

RESEARCH ARTICLE

Nonparametric Predictive Inference for Stock Returns

Rebecca M. Baker^a, Tahani Coolen-Maturi^{b*} and Frank P.A. Coolen^c

^a*Financial Strategy Group, Mercer, UK*; ^b*Durham University Business School, Durham University, Durham, DH1 3LB, UK*; ^c*Department of Mathematical Sciences, Durham University, Durham, DH1 3LE, UK*

In finance, inferences about future asset returns are typically quantified with the use of parametric distributions and single-valued probabilities. It is attractive to use less restrictive inferential methods, including nonparametric methods which do not require distributional assumptions about variables, and imprecise probability methods which generalise the classical concept of probability to set-valued quantities. Main attractions include the flexibility of the inferences to adapt to the available data and that the level of imprecision in inferences can reflect the amount of data on which these are based. This paper introduces nonparametric predictive inference (NPI) for stock returns. NPI is a statistical approach based on few assumptions, with inferences strongly based on data and with uncertainty quantified via lower and upper probabilities. NPI is presented for inference about future stock returns, as a measure for risk and uncertainty, and for pairwise comparison of two stocks based on their future aggregate returns. The proposed NPI methods are illustrated using historical stock market data.

Keywords: Imprecise probability; lower and upper probability; nonparametric predictive inference; pairwise comparison; stock returns.

1. Introduction

Stock return predictability is one of the most widely discussed topics in the finance literature, for a recent survey see Rapach and Zhou [32]. Stock return forecasting is a challenging research area, it is argued that stock returns consists of a large unpredictable component, in which best forecasting methods are able only to explain a small part of stock returns. Despite such concerns, many studies have been devoted to forecast stock returns, see e.g. Rapach and Zhou [32] and references therein. Recent studies provide strategies in order to improve stock return forecasting by taking into account model uncertainty and parameter instability [30], among these strategies are economically motivated model restrictions [20], combination of forecasts [31], diffusion indices [28], and regime shifts [23].

Approaches to predicting stock returns fall mainly into three broad categories: fundamental analysis, technical analysis and quantitative analysis [35]. Fundamental analysis involves analysing the business performance of the issuing company [5]. This includes investigation of macroeconomic factors that may affect the company's value (i.e. conditions within the economy and industry as a whole) as well as company-specific factors. This approach incorporates both quantitative and qualitative factors. Technical analysis is purely based on analysing historical market

*Corresponding author. Email: tahani.maturi@durham.ac.uk

data [5]. Stock prices observed on previous days are used to identify trends regarding the stock's future price. There are various approaches used when analysing historical data, including charting techniques and price-based indicators designed to identify trends. Quantitative analysis also makes use of historical market data, but employs mathematical and statistical modelling techniques. Models are developed and calibrated based on past data, and can then be used to make predictions about the future performance of a stock. This is most commonly done using Monte Carlo simulation: stochastic models are used to repeatedly simulate future stock returns, and key statistics such as the mean, median and standard deviation are calculated based on the output. The probability of achieving or exceeding a given return can also be calculated via Monte Carlo simulation.

Statistical inference about future stock returns is usually based on classical probability theory with precise probabilities satisfying Kolmogorov's axioms. Imprecise probability is a generalisation of classical probability theory enabling various less restrictive representations of uncertainty [1]. Recent research in this field has led to various new approaches to statistical inference, one of which is nonparametric predictive inference (NPI) [2, 6]. NPI is a frequentist statistics framework that uses lower and upper probabilities and has attractive properties from several perspectives [2, 6].

The NPI forecasting method presented in this paper is a quantitative analysis method which attempts to make only few assumptions in addition to available data, and provides an interesting alternative to Monte Carlo simulation as will be illustrated and discussed in this paper. The explicitly predictive nature of NPI, within the frequentist statistics framework, makes it a natural method for forecasting, where it is particularly important to emphasize that, when forecasting multiple future observations, these are explicitly interdependent, which corresponds to the idea of forecasting one observation, then adding that value to the data and forecast the next future observation based on the original data and the previous forecasted value. This sequential idea continues in the natural manner, and is implicitly done by the NPI approach. A major consequence of this is that the forecasts exhibit more variation than would be the case if the multiple future observations were assumed to be conditionally independent, given the original data. This important feature has also been studied in relation to an NPI-based alternative to bootstrapping, where the increased variability avoids the underestimation of variance from which the traditional bootstrap suffers [4].

This paper presents NPI for future stock returns and NPI-based risk measures, and illustrates these using historical stock market data. NPI for comparison of two stocks based on their future aggregate returns is also presented. In Section 2 relevant background literature is summarised. Section 3 presents the new NPI methods for stock returns, these are illustrated via an example in Section 4. Section 5 concludes the paper with discussion of several related issues.

2. Preliminaries

In classical probability theory, the probability for an event E is given by a precise value $p(E) \in [0, 1]$, where $p(\cdot)$ is a probability satisfying Kolmogorov's axioms. However, a precise probability is not always an appropriate measure when faced with incomplete information or knowledge. An alternative approach is to use imprecise probability, which is an umbrella term encompassing a wide range of generalizations of classical probability theory [1]. Imprecise probability is a well-established concept, a historical overview is presented by Hampel [22]. In the past two decades, interest in this field has increased and significant progress has been

made, both on theory and aspects of implementation [1]. In theory of imprecise probability, uncertainty about event E is quantified by lower probability $\underline{P}(E)$ and upper probability $\overline{P}(E)$ with $0 \leq \underline{P}(E) \leq \overline{P}(E) \leq 1$. Classical, precise probability is the special case with $\underline{P}(E) = \overline{P}(E)$, while the vacuous statement $\underline{P}(E) = 0$ and $\overline{P}(E) = 1$ reflects complete lack of knowledge about the event E . Of course, the aim is to assign meaningful and non-trivial lower and upper probabilities to events of interest, where $\underline{P}(E)$ can be interpreted as reflecting the evidence in favour of event E , and $\overline{P}(E)$ the evidence against E , hence in favour of the complementary event E^c , also reflected through the conjugacy property $\overline{P}(E) = 1 - \underline{P}(E^c)$ [1].

One frequentist statistical method which provides meaningful imprecise probabilistic inferences is nonparametric predictive inference (NPI), which is based on Hill's assumption $A_{(n)}$ [24]. NPI gives direct lower and upper probabilities for one or more future real-valued random quantities based on observed values of n related random quantities [2, 8]. Effectively, in NPI it is assumed that the rank of a future real-valued observation among n observed values is equally likely to have each possible value, so from 1 to $n + 1$. While this assumption is implied by exchangeability [16] before the n observations become available, $A_{(n)}$ keeps this property after the n observations have become available. Hence, $A_{(n)}$ can be regarded as a post-data exchangeability assumption, such that the future observation is equally likely to be in each interval of the partition of the real-line, or part of this if observations are known or assumed to certainly belong to it, as created by the n observations. For ease of presentation it is assumed throughout this paper that there are no tied observations; if these occur then they can be dealt with by assuming that such observations differ by a very small amount, a common method to break ties in statistics.

