwe, Kenya, India,	+ Lebanon,	+ Lebanon,	Baseline	Baseline
	Bangladesh, Nigeria, Vietnam,	Jamaica,	Jamaica,		
	Morocco, Philippines, Tunisia,	Colombia	Colombia		
	Egypt, Sri Lanka, Ukraine,				
	Indonesia, Jordan				
Emerging	Lebanon, Jamaica, Colombia, South	- Lebanon,	- Lebanon,	+ Bulgaria,	+ Bulgaria,
	Africa, Bosnia and Herzegovina,	Jamaica,	Jamaica,	Russia,	Russia,
	Peru, Botswana, Brazil, Thailand,	Colombia	Colombia	Malaysia,	Malaysia,
	Serbia,			China	China
	Mexico, Turkey, Argentina,	+ Bulgaria,	+ Bulgaria,		
	Mauritius, Kazakhstan	Russia,	Russia,		
		Malaysia,	Malaysia,		
		China	China		
Early-developed	Bulgaria, Russia, Malaysia, China,	- Bulgaria,	- Bulgaria,	- Bulgaria,	- Bulgaria,
	Romania, Chile, Croatia, Oman,	Russia,	Russia,	Russia,	Russia,
	Trinidad and Tobago, Poland,	Malaysia,	Malaysia,	Malaysia,	Malaysia,
	Hungary, Greece, Lithuania,	China	China	China	China
	Bahrain,				
	Portugal, Czech Republic, Estonia,		+ Korea,		+ Korea,
	Kuwait, Slovenia, Spain, Taiwan		Italy, United		Italy, United
			Arab Emirates		Arab Emirates
Developed	Korea, Italy, United Arab Emirates,	Baseline	- Korea, Italy,	Baseline	- Korea, Italy,
	France, Japan, United Kingdom,		United Arab		United Arab
	New Zealand, Canada, Israel,		Emirates		Emirates
	Belgium, Germany, Hong Kong,				
	Austria, Finland, Qatar, Australia,				
	Sweden, Netherlands, Singapore,				
	Denmark, USA, Norway, Ireland,				
	Switzerland				

Note: Countries are arranged in ascending order according to their GDP per capita. Categorization is based on 75 countries, and this categorization is used for all samples. The income category of a country does not change depending on the sample used. The “+” sign adds countries to the baseline list in the same category. The “-” sign removes countries from the list in the same category. “Baseline” means that the category remained unchanged compared to the baseline.

Table 3. Comparison of MSCI and Baseline Classifications

	Baseline classification, ED in D	MSCI classification	Differences
Frontier	Bangladesh, Egypt, India, Indonesia, Jordan, Kenya, Morocco, Nigeria, Pakistan, Philippines, Sri Lanka, Tunisia, Ukraine, Vietnam, Zimbabwe	Bahrain, Bangladesh, Benin, Burkina Faso, Croatia, Estonia, Iceland, Ivory Coast, Jordan, Kazakhstan, Kenya, Lithuania, Mauritius, Morocco, Nigeria, Oman, Pakistan, Romania, Senegal, Serbia, Slovenia, Sri Lanka, Tunisia, Vietnam	Egypt, India, Indonesia - emerging in MSCI classification; Ukraine and Zimbabwe, standalone
Emerging	Argentina, Bosnia and Herzegovina, Botswana, Brazil, Colombia, Jamaica, Kazakhstan, Lebanon, Mauritius, Mexico, Peru, Serbia, South Africa, Thailand, Turkey	Brazil, Chile, China, Colombia, Czech Republic, Egypt, Greece, Hungary, India, Indonesia, Korea, Kuwait, Malaysia, Mexico, Peru, Philippines, Poland, Qatar, Russia, Saudi Arabia, South Africa, Taiwan, Thailand, Turkey, United Arab Emirates	Argentina, Bosnia and Herzegovina, Botswana, Lebanon, Jamaica, standalone in MSCI classification; Kazakhstan, Mauritius, Serbia, frontier
Developed	Australia, Austria, Bahrain, Belgium, Bulgaria, Canada, Chile, China, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hong Kong, Hungary, Ireland, Israel, Italy, Japan, Korea, Kuwait, Lithuania, Malaysia, Netherlands, New Zealand, Norway, Oman, Poland, Portugal, Qatar, Romania, Russia, Singapore, Slovenia, Spain, Sweden, Switzerland, Taiwan, Trinidad and Tobago, United Arab Emirates, United Kingdom, USA	Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Hong Kong, Ireland, Israel, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Singapore, Spain, Sweden, Switzerland, United Kingdom, USA	Bahrain, Estonia, Oman, Romania, Slovenia, frontier; Chile, China, Czech Republic, Greece, Hungary, Korea, Kuwait, Malaysia, Poland, Qatar, Russia, Taiwan, United Arab Emirates, emerging; Bulgaria, standalone.
Standalone		Argentina, Bosnia and Herzegovina, Botswana, Bulgaria, Jamaica, Lebanon, Malta, Palestine, Panama, Trinidad and Tobago, Ukraine, and Zimbabwe	We chose not to use this classification, and instead added Early-developed market class.

Note: Categorization is based on 75 countries; this categorization is used for all samples. The income category of a country does not change depending on the sample used. The developed market in the table includes both early-developed and developed markets according to their own classification.

Annualized STD was calculated in the following way:

$$STD = \sigma\sqrt{252}, \quad (4)$$

where σ is the STD of daily returns in a given month; and STD is the annualized STD of a given country equity index in a given month.

Annualized DSTD is used to capture the variability of negative returns of country equity indices; it was calculated in the following way:

$$DSTD = \sigma_{r<0}\sqrt{252}, \quad (5)$$

where $\sigma_{r<0}$ is the STD of daily returns less than zero in a given month; and $DSTD$ is the annualized DSTD (annualized STD of negative returns in our case).

Drawdowns are calculated separately based on daily values of country indices for each month. They are calculated using the following formula:

$$DD_T = \left[\frac{\max_{t \in (0, T)} P(t) - P(T)}{\max_{t \in (0, T)} P(t)} \right]_+, \quad (6)$$

where $P(T)$ is the value of the country equity index on a given day “ T ” in a given month; and DD_T is the value of a drawdown of an equity index on a given day of a given country. Each month drawdowns are

Table 4. Comparison of MSCI and Baseline Classifications

	Baseline classification, ED in E	MSCI classification	Differences
Frontier	Bangladesh, Egypt, India, Indonesia, Jordan, Kenya, Morocco, Nigeria, Pakistan, Philippines, Sri Lanka, Tunisia, Ukraine, Vietnam, Zimbabwe	Bahrain, Bangladesh, Benin, Burkina Faso, Croatia, Estonia, Iceland, Ivory Coast, Jordan, Kazakhstan, Kenya, Lithuania, Mauritius, Morocco, Nigeria, Oman, Pakistan, Romania, Senegal, Serbia, Slovenia, Sri Lanka, Tunisia, Vietnam	Egypt, India, Indonesia, emerging in MSCI classification; Ukraine and Zimbabwe, standalone
Emerging	Argentina, Bahrain, Bosnia and Herzegovina, Botswana, Brazil, Bulgaria, Chile, China, Colombia, Croatia, Czech Republic, Estonia, Greece, Hungary, Jamaica, Kazakhstan, Kuwait, Lebanon, Lithuania, Malaysia, Mauritius, Mexico, Oman, Peru, Poland, Portugal, Romania, Russia, Serbia, Slovenia, South Africa, Spain, Taiwan, Thailand, Trinidad and Tobago, Turkey	Brazil, Chile, China, Colombia, Czech Republic, Egypt, Greece, Hungary, India, Indonesia, Korea, Kuwait, Malaysia, Mexico, Peru, Philippines, Poland, Qatar, Russia, Saudi Arabia, South Africa, Taiwan, Thailand, Turkey, United Arab Emirates	Bahrain, Croatia, Estonia, Kazakhstan, Lithuania, Mauritius, Romania, Serbia, Slovenia, frontier; Portugal, Spain, developed; Argentina, Bosnia and Herzegovina, Botswana, Bulgaria, Trinidad and Tobago, Lebanon, Jamaica, standalone
Developed	Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Hong Kong, Ireland, Israel, Italy, Japan, Korea, Netherlands, New Zealand, Norway, Qatar, Singapore, Sweden, Switzerland, United Arab Emirates, United Kingdom, USA	Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Hong Kong, Ireland, Israel, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Singapore, Spain, Sweden, Switzerland, United Kingdom, USA	Korea, Qatar, United Arab Emirates, emerging in MSCI classification
Standalone		Argentina, Bosnia and Herzegovina, Botswana, Bulgaria, Jamaica, Lebanon, Malta, Palestine, Panama, Trinidad and Tobago, Ukraine, Zimbabwe	We chose not to use this classification, and instead added early-developed market class.

