“Stock returns are not always from the same distribution: Evidence from the Great Recession”

AUTHORS	Nektarios A. Michail Marina Magidou
ARTICLE INFO	Nektarios A. Michail and Marina Magidou (2020). Stock returns are not always from the same distribution: Evidence from the Great Recession. *Investment Management and Financial Innovations*, *17*(3), 189-204. doi:10.21511/imfi.17(3).2020.15
DOI	http://dx.doi.org/10.21511/imfi.17(3).2020.15
RELEASED ON	Monday, 21 September 2020
RECEIVED ON	Thursday, 28 November 2019
ACCEPTED ON	Thursday, 20 February 2020
LICENSE	This work is licensed under a Creative Commons Attribution 4.0 International License
JOURNAL	"Investment Management and Financial Innovations"
ISSN PRINT	1810-4967
ISSN ONLINE	1812-9358
PUBLISHER	LLC “Consulting Publishing Company “Business Perspectives”
FOUNDER	LLC “Consulting Publishing Company “Business Perspectives”

NUMBER OF REFERENCES **32**

NUMBER OF FIGURES **9**

NUMBER OF TABLES **4**

© The author(s) 2026. This publication is an open access article.

BUSINESS PERSPECTIVES

LLC "CPC "Business Perspectives"
Hryhorii Skovoroda lane, 10,
Sumy, 40022, Ukraine

www.businessperspectives.org

Nektarios A. Michail (Cyprus), **Marina Magidou** (Cyprus)

STOCK RETURNS ARE NOT ALWAYS FROM THE SAME DISTRIBUTION: EVIDENCE FROM THE GREAT RECESSION

Abstract

Portfolio allocation strategies, and notably the mean-variance approach, use past returns to assign optimal weights. Even though both past and expected returns should come from the same distribution, a formal test of whether this holds in practice has not been conducted yet. Thus, the study examines if the daily returns of 242 companies with continuous trading in the S&P index come from the same distribution using the Kolmogorov-Smirnov, Cramér-Von Mises, and Wilcoxon rank-sum tests. The tests suggest that generally stock returns do come from the same distribution. However, the hypothesis is rejected during the Great Recession, with the rejection rate increasing as the forecast horizon increased. The rejection rate, using an array of macroeconomic variables, is found to record high levels of persistence. Although macroeconomic variables were not found to be statistically significant determinants of the rejection rate, market distress has a small but significant effect.

Keywords distribution, Kolmogorov-Smirnov, Cramér-von Mises, Wilcoxon rank-sum, returns, portfolio allocation

JEL Classification G11, C14, C58

Received on: 28th of November, 2019
Accepted on: 20th of February, 2020
Published on: 21st of September, 2020

© Nektarios A. Michail, Marina Magidou, 2020

Nektarios A. Michail, Ph.D., Department of Commerce, Finance and Shipping, Cyprus University of Technology; Economic Analysis and Research, Central Bank of Cyprus, Cyprus.

Marina Magidou, Ph.D. Candidate, Department of Commerce, Finance and Shipping, Cyprus University of Technology, Cyprus. (Corresponding author)

This is an Open Access article, distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

Conflict of interest statement:
Author(s) reported no conflict of interest

INTRODUCTION

The exact specification of the stock market returns distribution has intrigued both academics and practitioners as it holds a prominent role in portfolio construction. In the case of the most popular portfolio selection theory, one of Markowitz's basic assumptions in his 1952 seminal paper is that investors make base their decisions for portfolio selection based on expected risk and return. Implicit in this assumption is that these two factors are known, or at least they can be approximated with relative accuracy.

The most common way to gauge risk and return in the financial world is through past performance, which implies the use of an empirical distribution based on historical returns (Bodie, Kane, & Marcus, 2008). Given the large available sample of stock returns and, most importantly, if past and future returns belong to in the same distribution, practitioners can find the optimal weights for their desired portfolio mix of risk and return based on the Markowitz procedure. However, implicit in this procedure is the assumption that past and future returns come from the same distribution.

To examine whether this basic assumption holds, this study employs the nonparametric two-sample Kolmogorov-Smirnov test for the equality of continuous probability distributions. The test is sensitive to differences in both the location and shape of the empirical cumulative distribution functions of two samples. If returns are from the same distribution, then any portfolio allocation ap-

proach based on the use of past returns is justified. If not, then inference based on past returns can lead to erroneous conclusions.

The results, using the daily data for 242 companies in the S&P 500 index, show that the Markowitz procedure can be a good approximation during most of the sample period; however, it should not be viewed as a universal attribute. During the period of the financial crisis, the vast majority of stock-specific return distributions reject the equality hypothesis. The results are confirmed using two additional nonparametric tests, the Cramér-Von Mises and the Wilcoxon rank-sum test, which point to qualitatively similar conclusions. Testing for the macroeconomic drivers of the rejection rate using a variety of regression models with GARCH errors shows that, while the rate is highly persistent, common explanatory variables such as financial uncertainty and the interest rate are not significant drivers of its behavior. Overall market distress appears to have a statistically significant, albeit economically small, impact.

To our knowledge, this is the first study in the literature, which specifically examines this basic, albeit implicit, assumption[1]. Thus far, empirical studies, which aimed at examining the fit of statistical distributions on the data, do not tend to use individual stock data, nor do they distinguish between estimation and forecast samples. Existing studies (e.g., Egan, 2007; Malevergne, Pisarenko, & Sornette, 2005; Aparicio & Estrada, 2001) utilize aggregate indices and do not focus on portfolio allocation in individual stocks. As such, they do not require the distribution to be the same in the estimation and forecast samples.

The findings of this study provide a rationale for why previous research (e.g., DeMiguel, Garlappi, & Uppal, 2007; Kirby & Ostdiek, 2012) found that the mean-variance approach does not offer important out-of-sample benefits. Similarly, the findings support the view that portfolio performance is sensitive to changes in asset means (Chopra & Ziemba, 1993; Best & Grauer, 1991), which is a defining feature of the distribution.

The remainder of this paper is organized as follows: section 1 provides an overview of the related literature, section 2 presents the tests for the equality of distributions and the equality rejection with the macroeconomic conditions, section 3 presents the results for both the main analysis and the robustness tests, section 4 discusses the findings and final section provides a brief summary and conclusions.

1. LITERATURE REVIEW

Markowitz (1952) introduced the context of investment portfolios selection and construction. In his seminal paper, he supports that investors make base their decisions for portfolio selection based on expected risk and return. Therefore, investors choose to maximise the expected portfolio returns and simultaneously minimize the risk. Implicit in this assumption is that these two factors are known, or at least they can be approximated with relative accuracy.

The use of an empirical distribution based on historical returns enables the investment community to gauge risk and return through past performance (Bodie, Kane, & Marcus, 2008). Based on the Markowitz's procedure, investors can rely on the large available sample of stock returns to find the optimal weights for their desired portfolio mix of risk and return, only if past and future returns belong to the same distribution. Otherwise, inference based on past returns can lead to wrong investment decisions.

While prior empirical studies have focused primarily on examining the fit of statistical distributions on the data, the assumption that past and future returns come from the same distribution holds remains largely unexplored. For example, Egan (2007) and Malevergne, Pisarenko, and Sornette (2005) show that non-normal distribu-

1 The study most similar to this is Chae and Lee (2018) who examine the significance of differences of the return distribution (distribution uncertainty) in the cross-sectional pricing of stocks.

http://dx.doi.org/10.21511/imfi.17(3).2020.15

tions (Stretched Exponential and Pareto) are better at capturing returns data, using the data spanning up to 100 years for the NASDAQ and the S&P indices. Similarly, Aparicio and Estrada (2001) find that the hypothesis of normality is rejected for 13 European securities markets. Nonetheless, studies such as the above use aggregate indices and do not focus on portfolio allocation in individual stocks. As such, they do not require the distribution to be the same in the estimation and forecast samples.

The burgeoning literature on examining the equality of continuous probability distributions proposes the nonparametric two-sample Kolmogorov-Smirnov test. The form of the Kolmogorov-Smirnov test and its asymptotic distribution under the null hypothesis was first published by Kolmogorov (1933), while a table of the distribution was offered by Smirnov (1948). It was first presented to the English-speaking audience by Massey (1951). The implementation of the test, and its sensitivity to differences in both the location and shape of the empirical cumulative distribution functions, serve the purpose of the paper to examine whether two samples come from the same distribution.

Another stream of literature on examining the equality of distributions uses the Cramér-Von Mises criterion as an alternative test to Kolmogorov-Smirnov test, which is widely used for comparing two empirical distributions. The criterion is named after Harald Cramér and Richard Edler Von Mises who first proposed it in 1928–1930 (Cramér, 1928; Von Mises, 1928). The generalization to two samples is due to Anderson (Anderson, 1962).

