FIRST PASSAGE TIMES IN PORTFOLIO OPTIMIZATION: A NOVEL NONPARAMETRIC APPROACH

Gabriel Zsurkis | João Nicolau
Paulo M. M. Rodrigues
FIRST PASSAGE TIMES IN PORTFOLIO OPTIMIZATION: A NOVEL NONPARAMETRIC APPROACH

Gabriel Zsurkis | João Nicolau
Paulo M. M. Rodrigues
First passage times in portfolio optimization: a novel nonparametric approach

Gabriel Zsurkis
Banco de Portugal
ISEG-Universidade de Lisboa

João Nicolau
ISEG-Universidade de Lisboa
CEMAPRE

Paulo M. M. Rodrigues
Banco de Portugal
Nova School of Business and Economics

June 2023

Abstract

This paper introduces a portfolio optimization procedure that aims to minimize the intra-horizon (IH) risk subject to a minimum expected time to achieve a target cumulative return. To estimate the first passage probabilities and the expected time a novel nonparametric method and a new Markov chain order determination approach are developed. The optimization framework proposed allows us to include novel path-dependent measures of risk and return in the asset allocation problem. An empirical application to S&P 100 companies, a risk-free asset and stock indices is provided. Our empirical results suggest that the proposed framework exhibits more consistency between in-sample and out-of-sample performance than the mean-variance model and an alternative optimization problem that minimizes the MaxVaR measure of Boudoukh et al. (2004). Overall, the portfolio optimization approach we introduce results in higher out-of-sample annualized returns for relatively low levels of IH risk.

JEL: G11, G17, C14, C22, C41

Keywords: Portfolio Optimization, Markov Chains, Intra-horizon Risk, First-passage Probability.

Disclaimer: The analyses, opinions and findings of this paper represent the views of the authors, they are not necessarily those of the Banco de Portugal or the Eurosystem.

E-mail: gzsurkis@bportugal.pt; nicolau@iseg.ulisboa.pt; pmrodrigues@bportugal.pt
1. Introduction

One of the most influential asset allocation optimization methodology was proposed in the seminal work of Markowitz (1952, 1959). Markowitz’s approach associates profits and risk to the expectation and variance of returns, giving rise to the mean-variance (MV) portfolio optimization and to the concept of efficient frontier, the curve that represents the efficient portfolio for a given risk level (see Markowitz 2014 and Kolm et al. 2014 for reviews). Despite being a common workhorse model, the traditional MV approach has not been free of criticism, and the attempts to solve its limitations has fostered important advances in the portfolio optimization literature. An important drawback is related to the estimation error incurred when historical returns data are used to estimate the mean and covariance matrix of asset returns. Several authors proposed the use of shrinkage estimators, which are typically biased but exhibit less variance than the sample estimator; see, for instance, Ledoit and Wolf (2004) and Bodnar et al. (2023).

Furthermore, mean-variance portfolio theory is often rejected due to the growing empirical evidence that asset returns are non-gaussian (for instance, Mandelbrot 1963, Cont 2001, Nicolau and Rodrigues 2019, Gao et al. 2019, and Lassance et al. 2022, Section EC.2). Thus, significant deviations from the normal distribution may affect the mean-variance optimal portfolio allocation (see, for instance, Penev et al. 2019). Çelikyurt and Özekici (2007) consider the assumption of independence for the returns process too strong and propose a mean-variance optimization problem where the mean vector and covariance matrix depend on exogenous (e.g., economic or political) factors modulated through a Markov chain with perfectly observable states. More recently, Palczewski et al. (2015) relied on Markov chain approximations to allow the drift of a Geometric Brownian Motion (GBM) process to be state-dependent. Other works propose portfolio strategies that incorporate information about higher moments\(^1\), such as skewness and kurtosis, by including optimization constraints for these moments (de Athayde and Flôres 2004 and Briec et al. 2007), by proposing alternative portfolio performance measures (Joro and Na 2006) or by improving higher moments indirectly using factor diversification methods (Lassance et al. 2022). More recently, Lassance and Vrins (2023) introduces a novel portfolio selection framework that minimizes the distance between the empirical portfolio-return distribution and a target distribution that reflects some desired features. Another branch of this literature has been focused on the left tail risk minimization, such as value-at-risk (VaR) (e.g., Basak and Shapiro 2001, El Ghaoui et al. 2003 and Lwin et al. 2017) and the conditional value-at-risk or expected shortfall (Rockafellar and Uryasev 2000, Fábián 2008 and Nasini et al. 2022).

---

1. Notwithstanding, Schuhmacher et al. 2021 show that, in the presence of a risk-free asset, skewed returns do not justify a rejection of mean-variance analysis if asset returns follow a specific skew-elliptical distribution. Additionally, Markowitz (2014) argues that normal (Gaussian) return distribution is not a necessary condition for the applicability of modern portfolio theory.
All works reported so far focus on the risk measures’ values at the end of a specified trading horizon and ignore the portfolio’s value dynamic path. As an alternative to these static measures, some researchers consider dynamic drawdown measures (Chekhlov et al. 2005, Zabarankin et al. 2014, Drenovak et al. 2022 and Ding and Uryasev 2022), which are (at each time moment) drops of the portfolio cumulative return from its peak value since the beginning of the investment. Another closely related approach focuses on the dynamic path of a portfolio’s cumulative losses. Kritzman and Rich (2002), Boudoukh et al. (2004), Rossello (2008), Bakshi and Panayotov (2010) and Farkas et al. (2021) introduce methodologies that take into account the magnitude of potential losses incurred over the trading horizon, denoted intra-horizon (IH) risk.