Note: Categorization is based on 75 countries; this categorization is used for all samples. The income category of a country does not change depending on the sample used. Emerging markets in the table include both early-developed and emerging markets, according to their own classification.

recalculated (i.e. equity index values in a previous month do not affect the next month).

MDD is a measure of the magnitude of the maximum percentage decline of the portfolio value. In this study, MDD is the maximum value of all drawdowns of a given country equity index in a given month. It is calculated in the following way:

$$MDD = \max_{T \in (0, N)} DD_T, \tag{7}$$

where DD_T is the value of a drawdown on a given day in a month; N is the number of daily returns in a given month; and MDD is the maximum drawdown of an equity index in a given month.

The Ulcer index (UI) is an index developed by Peter Martin in 1987 and is one of the volatility metrics made to capture downside variability (Martin & MacCann, 1989). Unlike MDD, which focuses only on the greatest drawdown, UI takes into account all drawdowns in the period to measure the magnitude of the decline of the portfolio values. In our study, we used the following formula for calculation of UI:

$$UI = \frac{\sum_{i=1}^{n_{dd}} DD_i^2}{n_{dd}}, \tag{8}$$

where DD_i is the value of a drawdown on a day “ i ” in a given month; n_{dd} is the number of drawdowns in a given month; and UI is the value of the Ulcer index of an equity index in a given month.

Value-at-risk (VaR) is “a common consistent measure of risk across different positions and risk factors”; it is a measure of the magnitude of a possible loss of a portfolio according to Dowd (2005). ES is a natural development of VaR, which retains the benefits of VaR while avoiding its shortcomings (Dowd, 2005). VaR and ES are calculated using two methods: historical and Gaussian. The historical method uses empirical distribution of daily returns; the Gaussian method assumes that daily returns are normally distributed. Historical 97.5% VaR for a given month was calculated in the following way:

$$VaR.H = -(1 - \gamma)R_x - \gamma R_{x+1}, \quad (9)$$

where

$$x = \lfloor N(1 - 0.975) \rfloor; \quad (10)$$

$$g = N(1 - 0.975) - x, \quad (11)$$

where N is the number of observations of daily returns in a given month; 0.975 is the confidence interval; R_x is the x th order statistic of daily returns; daily returns are in ascending order; x is an observation number calculated using the formula above; $VaR.H$ is the historical 97.5% VaR in a given month of a given country equity index; value is calculated each month independently; and γ equals to 0.5 if $g = 0$, and 1 otherwise.

Historical 97.5% ES for a given month was calculated in the following way:

$$ES.H = -\frac{1}{x} \sum_{i=1}^x R_i, \quad (12)$$

where R_i is the i th order statistic of daily returns in a given month; daily returns are in ascending order; x is an observation number calculated using the formula above; $ES.H$ is the historical 97.5% ES in a given month of a given country equity index; value is calculated each month independently.

Gaussian 97.5% VaR for a given month was calculated in the following way:

$$VaR.N = -\Phi(0.025) \times \sigma - \frac{1}{N} \sum_{i=1}^N R_i, \quad (13)$$

where R_i is the i th order statistic of daily returns in a given month; N is the number of observations of daily returns in a given month; σ is the STD of daily return in a given month calculated using above mentioned formula; and $VaR.N$ is the Gaussian 97.5% VaR in a given month of a given country equity index; value is calculated each month independently.

Gaussian 97.5% ES for a given month was calculated in the following way:

$$ES.N = \frac{\varphi(\Phi(1-0.975))}{1-0.975} \times \sigma + \frac{1}{N} \sum_{i=1}^N R_i, \quad (14)$$

where R_i is the i th order statistic of daily returns in a given month; N is the number of observations of daily returns in a given month; σ is the STD of daily return in a given month calculated using the above mentioned formula; and $ES.N$ is the Gaussian 97.5% ES in a given month of a given country equity index; value is calculated each month independently.

After the calculation of volatility metrics for each country equity index, the time series of the metrics with monthly frequency for each country equity index is formed. Monthly time series for each income category is then calculated based on the mean volatility metrics of countries included in a category for each month.

$$VM_{i,m} = \frac{1}{C_i} \sum_{c=1}^{C_i} VM_{c,i,m}, \quad (15)$$

where $VM_{i,m}$ is the volatility metric (STD, DTST, MDD, UI, VaR.H, VaR.N, ES.H, ES.N) of an income category i (frontier, emerging, early-developed, developed) in a month m ; $VM_{c,i,m}$ is the volatility metric of a country c in the income category i in a month m ; C_i is the number of countries in the income category i , with changes depending on the sample and income categorization version; i is the income category: frontier, emerging, early-developed, developed; and m is the ranges, depending on the number of months in a sample.

Additionally, daily and monthly returns are calculated for each income category using means of returns of country indices. Monthly return is calculated as

$$R.M_{i,m} = \frac{1}{C_i} \sum_{c=1}^{C_i} R.M_{c,i,m}, \quad (16)$$

where $R.M_{i,m}$ is the monthly return of an income category i (frontier, emerging, early-developed, developed) in a month m ; $R.M_{c,i,m}$ is the monthly return of a country c in the income category i in a month m ; C_i is the number of countries in the income category i , with changes depending on the sample and income categorization version; i is the income category: frontier, emerging, early-developed, developed; and m is the ranges depending on the number of months in a sample. Daily return of a category is calculated in the following way:

$$R.D_{i,d} = \frac{1}{C_i} \sum_{c=1}^{C_i} R.D_{c,i,d}, \quad (17)$$

where $R.D_{i,d}$ is the daily return of an income category i (frontier, emerging, early-developed, developed) on a day d ; $R.D_{c,i,d}$ is the daily return of a country c in the income category i on a day d ; C_i is the number of countries in the income category i , with changes depending on the sample and income categorization version; i is the income category: frontier, emerging, early-developed, developed; and d is the ranges, depending on the number of days in a sample.

3.2. Kruskal–Wallis Rank Sum Test

The Kruskal–Wallis rank sum test or Kruskal–Wallis H test is a non-parametric test to check whether one or more groups in the tested data set originate from one distribution or not. Hollander and Wolfe (1973) state that the “[n]ull hypothesis of Kruskal–Wallis rank sum test is that the location parameters of the distribution of the tested dataset are the same in each group. The alternative is that they differ in at least one.”

The test is performed by first ranking all the values in the data set in ascending order. The data used in this study had no ties, and thus the following formula for H statistic was used:

$$H = \frac{12}{N_T(N_T+1)} \sum_{c=1}^4 n_i \left(\bar{r}_i - \frac{N_T+1}{2} \right)^2, \quad (18)$$

where N_T is the total number of observations in the data set (948 in the first sample and 540 in the second sample in case of volatility metrics, and 20,604 in the first sample and 11,736 in the second sample in the case of daily returns); n_i is the number of observations

in a group i (237 and 135 in the first and the second sample of volatility metrics, 5151 and 2934 in the case of daily returns); \bar{r}_i is the rank average of group i ; and H is the statistic used to compare distributions in the Kruskal–Wallis rank sum test.

The p value of the H-statistic is then approximated with chi-squared distribution with three degrees of freedom. In this study, if the p value of the H-statistic is lower than 0.05, we conclude that the null hypothesis is rejected.

3.3. Pairwise Wilcoxon Rank Sum Test

Usually, in research similar to this study, different kinds of panel regressions are performed. However, in the case of our data, we failed to pass tests for all the assumptions needed for the OLS panel regression, like cross-sectional independence, linearity, normality of the series and residuals, homoscedasticity, and stationarity. Moreover, in this study we aimed to divide countries into four categories based on the 2020 GDP per capita level, and this level of development remained the same in the whole period. This deemed such panel regression methods as least squares dummy variable (LSDV) and random and fixed effects unusable, as they all need independent variables that change over time. The Kruskal–Wallis rank sum test is superior to those methods because it does not have strict requirements for the data used, as it ranks values from smaller to larger, and based on the ranks, conclusions are drawn about whether the distributions are significantly different. A similar thing can be said about the pairwise Wilcoxon rank sum test; this was the reason we selected these tests to refer to our main hypotheses and questions.

The Wilcoxon rank sum test is a non-parametric test used to determine whether two groups originate from one distribution or not. The test is performed pairwise for four groups in total, creating six possible pairs. Values in each pair are ranked and a U statistic for each pair is calculated. The U statistic is calculated using the following formula:

$$U = \min\{r_1, r_2\} - k; \quad (19)$$

$$k = \frac{n(n+1)}{2}, \quad (20)$$

where n is the same as n_i in the Kruskal–Wallis test, a number of observations in a group i , (237 and 135 in the first and the second sample of volatility metrics, 5151 and 2934 in case of daily returns); r_1 and r_2 are the rank sums in a group one and group two in a pair, calculated separately for each pair; k is calculated based on the sample and equals to 237.5 and 135.5 in the first and the second sample, respectively; and U is the statistic based on which the difference of distributions is found.