To further strengthen the validity of the findings, the existing literature suggests another test to determine whether two independent samples, selected from populations, have the same distribution. In particular, one employs the Wilcoxon rank-sum test, a non-parametric statistical hypothesis test after Wilcoxon (1945) who proposed the rank-sum test for two independent samples. It tests the null hypothesis that it is equally likely that a randomly selected value from one sample will be less than or greater than a randomly selected value from a second sample.

The possible correlation of macroeconomic changes should also hold when using any portfolio strategy that uses the past to predict the future. Prior literature suggests that macroeconomic variables can affect future firm performance (Issah & Antwi, 2017). For instance, according to Humpe and Macmillan (2009), the long-term interest rate should contain an indicator of the long-term perception of the economy regarding the discount rate, while Andreou (2015) supports the view that the view regarding potential market default could potentially play a role. Therefore, to shed more light on the main findings of the paper, it was essential to further examine whether the series of rejection fares any relationship with the underlying macroeconomic conditions, following prior literature.

2. METHODS

2.1. The Kolmogorov-Smirnov test

In the Kolmogorov-Smirnov setup, supposing that there are two samples $X_{1,1},...,X_{1,m}$ and $X_{2,1},...,X_{2,n}$, where m and n are the respective sample sizes, the test examines whether differences exist in the cumulative distribution function (CDF) of the first population sample $F_1(x)$ and the CDF of the second population sample $F_2(x)$. The Kolmogorov-Smirnov statistic, denoted as KS, is found by

$$KS_{m,n} = \max_x \left| \hat{F}_1(x) - \hat{F}_2(x) \right| \qquad (1)$$

and is calculated by finding the maximum absolute value of the differences between two sample CDFs. The null hypothesis is rejected at level α if

$$KS_{m,n} = c(\alpha)\sqrt{\frac{m+n}{mn}}, \qquad (2)$$

where the value of $c(\alpha)$ is given b

y $c(\alpha) = \sqrt{-\frac{1}{2}\ln\left(\frac{\alpha}{2}\right)}$ and α denotes the significance level[2]. In other words, the null hypothesis (H_0) is that $\hat{F}_1(x) = \hat{F}_2(x)$, and we reject the null

2 The statistical values of the Kolmogorov distribution were obtained using the KSINV function in the Real Statistics Resource Pack, provided by Charles Zaiontz.

if $\left|\hat{F}_1(x)-\hat{F}_2(x)\right|>KS_{m,n}$

where $\left|\hat{F}_1(x)-\hat{F}_2(x)\right|$

is as specified above. If the null is true, $\left|\hat{F}_1(x)-\hat{F}_2(x)\right|$ has a distribution that does not depend on $\hat{F}_1(x)=\hat{F}_2(x)$, but depends on n and m.

Note that an important point of the two-sample test is that it does not specify the nature of that common distribution. In other words, the test does not examine whether the common distribution is Normal, Student's T, or Weibull but only whether the two samples come from the same distribution. As such, failure to reject the hypothesis does not ensure that a method for constructing the mean-variance portfolio under whichever distribution is correct. Not rejecting the hypothesis, however, suggests that since the two distributions will likely be the same, one could likely obtain good out-of-sample performance if inference is based on the estimation distribution.

In this study, the Kolmogorov-Smirnov statistic for equality of distributions is estimated using a sample of daily stock returns for 242 firms, which have been actively trading in the S&P 500 index from January 1, 2000 to December 31, 2014, with a total of 3773 observations for each stock. All data were obtained from the Center for Research in Security Prices (CRSP) database. For estimation, the in-sample window size is set at 1000 rolling observations with a daily step, in line with most other studies in the literature (e.g., Yilmaz, 2012; Kim, Shamsuddin & Lim, 2011).

The equality of the 1000-observation distribution is compared with the distribution obtained using a window of 100 observations after the end of the estimation sample. Notation-wise, the estimation sample, $F_1(x)$, starts at t and ends at t + 1000, while observations t + 1001 to t + 1101 comprise $F_1(y)$, i.e., the sample, which will be used for distribution equality. For simplicity, one will refer to $F_1(x)$ as the estimation sample and to $F_1(y)$ as the forecast sample. A similar-sized window for the evaluation of forecasts was also employed by Pesaran and Pick (2011).

For robustness, results for samples comprising of 50 and 200 observations are also estimated. The results from the estimations can be found in subsection 3.1.

2.2. The Cramér-Von Mises test

Supposing that one has the observed values for two samples in increasing order, i.e., $x_1, x_2, ..., x_M$ for the first sample and $y_1, y_2, ..., y_N$ for the second sample, respectively, where M and N are the respective sample sizes. The test examines whether two samples come from the same distribution. If $r_1, r_2, ..., r_M$ are the ranks of the x's in the combined sample, and $s_1, s_2, ..., s_N$ are the ranks of the y's in the combined sample, Anderson (1962) shows that

$$T=\frac{NM}{N+M}w^2=\frac{U}{NM(N+M)}-\frac{4MN-1}{6(M+N)}, \quad (3)$$

where U is defined as

$$U=M\sum_{i=1}^{M}(r_i-i)^2+N\sum_{j=1}^{N}(s_j-j)^2. \quad (4)$$

If the value of T is larger than the tabulated values, one can reject the hypothesis that two samples come from the same distribution. The results from the estimations can be found in subsection 3.2.

2.3. The Wilcoxon rank-sum test

The test involves the calculation of a statistic, denoted as U. Statistics equivalent to U can be considered the sum of ranks in one of the samples, rather than U itself. Its calculation, firstly, requires assigning numeric ranks to all the observations, and then adds the ranks for the observations which came from sample 1. The sum of ranks in sample 2 is now determinate since the sum of all the ranks equals $T(T+1)/2$, where T is the total number of observations. U is given by

$$U_1=MN+\frac{M(M+1)}{2}-r, \quad (5)$$

http://dx.doi.org/10.21511/imfi.17(3).2020.15

$$U_2 = MN + \frac{N(N+1)}{2} - s, \quad (6)$$

where M and N is the size for the first and second samples, respectively, and r and s is the sum of the ranks in the first and second samples, respectively.

Thus, smaller values of U support the research hypothesis, and larger values of U support the null hypothesis. For any U test, the theoretical range of U is from 0 (complete separation between groups, H_0 most likely false and H_1 most likely true) to $N*M$ (little evidence in support of H_1). In every test, we must determine whether the observed U supports the null or research hypothesis. This is done following the same approach used in parametric testing.

Specifically, one determines a critical value of U such that if the observed value of U is less than or equal to the critical value, one rejects H_0 in favor of H_1 and if the observed value of U exceeds the critical value, one does not reject H_0. To determine the appropriate critical value, one needs sample sizes (N and M) and two-sided level of significance. The results from the estimations can be found in subsection 3.3.

2.4. Equality rejection and macroeconomic conditions

To examine whether the series of rejection fares any relationship with the underlying macroeconomic conditions, one uses a regression model with GARCH errors, as first presented by Bollerslev (1986). Formally, the GARCH (p, q) model can be formulated through a mean equation:

$$Y_t = \alpha_0 + \alpha_1 M_t + \varepsilon_t, \quad (7)$$

where Y_t is the percentage of firms, which reject the hypothesis of equality on a given day (rejection rate), M_t is a vector of macroeconomic variables, and ε_t is the error term, which evolves according to the following process:

$$\varepsilon_t \left| \Omega_t \sim N(0, h_t), \right. \quad (8)$$

$$h_t = \beta_0 + \sum_{k=1}^{q} \beta_k \varepsilon_{t-k}^2 + \sum_{j=1}^{p} \gamma_j h_{t-j}. \quad (9)$$

In equation (3), ε_{t-k}^2 represents past squared values of the errors, which can be interpreted as the feed-through of shocks to the variance and h_{t-j} represents past values of the error variance (interpreted as the persistence of shocks to the error variance)3. p and q are the orders of ARCH and GARCH terms, respectively. As in other studies in the literature, one limits the scope of the estimation to the GARCH (1,1), which has been shown to perform well in financial markets (Andersen & Bollerslev, 1998; Hansen & Lunde, 2005). A similar setup for rejection rates has been used by Michail (2019).