The IH risk is critical in a market-to-market environment since sharp declines in asset values can affect trading strategies. However, there are few works that incorporate IH risk into asset allocation optimization; to the best of our knowledge Gupta et al. (2016) is possibly one of the first. This market risk component is crucial in practice, since portfolio managers may have thresholds that impose a stop-loss decision when broken. Thus, in this context, the optimization problem has to incorporate a constraint regarding the probability of breaching some pre-specified drawdown threshold somewhere during the investment horizon. The methodologies available to capture IH risk are mainly based on first passage probabilities, the probability that an event occurs for the first time within a finite horizon. Initially, this statistic was computed assuming that the asset price dynamics are well described by a GBM process (Karlin and Taylor 1975 and Boudoukh et al. 2004). However, since jumps or sudden price movements are believed to be an essential component of financial asset prices, more flexible models based on Lévy processes have been considered; see, for instance, Bakshi and Panayotov (2010), Leippold and Vasiljević (2020) and Farkas et al. (2021). As reported in Bakshi and Panayotov (2010), there are large variations in risk measures across different jump models, indicating substantial model risk.

In this work, we introduce a framework to formalize portfolio selection problems that allows us to incorporate the portfolio management issues referred to above, i.e., i) IH risk; and ii) the probability of breaching some pre-specified drawdown threshold. Our contribution is threefold: First, we propose a novel nonparametric approach to estimate the first hitting time probabilities that rely on the Markovian properties of returns. Thus, stationary Markov chains\(^2\) are used to estimate all relevant probabilities. This method remains valid even when the underlying price process exhibits an upward stochastic trend, as generally happens with security prices. Since the approach is nonparametric, the widely documented model uncertainty problem in portfolio analysis is avoided (see, for instance, Avramov and Zhou 2010, for a literature review). Second, a new Markov chain order selection

\(^2\) A Markov chain of order \(k\) is a stochastic model describing a sequence of states, where the probabilities of transition between states only depend on the information about the last \(k\) states already observed.
approach\(^3\) is introduced. Third, a novel portfolio optimization method, which aims to minimize the IH risk subject to a minimum expected time to achieve a cumulative target return, is proposed.

The remainder of the paper is organized as follows. Section 2 introduces the novel nonparametric approach and formalizes the optimization problem. Section 3 presents an empirical application of the proposed methodology to portfolio selection. Specifically, we consider the allocation of wealth across 87 companies belonging to the S&P 100 index, four stock indices and a risk-free asset. Finally, Section 4 concludes the paper. Proofs of the results reported in the paper are provided in the Appendix.

2. The proposed approach for portfolio selection

Consider an asset manager that is responsible for investing an initial wealth across a number of different securities (stocks, bonds, etc.) and faces two external parameters: a return target and a lower threshold that, once broken, forces the liquidation of the positions prematurely (Gupta et al. 2016). Therefore, when the target return is first attained, the asset manager’s mandate is successfully completed. On the other hand, when the portfolio value falls below the lower threshold for the first time, the investment process stops in order to avoid greater losses.

In this context, the first time a portfolio value, say \(y_t\), crosses some predefined threshold is crucial in practice. Assuming that \(\{y_t\}\) starts at \(x_0\), let the first time that \(y_t\) crosses the upper threshold, \(y^+ = (1 + r^+)x_0\), and the lower threshold, \(y^- = (1 + r^-)x_0\), be defined as,

\[
T^+ := T_{(1+r^+)x_0} = \inf\{t > 0 : y_t \geq y^+\}
\]

and

\[
T^- := T_{(1+r^-)x_0} = \inf\{t > 0 : y_t \leq y^-\},
\]

respectively, with \(t = 1, ..., n\), and where \(r^-\) and \(r^+\) are such that \(y^- < x_0 < y^+\). Thus, \(r^+\) represents the return target and \(r^-\) the cumulative loss that motivates a stop-loss decision.

2.1. Nonparametric estimation of first hitting time probabilities

The assumption that \(y_t\) crosses the threshold \(y^+ (y^-)\) an infinite number of times, as \(n \rightarrow \infty\), is required in order to estimate the first passage time distributions of \(T^+\) and \(T^-\), as well as all the relevant probabilities and expected times in a nonparametric way. As shown by Nicolau (2017), this aspect follows from the

---

3. Since, in practice, we do not know the most appropriate Markov chain order for a particular application, statistical methods are used to estimate it.
stationarity assumption and, in particular, from a positive Harris recurrence of \( \{ y_t \} \) (Harris 1956). However, since the focus here is on the price of a security, which may exhibit a strong positive trend, very much like a random walk with drift, \( y_t \) will not cross a fixed threshold a large enough number of times to allow for suitable estimation of the probabilities of interest. Nevertheless, in this nonstationary context, the events “\( y_t \) increases \( r\% \)” or “\( y_t \) decreases \( r\% \)” may actually occur an infinite number of times as \( n \to \infty \). Thus, the rationale behind the methodology that we introduce is that this type of events (“\( y_t \) increases by \( r\% \)” or “\( y_t \) decreases by \( r\% \)”) may be modeled through stationary Markov chains, which enables the estimation of the first passage time distributions of \( T^+ \) and \( T^- \).

**Assumption 1** (Conditional Homogeneity). \( P(T_{(1+r)x} \leq t | x_0 = x) = P(T_{(1+r)z} \leq t | x_0 = z) \) for any value \( x \) and \( z \).

In general, Assumption 1 is plausible if, for instance, \( y_t \) is the price of a security traded in an efficient market. In fact, if prices fully reflect all known information, as implied by the efficient markets hypothesis (Fama 1970), it is not possible for investors to purchase undervalued or sell overpriced stocks. Thus, in this context, the time that \( y_t \) takes to increase \( r\% \) is independent of the value at which the stock is currently traded, that is, it must be independent of \( x_0 \). Note that the well-known GBM is an example of a process that satisfies Assumption 1.