The p values of the U statistic are then adjusted using the Holm–Bonferroni method. The p values are adjusted to control the family-wise error rate. To adjust the p values, they are first ordered in ascending order, and then corresponding to each p value, a null hypothesis is calculated in the following way:

$$H_i = \frac{0.05}{5-i}, \quad (21)$$

where i ranges from 1 to 4; H_i is the null hypothesis number i of the adjustment method and is compared to the p values in ascending order.

Then p values are compared to the corresponding null hypothesis (i.e., P_1 to H_1). Null is rejected if the p value is greater than the corresponding H value.

After the null hypothesis is rejected, all subsequent p values are equal to the adjusted p value of the first rejection. p values are adjusted in the following way: where P_i is the p value number i of p values of the pairwise Wilcoxon test ordered in ascending order.

$$P. adj_i = P_i \times i. \quad (22)$$

As an example, four p values equal to 0.001, 0.005, 0.3, and 0.5 will be adjusted to 0.004, 0.015, 0.6, and 0.6.

3.4. Data Description

The research is based on two data sets containing net income indices for various countries. We have selected the countries for our research in order to have the longest possible data samples. In reality, we had to take into account two contradictory requirements: the highest possible number of countries in our full sample and the longest common data history for all the series selected. In order to accomplish this task, we decided to create two data samples:

- the first one with a longer historical time series but covering fewer countries (51 countries in the period from 31 May 2002 to 28 February 2022)
- the second one with shorter historical time series but covering more countries (75 countries in the period from 30 November 2010 to 28 February 2022)

The data frequency is daily. The data for each country equity index were cleaned and corrected for any significant outliers. IMI indices were chosen because they include large, medium, and small-cap stocks for each country. Inclusion of medium and small capitalization equity is necessary for studying

Table 5. Average Summary Statistics of Equity Indices Daily Returns

	GDP	R.A	Mean	Min	Median	Max	ASD	Skewness	Kurtosis
First sample: 51 Countries, 2002–2022									
Frontier	3334	9.24%	0.04%	-13.2%	0.05%	13.0%	22.2%	-0.72	18.2
Emerging	6852	9.73%	0.04%	-13.4%	0.06%	13.6%	29.1%	-0.25	12.6
Early_developed	18122	5.43%	0.04%	-14.9%	0.05%	13.1%	24.2%	-0.52	19.3
Developed	50603	7.69%	0.04%	-15.7%	0.04%	14.3%	22.7%	-0.24	12.4
Second sample: 75 Countries, 2010–2022									
Frontier	2804	0.52%	0.03%	-13.8%	0.03%	9.4%	21.7%	-1.77	45.4
Emerging	6921	0.52%	0.02%	-14.4%	0.02%	9.9%	24.8%	-1.20	38.4
Early_developed	17858	2.13%	0.02%	-15.1%	0.02%	10.8%	20.4%	-0.82	22.2
Developed	50603	6.80%	0.02%	-17.5%	0.01%	10.4%	19.7%	-0.54	12.3

Note. Summary statistics use only the baseline income categorization outcome. A full list of countries in the baseline outcome can be seen in Table 2.

Table 6. Summary Statistics for Metrics of the First Sample

	Frontier	Emerging	Early-developed	Developed	Frontier	Emerging	Early-developed	Developed
Min	R.D				R.M			
	-9.1	-12.5	-9.6	-10.6	-26.6	-31.8	-27.3	-24.7
Mean	0.0	0.1	0.0	0.0	1.0	1.2	0.7	0.8
Median	0.1	0.1	0.1	0.1	1.3	1.6	0.9	1.2
Max	5.6	9.7	9.0	7.6	15.7	20.6	17.2	15.4
Min	STD				DSTD			
	9.3	9.6	11.0	8.7	3.8	3.9	5.1	4.9
Mean	18.9	25.2	20.4	19.2	12.2	15.2	12.5	11.7
Median	17.2	22.6	17.8	16.5	10.9	13.7	10.7	9.9
Max	69.7	102.4	91.4	87.3	56.3	69.1	57.6	52.8
Min	VaR.H				VaR.N			
	0.7	0.6	0.9	0.9	1.1	0.9	1.2	1.0
Mean	2.1	2.8	2.3	2.1	2.3	3.1	2.5	2.3
Median	1.9	2.5	1.9	1.8	2.0	2.8	2.1	1.9
Max	10.1	12.9	10.6	9.8	9.7	14.1	12.5	11.9
Min	ES.H				ES.N			
	0.9	0.7	1.0	1.0	1.4	1.7	1.7	1.3
Mean	2.5	3.3	2.7	2.5	2.8	3.8	3.0	2.9
Median	2.3	2.9	2.2	2.1	2.6	3.4	2.7	2.5
Max	11.9	13.4	12.4	11.9	9.1	13.6	12.3	11.8
Min	MDD				UI			
	1.2	1.3	1.5	1.4	0.7	0.8	0.9	0.7
Mean	5.7	7.4	6.1	5.6	3.3	4.3	3.5	3.2
Median	4.7	5.9	5.0	4.4	2.6	3.3	2.8	2.4
Max	33.9	44.5	38.7	35.2	21.4	29.8	24.1	22.9

Note. R.D, daily return; R.M., monthly return; STD, annualized standard deviation; DSTD, annualized downside semi-deviation; VaR.H, historical 97.5% value-at-risk; VaR.N, Gaussian 97.5% value-at-risk; ES.H, historical 97.5% expected shortfall; ES.N, Gaussian 97.5% expected shortfall. For the detailed description of income categories, see Table 2.

financial markets in the earlier stages of development. Main summary statistics for both samples can be found in Table 5.

GDP, average GDP per capita in current USD in 2020 of countries inside a category; R.A, average annualized return of country equity indices inside a category; Mean, average of mean daily returns of country equity indices; Min, average of minimums of equity indices' daily returns; Median, average of median of equity indices' daily returns; Max, average of maximums of equity indices' daily returns; ASD,

average annualized STD of daily returns of equity indices; Skewness, average skewness of equity indices' daily returns; Kurtosis, average kurtosis of equity indices' daily returns. All the summary statistics are applied for the whole sample period. First, the summary statistics for each country equity index are calculated; then the average of summary statistics inside a category are calculated and shown in the table.

GDP per capita was taken from World Bank database of World Development Indexes. There is an exception: Taiwan. The data for GDP per capita was

Table 7. Summary Statistics for Metrics of the Second Sample

	Frontier	Emerging	Early-developed	Developed	Frontier	Emerging	Early-developed	Developed
Min	R.D				R.M			
	-6.4	-8.1	-9.3	-10.6	-22.9	-23.0	-18.4	-16.3
Mean	0.0	0.0	0.0	0.0	0.5	0.3	0.4	0.7
Median	0.1	0.0	0.0	0.1	0.7	0.9	0.3	0.9
Max	2.1	4.0	5.0	7.6	12.6	14.0	13.6	15.4
Min	STD				DSTD			
	11.1	13.8	10.9	8.7	5.7	7.3	5.6	5.1
Mean	17.4	21.3	17.5	17.1	11.4	13.3	11.0	10.7
Median	16.0	19.8	15.9	15.4	10.2	12.3	9.3	9.5
Max	62.3	74.4	69.7	72.9	52.1	53.3	55.9	52.8
Min	VaR.H				VaR.N			
	1.0	1.3	0.9	0.9	1.2	1.5	1.2	1.0
Mean	2.0	2.4	2.0	1.9	2.1	2.6	2.1	2.1
Median	1.8	2.2	1.7	1.7	1.9	2.4	1.9	1.8
Max	9.1	9.9	10.2	9.6	8.8	10.3	9.4	9.7
Min	ES.H				ES.N			
	1.1	1.5	1.2	1.0	1.8	2.1	1.6	1.3
Mean	2.4	2.9	2.3	2.3	2.6	3.2	2.6	2.6
Median	2.1	2.6	2.1	2.0	2.4	3.0	2.4	2.3
Max	10.9	10.7	11.9	11.9	8.1	9.9	9.4	10.0
Min	MDD				UI			
	2.2	2.3	1.7	1.5	1.3	1.3	0.9	0.8
Mean	5.5	6.3	5.3	4.9	3.2	3.6	3.1	2.8
Median	4.7	5.4	4.5	4.1	2.8	3.2	2.5	2.2
Max	30.0	32.6	30.7	31.3	20.1	22.6	21.5	21.0

Note. R.D, daily return; R.M., monthly return; STD, annualized standard deviation; DSTD, annualized downside semi-deviation; VaR.H, historical 97.5% value-at-risk; VaR.N, Gaussian 97.5% value-at-risk; ES.H, historical 97.5% expected shortfall; ES.N, Gaussian 97.5% expected shortfall. For the detailed description of income categories, see Table 2.

taken from the IMF's *World Economic Outlook, October 2021* (International Monetary Fund, 2021). Countries are divided into four groups based on the GDP per capita levels measured in current USD in 2020: frontier, emerging, early-developed, developed.