To account for potential asymmetries in volatility, Glosten, Jagannathan and Runkle (1993) introduced an additional term to the GARCH specification. The GJR-GARCH model replaces equation (3) with

$$h_t = \beta_0 + \sum_{k=1}^{q} \beta_k \varepsilon_{t-k}^2 + \sum_{j=1}^{p} \gamma_j h_{t-j} + \\ + \sum_{k=1}^{p} \delta_k I_{t-k} \varepsilon_{t-k}, \quad (10)$$

where $I_{t-j} = 1$ if $\varepsilon_{t-k} < 0$ and zero otherwise. p and q are set to one. In the case where the last term in equation (4) is statistically significant, negative shocks have a distinct impact on stock market returns. If the sign is negative, then positive shocks have a different impact on rejection volatility.

To examine whether macroeconomic developments play a role in the estimation, the estimation uses the daily 10-year bond yield (see also Humpe and Macmillan (2009) and Campbell and Thompson (2007) for more on the relationship between stock markets and the long-run bond yield), while the CBOE VIX index is used as a measure of market fear and risk as it uses option-implied volatility in its estimation. Besides, one also employs the Market Default Likelihood Index (MDLI) of Andreou (2015) to account for market-wide distress and the overall probability of default. All data were obtained from the St Louis Fed[4].

3 For an overview of the interpretation of ARCH and GARCH terms, see Campbell, Lo, and MacKinlay (1997, p. 483) or Alexander (2008, p. 283).

4 The MDL index was obtained from the personal website of Panayiotis C. Andreou (https://www.pandreou.com/).

To account for September 2008, the month of the Lehman collapse, a dummy variable taking the value of one during the period, is also used. Besides, to examine whether abnormal behavior, with regard to the rejection rate, took place during the whole recession period, a dummy, which takes the value of 1 from December 2007 till June 2009, and 0 otherwise was also included in the estimation. Recession dates were obtained from the National Bureau of Economic Research (NBER). To avoid issues related to causality through the use of contemporaneous terms, all macroeconomic variables are included with a lag. This is also more practical since their usefulness in predicting the rejection rate can also be examined. The results of the estimation can be found in Table 4, panels (a) to (d).

3. RESULTS

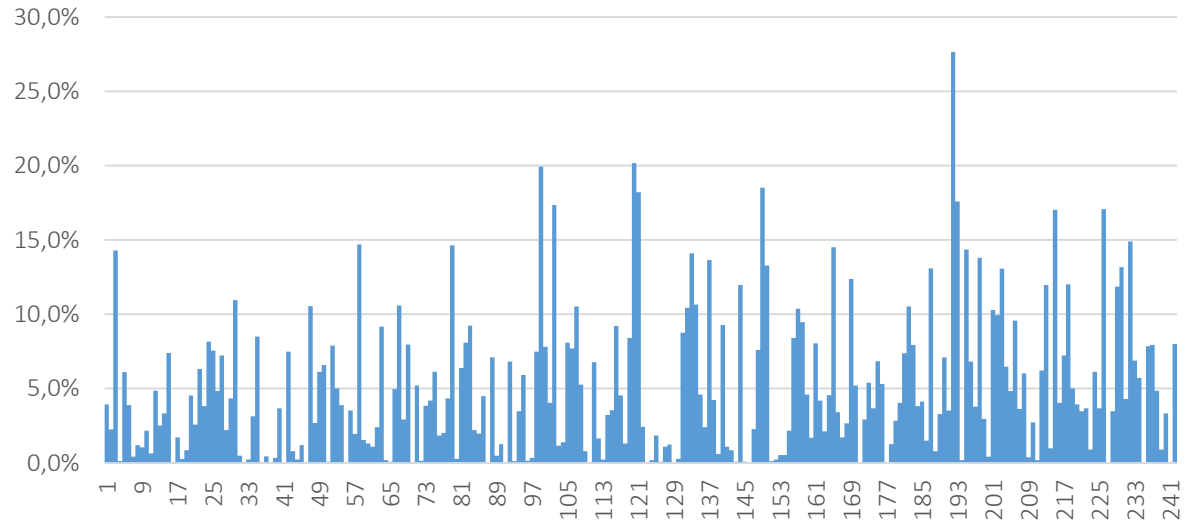
3.1. The Kolmogorov-Smirnov test

Figure 1 shows the test results on each of 242 firms under examination. In particular, Figure 1 indicates the percentage, out of a total of 2773 (3773 minus 1000 window) estimation days in which the estimation distribution was not the same as the forecast distribution, for each of 242 firms under study. At two ends of the spectrum, some forecast distributions were not equal to the estimation sample 25% of the times, while there also exist cases in which forecast distributions were always equal to the estimation ones. On average, estimation and forecast distributions were different, only 5% of total days (Table 1).

The fact that in only 5% of total trading days till distributions were unequal would support the use of the Markowitz procedure. However, this result hides an important caveat: the number of daily stock returns, which were different from their forecast distribution, has not been uniform across time. Figure 2 shows that the number of stock returns, which rejected the hypothesis of equality on each given day, as a percentage of total stocks, skyrocketed during the financial crisis.

In particular, starting from September 2008, when the number of stocks with different forecast distributions stood at 5%, the percentage skyrocketed to more than 60% in February 2009, returning to 5% a year after the upswing in September 2009. During that period, an average of 32.6% of stock returns rejected the hypothesis of equality (Table 1). Excluding these 12 months, the average hypothesis rejection drops to just 2.1%.

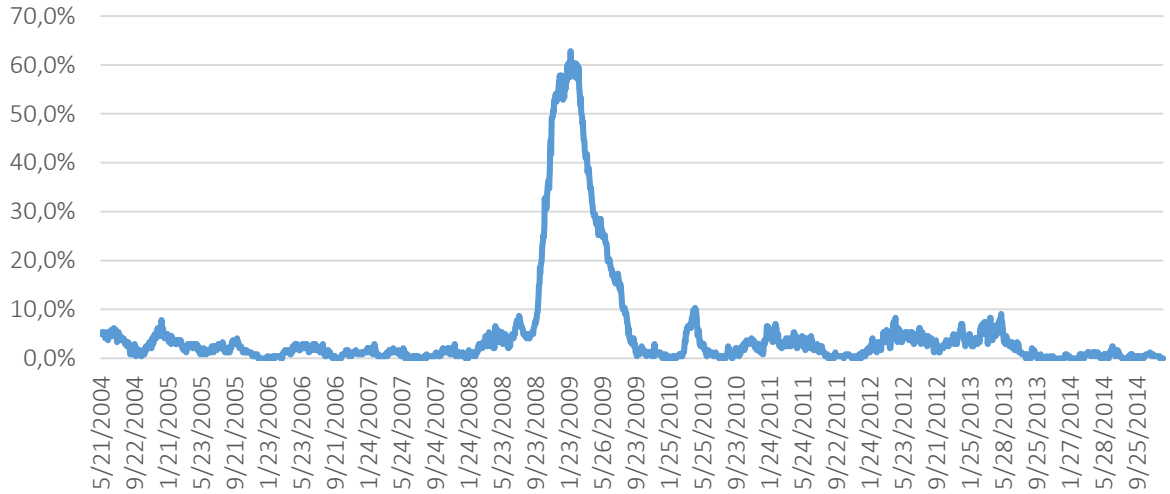
Robustness checks (Figure 3), using a forecast sample of 50 (panel (a)) and 200 (panel (b)) observations, show an image, by and large, similar to the one using the 100-observation forecast sample. In particular, panel (a) suggests that the impact is



Note: Figure 1 depicts the percentage of rejections of each stock, i.e., the number of samples for which the hypothesis of equal variance was rejected by the variance ratio test, over the total number samples for each stock. The x-axis reflects each stock available from January 1, 2000 till December 31, 2014.

Figure 1. Percentage of samples in which distributions were not equal

http://dx.doi.org/10.21511/imfi.17(3).2020.15



Note: Figure 2 depicts the percentage of individual stock returns which rejected the equality hypothesis at each given date, measured as the number of stocks for which the hypothesis was rejected at every given sample, over the total number of stocks. The rolling sample size was set at 1,000 daily observations and the forecast sample at 100. The dates depict the end-date of each rolling sample.

Figure 2. One-hundred (100) observations forecast sample

more protracted using the shorter forecast period: even though the hypothesis of equality is rejected only 3.2% of the times on average, it is rejected in more than 76% of individual firms at its peak (September 2008). While the peak is larger than the one for other forecast horizons, one in five stocks rejected the hypothesis on average during September 2008 – September 2009 period, resulting in an average rejection rate of 21.8%. Excluding this period, the average rejection rate goes down to 1.3%. A similar image can be viewed in panel b: the hypothesis of equality is rejected by 7.9% of the total estimation days, while excluding the crisis period, it is rejected only by 4.4%. During the crisis period, an average of 40.8% of all stock returns rejected the equality hypothesis. While the peak is significantly lower, the rejection rate is larger for September 2008 to September 2009 period.