Closed-form solutions for first-passage probabilities are only available for few specifications. One of them is the GBM, which is widely used in modeling market prices due to its mathematical tractability (Karlin and Taylor 1975). However, asset prices often display sudden changes by a very large amount (Eraker et al. 2003) and this feature is difficult to reproduce within continuous-path models based on Brownian motions (Huang 1985 and Johannes 2004). This has motivated the use of Lévy models to capture the short-run behavior of security prices. Beyond the lack of analytical explicitness (see, for instance, Kyprianou 2006), choosing the most appropriate model from a large number of Lévy-type models available in the literature (see, e.g., Madan and Seneta 1990 and Carr et al. 2002) may be a challenging task. Therefore, the nonparametric method that we introduce to estimate the first hitting time probability function may be a useful approach, since it is flexible enough to accommodate jumps and other nonnormalities of asset prices.

### 2.1.1. The auxiliary processes.

The estimation of the first hitting time probabilities \( T^+ \) and \( T^- \) is based on two steps. For illustration purposes, we will focus only on the \( T^+ \) statistic in (1). In the first step the event “\( y_t \) increase \( r^+\% \)” is translated into an auxiliary process using an auxiliary process algorithm.

---

4. For instance, Bakshi and Panayotov (2010) reveal large variations in the risk measures’ estimates across different Lévy jump models.
The auxiliary process algorithm:\footnote{The auxiliary process algorithm is easily adapted to the case where \( r^- < 0 \), i.e., the auxiliary process \( S^-_t \) is obtained by running the algorithm above with \( x_0, y^+ \) and \( y_t \) replaced by \( -x_0, y^- \) and \( -y_t \), respectively.}

1. Set \( i = 1, x_0 = y_1 \), where \( y_1 \) is the first observation of \( y_t \);
2. Let \( y^+ = x_0(1 + r^+) \), with \( r^+ > 0 \), where \( r^+ \% \) can be understood as the cumulative growth rate (in percentage);
3. Define
   \[
   S^+_t := \begin{cases} 
   1 & \text{if } y_t < y^+, \ y_{t-1} < y^+, \ldots, y_i < y^+; \\
   0 & \text{otherwise} 
   \end{cases}
   \]
4. Run \( S^+_t \) for \( t = i, i + 1, \ldots, \min(n, t^*) \), where \( t^* \) is such that \( S^+_{t^*} = 0 \) (a threshold has been reached) and \( S^+_{0} = 0 \) is used for initialization;
5. If \( t = n \) the procedure is stopped;
6. Set \( x_0 = y_{t^*} \) and \( i = t^* + 1 \). Clear the value \( t^* \) and return to step 2.

In the second step all relevant information regarding the estimation of the parameters of interest is extracted from the auxiliary process \( S^+_t \) (or \( S^-_t \)).

Assumption 2 (Markovian Property and Stationarity). The processes \( S^+_t \) and \( S^-_t \) are stationary discrete-time Markov processes of finite order \( k \).

The estimation approach treats \( S^+_t \) and \( S^-_t \) as Markov chains with state space \{0, 1\} and estimates the transition probabilities between the two states. A consequence of Assumption 2 is that the events “\( y_t \) increases \( r\%\)” or “\( y_t \) decreases \( r\%\)” may occur an infinite number of times as \( T \to \infty \). If, for instance, we set \( r \) to be negative (positive) and \( y_t \) exhibits a strong positive (negative) trend over time, \( S^+_t, \tau = +, - \), will be formed mostly by ones and no relevant information regarding the transition probabilities can be extracted from this sequence. That is, the probability of transition from state \( S^+_{t-1} = 1 \) to state \( S^+_t = 0 \) will tend to zero.

Given the auxiliary process \( S^+_t, \tau = +, - \), the probability that \( y_t \) crosses a threshold \( y^\tau \), \( \tau = + \) or \( - \), for the first time at time \( t \), starting from \( x_0 \), is
\[
P(T^\tau = t) = P(S^+_t = 0, S^-_{t-1} = 1, S^-_{t-2} = 1, \ldots, S^-_1 = 1, S^-_0 = 1), \quad \tau = +, -
\]
which is equivalent to
\[
P(T^\tau = t) = (1 - p_t) \prod_{i=1}^{t-1} p_i, \quad \tau = +, - \tag{3}
\]
where \( p_t := P(S^+_t = 1|S^-_{t-1} = 1, S^-_{t-2} = 1, \ldots, S^-_0 = 1), \ x_0 < y^+ \) for \( S^+_t \) and \( x_0 > y^- \) for \( S^-_t \).
In view of the Markovian property in Assumption 2, if \( i > k \) then \( p_i = p_k \), where \( k \) is the Markov chain’s order. Specifically, from Assumption 2 we have that,

\[
\begin{align*}
p_i &= P(S^*_i = 1 | S^*_{i-1} = 1, S^*_{i-2} = 1, \ldots, S^*_0 = 1) \\
&= P(S^*_i = 1 | S^*_{i-1} = 1, S^*_{i-2} = 1, \ldots, S^*_0 = 1) \\
&= p_k.
\end{align*}
\]

Then,

\[
P(T^* = t) = \begin{cases} 
(1 - p_t) \prod_{i=1}^{t-1} p_i & \text{for } t \leq k \\
(1 - p_k) \prod_{i=1}^{k-1} p_i p_t^{t-k} & \text{for } t > k
\end{cases}
\]

and consequently,

\[
E(T^*) = \sum_{t=1}^{k} t (1 - p_t) \prod_{j=1}^{t-1} p_j + \left(1 - p_k \right) \prod_{j=1}^{k-1} p_j \left( \sum_{t=r+1}^{\infty} t p_k^{t-k} \right) = \sum_{t=1}^{k} t (1 - p_t) \prod_{j=1}^{t-1} p_j + \prod_{j=1}^{k-1} p_j \frac{p_k (1 + k - k p_k)}{1 - p_k}. \quad (5)
\]

Hence, to compute (4) and (5) we only need the probabilities \( p_1, p_2, \ldots, p_k \), which can be easily estimated from standard Markov chain methods. For instance, when \( k = 1 \), the maximum likelihood estimate of \( p_1 \) is \( \hat{p}_1 = \eta_{11}/\eta_1 \), where \( \eta_{11} \) is the number of transitions of the type \( \{S^*_{i-1} = 1, S^*_i = 1\} \) and \( \eta_1 \) is the number of cases for which \( \{S^*_{i-1} = 1\} \). As another example, for \( k = 3 \), \( \hat{p}_3 = \eta_{111}/\eta_{111} \), where \( \eta_{111} \) and \( \eta_{111} \) are the number of transitions of the type \( \{S^*_{i-3} = 1, S^*_{i-2} = 1, S^*_{i-1} = 1, S^*_i = 1\} \) and \( \{S^*_{i-3} = 1, S^*_{i-2} = 1, S^*_{i-1} = 1\} \), respectively.