Inferences based on $A_{(n)}$ are predictive and nonparametric, and can be considered suitable if there is hardly any knowledge about the random quantity of interest, other than the n observations, or if one does not want to use such information, e.g. to study effects of additional assumptions underlying other statistical methods. Such inferences are exactly calibrated [27], which strongly justifies their use from frequentist statistics perspective. $A_{(n)}$ is not sufficient to derive precise probabilities for many events of interest, but optimal bounds for probabilities for all events of interest can be derived via the 'fundamental theorem of probability' [16]. These optimal bounds are lower and upper probabilities [1, 2], and are applied in NPI to a range of statistical applications, where through the use of latent variable representations also methods for Bernoulli and multinomial data have been developed [3, 6, 7, 9, 10]. A generalization of $A_{(n)}$ in order to deal with right-censored observations has also been presented [12] and was e.g. used in the development of NPI-based methods for opportunity-based replacement models in operational research [14, 15]

NPI has also been developed for multiple future observations, say m future observations, based on data consisting of n observations. This is based on $A_{(n+m-1)}$, which implies $A_{(n+k)}$ for all $k = 0, \dots, m - 2$ [6, 24]. This can also be viewed as a post-data version of a finite exchangeability assumption for $n + m$ random quantities. The implication is that each future observation is equally likely to fall in any interval I_j , and all possible orderings of the n data observations and the m future observations are equally likely. There are $\binom{n+m}{n}$ possible orderings in total, so the probability of any specific ordering of the m future observations among the n data observations is $\binom{n+m}{n}^{-1}$. It is important to emphasize that this inference for m future observations implicitly takes the interdependence of these future observations into account, so such simultaneous inference for m observations is identical to sequential inference where one first considers a single future observation based

only on the n available data observations, then a second one based on the n data and one earlier future observation, and so on. As mentioned in the introduction, this is crucial as it provides a correct view on variability in forecasting and ensures that strong frequentist consistency properties hold. For any event involving the m future observations, the numbers of orderings for which this event must hold or can hold are of interest. Generally in the NPI framework, the lower probability for an event of interest is derived by counting all orderings for which this event must hold, and the corresponding upper probability is derived by counting all orderings for which this event can hold.

3. NPI-based inferences about stock returns

This section presents the novel application of NPI for stock returns. Prediction of future stock prices is presented, followed by consideration of measuring risk and uncertainty. Finally, comparison of future returns of two stocks is presented. The novel methods will be illustrated and discussed in an example in Section 4.

3.1 Predicting stock returns

Let V_t be the value of a stock at time t . Throughout this paper time is considered to be discrete with intervals of equal length between observations. Assume that a time series of historical stock prices $V_t = v_t$, $t = 0, \dots, n$, is available, so the historical returns r_t can be calculated as

$$r_t = \frac{v_t - v_{t-1}}{v_{t-1}} \text{ for } t \in \{1, \dots, n\}$$

In order to use the NPI framework for inferences about future returns, the explicit assumption is henceforth made that the order of the n observed returns r_1, \dots, r_n is irrelevant. Of course, this assumption excludes aspects of time series beyond a general trend to be taken into account; developing NPI for such situations provides an interesting and important topic for future research. For ease of notation and without loss of generality, these observations are relabeled such that $r_1 < r_2 < \dots < r_n$, and it is assumed that there are no ties within the set of observed returns (see the comment on NPI in case of ties in Section 2). Assume a lower bound r_0 and an upper bound r_{n+1} for the range of possible returns, such that all observed returns are, and all future returns are assumed to be, in the interval $[r_0, r_{n+1}]$. The choice of these values will have some effect on the inferences so they must be considered with some care, but from practical perspective this does not really impact on the applicability of the presented method. The range of possible future returns is now partitioned into $n + 1$ intervals $I_i = (r_{i-1}, r_i)$, for $i = 1, \dots, n + 1$, and NPI is used for inference about m future returns R_t , $t = n + 1, \dots, n + m$.

Applying the general theory of NPI for multiple real-valued observations [6], there are $\binom{n+m}{m}$ orderings of the m future returns within the $n + 1$ intervals I_i , $i = 1, \dots, n + 1$, and all these orderings are equally likely. This enables inference about future returns by counting the number of orderings which satisfy a chosen stock selection criterion. The NPI lower probability for the event that the criterion is satisfied is derived by counting the number of orderings for which this criterion must be satisfied, while the corresponding NPI upper probability is derived by counting the number of orderings for which this criterion can be satisfied.

There are various stock selection criteria that could be used, depending on an investor's aims and preferences. These criteria relate to the future value V_{n+m} of the

stock, so m time units from now, or equivalently to the aggregate return achieved over the m future periods. One could also define criteria considering returns at multiple future time points, this is not considered here but NPI methods for such criteria could also be developed and would involve similar counting arguments. Restricting attention to the value m time units into the future, let the aggregate future return be denoted by \tilde{R} , then

$$\tilde{R} = \left[\prod_{t=n+1}^{n+m} (1 + R_t) \right]^{\frac{1}{m}} - 1 \quad (1)$$

Note that, for simplicity, m is not reflected in the notation \tilde{R} .

Next, the NPI lower and upper probabilities are derived for some events of interest involving the aggregate return \tilde{R} and a target return T_R . The main idea for these results is straightforward. Consider a particular ordering of the m future returns, $R_{n+j} \in I_{i_j} = (r_{i_j-1}, r_{i_j})$, where $i_j \in \{1, 2, \dots, n+1\}$ and $j = 1, \dots, m$. The maximum lower bound for \tilde{R} , denoted by \tilde{R}_L , is derived by setting each future return in Equation (1) to be equal to the lower bound of the interval in which it falls, that is $R_{n+j} = r_{i_j-1}$, where $i_j \in \{1, 2, \dots, n+1\}$ and $j = 1, \dots, m$. Strictly there is a limit argument here, as the I_{i_j} are open intervals, but the presented bound is indeed optimal and the chosen presentation is far easier than to consider and denote values only infinitesimally greater than the interval bounds. Similarly, the minimum upper bound for \tilde{R} , denoted by \tilde{R}_U , is derived by setting each future return in Equation (1) to be equal to the upper bound of the interval in which it falls, that is $R_{n+j} = r_{i_j}$, where $i_j \in \{1, 2, \dots, n+1\}$ and $j = 1, \dots, m$. Then

$$\underline{P}(\tilde{R} > T_R) = \frac{1}{\binom{n+m}{m}} \sum_O \mathbf{1}\{\tilde{R}_L > T_R\} \quad (2)$$

$$\overline{P}(\tilde{R} > T_R) = \frac{1}{\binom{n+m}{m}} \sum_O \mathbf{1}\{\tilde{R}_U > T_R\} \quad (3)$$

where \sum_O is the summation over all the $\binom{n+m}{m}$ possible orderings of the m future returns within the $n+1$ intervals, and $\mathbf{1}\{A\}$ is an indicator function which is equal to 1 if A is true and 0 otherwise.