Table 1. Average rejection rate in each period

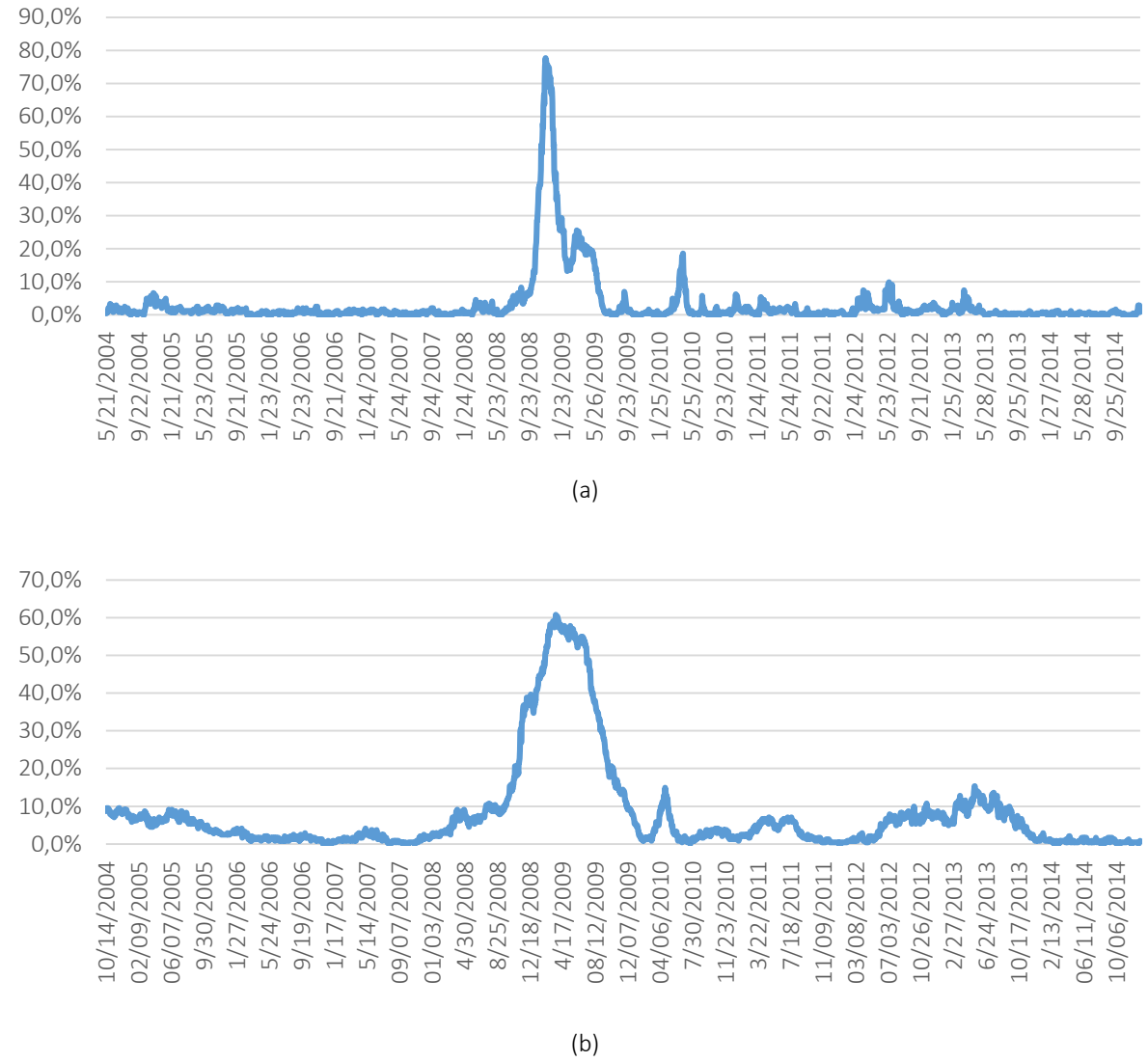
Period	Forecast sample size		
	50	100	200
Full sample	3.2%	5.0%	7.9%
Excluding crisis	1.3%	2.1%	4.4%
Crisis period	21.8%	32.6%	40.8%

Note: Table 1 presents the average rejection percentage out of total stocks (242) at the specified point in time. Full sample covers from January 1, 2000 till December 31, 2014 while the crisis period refers to September 2008 – September 2009 period.

An interesting complication of the test results is that the longer the forecast horizon, the larger the average rejection rate appears to be. While it makes intuitive sense, since a larger horizon allows for the existence of more values, which can change the empirical distribution of the returns, it nevertheless provides a rationale against long-run forecasts. Practitioners can expect that short-term realisations will come from the same distribution as past returns, in the case of statistical modeling. However, it could be the case that, as the horizon lengthens, realizations will most likely not belong to the same empirical distribution, with all the implications this entails for stock market forecasting.

3.2. The Cramér-von Mises test

Robustness tests are conducted using the Cramér-Von Mises test. Figure 4 shows the Cramér-Von Mises test results on each of 242 firms under study. In particular, Figure 4 indicates the percentage, out of a total of 2,773 (3,773 minus 1,000 window) estimation days in which the estimation distribution was not the same as the forecast distribution, for each of the 242 firms under study. At the two ends of the spectrum, some forecast distributions were not equal to the estimation sample 79% of the times, while there also exist cases in which forecast distributions were not equal to the estimation ones only 14% of the times. On average, estimation and forecast distributions were different 38% of total days.



Note: Figure 3 depicts the percentage of individual stock returns, which rejected the equality hypothesis at each given date, measured as the number of stocks for which the hypothesis was rejected at every given sample, over the total number of stocks. The rolling sample size was set at 1,000 daily observations and the forecast sample at 50 and 200 for panels (a) and (b), respectively. The dates depict the end-date of each rolling sample.

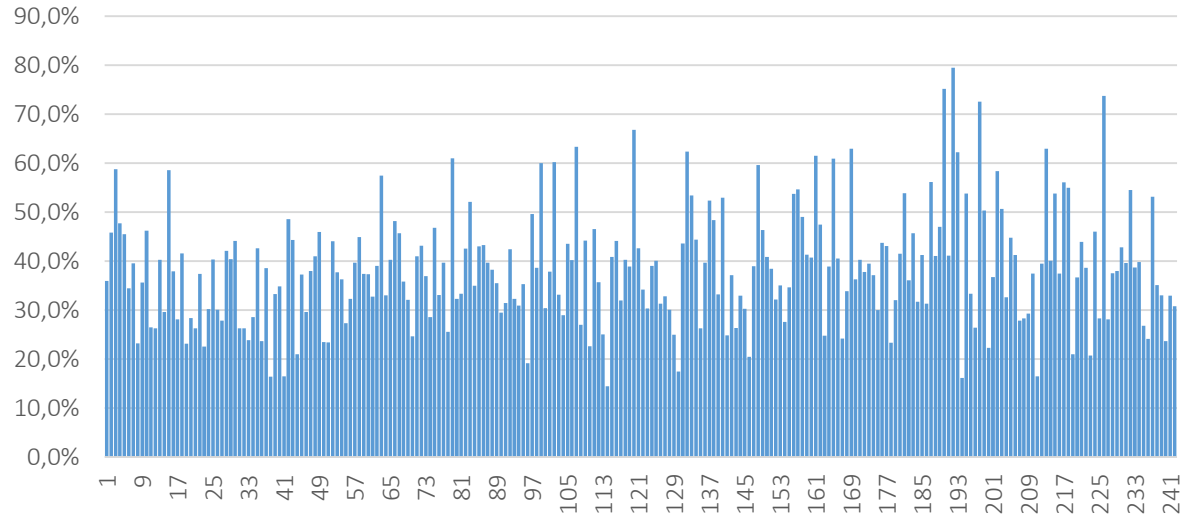
Figure 3. Fifty (50) and two-hundred (200) observation window

The results of Cramér-Von Mises test, strengthen the caveat presented by the Kolmogorov-Smirnov that the number of daily stock returns, which were different from their forecast distribution has not been uniform across time. Figure 5 shows that the number of stock returns, which rejected the hypothesis of equality on each given day, as a percentage of total stocks, increased rapidly during the financial crisis.

Particularly, before September 2008, an average of 38% of stock returns rejected the hypothesis of equality. Between September 2008 and 2009, on average, 87% of stock returns rejected the hypothesis of equality. During 12 months, 100% of stock returns rejected the hypothesis of equality, with 85 days. The percentage returns to an average of 29% of stock returns, which rejected the hypothesis of equality after the crisis.