In sensitivity analysis, we changed the sample periods to May 2007 to February 2022 in the first data set and November 2015– to February 2022 in the second data set. We also slightly changed the development categorization in order to check the

robustness of our results to initial assumption and categorization.

Summary statistics (not averaged for all the countries in the given group) of daily and monthly returns as well as the volatility metrics of the first and second sample are analyzed in Table 6 and Table 7.

Each risk measure (VaR, ES, SD, DSTD, UI, MDDs) was calculated based on the data with daily frequency each month. Thus, the first sample of 51 countries has 237 monthly observations of risk measures for

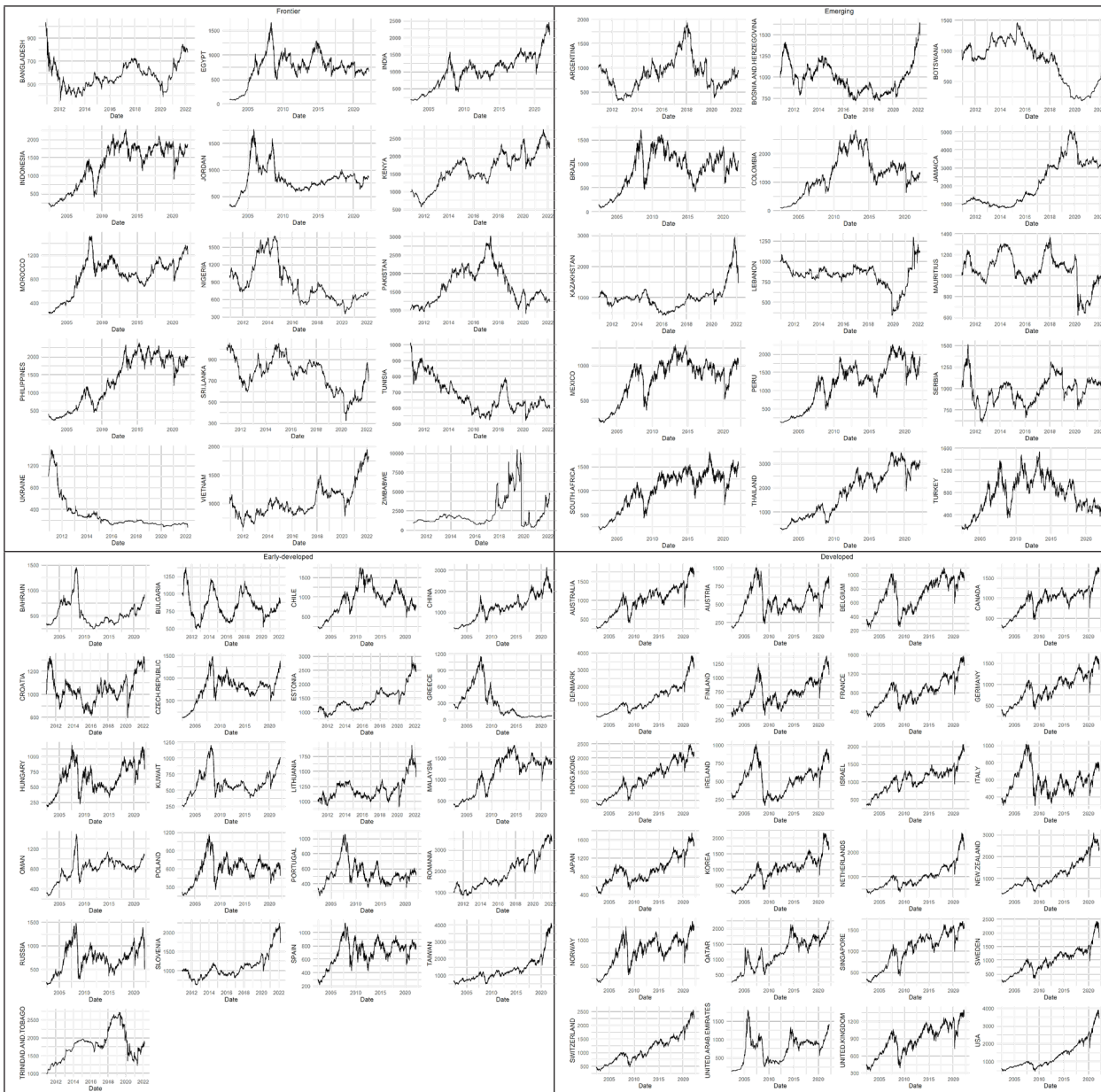


Figure 1. Daily Fluctuations of the Country Indices

Note. This four-panel figure shows the fluctuations of the prices of equity indices divided into four groups described based on baseline classification and presented in Table 2. For each country, we present the longest time period that was selected for this research.

each country, making, in total, 12,087 observations. The second sample of 75 countries has 135 monthly observations of risk measures for each country, making, in total, 10,125 observations. Figure 1 depicts the daily fluctuations of the country indices, showing the direction of the main trend during the research period and the magnitude of drawdowns encountered during the Great Financial Crisis (2007–2009) and the COVID 2020 crisis.

Additionally, the value of the MDD divided into two sample periods is visualized in Table 8. These numbers show very significant MDDs in the analyzed period, and at the same time, they show that the most severe drawdown was connected with the Great Financial Crisis and that only for some of the countries (indicated in red) the COVID 2020 crisis was connected with larger turmoil.

The results in Table 8 show quite an important difference between the drawdowns of each country's

Table 8. MDD for All Countries Under Investigation from Baseline Classification

Country	Class	First (%)	Second (%)	Country	Class	First (%)	Second (%)
ARGENTINA	Emerging		80.6	LITHUANIA	Early-developed		33.5
AUSTRALIA	Developed	66.8	46.5	MALAYSIA	Early-developed	51.9	47.0
AUSTRIA	Developed	76.8	59.3	MAURITIUS	Emerging		54.7
BAHRAIN	Early-developed	83.3	34.4	MEXICO	Emerging	64.8	60.2
BANGLADESH	Frontier		64.9	MOROCCO	Frontier	55.8	44.3
BELGIUM	Developed	75.1	47.0	NETHERLANDS	Developed	64.2	34.7
BOSNIA AND HERZEGOVINA	Emerging		48.9	NEW ZEALAND	Developed	65.7	36.5
BOTSWANA	Emerging		87.3	NIGERIA	Frontier		79.2
BRAZIL	Emerging	75.7	74.2	NORWAY	Developed	75.2	53.2
BULGARIA	Early-developed		64.2	OMAN	Early-developed	66.2	36.5
CANADA	Developed	61.5	43.3	PAKISTAN	Frontier		70.2
CHILE	Early-developed	72.7	72.7	PERU	Emerging	67.9	57.8
CHINA	Early-developed	73.2	41.7	PHILIPPINES	Frontier	61.7	49.9
COLOMBIA	Emerging	77.6	77.6	POLAND	Early-developed	78.2	59.9
CROATIA	Early-developed		40.3	PORTUGAL	Early-developed	67.4	50.4
CZECH REPUBLIC	Early-developed	67.2	61.1	QATAR	Developed	64.2	43.9
DENMARK	Developed	64.2	34.7	ROMANIA	Early-developed		44.3
EGYPT	Frontier	70.3	58.2	RUSSIA	Early-developed	79.8	66.5
ESTONIA	Early-developed		34.5	SERBIA	Emerging		60..2
FINLAND	Developed	73.4	47.4	SINGAPORE	Developed	64.3	39.5
FRANCE	Developed	60.7	39.8	SLOVENIA	Early-developed		40.1
GERMANY	Developed	62.9	46.4	SOUTH AFRICA	Emerging	63.2	60.8
GREECE	Early-developed	97.4	91.1	SPAIN	Early-developed	62.7	51.9
HONG KONG	Developed	64.4	32.5	SRI LANKA	Frontier		69.3
HUNGARY	Early-developed	81.3	62.5	SWEDEN	Developed	68.3	38.2
INDIA	Frontier	74.3	46.2	SWITZERLAND	Developed	52.7	26.6