For other events of interest the NPI lower and upper probabilities can be derived with similar counting arguments. For example, an investor might instead focus on minimising downside returns. In this situation, an appropriate selection criterion could be to set an upper limit p^* for the upper probability that the return drops below a given value, i.e. to select stocks such that $\overline{P}(\tilde{R} < T_R) \leq p^*$. The use of the upper probability leads to a prudent selection criterion in line with the risk averse nature of most investors. Another investor might wish to be quite certain to achieve a specified return, in which case a possible selection criterion would be to set a lower limit p_* on the lower probability that the return exceeds a given value, i.e. to select stocks such that $\underline{P}(\tilde{R} > T_R) \geq p_*$. In this case, use of the lower probability leads to a prudent criterion.

3.2 Measuring risk and uncertainty

The NPI approach can be used to assess the investment risk of a stock. Investors often have investment aims or restrictions for which they wish to take the risk of an investment into account, which is commonly done using statistics such as the

standard deviation and Value-at-Risk (VaR). The NPI method also enables such considerations, actually there are various ways that NPI predictive lower and upper probabilities can be used to measure risk and uncertainty. Two methods are proposed: the first compares two (or more) investments and assesses their relative levels of risk, the second method can be used to calculate a quantitative risk measure which assesses the level of uncertainty regarding the aggregate future return \tilde{R} .

Two investments can be compared by considering the predictive probability interval $\left[\underline{P}(\tilde{R} > T_R), \overline{P}(\tilde{R} > T_R) \right]$ for each of them, either just for a single target return or for several target return levels. If these intervals overlap for the same target return, it can be interpreted as an inconclusive comparison of the investments at this target return level. If one interval is fully to the right of the second on the real line, then the first investment can be said to dominate the second at this target return level.

When considering a high target return, the predictive probability interval for a relatively risky investment is likely to dominate that for a safer or less volatile investment. However, when considering a low target return, a risky investment is likely to be dominated by a safer investment. This is due to the fact that although risky investments have the potential for high returns, they also tend to experience more severe downside returns. This method of comparing investments provides a more detailed picture of the predicted future returns than the single selection criterion $\underline{P}(\tilde{R} > T_R) \geq p_*$, and it enables conclusions about the relative levels of risk and uncertainty for different investments.

The level of uncertainty regarding future returns can also be investigated by calculating various quantiles of the range of possible values for \tilde{R} . For example, one can focus on the first and third quartiles corresponding to the lower probabilities for \tilde{R} , i.e. Q_1 and Q_3 such that $\underline{P}(\tilde{R} > Q_1) = 0.75$ and $\underline{P}(\tilde{R} > Q_3) = 0.25$. The range of returns spanned by the interval $[Q_1, Q_3]$ gives an indication of the level of uncertainty when predicting future returns, and $Q_3 - Q_1$ can be used as a quantitative uncertainty measure, which will be referred to as the return range uncertainty measure. Of course, one can use different quantiles and also focus on upper probabilities, or both lower and upper probabilities, in such considerations.

3.3 Comparing two stock returns

An attractive event to consider in order to compare two stocks is that the future return from one stock will exceed, by at least some constant δ , the return from the other stock at m units of time from now. For ease of presentation, suppose that n historical returns are available from both stock A and stock B, this can straightforwardly be generalized to different numbers of data for the two stocks. Let O_A and O_B be the possible orderings of the m future returns from stocks A and B, respectively, among their n observed returns. Assuming that the returns of these two stocks are independent random quantities, the NPI lower and upper probabilities for the event that the future aggregate return from stock A will exceed, by at least δ , the future aggregate return from stock B at time m , are

$$\underline{P}(\tilde{R}^A > \tilde{R}^B + \delta) = \frac{1}{\binom{n+m}{m}^2} \sum_{l_A \in O_A} \sum_{l_B \in O_B} \mathbf{1}\{\tilde{R}_{L,l_A}^A > \tilde{R}_{U,l_B}^B + \delta\} \quad (4)$$

$$\overline{P}(\tilde{R}^A > \tilde{R}^B + \delta) = \frac{1}{\binom{n+m}{m}^2} \sum_{l_A \in O_A} \sum_{l_B \in O_B} \mathbf{1}\{\tilde{R}_{U,l_A}^A > \tilde{R}_{L,l_B}^B + \delta\} \quad (5)$$

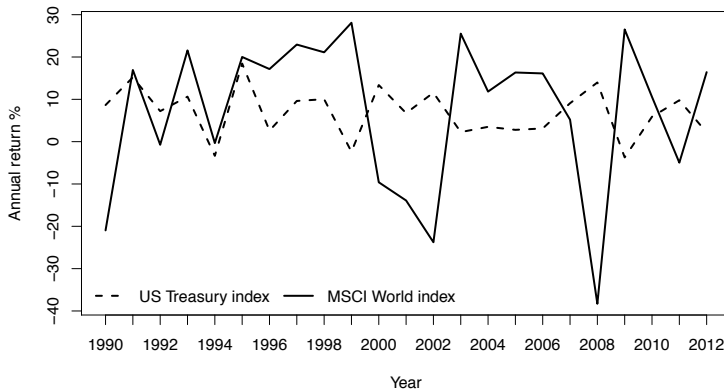


Figure 1. Annual returns for MSCI World index and US Treasury Master index

The proofs of these lower and upper probabilities are given in the Appendix, as usual in NPI these are based on counting combinations of orderings from O_A and O_B for which the event of interest must occur, to derive the lower probability (4), and for which it can occur, to derive the upper probability (5).

Based on these lower and upper probabilities, one could interpret $P(\tilde{R}^A > \tilde{R}^B + \delta) \geq 0.5$ as strong evidence that the future aggregate return from stock A will exceed by at least δ the future aggregate return from stock B , although of course one could choose different values or include the upper probability in more detailed considerations. The results presented in this section can be extended to allow comparison of several stocks and for different events of interest, along the lines of the multiple comparisons methodology developed by Coolen and van der Laan [11] and Maturi [29].

We should emphasize that the method presented here is not explicitly aimed at optimal investment decisions or these two stocks. It will be of interest to develop the NPI approach further for such aims, for example to see how it could be applied in scenarios such as pairs trading [18, 19, 21].