**2.1.2. Monte Carlo simulations.** Since GBM has a closed form solution for the expected time, we may assess the quality of our estimator by comparing our estimates with results from the exact expression of the GBM. Specifically, the GBM is governed by the stochastic differential equation,

\[
dy_t = \mu y_t dt + \sigma y_t dW_t, \quad y_0 = x_0 > 0,
\]

where \( W_t \) is a (standard) Brownian motion. Defining \( T := \inf \{ t \geq 0 : y_t \geq x_0 (1 + r) \} \), it can be shown (see, for example, Wilmott 1998, Section 14.5) that,

\[
E(T) = \frac{1}{\mu - \frac{1}{2} \sigma^2} \ln \left( \frac{y^+}{x_0} \right) = \frac{1}{\mu - \frac{1}{2} \sigma^2} \ln (1 + r), \text{ with } \mu > \frac{1}{2} \sigma^2. \quad (6)
\]
In the simulation exercise we set $\mu = \sigma = 0.01$. Each trajectory of $y_t$ is simulated according to the solution of the stochastic differential equation, which can be expressed in discrete-time as

$$y_{i}\Delta = y_{(i-1)}\Delta \exp \left(\mu - \frac{\sigma^2}{2}\right) \Delta + \sigma \sqrt{\Delta} \epsilon_i \Delta, \quad i = 1,2,\ldots,n,$$

where we set $\Delta = 1$ and $\epsilon_i \Delta \overset{i.i.d.}{\sim} N(0,1)$.

Given the sequence $\{y_1,y_2,\ldots,y_n\}$, the objective of this section is to evaluate how accurate the estimator in (5) is in the approximation of the true expected time in (6). To reduce the sampling variability we simulate $n_s = 500$ trajectories of $\{y_i; i = 1,\ldots,n\}$; and for each case we compute the expected time, $\hat{E}(T)_j$, $j = 1,\ldots,500$, after fixing the values of $n$ and $r$. To obtain $\hat{E}(T)_j$, we proceed as described previously: given $r$ and $n$, we build the sequence $\{S_t\}$ from which we compute $\hat{p}_1$. Since GBM is a Markov process of order one ($k = 1$), no additional probabilities $p_k$, $k \geq 2$ are needed. Thus, $\hat{E}(T)_j = (1 - \hat{p}_1)^{-1}$.

Then we compute the global mean as $\overline{E(T)} = \frac{1}{n_s}\sum_{j=1}^{n_s} \hat{E}(T)_j$, which is compared to the true expected value $E(T)$ in (6). We analyse the proposed estimator as a function of the sample size $n_s$, and return $r$. Specifically, we considered $n \in \{500,1000,1500,2000\}$ and $r \in \{0.05,0.10,0.15,0.20,0.25,0.30,0.35\}$. Figure 1 compares $\overline{E(T)}$ with the values of $E(T)$ in (6), across the different values of $n$ and $r$ (one can see, for example, that the process takes an average of about 25 periods for the initial investment to grow 30% ($r = 0.3$)). Figure 1 shows that the proposed estimator is unbiased for $E(T)$, even when $n = 500$. Notice, however, that other combinations of $n$, $\mu$ or $\sigma$ may produce worse results (for example, if $n$, $\mu$ or $\sigma$ are such that $y_t$ never crosses $y^+ = x_0(1+r)$, or if it crosses only a small number of times).
2.1.3. Estimation, inference and Markov chain order determination. To calculate the standard errors of $E(T)$ and $P(T = t)$ we first need the variance-covariance matrix of $\hat{p} = (\hat{p}_1, \hat{p}_2, ..., \hat{p}_k)'$. Since the individual estimators of $p_i$, $i = 1, ..., k$, are obtained from different likelihood functions, and there is no obvious way to calculate the variance-covariance matrix, in what follows, a simple method is proposed. In order not to overload the notation, we will omit the superscript on $S$. Henceforth, $S$ can be understood as either $S_-$ or $S^+$.

Consider the multivariate linear model,

$$
\begin{align*}
    y_{t1} &= p_1 x_{t1} + \varepsilon_{t1} \\
    y_{t2} &= p_2 x_{t2} + \varepsilon_{t2} \\
    \vdots \\
    y_{tk} &= p_k x_{tk} + \varepsilon_{tk}
\end{align*}
$$

where $y_{ti} = S_t S_{t-1} ... S_{t-(i-1)}$ and $x_{ti} = S_{t-1} S_{t-2} ... S_{t-i}$. The system of equations in (7) can be expressed in matrix notation as

$$
    y_t = X_t p + \varepsilon_t
$$

where $y_t = (y_{t1}, y_{t2}, ..., y_{tk})'$, $X_t = \text{diag}(x_{t1}, x_{t2}, ..., x_{tk})$, $p = (p_1, p_2, ..., p_k)'$ and $\varepsilon_t = (\varepsilon_{t1}, \varepsilon_{t2}, ..., \varepsilon_{tk})'$.

Considering (8) we focus on the OLS estimators,

$$
    \hat{p} = \left( \sum_t X_t X_t' \right)^{-1} \sum_t X_t y_t,
$$

Figure 1: $E(T)$ versus the new nonparametric estimator
because these are equivalent to maximum likelihood estimates obtained for each
\(i = 1, \ldots, k\) separately. Inversion of the block diagonal matrix, \(\sum_t^n X_t X_t'\), leads to
the simple estimators:

\[
\hat{p}_i = \frac{\sum_t^n x_{ti}y_{ti}}{\sum_t^n x_{ti}^2} = \frac{\sum_t^n x_{ti}y_{ti}}{\sum_t^n x_{ti}}, \quad i = 1, 2, \ldots, k.
\]

Hence, the system OLS estimator of (8) is equivalent to the corresponding OLS
equation by equation estimators.