Continued **Table 8.** MDD for All Countries Under Investigation from Baseline Classification

Country	Class	First (%)	Second (%)	Country	Class	First (%)	Second (%)
INDONESIA	Frontier	72.1	60.5	TAIWAN	Early-developed	60.6	30.8
IRELAND	Developed	83.5	41.0	THAILAND	Emerging	62.3	46.5
ISRAEL	Developed	41.4	39.4	TRINIDAD AND TOBAGO	Early-developed		55.0
ITALY	Developed	70.7	50.5	TUNISIA	Frontier		48.6
JAMAICA	Emerging		45.5	TURKEY	Emerging	75,6%	75.6
JAPAN	Developed	53.3	31.2	UKRAINE	Frontier		95.1
JORDAN	Frontier	67.0	37.2	UNITED ARAB EMIRATES	Developed	86.1	58.1
KAZAKHSTAN	Emerging		66.6	UNITED KINGDOM	Developed	63.7	43.2
KENYA	Frontier		47.2	USA	Developed	55.7	35.0
KOREA	Developed	72.1	50.0	VIETNAM	Frontier		49.9
KUWAIT	Early-developed	68.5	44.6	ZIMBABWE	Frontier		96.5
LEBANON	Emerging		68.7				

Note. The first drawdown is the MD from the first sample, and the second one is the MD from the second sample. Red numbers indicate that the higher MD was in the second sample, which is covered by the first sample.

equity index. Additionally, we know that some countries dominated in some periods, for example, the US, from 2009 until 2022. However, taking into account that we aggregated results in each group using the equal weighting schemes, the issue of excluding or including any does not affect final results significantly because the portfolio of any group of countries is not market-cap-weighted.

4. Empirical Results and Discussion

RHI: Do daily and monthly return distributions of country equity indices differ with regard to the level of economic development?

In the verification of this hypothesis, the Kruskal–Wallis rank-sum test was used to find whether daily and monthly returns distributions are different between income categories. This particular test was used here, as distributions of daily and monthly returns have extreme values and are not normally

distributed, as can be seen in Figure 2 and Figure 3 for 51 and 75 country indices samples, respectively.

The null hypothesis of the Kruskal–Wallis test is the equality of the location parameters of the distribution. The p value of the Kruskal–Wallis test of daily returns for the 51-country sample is equal to 0.201, and for the 75-country sample, it is 0.524, and thus the null hypothesis is not rejected in both samples, and there exists no such pair that have location parameters that are significantly different; see Table 9.

The p value of the Kruskal–Wallis test of monthly returns for the 51-country sample is equal to 0.767, and for the 75-country sample it is 0.91, and thus the null hypothesis is not rejected in both samples, and there exists no such pair that have location parameters that are significantly different.

Null is rejected if the p value is lower than the 0.05 significance level. If the null in a particular pair is rejected, then that pair has significant differences from each other.

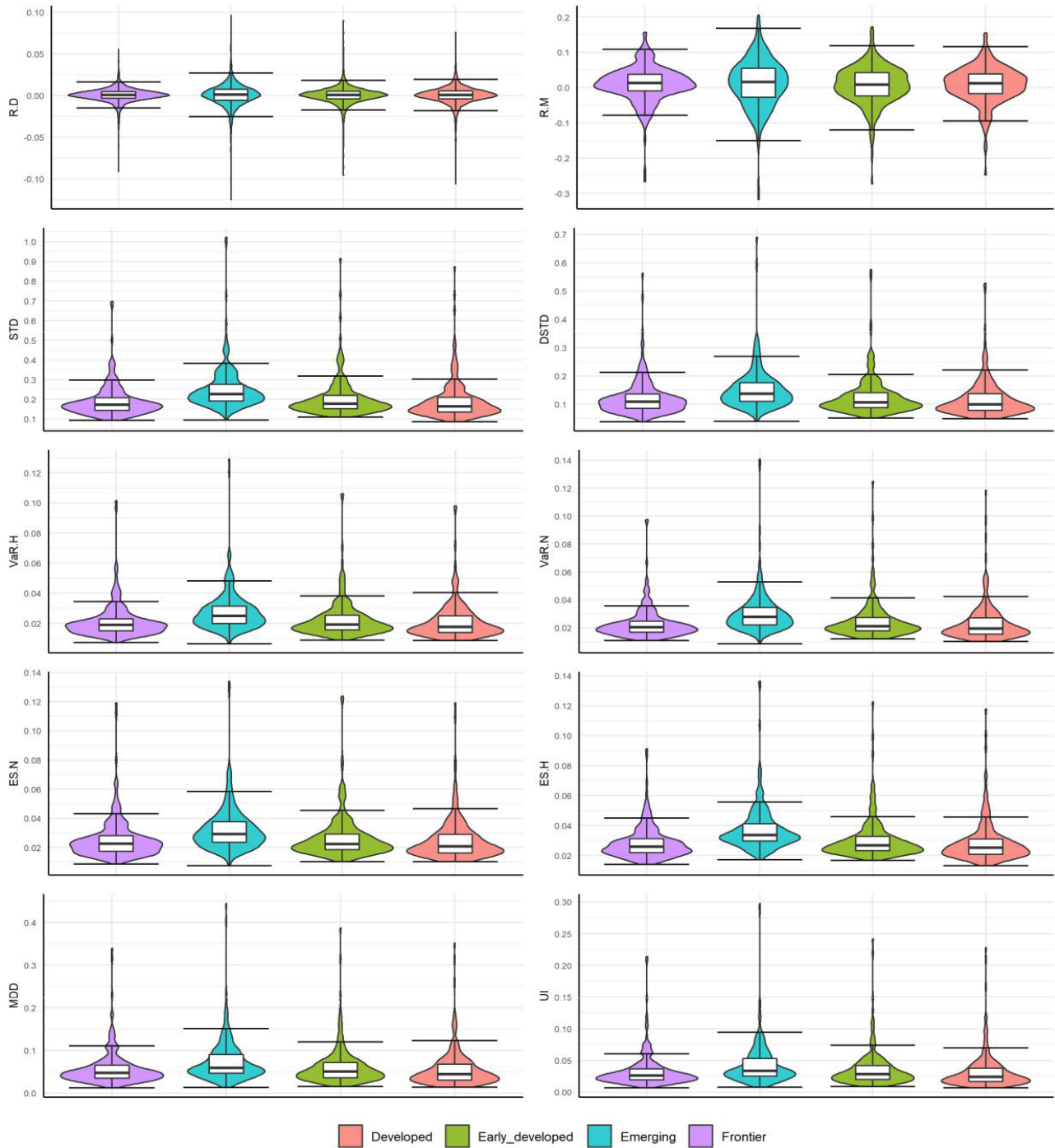


Figure 2. First Sample Distributions

Note. R.D, daily return of four groups; R.M, monthly return; STD, annualized standard deviation; DSTD, annualized downside semi-deviation; VaR.H, historical 97.5% value-at-risk; VaR.N, Gaussian 97.5% value-at-risk; E.S.H, historical 97.5% expected shortfall; E.S.N, Gaussian 97.5% expected shortfall.