4. An illustrative example

The NPI methods for future stock returns, presented in this paper, are illustrated using two real-world data sets, considering indices instead of single stocks. The first data set consists of the total returns on the MSCI World Index, which comprises 1,606 stocks from 23 developed markets across the world, providing an indication of stock market performance for global developed markets as a whole. The second data set consists of the BoA US Treasury Master index (USTre). This index measures the total return on a universe of US government bonds of all maturities, and has been chosen because it represents a comparatively less risky investment prospect than the MSCI World index. As data in this example, the annual returns on both indices spanning the period from 1990 until 2012 are used, the 23 observations for each index over this period are presented in Figure 1. The mean annual return over the period was about 7% on both indices. However, the standard deviation of MSCI World returns was about 18% while the standard deviation of US Treasury returns was about 6%. In the following illustration of the new NPI methods presented in this paper, attention is restricted to future returns over the next 3 and 6 years, of course other or more future time points could be considered.

The 23 annual returns (per index) are ordered from the smallest to the largest, and labelled $r_1 < \dots < r_{23}$. Appropriate values for the lower and upper bounds for the range of possible returns must be chosen, denoted by r_0 and r_{24} . The interval $[r_0, r_{24}]$ is partitioned into 24 intervals (r_{i-1}, r_i) for $i = 1, \dots, 24$, where open intervals are used in line with the general NPI approach [6] but, for continuous data, this is of little practical relevance and one could assume half-open or closed intervals without affecting the inferences. Consider future returns over the next $m = 3$ and $m = 6$ years, this leads to $\binom{26}{3} = 2,600$ and $\binom{29}{6} = 475,020$ possible arrangements of the 3 and 6 future observations among the 23 data observations, respectively, for each index. To illustrate the use of NPI for the event that the aggregate future return over 3 and over 6 years will exceed one or more specific targets, consider target return levels of $T_R = 0, 1, \dots, 10\%$. The method presented in Section 3.1 was implemented with the statistical software R, which was used for all computations in this example. First, the NPI method described in Section 3.2 was applied in order to compare the relative levels of risk of these two indices. The NPI lower and upper probabilities were calculated for events $\tilde{R} > T_R$, for each index, for target return levels $T_R = 0, 1, \dots, 10\%$. In order to allow a fair comparison of the indices, for both data sets the lower and upper limits of the range of possible returns are set at $r_0 = -40\%$ and $r_{24} = 30\%$.

The results of the analysis are shown in Table 1, they show several interesting features. First, imprecision (the difference between corresponding upper and lower probabilities) tends to be large for the larger horizon ($m = 6$) than for the smaller horizon ($m = 3$). Typically, in NPI imprecision tends to be a decreasing function of the number of available data observations, and an increasing function of the prediction horizon, this is in line with intuition. A further aspect of the NPI approach that must be emphasized, and in which it differs crucially from other approaches such as Monte Carlo simulation, is that the m future observations which are jointly considered for inference are mutually dependent [6]. Perhaps the easiest way to think about this is that the results based on all $n + m$ orderings of data observations and future observations being equally likely, are identical to sequential application of NPI for a single observation each time, where a new observation is added to the data set before prediction of the next observation. Hence, there is more variation in the NPI approach with multiple future observations than for Monte Carlo simulation or similar approaches. Note that this increased variation is fully in line with frequentist theory as the NPI approach is exactly calibrated [27]. A more detailed study of this feature, in particular in comparison to standard bootstrap approaches and with application to a range of inferences in finance, has been initiated, we hope to report on progress in the near future.

Some aspects of the results reported in Table 1 that are worth noticing are as follows. Consider a potential investor who is interested in the events that the aggregate future return will exceed 0% (i.e. that the return will be positive), 7% (i.e. that the return will exceed the long-term average) and 10% (which would represent a strong return on this equity index). When considering the future returns over 6 years, the NPI lower and upper probabilities for exceeding an annual return of 7% and 10% in aggregate tend to be smaller than the corresponding ones over 3 years, even though this longer time horizon has led to increased imprecision. To get a positive return in $m = 6$ years has higher lower and upper probabilities for the MSCI than for $m = 3$ while for the USTRe they are quite similar when the increased imprecision is taken into account. There are several forces at play in this method of prediction, these are directly related to the observed data. For example, one or more rather extreme observed returns are likely to influence prediction more strongly for a shorter future time horizon, and also the assumed bounds r_0 and

$T_R\%$	$m = 3$			$m = 6$		
	MSCI index	USTRe index	Dominant	MSCI index	USTRe index	Dominant
0	[0.6277, 0.7308]	[0.8631, 0.9785]	USTRe	[0.6554, 0.7837]	[0.8015, 0.9958]	USTRe
1	[0.6119, 0.7127]	[0.8327, 0.9481]	USTRe	[0.6195, 0.7525]	[0.7856, 0.9880]	USTRe
2	[0.5854, 0.6854]	[0.8065, 0.9219]	USTRe	[0.5822, 0.7184]	[0.7649, 0.9705]	USTRe
3	[0.5619, 0.6608]	[0.7500, 0.8654]	USTRe	[0.5435, 0.6820]	[0.7302, 0.9366]	USTRe
4	[0.5308, 0.6277]	[0.6931, 0.8085]	USTRe	[0.5039, 0.6431]	[0.6732, 0.8792]	USTRe
5	[0.5054, 0.6004]	[0.6092, 0.7246]	USTRe	[0.4634, 0.6022]	[0.5899, 0.7940]	Overlap
6	[0.4746, 0.5669]	[0.5162, 0.6315]	Overlap	[0.4229, 0.5596]	[0.4828, 0.6827]	Overlap
7	[0.4458, 0.5350]	[0.4192, 0.5331]	Overlap	[0.3825, 0.5155]	[0.3636, 0.5550]	Overlap
8	[0.4227, 0.5081]	[0.3196, 0.4327]	Overlap	[0.3430, 0.4708]	[0.2483, 0.4254]	Overlap
9	[0.3877, 0.4704]	[0.2338, 0.3435]	MSCI	[0.3056, 0.4262]	[0.1515, 0.3074]	Overlap
10	[0.3677, 0.4465]	[0.1596, 0.2650]	MSCI	[0.2697, 0.3817]	[0.0818, 0.2109]	MSCI

Table 1. NPI lower and upper probabilities $[P, \bar{P}]$ for the event $(\tilde{R} > T_R)$

Index	$m = 3$			$m = 6$		
	Q_1	Q_3	$Q_3 - Q_1$	Q_1	Q_3	$Q_3 - Q_1$
MSCI World	-4.96%	14.05%	19.01%	-2.96%	10.56%	13.52%
US Treasury Master	3.01%	8.81%	5.80%	2.51%	7.98%	5.47%