To exemplify, consider the cases \(k = 1\) and \(k = 2\). For \(k = 1\), we have that,

\[
\hat{p}_1 = \frac{\sum_t^n x_{t1}y_{t1}}{\sum_t^n x_{t1}} = \frac{\sum_t^n S_{t-1}S_t}{\sum_t^n S_{t-1}} = \frac{\eta_{11}}{\eta_1}, \quad S_t \in \{0, 1\},
\]

and for \(k = 2\) it follows that,

\[
\hat{p}_2 = \frac{\sum_t^n x_{t2}y_{t2}}{\sum_t^n x_{t2}} = \frac{\sum_t^n S_tS_{t-1}S_{t-2}}{\sum_t^n S_{t-1}S_{t-2}} = \frac{\eta_{111}}{\eta_{11}}.
\]

The statistic \(\hat{p}_1 = \eta_{11}/\eta_1\) is the maximum likelihood estimate of \(p_1\), when \(k = 1\),
as we have mentioned previously (see, for example, Basawa and Prakasa Rao 1980,
Chapter 4).

In what follows we make the weaker assumption that the Markov chain \(\{S_t\}\)
is stationary (Assumption 2). If the sequence \(\{S_t\}\) is not formed exclusively by
ones or zeros, that is, if the threshold is crossed many times, then the Markov
chain is irreducible, and it has a unique stationary distribution (see, for instance,
Douc et al. 2018, Corollary 7.2.3). Hence, it follows that the process \(\{S_t\}\) is weakly
dependent, all moments are finite, and the law of large numbers, and the central
limit theorem hold. Furthermore, \(E(X_tX_t')\) is non-singular as \(E(x_{tk}^2) = E(x_{rk}) =
P(S_{t-1} = 1, S_{t-2} = 1, \ldots, S_{t-k} = 1) = \pi_k > 0\) (since all states must be positive
recurrent).

Under the assumption of stationarity of \(\{S_t\}\), it follows that the OLS estimator
\(\hat{p}_k\) is consistent. In fact,

\[
\hat{p}_k = \frac{\sum_t^n x_{kt}y_{kt}}{\sum_t^n x_{kt}^2} = \frac{\sum_t^n S_tS_{t-1}S_{t-2} \ldots S_{t-k}}{\sum_t^n S_{t-1}S_{t-2} \ldots S_{t-k}}
\]

\[
P \rightarrow \frac{P(S_t = 1, S_{t-1} = 1, \ldots, S_{t-k} = 1)}{P(S_{t-1} = 1, S_{t-2} = 1, \ldots, S_{t-k} = 1)} = \frac{\pi_{k+1}}{\pi_k} = \frac{p_k\pi_k}{\pi_k} = p_k,
\]

where \(\pi_{k+1} = P(S_t = 1, S_{t-1} = 1, \ldots, S_{t-k} = 1)\) and \(\pi_k = P(S_{t-1} = 1, S_{t-2} = 1, \ldots, S_{t-k} = 1)\).
Note also that the orthogonality condition \(E(X_t\epsilon_t) = 0\) holds. For illustration,
considering the $k$-th element of $E(X_i \varepsilon_i)$, we note that,

$$E(x_{ik} \varepsilon_{tk}) = E \left( S_{i-1} S_{i-2} ... S_{i-k} \left( S_t S_{t-1} ... S_{t-(k-1)} - p_k S_{i-1} S_{i-2} ... S_{i-k} \right) \right)$$

$$= E \left( S_{i-1} S_{i-2} ... S_{i-k} - p_k S_{i-1} S_{i-2} ... S_{i-k} \right)$$

$$= \pi_{k+1} - p_k \pi_k = 0.$$

Although $\varepsilon_{tk}$ is orthogonal to $x_{ik}$, $\varepsilon_{tk}$ is serially correlated. For example, consider again the case when $k = 1$. First, note that,

$$E(S_i | S_{i-1}) = p_1 S_{i-1} + p_{01} (1 - S_{i-1}), \quad S_{i-1} \in \{0, 1\}, \quad (10)$$

where $p_{01} := P(S_i = 1 | S_{i-1} = 0)$. From (10) we may obtain an estimate of $p_1$ (and $p_{01}$) from the regression,

$$S_i = p_1 S_{i-1} + p_{01} (1 - S_{i-1}) + u_i, \quad (11)$$

where $u_i := S_i - E(S_i | S_{i-1})$, is by construction a martingale difference, and so serially uncorrelated. Moreover, $S_{i-1}$ is algebraically orthogonal to $(1 - S_{i-1})$, i.e. $\sum_{i=1}^n S_{i-1} (1 - S_{i-1}) = 0$, and therefore the same estimate $\hat{p}_1$ can be obtained from the more parsimonious regression $y_{i1} = p_1 x_{i1} + \varepsilon_{i1}$, or

$$S_i = p_1 S_{i-1} + \varepsilon_{i1}. \quad (12)$$

However, $\{\varepsilon_{i1}\}$ in (12) is no longer a martingale difference, since $\varepsilon_{i1} = p_{01} (1 - S_{i-1}) + u_i$, it will be serially correlated. $S_{i-1}$ is still a predetermined variable in the sense that $E(S_{i-1} \varepsilon_{i1}) = E(S_{i-1} (p_{01} (1 - S_{i-1}) + u_i)) = 0$. Given this result, we may ask why not just use regression (11), which has spherical disturbances? The reason is that as $k$ increases, the number of control variables increases exponentially. Since the OLS estimates from the more parsimonious regressions $y_{it} = p_t x_{it} + \varepsilon_{it}$, $i = 1, ..., k$, provide identical numerical results for $p_t$, as all the control variables are algebraically orthogonal to $x_{it}$, we prefer to use the system in (7) with just one explanatory variable per equation.