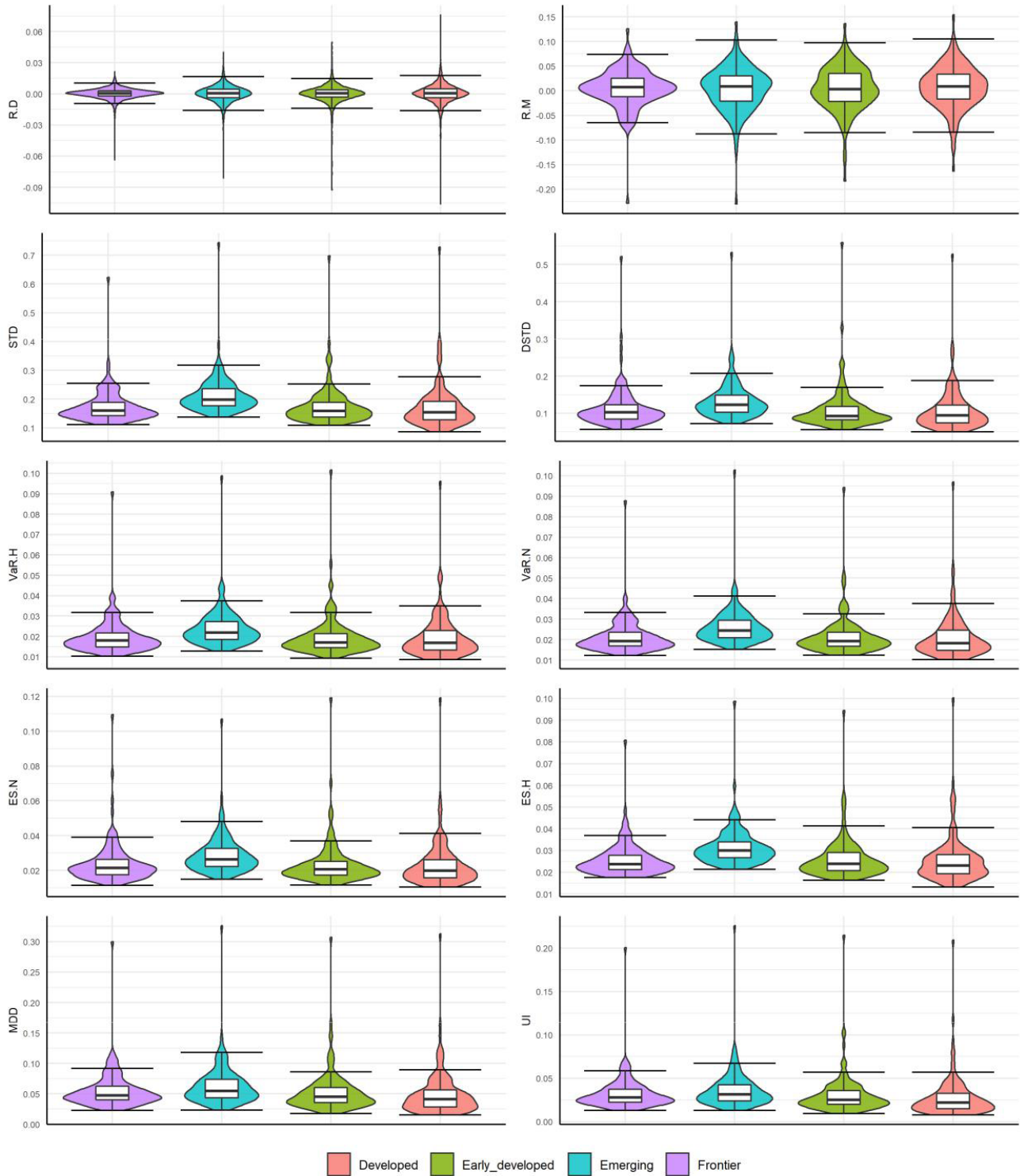


Figure 3. Second Sample Distributions

Note. R.D, daily return of four groups; R.M, monthly return; STD, annualized standard deviation; DSTD, annualized downside semi-deviation; VaR.H, historical 97.5% value-at-risk; VaR.N, Gaussian 97.5% value-at-risk; ES.H, historical 97.5% expected shortfall; ES.N, Gaussian 97.5% expected shortfall.

Table 9. Results of Baseline Income Categorization

	R.D	R.M	STD	DSTD	VaR.H	VaR.N	ES.H	ES.N	MDD	UI
Kruskal–Wallis rank sum test p values of H statistic:										
First sample	0.201	0.767	0	0	0	0	0	0	0	0
Pairwise Wilcoxon rank sum test p values of U statistic										
F_E	-	-	0	0	0	0	0	0	0	0
F_ED	-	-	0.077	0.479	0.306	0.077	0.515	0.096	0.265	0.185
F_D	-	-	0.34	0.222	0.453	0.341	0.515	0.365	0.265	0.185
E_ED	-	-	0	0	0	0	0	0	0	0
E_D	-	-	0	0	0	0	0	0	0	0
ED_D	-	-	0.008	0.044	0.078	0.016	0.06	0.009	0.027	0.009
Kruskal–Wallis rank sum test p values of H statistic										
Second sample	0.524	0.91	0	0	0	0	0	0	0	0
Pairwise Wilcoxon rank sum test p values of U statistic										
F_E	-	-	0	0	0	0	0	0	0.023	0.075
F_ED	-	-	0.761	0.417	0.513	0.804	0.463	0.729	0.097	0.075
F_D	-	-	0.272	0.086	0.273	0.159	0.146	0.457	0.003	0
E_ED	-	-	0	0	0	0	0	0	0	0
E_D	-	-	0	0	0	0	0	0	0	0
ED_D	-	-	0.282	0.417	0.513	0.186	0.463	0.501	0.089	0.033

Note.: R.D, daily return of the four groups; R.M, monthly return; STD, annualized standard deviation; DSTD, annualized downside semi-deviation; VaR.H, historical 97.5% value-at-risk; VaR.N, Gaussian 97.5% value-at-risk; ES.H, historical 97.5% expected shortfall; ES.N, Gaussian 97.5% expected shortfall. F_E, frontier and emerging market pair; F_ED, frontier and early-developed market pair; F_D, frontier and developed market pair; E_ED, emerging and early-developed market pair; E_D, emerging and developed market pair; ED_D, early-developed and developed market pair.

RH2: Do distributions of volatility metrics of country equity indices differ with regard to the level of economic development?

Volatility of returns was measured using six volatility metrics: annualized STD, annualized DSTD, UI, MDD, 97.5% value-at-risk (VaR.N, VaR.H), 97.5% expected shortfall (ES.N, ES.H). To test this hypothesis, the Kruskal–Wallis rank-sum test was used due to the non-normal distribution of volatility metrics and their extreme values. After which, if the null of the Kruskal–Wallis test is rejected, the pairwise Wilcoxon rank sum test is used to find which pairs caused the rejection of the null. Volatility metrics and their distributions are illustrated in Figure 2 for the sample of 51 country indices in the period from June 2002 to February 2022.

We can expect, based on the graphical analysis, that location parameters of STD of frontier, early-developed, and developed markets are similar, while location parameters differ from other markets in the case of emerging markets.

In the case of the STD in the first sample with 51 country indices in the period of 2002–2022, the null hypothesis of Kruskal–Wallis is rejected, and there exists a pair that has location parameters that are significantly different. To check which pair it is, the pairwise Wilcoxon test is used, which has its p values adjusted by the Holm–Bonferroni method. Emerging markets are significantly different from other markets. Frontier and developed markets are not significantly different. Frontier and early-developed markets are not significantly different, while early-developed and developed markets are significantly different.

The Kruskal–Wallis test rejected the null of the equality of the location parameters of the distribution of DSTD, historical and Gaussian VaR, historical and Gaussian ES in the first sample. According to the results of the pairwise Wilcoxon test, in the first sample, in the case of VaR.H and ES.H, emerging markets are significantly different from other markets, while other markets are not.

In the case of DSTD, VaR.N, ES.N, MDD, and UI, emerging markets are significantly different from other markets, and frontier markets are not significantly different from other markets (besides the emerging ones); however, the null of equality of location parameters is rejected for developed and early-developed markets. To see the exact results of the Kruskal–Wallis and pairwise Wilcoxon tests, see Table 9.

The distribution of volatility metrics and returns of second sample are shown in Figure 3. Based on the graphical analysis, we find the second sample is similar to the first sample and expect emerging markets to be different from other markets; however, it is unclear about the similarity of other income categories.

Similar to the first sample, the annualized STD in the case of the 75-country indices in the period from December 2010 to February 2022, emerging markets differ from all other markets; however, in a difference from the first sample, the pairwise Wilcoxon test did not reject the null of equality location parameters for pairs of frontier, early-developed, and developed markets.

Regarding DSTD, VaR(H and N), and ES(H and N) in the second sample, emerging markets are significantly different from other markets, while other markets have pairwise equal location parameters.

In the second sample, the results of the pairwise Wilcoxon test for MDD are similar in the case of emerging markets, but they are significantly different from all other markets. Early-developed markets are not significantly different from frontier and developed markets, while the frontier and developed market pair has significantly different location parameters.

In the case of the UI in the second sample, the picture is quite different. Developed markets do not have significantly equal location parameters as other markets. Frontier markets are similar to both emerging and early-developed markets, and early-developed markets are different from emerging markets.

4.1. Robustness Tests

Two robustness tests are used to check the quality of the results. In the first test, we change the time frames of the sample five years forward; therefore, such a first new sample is in the period from 2007 to 2022, while the second sample is in the period from 2015 to 2022. In the second robustness test, we use four other possible ways of grouping countries based on the GDP per capita level.

RQ1: Is the result obtained robust to the change in time period used?

Time period change did not affect the results of the Kruskal–Wallis test but did affect the results of the pairwise Wilcoxon test. In the case of the pairwise Wilcoxon test, the results for emerging market pairs stayed the same in both samples. However, the results were changed for almost all risk metrics for other pairs, which can be seen in Table 10.

RQ2: Is the result obtained robust to the change in the income categories of countries?

In this robustness test, four other possible income categorizations according to our method are used to compare with the results of the baseline income category. The exact income category of each country is shown in Table 2.