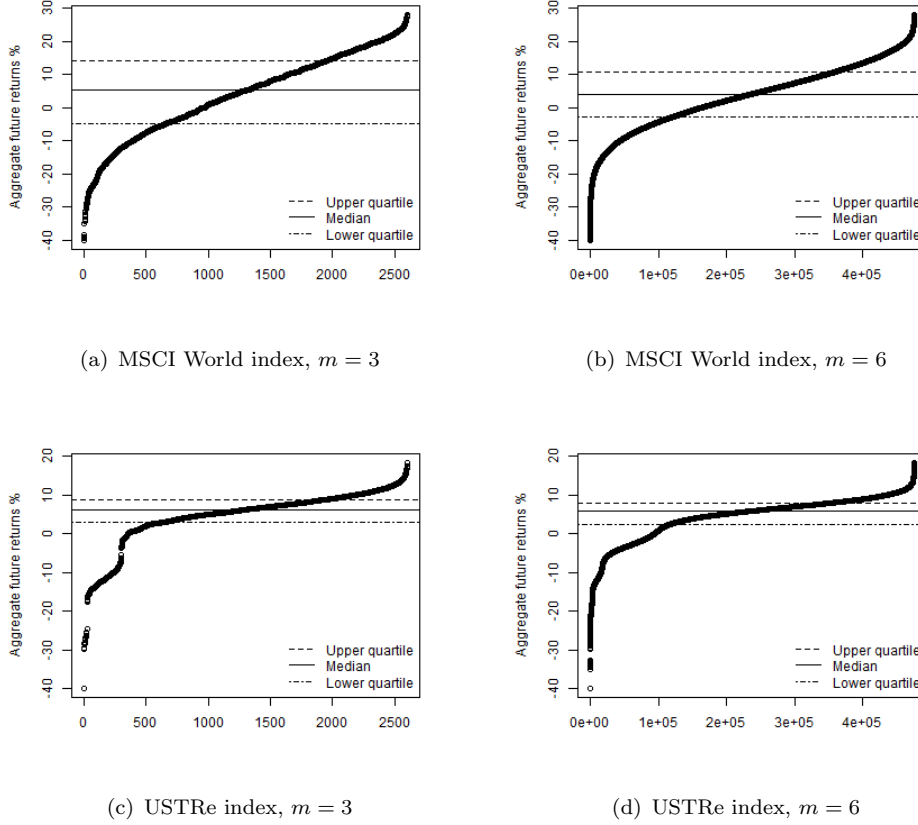
Table 2. Return range uncertainty measure

r_{24} for the range of possible returns have more influence on a shorter time period. A further force is the aforementioned dependence of the m future observations, which causes greater variability which particularly affects the MSCI index as it had substantially more variability in the observed data.

Table 1 also shows, for each individual case, the results of comparison of the MSCI and USTRe indices as previously discussed, so depending on whether or not the respective intervals are overlapping. This clearly reflects that if one aims at high return, MSCI is the best investment in the sense of having substantially higher lower and upper probabilities of achieving high return. USTRe is the best option in order to achieve low but positive returns, effectively because it is unlikely to lead to negative returns which are substantially more likely for MSCI. For values around the historic average of about 7% annual return, achieved by both indices over the last 23 years, the intervals are overlapping and one could not base a strong preference for either index on these data when using this specific NPI method.

The NPI methodology for calculating an uncertainty measure based on quantiles of the range of possible values for \tilde{R} , as presented in Section 3.2, leads to the results shown in Table 2, with attention restricted to the lower and upper quartiles for the lower probabilities. The return range uncertainty measure is much larger for the MSCI World index than for the US Treasury Master index, reflecting the substantially greater variability in the data for the former index and hence that the latter is less risky. This is illustrated further by the plots in Figure 2, which show the range of aggregate future lower returns, \tilde{R}_L , for both indices, and for $m = 3, 6$, as well as the lower and upper quartiles Q_1 and Q_3 , and the median. As mentioned before, these \tilde{R}_L are derived by taking all possible orderings of m future returns with the n observed return values, and then putting each future value at the left end point of the interval that is part of the partition created by the n observed return values. Note that particularly the tails of these distributions are quite strongly affected by the choice of bounds for the range of the return values, set at $r_0 = -40\%$ and $r_{24} = 30\%$. The area of particular interest here, namely between Q_1 and Q_3 , will be hardly affected by this choice, as orderings in which one out of 3 or 6 values occur in an end interval will mostly lead to aggregate lower return outside this range. A similar plot can be made for the aggregate future upper returns, \tilde{R}_U , this is not presented here as the idea is similar and focus is on the lower returns, which is more interesting with a view to risk assessment and management.

An explicit way to compare two indices was presented in Section 3.3, namely by considering the event that the future aggregate return of one index will exceed,

Figure 2. Aggregate future lower returns values \tilde{R}_L

by at least some amount δ , the future aggregate return of the other index. This is illustrated for the case considered in this example, again with $r_0 = -40\%$ and $r_{24} = 30\%$ for both indices and with attention restricted to future returns over the next $m = 3$ years. The NPI lower and upper probabilities for the event that the future aggregate return from the US Treasury index (MSCI World index) will exceed by at least $\delta = 0, 1, \dots, 30\%$ the future aggregate return from the MSCI World index (US Treasury index) are given in Figure 3(a) (Figure 3(b)). Of course, these lower and upper probabilities are decreasing as function of δ . The basic case of $\delta = 0$ shows lower and upper probabilities which are nearly symmetric around 0.5, which is due to both indices historically having average annual returns of about 7%. The lower and upper probabilities in Figure 3(b) decrease faster than those in Figure 3(a), which is a result of the detailed orderings of the data, and hence of all future orderings per index considered; it does not necessarily reflect the greater variability of the data for the MSCI index because this variability occurs in both tails, hence there are two forces at play in such comparisons which tend to cancel each other out. These NPI lower and upper probabilities are attractive, as they address directly events of interest to an investor. The nature of these inferences is very different from established frequentist statistics methods, where for example a hypothesis of equal future returns might be tested, which gives less flexibility in focussing on the practical questions of real interest and is also harder to interpret for many people.