Checks, using a forecast sample of 50 (Figure 6, panel (a)) and 200 (Figure 6, panel (b)), observations show an image, by and large, similar to the one using the 100-observation forecast sample. In particular, panel (a) suggests that the impact is more concentrated between October 2008 and

http://dx.doi.org/10.21511/imfi.17(3).2020.15



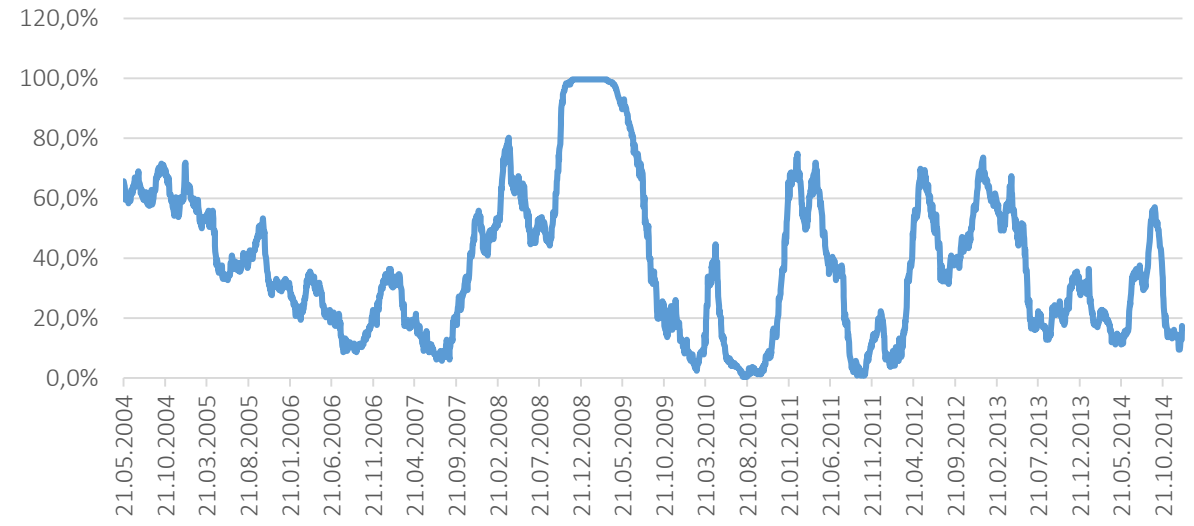
Note: Figure 4 depicts the percentage of rejections of each stock, i.e., the number of samples for which the hypothesis of equal variance was rejected by the variance ratio test, over the total number samples for each stock. The x-axis reflects each stock available from January 1, 2000 till December 31, 2014.

Figure 4. Percentage of samples in which distributions are not equal using Cramér-von Mises test

May 2009, using the shorter forecast period, while in panel (b) is more prolonged until December 2009. However, using the previous comparable period, the effect is still there. For panels (a) and (b), respectively, before September 2008, an average of 20% and 18% of stock returns rejected the hypothesis of equality. Between September 2008 and 2009, on average, 69% and 57% of stock returns rejected the hypothesis of equality, with 42 and 43 days of 100% of equality hypothesis rejection. The post-crisis percentage returns to an average of 14% and 19% of stock returns, which rejected the hypothesis of equality.

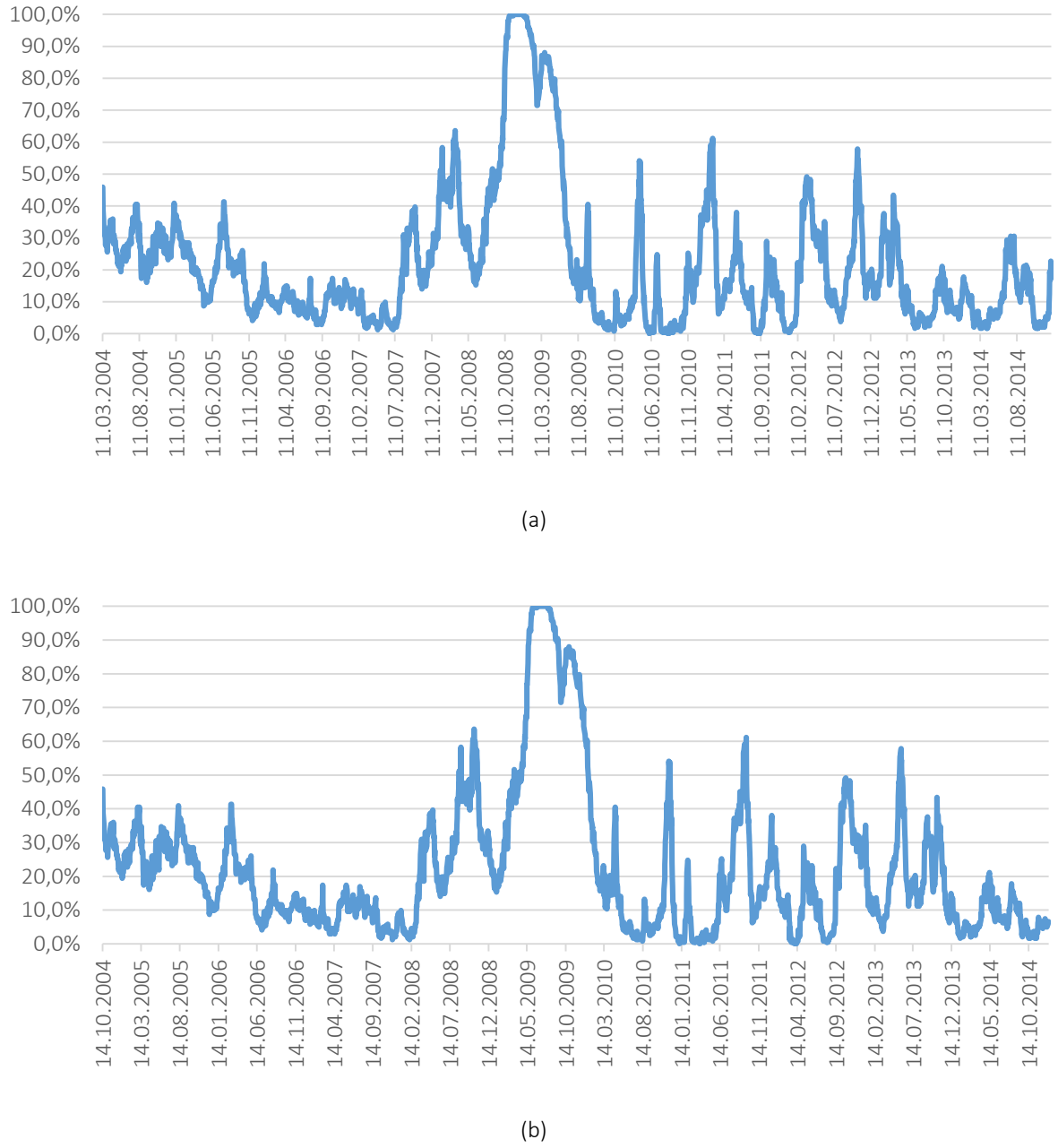
Table 2. Average rejection rate in each period

Period	Forecast sample size		
	50	100	200
Full sample	34.3%	51%	31.3%
Excluding crisis	17%	33.5%	18.5%
Crisis period	69%	87%	57%



Note: Figure 5 depicts the percentage of individual stock returns, which rejected the equality hypothesis at each given date, measured as the number of stocks for which the hypothesis was rejected at every given sample, over the total number of stocks. The rolling sample size was set at 1,000 daily observations and the forecast sample at 100. The dates depict the end-date of each rolling sample.

Figure 5. One-hundred (100) observation forecast sample using Cramér-von Mises test



Note: Figure 6 depicts the percentage of individual stock returns, which rejected the equality hypothesis at each given date, measured as the number of stocks for which the hypothesis was rejected at every given sample, over the total number of stocks. The rolling sample size was set at 1,000 daily observations and the forecast sample at 50 and 200 for panels (a) and (b), respectively. The dates depict the end-date of each rolling sample.