In the second and third versions of income categorization, the null of the Kruskal–Wallis test is not rejected for all volatility metrics as well as for daily and monthly returns, which can be seen in Table 11.

In the fourth version of income categorization, the null of the Kruskal–Wallis test was rejected in STD, VaR.N, and ES.N in the 51-country sample. In this first sample, the null hypothesis of pairwise Wilcoxon test was rejected in the frontier-emerging pair for STD, VaR.N, and ES.N and for the emerging and early-developed pair for STD and ES.N. Thus, there is evidence that frontier and emerging markets do not originate from the same distribution, and similarly, emerging and early-developed markets have different distributions of STD and ES.N. The Kruskal–Wallis test rejected the null of equal location parameters for STD and ES.N in the 75-country sample, and the pairwise Wilcoxon test for these metrics determined that frontier markets are stochastically different from developed markets.

Table 10. Results of the Time Period Change

	R.D	R.M	STD	DSTD	VaR.H	VaR.N	ES.H	ES.N	MDD	UI
Kruskal–Wallis rank sum test p values of H-statistic:										
First sample	0.695	0.945	0	0	0	0	0	0	0	0
Pairwise Wilcoxon rank sum test p values of U statistic:										
F_E	-	-	0	0	0	0	0	0	0	0
F_ED	-	-	0	0.096	0.015	0.001	0.034	0	0.017	0.015
F_D	-	-	0.292	0.966	0.534	0.451	0.725	0.236	0.83	0.588
E_ED	-	-	0	0	0	0	0	0	0.006	0.015
E_D	-	-	0	0	0	0	0	0	0	0
ED_D	-	-	0.021	0.096	0.083	0.02	0.077	0.027	0.02	0.011
Kruskal–Wallis rank sum test p values of H-statistic:										
Second sample	0.833	0.995	0	0	0	0	0	0	0	0
Pairwise Wilcoxon rank sum test p values of U statistic:										
F_E	-	-	0	0	0	0	0	0	0.018	0.093
F_ED	-	-	0.287	0.224	0.548	0.362	0.447	0.189	0.17	0.093
F_D	-	-	0.049	0.098	0.299	0.072	0.156	0.03	0.018	0.002
E_ED	-	-	0	0	0	0	0	0	0	0.001
E_D	-	-	0	0	0	0	0	0	0	0
ED_D	-	-	0.287	0.437	0.548	0.362	0.447	0.203	0.17	0.093

Note. R.D, daily return of the four groups; R.M, monthly return; STD, annualized standard deviation; DSTD, annualized downside semi-deviation; VaR.H, historical 97.5% value-at-risk; VaR.N, Gaussian 97.5% value-at-risk; ES.H, historical 97.5% expected shortfall; ES.N, Gaussian 97.5% expected shortfall; F_E, frontier and emerging market pair; F_ED, frontier and early-developed market pair; F_D, frontier and developed market pair; E_ED, emerging and early-developed market pair; E_D, emerging and developed market pair; ED_D, early-developed and developed market pair. Null is rejected if p value is lower than 0.05 significance level. If the null in a particular pair is rejected, then that pair has significant differences from each other.

Table 11. p Values of the Kruskal–Wallis Test for All Income Categorization Outcomes

	R.D	R.M	STD	DSTD	VaR.H	VaR.N	ES.H	ES.N	MDD	UI
First sample: 51 Countries, 2002–2022										
Version 2	0.279	0.819	0.153	0.628	0.419	0.248	0.548	0.151	0.32	0.594
Version 3	0.428	0.833	0.121	0.805	0.432	0.181	0.543	0.138	0.277	0.468
Version 4	0.231	0.81	0.014	0.53	0.14	0.027	0.305	0.014	0.095	0.24
Version 5	0.358	0.819	0.01	0.697	0.138	0.018	0.293	0.012	0.081	0.177
Second sample: 75 Countries, 2010–2022										
Version 2	0.484	0.867	0.145	0.275	0.233	0.191	0.238	0.161	0.156	0.075
Version 3	0.583	0.843	0.097	0.127	0.164	0.137	0.156	0.121	0.111	0.051
Version 4	0.507	0.819	0.022	0.11	0.067	0.051	0.082	0.02	0.289	0.222
Version 5	0.622	0.85	0.03	0.225	0.095	0.067	0.117	0.026	0.384	0.302

Note. R.D, daily return of the four groups; R.M, monthly return; STD, annualized standard deviation; DSTD, annualized downside semi-deviation; VaR.H, historical 97.5% value-at-risk; VaR.N, Gaussian 97.5% value-at-risk; ES.H, historical 97.5% expected shortfall; ES.N, Gaussian 97.5% expected shortfall. The null is rejected if the p value is lower than the 0.05 significance level. For the detailed description of income categories, see Table 2.

Table 12. Pairwise Wilcoxon Test p Values for Income Versions 4 and 5

	STD	VaR.N	ES.N	STD	VaR.N	ES.N
	Version 4			Version 5		
First sample: 51 Countries, 2002–2022						
F_E	0.034	0.037	0.044	0.031	0.037	0.044
F_ED	0.857	1	1	0.507	0.357	0.571
F_D	0.857	0.669	1	0.73	0.57	0.885
E_ED	0.034	0.097	0.023	0.507	0.555	0.571
E_D	0.151	0.315	0.14	0.020	0.056	0.014
ED_D	0.857	1	1	0.507	0.487	0.571
Second sample: 75 Countries, 2010–2022						
F_E	0.328	-	0.253	0.328	-	0.253
F_ED	0.716	-	0.803	0.707	-	0.868
F_D	0.033	-	0.034	0.031	-	0.028
E_ED	0.716	-	0.554	0.707	-	0.868
E_D	0.716	-	0.803	0.707	-	0.868
ED_D	0.093	-	0.094	0.223	-	0.205

Note. STD, annualized standard deviation; VaR.N, Gaussian 97.5% value-at-risk; ES.N, Gaussian 97.5% expected shortfall; F_E, frontier and emerging market pair; F_ED, frontier and early-developed market pair; F_D, frontier and developed market pair; E_ED, emerging and early-developed market pair; E_D, emerging and developed market pair; ED_D, early-developed and developed market pair. The null is rejected if the p value is lower than 0.05 significance level. If the null in a particular pair is rejected, then that pair has significant differences from each other. For the detailed description of income categories, see Table 2.

In the fifth version of income categorization, the results of the Kruskal–Wallis test are the same as in the fourth version; however, pairwise Wilcoxon is different in the first sample. Instead of rejecting the null in the emerging and early-developed pair, the null is rejected in the emerging and developed pair. Thus, here the emerging and developed markets have different distributions of STD and ES.N, which can be seen in Table 12.

5. Conclusions

This paper aimed to find evidence of differences among markets of four income categories: frontier, emerging, early-developed, and developed. The main hypotheses are: RH1, whether daily and monthly return of country equity indices differ with regard to the level of economic development; and RH2, whether the volatility metrics of country equity indices differ with regard to the level of economic development.

Based on these hypotheses, the following research questions need to be answered: whether the results are robust to: RQ1, the change in time period used; and RQ2, the change in the income categories of countries.

The data set used in this study consists of MSCI IMI indices of 75 countries. The data set was divided into two samples based on the availability of data: the 51-country sample of daily values of MSCI IMI indices over the period from 31 May 2002 to 28 February 2022 and the 75-country sample of daily values of MSCI IMI indices over the period from 30 November 2010 to 28 February 2022. Additionally, GDP per capita in current USD in 2020 taken from the World Bank database was used, while GDP per capita of Taiwan in 2020 was taken from the projection of the IMF in the *World Economic Outlook, October 2021* (International Monetary Fund, 2021).

Countries were categorized into four income levels based on the GDP per capita in current USD in 2020: frontier, emerging, early-developed and developed.

Table 13. Reference to Research Hypotheses and Questions

RH/RQ	Verification	Details
RH1	Rejected	Daily and monthly return do not depend on the level of economic development
RH2	Not rejected	Volatility metrics depend on the level of economic development
RQ1	Rejected	The obtained result is sensitive to varying time periods
RQ2	Rejected	The obtained result is sensitive to varying income categorizations

Note. A detailed description of each hypothesis and research question may be found in the Introduction.

Six volatility metrics were calculated in monthly subsections: annualized STD, annualized DSTD, UI, MDD, 97.5% value-at-risk (VaR.N, VaR.H), and 97.5% expected shortfall (ES.N, ES.H). The Kruskal–Wallis rank sum test was used to determine if there existed a difference among the four income categories in their daily and monthly returns and volatility metrics. Then, the pairwise Wilcoxon test was used to find which market pair was significantly different from the other.