As mentioned before, the choice of the lower and upper limits for the range of possible return values, set at $r_0 = -40\%$ and $r_{24} = 30\%$ for both indices in this example thus far, will have some influence on the inferences. To explore this fur-

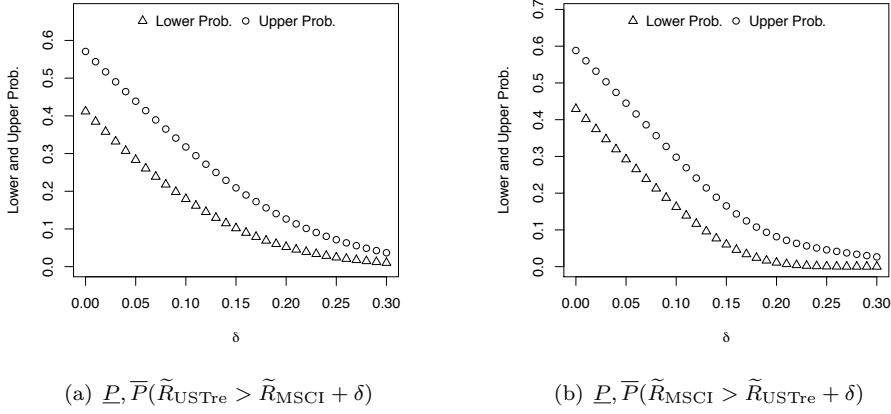


Figure 3. NPI lower and upper probabilities

	(1)	(2)	(3)	(4)	(5)
MSCI	$r_0 = -40\%, r_{24} = 30\%$	$r_0 = -40\%, r_{24} = 30\%$	$r_0 = -40\%, r_{24} = 40\%$	$r_0 = -50\%, r_{24} = 30\%$	$r_0 = -40\%, r_{24} = 30\%$
USTre	$r_0 = -40\%, r_{24} = 30\%$	$r_0 = -10\%, r_{24} = 30\%$	$r_0 = -10\%, r_{24} = 30\%$	$r_0 = -10\%, r_{24} = 30\%$	$r_0 = -10\%, r_{24} = 20\%$
δ	$P, \bar{P}(\tilde{R}_{MSCI} > \tilde{R}_{USTre} + \delta)$				
0	[0.4293, 0.5882]	[0.4293, 0.5631]	[0.4293, 0.5695]	[0.4293, 0.5631]	[0.4395, 0.5631]
0.05	[0.2922, 0.4449]	[0.2922, 0.4108]	[0.2922, 0.4198]	[0.2922, 0.4108]	[0.3010, 0.4108]
0.10	[0.1625, 0.2978]	[0.1625, 0.2565]	[0.1625, 0.2678]	[0.1625, 0.2565]	[0.1675, 0.2565]
0.15	[0.0604, 0.1653]	[0.0604, 0.1191]	[0.0604, 0.1334]	[0.0604, 0.1191]	[0.0615, 0.1191]
0.20	[0.0110, 0.0815]	[0.0110, 0.0347]	[0.0110, 0.0480]	[0.0110, 0.0347]	[0.0111, 0.0347]
δ	$P, \bar{P}(\tilde{R}_{USTre} > \tilde{R}_{MSCI} + \delta)$				
0	[0.4118, 0.5707]	[0.4369, 0.5707]	[0.4305, 0.5707]	[0.4369, 0.5708]	[0.4369, 0.5605]
0.05	[0.2832, 0.4387]	[0.2998, 0.4387]	[0.2957, 0.4387]	[0.2998, 0.4403]	[0.2998, 0.4283]
0.10	[0.1794, 0.3175]	[0.1889, 0.3175]	[0.1865, 0.3175]	[0.1889, 0.3285]	[0.1889, 0.3078]
0.15	[0.1019, 0.2089]	[0.1069, 0.2089]	[0.1059, 0.2089]	[0.1069, 0.2332]	[0.1069, 0.1999]
0.20	[0.0523, 0.1265]	[0.0546, 0.1265]	[0.0542, 0.1265]	[0.0546, 0.1523]	[0.0546, 0.1187]

Table 3. NPI lower and upper probabilities

ther, the pairwise comparison method above was repeated for a variety of different choices for these limits, differing for the two indices, which seems reasonable as the variabilities of the returns differ noticeably. Table 3 presents the results for five cases with different lower and upper limits, and only considering the values $\delta = 0, 5, 10, 15, 20\%$. It is clear that, for these specific inferences, hence for each specific value of δ and comparing the entries in the same row of the table, the choice of the limits does not have a substantial effect on the NPI lower and upper probabilities. As is easily understood, the influence of these limits is also decreasing as a function of the number of available data, this follows directly from the number of intervals in the partition formed by the data, it is not illustrated in this example.

To end this example, which illustrates the new theoretical methods presented in this paper, it is interesting to compare the NPI approach with Monte Carlo simulation, an established and indeed commonly used statistical method for predictive inferences related to stock returns. As stock market returns are often modelled by a lognormal distribution, this was assumed for the MSCI index in this example, with the two parameters estimated on the basis of the 23 observations for this index. The resulting lognormal distribution was used to simulate 2,500 m -year future returns. In order to compare the results with the earlier reported NPI method for this specific index, attention is restricted to predictions for $m = 3$ and $m = 6$ years ahead, and the events that the aggregate future returns over the next 3 and 6 years will exceed 0, 7, 10%. The results from this simulation study are presented in Table 4.

The Monte Carlo simulation method gives a higher probability for achieving a positive return over the 6-year period than the 3-year period, but a smaller probability for exceeding an aggregate annual return of 10% with the longer time

Event	$m = 3$	$m = 6$
$\tilde{R} > 0\%$	0.7268	0.8024
$\tilde{R} > 7\%$	0.4392	0.4220
$\tilde{R} > 10\%$	0.3208	0.2652

Table 4. Predictive probabilities based on 2,500 Monte Carlo simulations (MSCI index)

horizon. This is likely to result from the larger impact of relatively extreme values in the simulation over the shorter period than over the longer period. The comparison with the NPI results, as presented in Table 1, is not straightforward, as there are two main differences. First, the NPI approach does not make any assumptions about an underlying probability distribution, while the Monte Carlo method explicitly used the assumed lognormal distribution. Secondly, and this is a very important difference between the two methods, the future observations in the NPI approach are dependent, which in relation to a simulation as done for the Monte Carlo method means that, if the first sampled future return value were large, it becomes slightly more likely for the next one to also be a bit larger. This leads to more variability for the future values in the NPI method than for the Monte Carlo method, however it should be emphasized that, from theoretical perspective, this greater variability is fully justified [6]. More detailed study of this effect, in particular also by comparing the NPI approach to bootstrap methods, is ongoing, results are expected to be reported in the near future. In this example, this comparison of the results from the Monte Carlo simulation and the NPI method shows that NPI tends to lead to somewhat greater lower and upper probabilities for the events that the aggregate future return exceeds 7% and 10%, compared to the Monte Carlo probability, but mostly smaller lower and upper probabilities for the event that the aggregate future return will be positive. As mentioned, there are several aspects that influence these results, with the greater variability in the NPI method likely to be the cause for the results for the larger positive aggregate future return levels, but mostly the results from the two methods are not too far apart, with the Monte Carlo probability either inside or close to the interval created by the corresponding NPI lower and upper probabilities. One could also use this as an argument to support the assumption of a lognormal distribution as used for the Monte Carlo simulation. A similar comparison for the USTre index led to the same insights and is not reported here.