To obtain robust standard errors we use a heteroskedasticity and autocorrelation consistent (HAC) covariance matrix estimator. Formally,

$$\sqrt{n} (\hat{p} - p) = \left( \frac{1}{n} \sum_{i=1}^n X_i X'_i \right)^{-1/2} \left( \frac{1}{n} \sum_{i=1}^n X_i \varepsilon_i \right) \xrightarrow{d} N \left( 0, Q^{-1} SQ^{-1} \right) \quad (13)$$

where $Q = E(X_i X'_i)$ and $S = \lim_{n \to \infty} \frac{1}{n} \text{Var} \left( \sum_{i=1}^n X_i \varepsilon_i \right)$. Matrix $S$ is estimated using a HAC estimator.

Statistical inference regarding $E(T)$ or $P(T = t)$, which are functions of $p = (p_1, ..., p_k)'$, can be easily performed by the Delta method. For example, in the case $E(T) := h(p_1, ..., p_k)$, we have

$$\sqrt{n} (h(\hat{p}) - h(p)) \xrightarrow{d} N \left( 0, D' \Sigma D \right),$$

where $D = \partial h(p) / \partial p$ and $\Sigma = Q^{-1} S Q^{-1}$.
2.1.4. Markov chain order determination. There are several approaches to estimate the order of a Markov chain (see Zhao et al. 2001 and Katz 1981). Most of them rely on the Akaike or the Bayesian information criteria. These methods are not only cumbersome, as they require building several Markov chains of different orders, but they are probably also inefficient because our object of interest centers exclusively on the probabilities,

$$p_k := P(S_t = 1|S_{t-1} = 1, S_{t-2} = 1, ..., S_{t-k} = 1)$$  \hspace{1cm} (14)

and not on the entire transition matrix. We propose a new efficient and straightforward method to estimate the order, $k$, of (14). For that purpose, consider the auxiliary regression,

$$S_t = \beta_{k-1} x_{tk} + \beta_k x_{t,k-1} + \varepsilon_t, \quad t = k + 1, k + 2, ..., n,$$

(15)

where the error term, $\varepsilon_t$, is uncorrelated with $x_{tk}$ and $x_{t,k-1}$. Both the regressors and the dependent variable are binary. The estimator of $\beta_k$ is,

$$\hat{\beta}_k = \frac{p_k - p_{k-1}}{1 - p_{k-1}},$$

(16)

where $p_k = P(S_t = 1|S_{t-1} = 1, ..., S_{t-k} = 1)$: see the Appendix for details. Hence, from (16) we can assess whether $p_k$ is equal to $p_{k-1}$ by a standard t-test using the ratio $\hat{\beta}_k/\hat{\sigma}_k$, where $\hat{\sigma}_k$ is the estimated standard error of $\hat{\beta}_k$. If $H_0 : \beta_k = 0$ is not rejected then $p_k$ is equal to $p_{k-1}$ and the true order $k^*$ is smaller than $k$. For example, considering for simplicity of presentation that $k = 2$, if $P(S_t = 1|S_{t-1} = 1, S_{t-2} = 1) = P(S_t = 1|S_{t-1} = 1)$ then the event $\{S_t = 1\}$ is independent of $\{S_{t-2} = 1\}$ given $\{S_{t-1} = 1\}$ and the order of the resulting Markov chain is smaller than $k = 2$. Figure 2 illustrates the sequential Markov chain order, $k^*$, determination methodology that we propose. In this example, the maximum order is set to $k^{\text{max}} = 3$. The choice of the maximum order in empirical work can be determined using, for instance, Schwert’s rule, $k^{\text{max}} = \lfloor 4(n/100)^{1/4} \rfloor$, where $\lfloor z \rfloor$ denotes the integer part of $z$ (Schwert 1989).
First passage times in portfolio optimization: a novel nonparametric approach

Estimate: \( S_t = \beta_2 x_{t2} + \beta_3 x_{t3} + \varepsilon_t \)

Test: \( H_0: \beta_3 = 0 \) vs \( H_1: \beta_3 \neq 0 \)

\( k^* = 3 \)

Not Reject \( H_0 \)

Estimate: \( S_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \varepsilon_t \)

Test: \( H_0: \beta_2 = 0 \) vs \( H_1: \beta_2 \neq 0 \)

\( k^* = 2 \)

Reject \( H_0 \)

\( k^* = 1 \)

Not Reject \( H_0 \)

Figure 2: Illustration of Markov Chain Order Determination

In the case of a regression with only one “regressor”, i.e.,

\[ S_t = \theta x_{tk} + \varepsilon_t, \quad t = k + 1, k + 2, \ldots, n, \quad (17) \]

\( \theta \) corresponds to \( p_k \) (see Appendix for details).

2.2. The optimization problem

In order to emphasize the allocation problem let us redefine (1) and (2) as functions of the vector of portfolio weights \( \omega = (\omega_1, \ldots, \omega_m)' \), that is,

\[ T^+(\omega) = \inf \{ t > 0 : y_t(\omega) > (1 + r^+) x_0 \}, \]

and

\[ T^-(\omega) = \inf \{ t > 0 : y_t(\omega) < (1 + r^-) x_0 \}, \]

where \( x_0 \) is the initial wealth available, \( r^+ \) is the desired portfolio’s cumulative return, \( r^- \) is the cumulative loss rate that defines the lower threshold, and

\[ y_t(\omega) = \prod_{i=1}^{t} \left( 1 + \sum_{j=1}^{m} \omega_j r_{ji} \right) x_0, \]

with \( r_{ji} \) being the simple (one-period) return of asset \( j \) at time \( i \).
We propose the following optimization of a portfolio of $m$ assets,
\[
\min_{\omega_1, \ldots, \omega_m} P\left( T^- (\omega) \leq n^- \right)
\]
subject to the constraints
\[
E\left( T^+ (\omega) \right) \leq n^+ \tag{19}
\]
\[
\omega' \mathbf{1} = 1, \tag{20}
\]
where $\mathbf{1}$ is an $m$-dimensional vector of ones. Thus, (18) minimizes the probability that a cumulative loss rate of more than $r^- 100\%$ occurs in a time horizon of $n^-$ trading days. Additionally, the constraint in (19) requires the expected time to achieve a cumulative return target to be less than or equal to $n^+$ trading days.