Overall, the results of the Kruskal–Wallis and pairwise Wilcoxon tests based on the baseline income categorization (presented in Table 9) show that there are differences between markets depending on the level of economic development. There is no evidence that there are differences among markets in daily and monthly returns (RH1), while there is evidence of differences of volatility metrics of country equity indices depending on the level of economic development (RH2). However, the results are sensitive to the time period and income categorization method. Although time period change does not affect the results of the Kruskal–Wallis test, it slightly altered the results of the pairwise Wilcoxon test. Income categorization changes completely alter the results for volatility metrics in versions 2 and 3. There is still evidence of significant differences between markets in some volatility metrics in versions 4 and 5, but it seems that results depend mostly on the choice of frontier markets and less so on the categorization of other countries. We noticed this inconsistency within the existing literature, and it was one of the reasons that we decided to focus on this research. We think that our main contribution to the literature lies in revealing results contrary to the established literature and in adding to the discussion of this phenomenon. We have numerous results in the quantitative finance literature focused on asset allocation and portfolio management showing that the premise that higher volatility is closely connected with higher returns does

not necessarily work. Examples of such are the results of high-volatility portfolios versus low-volatility portfolios, and low-beta stocks versus high-beta stocks, which rather contradicts the information from Markowitz theory or the single-index Sharpe model and can be based on premises similar to our results.

To conclude, we can state that (RH1) there is no significant difference in daily and monthly return in the four markets, and (RH2) there exists a difference between the volatility metrics of equity indices depending on the level of economic development of countries. Additionally, the obtained results are (RQ1) somewhat sensitive to changes in time period and (RQ2) very sensitive to the categorization of country level of development. The results are summarized in Table 13.

Before we move to some extensions of this paper, it is important to indicate some policy investment implications of our results. Based on the presented research, we are able to refer to the established literature that acknowledged that emerging or frontier markets typically have a higher level of average returns and an accompanying higher level of volatility. Our paper shows that even when taking into account several different kinds of categorization, these characteristics do not describe analyzed markets in the proper way. First of all, the differences among average returns are not statistically significant (Figure 2, Figure 3, and Table 9), while in the case of the difference in the level of volatility measured based on six different volatility metrics (Table 9), our results are not consistent with the literature. The inconsistency is connected mainly with this: that we show that frontier markets regarded as highly volatile markets had significantly higher volatility than emerging markets, and at the same time, did not experience significantly different volatility than developed markets. This conclusion is confirmed for almost all volatility metrics under investigation. Additionally, the contribution to

knowledge coming from this research can have straightforward implications for asset allocation strategies in which researchers and practitioners too often assume, based on the existing literature, that the order of volatility level for the countries grouped based on their economic development from the highest to the lowest is: frontier, emerging, early-developed, and developed, while based on our results, this does not have to be necessarily true. Moreover, the benefits for investors could be quite significant for the investment decisions of individual and institutional investors. If many kinds of investment products prepared for investment banks used such unconfirmed or short-lived assumptions about the specific relation between the level of economic development and average return and volatility of the given countries, while, in reality, these assumptions are not valid, then the rationale for the existence of such investment products can be challenged. The simplest way to utilize the results of this paper in the real world is to release the assumption about the specific relation between the level of economic development and average return and volatility of the given countries and build portfolios without any unconfirmed relationships.

There are some limitations to this paper that can be developed in future work. First of all, the data available for some of the frontier and emerging markets is only for the period of the last eleven and one-half years. Second, although our methodology is able to show that there are differences among markets, we could not find how the markets are different from each other in particular. Third, our results only provide limited insight into the causes for these differences, though we suspect that the differences exist because of the liquidity differences in the markets. In future studies, liquidity-adjusted volatility metrics such as the liquidity-adjusted VaR proposed by Snoussi and El-Aroui (2012) and market liquidity-based categorization can be used to address this.

Acknowledgements

The article was prepared within the framework of a non-competitive project of the Ministry of Education “School of Eagles” based on the agreement between the Ministry of Education and the University of Warsaw, contract no. MNISW/2020/142/DIR/KH, signed on 26 May 2020. Project co-financed from the European Social Fund under the Operational Programme Knowledge Education Development 2014-2020.

References

- Ali, H. (2019). Does downside risk matter more in asset pricing? Evidence from China. *Emerging Markets Review*, 39, 154–174.
- Al-Tamimi, H. A. H., Alwan, A. A., & Abdel Rahman, A. (2011). Factors affecting stock prices in the UAE financial markets. *Journal of Transnational Management*, 16(1), 3–19.
- Amaresh, M., Anandasayanan, S., & Ramesh, S. (2020). Macro-economic variables and stock market performance: Empirical evidence from Colombo stock exchange. *Vidyodaya Journal of Humanities and Social Sciences*, 20(2), <http://dr.lib.sjp.ac.lk/handle/123456789/9147>
- Yartey, A. C. (2008). *The determinants of stock market development in emerging economies: Is South Africa different?* International Monetary Fund, African Dept.
- Amtiran, P., Indrastuti, R., Nidar, S., & Masyita, D. (2017). Macroeconomic factors and stock returns in APT framework. *International Journal of Economics and Management*, 11(S1), 197–206.
- Ang, A., Chen, J., & Xing, Y. (2006). Downside risk. *The Review of Financial Studies*, 19(4), 1191–1239.
- Atilgan, Y., & Demirtas, K. O. (2016). Risk-adjusted performances of world equity indices. *Emerging Markets Finance and Trade*, 52(3), 706–721.
- Cherif, M., & Gazdar, K. (2010). Institutional and macroeconomic determinants of stock market development in MENA region: New results from a panel data analysis. *International Journal of Banking and Finance*, 7(1), 139–159.
- Dellas, H., & Hess, M. (2005). Financial development and stock returns: A cross-country analysis. *Journal of International Money and Finance*, 24(6), 891–912.
- Dowd, K. (2005). *Measuring market risk* (2nd ed.). John Wiley & Sons.
- Estrada, J. (2007). Mean-semivariance behavior: Downside risk and capital asset pricing. *International Review of Economics & Finance*, 16(2), 169–185.
- Giri, A. K., & Joshi, P. (2017). The impact of macroeconomic indicators on indian stock prices: An empirical analysis. *Studies in Business & Economics*, 12(1), 61–78. <https://doi.org/10.1515/sbe-2017-0005>
- Ho, S.-Y., & Iyke, B. N. (2017). Determinants of stock market development: A review of the literature.

Studies in Economics and Finance, 34(1), 143–164. <https://doi.org/10.1108/SEF-05-2016-0111>

Hollander, M., & Wolfe, D. A. (1973). *Nonparametric statistical methods*. John Wiley & Sons.

International Monetary Fund (2021). *World economic outlook, October 2021: Recovery during a pandemic—health concerns, supply disruptions, price pressures*. International Monetary Fund. <https://www.imf.org/en/Publications/WEO/Issues/2021/10/12/world-economic-outlook-october-2021>

James, G. A., & Karoglou, M. (2010). Financial liberalization and stock market volatility: The case of Indonesia. *Applied Financial Economics*, 20(6), 477–486.

Kalam, K. K. (2020). The effects of macroeconomic variables on stock market returns: Evidence from Malaysia's stock market return performance. *Journal of World Business*, 55(8), 1–13.

Martin, P. G., & MacCann, B. B. (1989). *The investor's guide to fidelity funds*. John Wiley & Sons.

Montes, G. C., & Tiberto, B. P. (2012). Macroeconomic environment, country risk and stock market performance: Evidence for Brazil. *Economic Modelling*, 29(5), 1666–1678.

Osamwonyi, I. O., & Evbayiro-Osagie, E. I. (2012). The relationship between macroeconomic variables and stock market index in Nigeria. *Journal of Economics*, 3(1), 55–63.

Pradhan, R. P., Arvin, M. B., Hall, J. H., & Bahmani, S. (2014). Causal nexus between economic growth, banking sector development, stock market development, and other macroeconomic variables: The case of ASEAN countries. *Review of Financial Economics*, 23(4), 155–173.

Singh, T., Mehta, S., & Varsha, M. (2011). Macroeconomic factors and stock returns: Evidence from Taiwan. *Journal of Economics and International Finance*, 3(4), 217–227.

Snoussi, W., & El-Aroui, M.-A. (2012). Value-at-risk adjusted to the specificities of emerging markets: An analysis for the Tunisian market. *International Journal of Emerging Markets*, 7(1), 86–100. <https://doi.org/10.1108/17468801211197905>

Umutlu, M., Akdeniz, L., & Altay-Salih, A. (2010). The degree of financial liberalization and aggregated stock-return volatility in emerging markets. *Journal of Banking & Finance*, 34(3), 509–521.