5. Concluding remarks

This paper presents the first application of NPI to aspects of finance, in particular prediction of stock returns. Of course, in order to derive at meaningful inferences about future stock returns, the underlying data set of historical stock returns must be appropriate. Key considerations include: the time period spanned by the historical data set to be used; the time period between observations, i.e. whether one should use daily returns, weekly returns, etc; and the values assigned to the lower and upper bounds r_0 and r_{n+1} .

The time period spanned by the data set should be considered carefully, since this has implications for the reasonableness of the exchangeability assumption discussed in Section 2. If the time period is excessively long, it may not be reasonable to assume that all past and future returns are exchangeable, since there may have been significant fluctuations in the economy or industry as a whole affecting the level of stock returns. However, the data set must be large enough to provide meaningful inferences about future returns, small values of n compared to m are likely to lead to very imprecise inferences which may not be of much use.

The time period between observations is also an important consideration with regard to the exchangeability assumption. It has been proposed that certain days of the week or times of the year experience significant differences with regard to stock returns and changes in market indices. These anomalies are known as calendar effects, the most important of which are the weekend effect and the January effect [33]. The weekend effect refers to the fact that stocks tend to exhibit relatively large daily returns on Fridays compared with the daily returns on Mondays. The January effect refers to the larger stock returns experienced in January compared with those seen in other months. It may be important to account for these effects when assuming exchangeability within a data set, but adaptation of the NPI method through data manipulation techniques in order to deal with such effects is beyond the scope of this introductory paper. In connection to this topic, and as commented on earlier, the assumption that the original order of the historical returns is irrelevant may not be satisfactory in some real-world applications, for example if there are time series aspects beyond a general trend which are likely to have a substantial impact on the future stock returns. It may be possible to apply NPI in such cases together with some more detailed modelling, investigating this is an important topic for future research. As bootstrap methods are important for many data-based inferences in finance, it is of interest to mention the recently developed NPI approach to bootstrapping [4]. This provides bootstrap samples with more variability than traditional bootstrap methods, and it will be of interest to develop it further for finance applications and to compare it to other variations such as the wild bootstrap [25].

The choice of the values r_0 and r_{n+1} also affects the inferences about future returns, this has been discussed and illustrated in the example. For the analysis undertaken in this paper, r_0 was set slightly below the minimum value in the set of historical returns and r_{n+1} slightly above the maximum value in this set. However, detailed investigation of the impact of these values on specific inferences and in specific applications is strongly recommended, more general guidance could be achieved in future research.

The NPI methods presented in this paper provide already a relatively straightforward approach to a variety of important topics related to future stock returns, and the frequentist statistics properties of NPI provide strong justification for the use of these methods, in particular when there is little further information about the stocks in addition to the historical data, or if one explicitly wishes not to use any additional information. There are many opportunities for extending the results in future research. It will be of interest to further investigate the impact on predictive probabilities of varying the timeframe and frequency of the historical returns data set, the number of future returns considered and the values assigned to the lower and upper limits on future returns. A wider range of stock selection criteria could be investigated in addition to those presented in Section 3.1 and these could be tested on a more extensive range of data sets.

Considering the dependence structure between the stock returns via copula [17, 26, 34] is another interesting research topic. A first approach to develop NPI in order to take such dependence into account has recently been published [13] and research on this topic is ongoing. NPI for multivariate data including data with additional predictors, is also an important topic for future research. Furthermore, the NPI methodology was presented for assessing the investment risk of a stock and for comparing investments with regard to their risk and uncertainty characteristics. This initial investigation into NPI-based risk measures can be extended in several ways, for example by considering different risk measures and linking it to selection of stocks for a portfolio. The application of NPI to financial analysis is a large area of

potential research to which this paper has made an important initial contribution.

Acknowledgement

The authors would like to thank three anonymous reviewers for supporting our work and for their valuable comments that helped us to improve the paper.

Appendix

The lower probability (4) is obtained by deriving the maximum lower bound for $P(\tilde{R}^A > \tilde{R}^B + \delta)$, as follows:

$$\begin{aligned}
& P(\tilde{R}^A > \tilde{R}^B + \delta) \\
&= \frac{1}{\binom{n+m}{m}} \sum_{l_B \in O_B} P(\tilde{R}^A > \tilde{R}^B + \delta | R_{n+j}^B \in (r_{i_j-1}^B, r_{i_j}^B), l_B \in O_B, i_j \in \{1, \dots, n+1\}, j = 1, \dots, m) \\
&\geq \frac{1}{\binom{n+m}{m}} \sum_{l_B \in O_B} P(\tilde{R}^A > \tilde{R}^B + \delta | R_{n+j}^B = r_{i_j}^B, l_B \in O_B, i_j \in \{1, \dots, n+1\}, j = 1, \dots, m) \\
&= \frac{1}{\binom{n+m}{m}} \sum_{l_B \in O_B} P(\tilde{R}^A > \tilde{R}_{U,l_B}^B + \delta) \\
&\geq \frac{1}{\binom{n+m}{m}^2} \sum_{l_A \in O_A} \sum_{l_B \in O_B} \mathbf{1}\{\tilde{R}_{L,l_A}^A > \tilde{R}_{U,l_B}^B + \delta\} = \underline{P}(\tilde{R}^A > \tilde{R}^B + \delta)
\end{aligned}$$

The first (second) inequality follows by setting each future return from stock B (stock A) in Equation (1) to be equal to the upper (lower) bound of the interval in which it falls.

The upper probability (5) is obtained by deriving the minimum upper bound for $P(\tilde{R}^A > \tilde{R}^B + \delta)$, as follows:

$$\begin{aligned}
& P(\tilde{R}^A > \tilde{R}^B + \delta) \\
&= \frac{1}{\binom{n+m}{m}} \sum_{l_B \in O_B} P(\tilde{R}^A > \tilde{R}^B + \delta | R_{n+j}^B \in (r_{i_j-1}^B, r_{i_j}^B), l_B \in O_B, i_j \in \{1, \dots, n+1\}, j = 1, \dots, m) \\
&\leq \frac{1}{\binom{n+m}{m}} \sum_{l_B \in O_B} P(\tilde{R}^A > \tilde{R}^B + \delta | R_{n+j}^B = r_{i_j-1}^B, l_B \in O_B, i_j \in \{1, \dots, n+1\}, j = 1, \dots, m) \\
&= \frac{1}{\binom{n+m}{m}} \sum_{l_B \in O_B} P(\tilde{R}^A > \tilde{R}_{L,l_B}^B + \delta) \\
&\leq \frac{1}{\binom{n+m}{m}^2} \sum_{l_A \in O_A} \sum_{l_B \in O_B} \mathbf{1}\{\tilde{R}_{U,l_A}^A > \tilde{R}_{L,l_B}^B + \delta\} = \overline{P}(\tilde{R}^A > \tilde{R}^B + \delta)
\end{aligned}$$

The first (second) inequality follows by setting each future return from stock B (stock A) in Equation (1) to be equal to the lower (upper) bound of the interval in which it falls.