Asset allocation is generally decided based on return objectives, risk tolerance and time horizon. The portfolio optimization framework presented in (18) - (20) provides a flexible way to manage investor’s risk-return preferences in the asset allocation problem since it minimizes the IH risk subject to a return-based constraint. The IH risk can be controlled by defining the cumulative loss rate $r^-$ and the number of trading days $n^-$ during which the probability that the portfolio value crosses the lower threshold is minimized. Regarding the return target, investors have to choose the desired cumulative return $r^+$ and the expected maximum number of days considered reasonable to achieve $r^+$. Note that, for a fixed $r^+$, a smaller $n^+$ implies a higher annualized return. Thus, a smaller $n^+$ is expected to be associated with a higher IH risk, since a riskier portfolio may be needed in order to attain $r^+$ faster. On the other hand, a large $n^+$ value and a small $r^-$ means that the main concern is to avoid cumulative losses even if it implies very low annualized returns.

As mentioned in the Introduction, IH risk can have significant impact on trading strategies. For instance, sharp decreases in asset values may trigger stop-loss decisions. In practice, the use of stop-loss orders to limit the size of losses potentially reduce the overall drawdown risk of the portfolio. If agents are risk-seeking over losses (in the sense that they prefer chances of losing a larger amount to losing for sure much less) as stated by prospect theory (see Kahneman and Tversky 1979, Tversky and Kahneman 1992, Barberis 2013 and Barberis et al. 2021), the existence of an *a priori* drawdown threshold of cumulative losses implying the sale of the entire portfolio avoids making impulsive decisions which may result in even greater losses. Furthermore, Richards et al. (2017) find that investors who use stop losses exhibit less disposition effects\(^6\) than investors who do not.

---

\(^6\) Disposition effect refers to investor’s decision-making bias which led investors to hold on longer when investments have depreciated in value than when investments have appreciated in value (Shefrin and Statman 1985), a concept closely related to prospect theory’s asymmetric risk aversion, according to which investors are risk-averse when faced with gains and risk-seeking when faced with losses (see, for instance, Dacey and Zielonka 2008 and Barberis and Wei 2009).
2.3. The optimization algorithm

Regarding the optimization algorithm we propose that the objective function to be minimized is,

$$
\Psi(\omega) = P(T^-(\omega) \leq n^-) + \gamma I\left(E\left(T^+(\omega) \right) \right),
$$

(21)

where

$$
I\left(E\left(T^+(\omega) \right) \right) := \begin{cases}
0 & \text{if } E\left(T^+(\omega) \right) \leq n^+ \\
E\left(T^+(\omega) \right) & \text{otherwise}
\end{cases},
$$

(22)

and $\gamma$ assumes a large positive value (for instance, 1,000,000) in order to discard solutions that do not respect the constraint $E\left(T^+(\omega) \right) \leq n^+$.

As is evident from (21) and (22), classical optimization methods (such as, linear and quadratic programming) based on exploiting the derivatives of the objective function are not applicable here since, for a particular vector $\omega$, the following steps have to be iteratively performed:

1. Build the auxiliary processes $S^+$ and $S^-$;
2. Estimate the Markov chain orders for $S^+$ and $S^-$;
3. Obtain the probability functions for the first passage time processes $T^+$ and $T^-$;
4. Compute $E(T^+(\omega))$ and $P(T^-(\omega) \leq n^-)$.

Thus, in order to minimize the objective function in (21), heuristic iterative stochastic search methods should be considered. We will employ the Threshold Accepting (TA) algorithm introduced by Dueck and Scheuer (1990) which was one of the first heuristic approaches applied to portfolio selection problems. Roughly speaking, this heuristic algorithm consists of starting with a randomly chosen feasible solution and successively picking (randomly) new solutions. Each of the new solutions, called neighbor solutions, is evaluated. If the new solution is better or as long as its deviation from the previous solution does not exceed certain thresholds, even though it is worse, it is accepted. In order to implement this method we need to define, in addition to the objective function and constraints handling functions, a neighborhood function and the thresholds.

Since several specifications can be considered for the neighborhood function and for the thresholds, many variations of the TA algorithm are possible. We will use a variant proposed in Gilli and Schumann (2010) and Gilli et al. (2011), as

---

7. Since under Assumption 1 the portfolio's initial value has no impact on the time that $y_t$ takes to cross $r^-$, the optimization problem considers the same $x_0$ values to compute both $S^+$ and $S^-$. 
they have shown that it performs well in portfolio optimization problems. We will make use of a neighborhood algorithm (Gilli et al. 2011, p.394), which defines how we move from one solution to the next (see detailed description of Algorithm 1: Neighborhood Algorithm in the Appendix). Moreover, the threshold sequence consists of an ordered vector of positive numbers that decrease to zero or at least become very small. In order to compute the threshold sequence we resort to the Threshold Sequence algorithm proposed by Gilli et al. (2011, p.392) (for a detailed description of this algorithm see the Appendix). Finally, TA is the main optimization algorithm (Gilli et al. 2011, p.390), which employs the neighborhood and threshold sequence algorithms discussed above (see Appendix for details).