Figure 6. Fifty (50) and two-hundred (200) observation forecast sample using Cramér-von Mises test

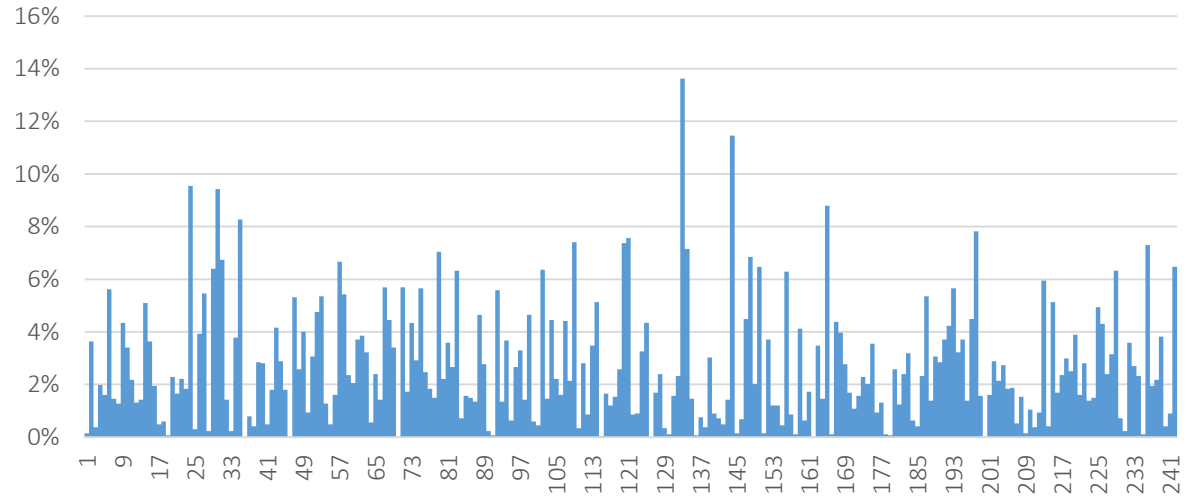
Table 2 presents the average rejection percentage out of total stocks (242) at the specified point in time. Full sample covers from January 1, 2000 till December 31, 2014, while the crisis period refers to the September 2008 – September 2009 period.

While the peak is significantly lower, the rejection rate is larger for September 2008 to September 2009 period. Therefore, in both alternative estimation windows used for robustness, i.e., 50 and 200-observation forecast sample, one can view a similar image, which is consistent with the initial 100-observation forecast sample.

3.3. The Wilcoxon rank-sum test

Robustness tests are conducted using the Wilcoxon rank-sum test. Figure 7 shows the Wilcoxon rank-

http://dx.doi.org/10.21511/imfi.17(3).2020.15

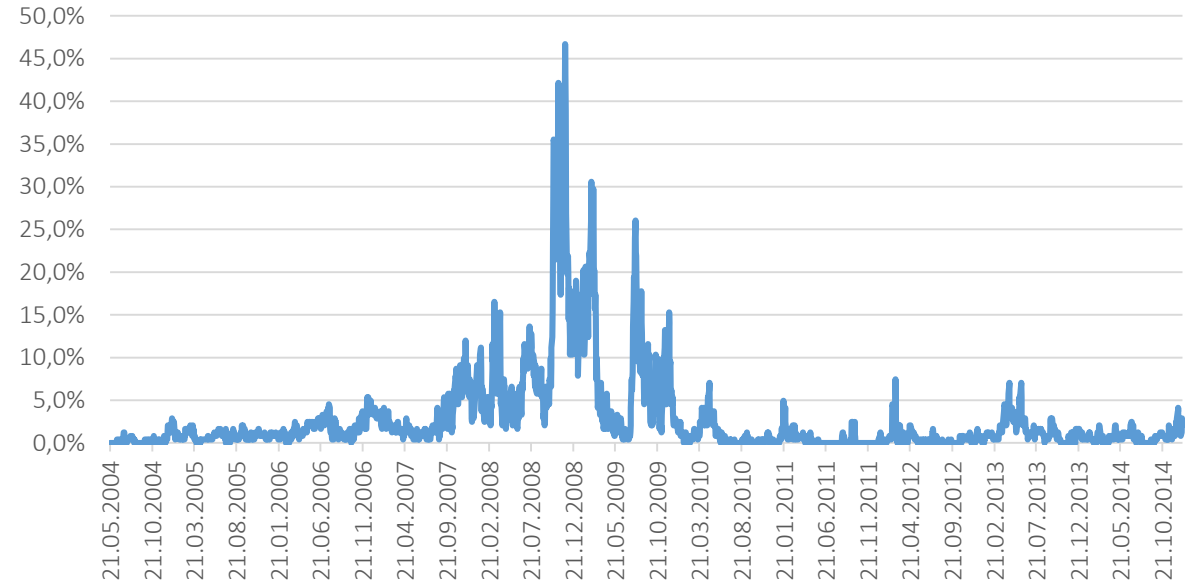


Note: Figure 7 depicts the percentage of rejections of each stock, i.e., the number of samples for which the hypothesis of equal variance was rejected by the variance ratio test, over the total number samples for each stock. The x-axis reflects each stock available from January 1, 2000 till December 31, 2014.

Figure 7. Percentage of samples in which distributions are not equal using Wilcoxon rank-sum test

sum test results on each of the 242 firms under examination. Specifically, Figure 7 specifies the percentage, out of a total of 2,773 (3,773 minus 1,000 window) estimation days in which the estimation distribution was not the same as the forecast distribution, for each of the 242 firms under study. At the two ends of the spectrum, some forecast distributions were not equal to the estimation sample 14% of the times, while there also exist cases in which forecast distributions were always equal to the estimation ones. On average, estimation and forecast distributions were different 3% of total days.

While the overall percentages are weaker using this statistic test, the number of stock returns which rejected the hypothesis of equality on each given day, as a percentage of total stocks, rise steeply during financial crisis. In particular, as present-



Note: Figure 8 depicts the percentage of individual stock returns, which rejected the equality hypothesis at each given date, measured as the number of stocks for which the hypothesis was rejected at every given sample, over the total number of stocks. The rolling sample size was set at 1,000 daily observations and the forecast sample at 100. The dates depict the end-date of each rolling sample.

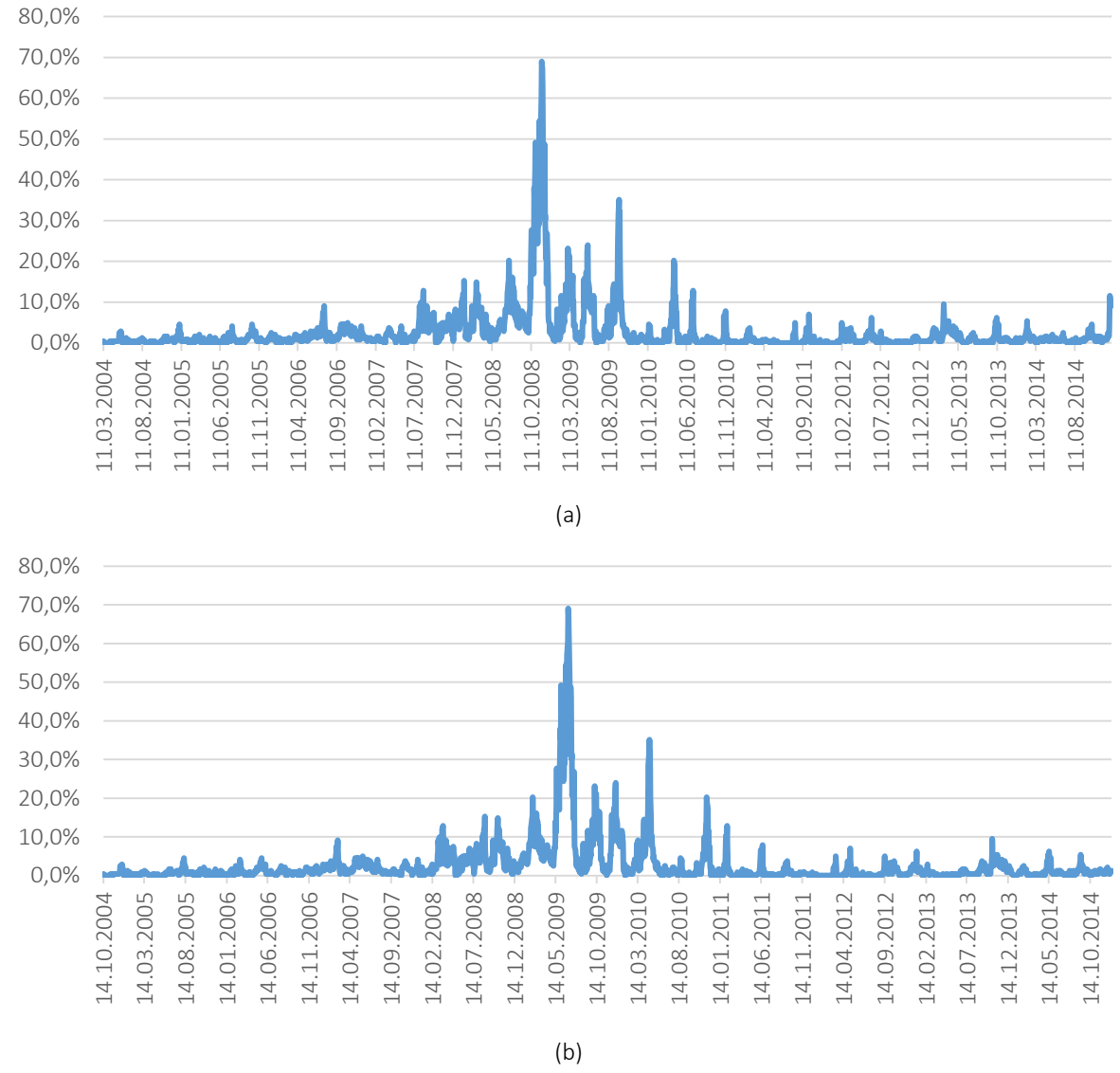
Figure 8. One-hundred (100) observation forecast sample using Wilcoxon rank-sum test