References

- [1] Augustin, T., Coolen, F.P.A., de Cooman, G., Troffaes, M.C.M., 2014. *Introduction to Imprecise Probabilities*. Wiley, Chichester.
- [2] Augustin, T., Coolen, F.P.A., 2004. Nonparametric predictive inference and interval probability. *Journal of Statistical Planning and Inference* 124, 251–272.
- [3] Baker, R.M., Coolen, F.P.A., 2010. Nonparametric predictive category selection for multinomial data. *Journal of Statistical Theory and Practice* 4, 509–526.
- [4] BinHimd, S., 2014. *Nonparametric Predictive Methods for Bootstrap and Test Reproducibility*. PhD thesis, Durham University, www.npi-statistics.com.
- [5] Bodie, Z., Kane, A., Marcus, A.J., 2011. *Investments and Portfolio Management*. 9 ed., McGraw-Hill, New York.
- [6] Coolen, F.P.A., 2011. Nonparametric predictive inference, in: Lovric, M. (Ed.), *International Encyclopedia of Statistical Science*. Springer, pp. 968–970.
- [7] Coolen, F.P.A., 1998. Low structure imprecise predictive inference for bayes' problem. *Statistics & Probability Letters* 36, 349–357.
- [8] Coolen, F.P.A., 2006. On nonparametric predictive inference and objective bayesianism. *Journal of Logic, Language and Information* 15, 21–47.
- [9] Coolen, F.P.A., Augustin, T., 2005. Learning from multinomial data: a nonparametric predictive alternative to the imprecise dirichlet model, in: *Proceedings of the Fourth International Symposium on Imprecise Probability: Theories and Applications*, pp. 125–134.
- [10] Coolen, F.P.A., Augustin, T., 2009. A nonparametric predictive alternative to the imprecise dirichlet model: the case of a known number of categories. *International Journal of Approximate Reasoning* 50, 217–230.
- [11] Coolen, F.P.A., van der Laan, P., 2001. Imprecise predictive selection based on low structure assumptions. *Journal of Statistical Planning and Inference* 98, 259–277.
- [12] Coolen, F.P.A., Yan, K.J., 2004. Nonparametric predictive inference with right-censored data. *Journal of Statistical Planning and Inference* 126, 25–54.
- [13] Coolen-Maturi, T., Coolen, F.P.A., Muhammad, N., 2016. Predictive inference for bivariate data: Combining nonparametric predictive inference for marginals with an estimated copula. *Journal of Statistical Theory and Practice* , DOI: 10.1080/15598608.2016.1184112.
- [14] Coolen-Schrijner, P., Coolen, F.P.A., Shaw, S.C., 2006. Nonparametric adaptive opportunity-based age replacement strategies. *Journal of the Operational Research Society* 57, 63–81.
- [15] Coolen-Schrijner, P., Shaw, S.C., Coolen, F.P.A., 2009. Opportunity-based age replacement with a one-cycle criterion. *Journal of the Operational Research Society* 60, 1428–1438.
- [16] De Finetti, B., 1974. *Theory of Probability*. Wiley, London.
- [17] De Luca, G., Riveccio, G., 2009. Archimedean copulae for risk measurement. *Journal of Applied Statistics* 36, 907–924.
- [18] Do, B., Faff, R., 2010. Does simple pairs trading still work?. *Financial Analysts Journal* 66, 83–95.
- [19] Do, B., Faff, R., 2012. Are pairs trading profits robust to trading costs?. *Journal of Financial Research* 35, 261–287.
- [20] Ferreira, M., Santa-Clara, P., 2011. Forecasting stock market returns: the sum of the parts is more than the whole. *Journal of Financial Economics* 100, 514–537.

- [21] Gatev, E., Goetzmann, W., Rouwenhorst, K. 2006. Pairs trading: Performance of a relative-value arbitrage rule. *Review of Financial Studies* 19, 797-827.
- [22] Hampel, F., 2009. Nonadditive probabilities in statistics. *Journal of Statistical Theory and Practice* 3, 11–23.
- [23] Henkel, S., Martin, J., Nadari, F., 2011. Time-varying short-horizon predictability. *Journal of Financial Economics* 99, 560-580.
- [24] Hill, B.M., 1968. Posterior distribution of percentiles: Bayes' theorem for sampling from a population. *Journal of the American Statistical Association* 63, 677–691.
- [25] Kim, J., 2006. Wild bootstrapping variance ratio tests. *Economics Letters* 92, 38-43.
- [26] Krupskii, P., Joe, H., 2015. Tail-weighted measures of dependence. *Journal of Applied Statistics* 42, 614-629.
- [27] Lawless, J.F., Fredette, M., 2005. Frequentist prediction intervals and predictive distributions. *Biometrika* 92, 529–542.
- [28] Ludvigson, S., Ng, S., 2007. The empirical risk-return relation: a factor analysis approach. *Journal of Financial Economics* 83, 171-222.
- [29] Maturi, T.A., 2010. Nonparametric Predictive Inference for Multiple Comparisons. PhD thesis, Durham University, www.npi-statistics.com.
- [30] Pesaran, M.H., Timmermann, A., 1995. Predictability of stock returns: Robustness and economic significance. *The Journal of Finance* 50, 1201–1228.
- [31] Rapach, D., Strauss, J., Zhou, G., 2010. Out-of-sample equity premium prediction: combination forecasts and links to the real economy. *Review of Financial Studies* 23, 821-862.
- [32] Rapach, D., Zhou, G., 2013. Chapter 6 - forecasting stock returns, in: Elliott, G., Timmermann, A. (Eds.), *Handbook of Economic Forecasting*. Elsevier. volume 2, Part A of *Handbook of Economic Forecasting*, pp. 328 – 383.
- [33] Sullivan, R., Timmermann, A., White, H., 2001. Dangers of data mining: the case of calendar effects in stock returns. *Journal of Econometrics* 105, 249–286.
- [34] Trede, M., Savu, C., 2013. Do stock returns have an Archimedean copula?. *Journal of Applied Statistics* 40, 1764-1778.
- [35] Wilmott, P., 2007. *Paul Wilmott introduces quantitative finance*. 2 ed., Wiley, Chichester.