ed in Figure 8, before September 2008, an average of 2% of stock returns rejected the hypothesis of equality. Between September 2008 and 2009, on average, 12% of stock returns rejected the hypothesis of equality. During 12 months, there were days that 47% of stock returns rejected the hypothesis of equality. After the financial crisis, the percentage returns to 1% of stock returns, which rejected the hypothesis of equality.

The robustness checks (Figure 9) that have been implemented using a forecast sample of 50 (panel (a)) and 200 (panel (b)) observations almost replicates exactly the previous image. In both panel (a) and panel (b) before September 2008, an average of 2% of stock returns rejected the hypothesis of equality. Between September 2008 and 2009, on average, 11% of stock returns rejected the hypothesis of equality. The post-crisis percentage returns to an average of 1% of stock returns, which rejected the equality hypothesis.

Table 3. Average rejection rate in each period

Period	Forecast sample size		
	50	100	200
Full sample	4.67%	5%	4.67%
Excluding crisis	1.5%	1.5%	1.5%
Crisis period	11%	12%	11%



Note: Figure 9 depicts the percentage of individual stock returns, which rejected the equality hypothesis at each given date, measured as the number of stocks for which the hypothesis was rejected at every given sample, over the total number of stocks. The rolling sample size was set at 1,000 daily observations and the forecast sample at 50 and 200 for panels (a) and (b), respectively. The dates depict the end-date of each rolling sample.

Figure 9. Fifty (50) and two-hundred (200) observation forecast sample using Wilcoxon rank-sum test

http://dx.doi.org/10.21511/imfi.17(3).2020.15

Table 3 presents the average rejection percentage out of total stocks (242) at the specified point in time. Full sample covers from January 1, 2000 to December 31, 2014, while the crisis period refers to September 2008 – September 2009 period.

3.4. Equality rejection and macroeconomic conditions

As the estimation results suggest, the macroeconomic variables do not appear to have had an impact on the level of rejection rate, since neither the interest rate nor the VIX index found to be statistically significant. In contrast, the Lehman dummy's significance illustrates that the increase in the percentage of inequality of distributions during that period cannot be associated with any other development.

Table 4. Estimation results

Mean equation				
	(a)	(b)	(c)	(d)
Recession	0.04 (0.04)	0.03 (0.04)		
Housing	−0.03 (0.02)	−0.01 (0.02)	−0.01 (0.02)	
Lehman	0.24** (0.04)	0.21** (0.09)	0.24** (0.12)	0.23** (0.11)
VIX_{t-1}	−0.02 (0.11)	0.06 (0.12)	0.10 (0.12)	0.14 (0.11)
$Interest\ Rate_{t-1}$	−0.10 (0.08)	0.06 (0.09)	0.09 (0.09)	0.05 (0.07)
$MDLI_{t-1}$	0.03** (0.01)	0.03** (0.01)	0.03*** (0.01)	0.03*** (0.01)
AR	0.98	0.98	0.98	0.98
Constant	−0.02 (0.03)	−0.01 (0.04)	−0.03 (0.04)	0.02 (0.04)
Volatility equation				
	(a)	(b)	(c)	(d)
Constant	0.00*** (0.00)	0.00*** (0.00)	0.00*** (0.00)	0.00*** (0.00)
ARCH	0.16*** (0.01)	0.21*** (0.02)	0.21*** (0.01)	0.21*** (0.02)
GARCH	0.84*** (0.01)	0.86*** (0.01)	0.86*** (0.01)	0.86*** (0.01)
δ_k		−0.15*** (0.02)	−0.15*** (0.02)	−0.15*** (0.02)
Recession			0.01** (0.00)	0.01** (0.00)
Housing				−0.01* (0.00)
R-squared	0.99	0.99	0.99	0.99
DW statistic	1.84	1.86	1.88	1.88

Notes: The dependent variable is the number of rejected hypotheses each day as a percentage of the 242 firms in the sample, using the 100-observation forecast sample. Lehman is a dummy variable taking the value of 1 for the whole of September 2008 when Lehman Brothers collapsed. Housing is a dummy variable taking the value of 1 from January 2005 till December 2006 (dates were obtained by Kim et al., 2011). Recession is a dummy variable taking the value of 1 from December 2007 till June 2009 (dates from NBER). VIX is the Chicago Board Options Exchange volatility index (divided by 100 for comparability) and Interest rate is defined as the 10-year government bond yield (divided by 10 for comparability). MDLI is the Market Default Likelihood Index of Andreou (2015). AR is the sum of the coefficients for the lagged dependent variables while ARCH and GARCH denote the ARCH and GARCH terms, respectively. The Ljung-Box test for serial correlation rejects the hypothesis of autocorrelated errors in 35 lags. The ARCH LM test does not reject the hypothesis of no ARCH effects in any of the above models. Standard errors in parentheses. *, **, *** denotes significance at the 10%, 5% and 1% level, respectively.

The results also underline the connection between the overall market distress and several rejections. The impact is statistically significant throughout all the estimations, while the positive sign suggests that as market distress increases the rejections of the equality hypothesis also increase. As the standard deviation of the MDLI is approximately 3.67, a one-standard-deviation shock would mean an increase of 0.11 percentage points in the rejection rate. Although the number is small, it should be borne in mind that given the magnitude of the AR terms, the shock will be highly persistent[5].

Moving to the volatility equation, ARCH and GARCH terms are statistically significant throughout all panels. Moreover also appears to be statistically significant and negative, suggesting that positive shocks have a different impact on volatility. Finally, while the housing boom and recession dummies are not significant in the mean equation, the conditional volatility of the rejection rate decreased during the period of the housing boom and increased during the recession (panel (d)), again suggesting that macroeconomic conditions have at least some impact on the rejection rate.

4. DISCUSSION

This study examines if the daily returns of 242 companies with continuous trading in the S&P index during 2000-2014 come from the same distribution using the Kolmogorov-Smirnov, Cramér-Von Mises, and Wilcoxon rank-sum tests. The findings suggest that although the dis-

5 Although not reported here, similar results were obtained using a panel structure for the estimates.

tribution of returns generally tends to be equal, during the Great Recession, the equality hypothesis of stock returns is frequently rejected, and the rejection rate tends to increase with the forecast horizon increased. Therefore, the longer the forecast horizon, the larger the average rejection rate appears to be, which opposes the long-run forecasts. The choice of a shorter forecast period is not a panacea, though, in times of distress, the percentage of firms, which records the returns where the forecast sample does not belong to the same distribution as the estimation sample increases by much more when the horizon is shorter. The reaction is nonetheless shorter-lived compared to longer horizons. As such, while short-term forecasting can likely be more accurate most of the times, it can be more prone to errors in the periods of turmoil.

The results of the Kolmogorov-Smirnov test present the caveat that the number of daily stock returns, which were different from their forecast distribution has not been uniform across time. Specifically, the rejection of the equality hypothesis increased rapidly during the financial crisis. The robustness checks conducted using both the Cramér-Von Mises test and the Wilcoxon rank-sum test qualitatively replicates the previous image.

The results using all three different statistic tests and the different estimation windows are robust. We can conclude that our results are not driven by the method used to either test the distribution equality, or several observations used to the forecast sample. Regardless of the strength of the percentage of rejection, the equality hypothesis or the peak of the stock returns that reject it, the image is identical and consistent. The findings suggest that while, in general, the equality of distribution hypothesis holds, the percentage of rejection increases rapidly during of financial crisis.

A further complication of the results is that macroeconomic developments should play an important role since the percentage of rejection rises significantly during the Great Recession. The possible correlation of macroeconomic changes and the equality of distributions should hold when using any portfolio strategy, which uses the past to predict the future. As such, it is not limited to the mean-reversion optimal portfolio procedure.

The findings suggest that the rejection rate is highly persistent and does not appear to have a relationship with usual macroeconomic variables. Nonetheless, a market distress index can capture, to some extent, the developments in the rejection rate. This finding, along with the result that conditional volatility being related to macroeconomic states such as the housing boom and the recession, suggests that the rejection of equality of distribution, is to some extent, correlated with changes in the macroeconomic environment.

CONCLUSION

The main conclusion is that the usual assumption that the distribution of past returns equals that of future returns holds in general. However, timing is important when it comes to using the mean-variance portfolio in practice: for example, during the Great Recession, stock returns rejected the hypothesis of equality more than one out of five times, with the rejection rate increasing as the forecast horizon increased. The rejection rate is increasing in the forecast horizon in normal conditions, while it increases substantially more in the short run in times of market distress. Macroeconomic variables were not found to be statistically significant determinants of the rejection rate, with a large unexplained part remaining during the month of the Lehman collapse, while market distress has a small but significant effect.

The above findings bear important implications. First, the results from this empirical study on the goodness-of-fit of the "past returns can be used for optimal future allocation of resources" doctrine suggest that while this holds in general, the profitability of the mean-variance portfolio (or any other strategy which allocates weights based on past returns) will be time-dependent. As the results further suggest, it could be the case that the optimality of portfolio selection may even be macro-dependent. Second, the

http://dx.doi.org/10.21511/imfi.17(3).2020.15

findings point to the fact that past returns could be useful when the economy is not in turbulence but not otherwise. As such, other successful portfolio allocation strategies may arise when the results are taken into consideration.

REFERENCES

1. Alexander, C. (2008). *Market Risk Analysis: Pricing, Hedging and Trading Financial Instruments*. John Wiley & Sons. Retrieved from https://www.amazon.com/Analysis-Pricing-Hedging-Financial-Instruments/dp/0470997893
2. Andersen, T. G., & Bollerslev, T. (1998). Answering the Skeptics: yes, standard volatility models do provide accurate forecasts. *International Economic Review, 39*(4), 885-905. Retrieved from https://www.jstor.org/stable/2527343?seq=1
3. Anderson, T. W. (1962). On the distribution of the two-sample Cramer-von Mises criterion. *Annals of Mathematical Statistics,* 1148-1159. Retrieved from https://projecteuclid.org/download/pdf_1/euclid.aoms/1177704477
4. Andreou, P. C. (2015). Effects of market default risk on index option risk-neutral moments. *Quantitative Finance, 15*(2), 2021-2040. Retrieved from http://dro.dur.ac.uk/17181/1/17181.pdf
5. Aparicio, F. M., & Estrada, J. (2001). Empirical distributions of stock returns: European securities markets, 1990–95. *European Journal of Finance, 7*(1), 1-21. Retrieved from https://www.tandfonline.com/doi/abs/10.1080/135184701217 86?-journalCode=rejf20
6. Best, M. J., & Grauer, R. R. (1991). On the sensitivity of mean-variance-efficient portfolios to changes in asset means: some analytical and computational results. *Review of Financial Studies, 4*(2), 315-342. Retrieved from https://pdfs.semanticscholar.org/10bf/02a77f8d99d987bb80e27ef462c19fc1e0c6.pdf
7. Bodie, Z., Kane, A., & Marcus, A. J. (2008). *Investments* (7th ed.). Boston: McGraw-Hill.
8. Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics, 31*(3), 307-327. https://doi.org/10.1016/0304-4076(86)90063-1
9. Campbell, J. Y., & Thompson, S. B. (2007). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies, 21*(4), 1509-1531. Retrieved from https://academic.oup.com/rfs/article-abstract/21/4/1509/1567518
10. Campbell, J. Y., Lo, A. W. C., & MacKinlay, A. C. (1997). *The econometrics of financial markets*. Princeton, NJ: Princeton University Press. Retrieved from https://www.amazon.com/Econometrics-Financial-Markets-John-Campbell/dp/0691043019
11. Chae, J., & Lee, E. J. (2018). Distribution uncertainty and expected stock returns. *Finance Research Letters, 25,* 55-61. https://doi.org/10.1016/j.frl.2017.10.006
12. Chopra, V. K., & Ziemba, W. T. (1993). The effect of errors in means, variances, and covariances on optimal portfolio choice. *Journal of Portfolio Management, 19,* 6-11. Retrieved from https://jpm.pm-research.com/content/19/2/6
13. Cramér, H. (1928). On the composition of elementary errors: First paper: Mathematical deductions. *Scandinavian Actuarial Journal, 1,* 13-74. Retrieved from https://www.tandfonline.com/doi/abs/10.1080/03461238.1928.10416862
14. DeMiguel, V., Garlappi, L., & Uppal, R. (2007). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies, 22*(5), 1915-1953. Retrieved from https://academic.oup.com/rfs/article-abstract/22/5/1915/1592901
15. Egan, W. (2007). *The distribution of S&P 500 index returns* (Working Paper). Retrieved from https://pdfs.semanticscholar.org/5c90/10b27967829e21be420d549ef632c33ae755.pdf
16. Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance, 48*(5), 1779-1801. https://doi.org/10.1111/j.1540-6261.1993.tb05128.x
17. Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models: does anything beat a GARCH (1, 1)? *Journal of Applied Econometrics, 20,* 873-889. https://doi.org/10.1002/jae.800
18. Humpe, A., & Macmillan, P. (2009). Can macroeconomic variables explain long-term stock market movements? A comparison of the US and Japan. *Applied Financial Economics, 19*(2), 111-119. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1026219
19. Issah, M., & Antwi, S. (2017). Role of macroeconomic variables on firms' performance: Evidence from the UK. *Cogent Economics & Finance, 5*(1), 1405581. Retrieved from https://www.tandfonline.com/doi/pdf/10.1080/23322039.2017.1405581
20. Kim, J. H., Shamsuddin, A., & Lim, K. P. (2011). Stock return predictability and the adaptive markets hypothesis. Evidence from century-long US data. *Journal of Empirical Finance, 18*(5), 868-879. https://doi.org/10.1016/j.jempfin.2011.08.002
21. Kirby, C., & Ostdiek, B. (2012). It's all in the timing: simple active portfolio strategies that outperform naive diversification. *Journal of Financial and Quantitative Analysis, 47*(2), 437-467. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1530022
22. Kolmogorov, A. N. (1933). *Sulla determinazione empirica di una*

legge di distribuzione (pp. 83-91). G. Ist. Ital. Attuari. Retrieved from https://www.scienceopen.com/document?vid=c3c08573-63b2-4153-a72e-97bd1b3663a0

23. Malevergne, Y., Pisarenko, V., & Sornette, D. (2005). Empirical distributions of stock returns: between the stretched exponential and the power law? *Quantitative Finance, 5*(4), 379-401. https://doi.org/10.1080/14697680500151343
24. Markowitz, H. (1952). Portfolio selection. *Journal of Finance, 7*(1), 77-91. https://doi.org/10.1111/j.1540-6261.1952.tb01525.x
25. Massey, Jr. F. J. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association, 46*(253), 68-78. Retrieved from https://www.tandfonline.com/doi/abs/10.1080/01621459.1951.10500769
26. Michail, N. A. (2019). Stock market efficiency 2000–2014: the role of the Great Recession. *International Journal of Banking, Accounting and Finance, forthcoming.*
27. Pesaran, M. H., & Pick, A. (2011). Forecast combination across estimation windows. *Journal of Business & Economic Statistics, 29*(2), 307-318. Retrieved from https://www.jstor.org/stable/25800802?seq=1
28. Smirnov, N. (1948). Table for estimating the goodness of fit of empirical distributions. *Annals of Mathematical Statistics, 19*(2), 279-281. Retrieved from https://www.jstor.org/stable/2236278?seq=1
29. Von Mises, R. (1928). *Statistik und wahrheit*. Julius Springer.
30. Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin, 1*(6), 80-83. Retrieved from https://www.jstor.org/stable/3001968?seq=1
31. Yilmaz, K. (2012). Martingale property of exchange rates and central bank interventions. *Journal of Business & Economic Statistics, 21*(3), 383-395. Retrieved from https://www.jstor.org/stable/1392587?seq=1
32. Zaiontz, C. (2015). *Real Statistics Resource Pack*. Retrieved from http://www.real-statistics.com

http://dx.doi.org/10.21511/imfi.17(3).2020.15