3. Empirical Application

In order to illustrate the proposed methodology for portfolio selection, we consider stocks from the S&P 100 index, since this index includes a diverse mix of industries and provides a good representation of the overall U.S. market. We use daily data (adjusted closing prices) covering the period from January 2005 to December 2022 (4531 observations) from Yahoo! Finance. Stocks with missing values in this period were excluded, leaving us with 87 companies. Additionally, we also included a risk-free asset (US 13 week treasury bill) and four stock indices (Dow Jones, S&P 500, Nasdaq Composite and Nasdaq 100) in the analysis.

Subsection 2.3 presents the optimization problem and describes the steps that have to be iteratively performed. For a particular vector of portfolio weights, we face two main estimation issues. First, the determination of the Markov chain order for the auxiliary processes $S^+$ and $S^-$ is based on the approach illustrated in Figure 2 with $k_{\text{max}} = 5$, where standard normal critical values are used to test $H_0: \beta_k = 0$ against $H_0: \beta_k \neq 0$ in (15), for $k = 1, \ldots, 5$, at a 5% significance level. Second, for the estimation of the transition probabilities we considered the expression in (9). Thus, having estimated the Markov chain order and the transition probabilities of interest for the two auxiliary processes, we easily obtain estimates for expressions (4) and (5), which are included in the objective function to be minimized (expressions (21) and (22)).

The solutions to the proposed optimization problem were obtained using the TA algorithm (see Appendix for a detailed description of this algorithm). We considered $w_{\text{min}} = 0$ and $w_{\text{max}} = 1$ in the Neighborhood Algorithm, which corresponds to considering that short-selling is not allowed. However, short-selling can be straightforwardly incorporated in this approach by choosing a negative value for $w_{\text{min}}$. Moreover, we choose $n_{\text{rounds}} = 10$ (since Gilli and Schumann 2010, report that the performance of the algorithm stays roughly the same for more than 10

---

8. Some literature promotes index investing due to the inability of the active management to outperform the market (see, for instance, Fama and French 2010).
thresholds), and \( n_{\text{steps}} = 1000 \) and \( n_{\text{deltas}} = 2000 \), which corresponds to a number of iterations of \( n_{\text{rounds}} \times n_{\text{steps}} = 10000 \).\(^9\) To initiate the algorithm, we consider that all the wealth is allocated to the risk-free asset. Since the risk-free asset return is quite low in practice, the restriction \( E(T^+(\omega)) \leq n^+ \) may not hold for these starting portfolio weights. Thus, the optimization algorithm will change the portfolio composition towards riskier assets until the return constraint is satisfied with the least possible IH risk.

Unfortunately, in the parametric literature that also addresses IH risk, such as Bakshi and Panayotov (2010) and, more recently, Leippold and Vasiljević (2020) and Farkas \textit{et al.} (2021), it is difficult to incorporate IH risk restrictions in a portfolio optimization problem in order to compare it with our nonparametric approach. This is mainly because there is substantial model risk associated with these methodologies, with large variations in risk measures across different jump models (Bakshi and Panayotov 2010). Moreover, the computational burden to find the optimal portfolio allocation would be substantial since the first passage probabilities do not have closed-form solutions for most of the jump models and it would be necessary to compute them numerically. Thus, we will only consider the MaxVaR measure proposed by Boudoukh \textit{et al.} (2004), which assumes that asset prices follow a GBM and, consequently, provides a closed-form analytical expression for the first passage time probability.

Summing up, we illustrate the proposed methodology by comparing it with the well-known mean-variance optimization\(^{10}\) (Markowitz 1959) and an alternative optimization problem that also considers IH risk by minimizing the MaxVaR measure introduced by Boudoukh \textit{et al.} (2004), subject to a constraint based on the empirical mean of daily returns. The solutions to this portfolio selection approach were also obtained considering the TA algorithm. Therefore, all three optimization problems minimize a risk measure subject to a return constraint. Four annualized return targets were considered, which can be related with different investor profiles: 5\% (conservative), 6.8\% (moderate - this corresponds to the annualized S&P 500 returns between 2005 and 2022), 10\% (moderately aggressive) and 20\% (aggressive).

The proposed portfolio optimization framework presented in (18) - (20) allows for the inclusion of the investor’s risk-return preferences in the asset allocation problem. To this end four parameters have to be chosen: the cumulative target return \( r^+ \); the cumulative loss rate \( r^- \) that determines the drawdown threshold; the number, \( n^- \), of trading days during which the probability that the portfolio value crosses the lower threshold or IH risk needs to be controlled; and finally, \( n^+ \), the maximum number of days to achieve the target return. Since there is not

\(^9\) The TA algorithm was run in R, using the TAopt:Optimisation with Threshold Accepting code contained in the NMOF: Numerical Methods and Optimization in Finance package (see Schumann 2011-2020).

\(^{10}\) The covariance matrix estimation was performed using the shrinkage method introduced by Ledoit and Wolf (2004) available in the R package \texttt{cvCovEst}.
much information in the literature to guide the selection process of these parameter values, we consider $r^+ = 5\%$, $r^- = -5\%$, $n^- = 20$ and some different values for $n^+$ in order to illustrate the proposed portfolio optimization methodology. Thus, the optimization problem will minimize the probability that the cumulative loss over 20 trading days is greater than 5%. A smaller $n^+$ is expected to be associated with a higher IH risk, since a riskier portfolio must be needed if we want to attain $r^+$ faster. In other words, a negative relationship is expected between $P\left(T^- (\omega) \leq n^- \right)$ and $E(T^+(\omega))$. In order to ensure some comparability, we compute the annualized target return by assuming that the $r^+ 100\%$ return objective is obtained in $n^+$ trading days. When $n^+$ is low, the annualized return is higher, which tends to be associated with higher values for IH risk.

3.1. In-sample analysis

Table 1 presents the chosen portfolios, with the stocks aggregated by sectors, considering the optimization problem in (18) - (20) and those that are obtained by employing the mean-variance optimization and the MaxVaR approach introduced by Boudoukh et al. (2004). The annualized log return statistics suggests that higher-mean/higher volatility portfolios are associated with higher intra-horizon risk (higher $P(T^-(\omega) \leq 20)$) and, naturally, lower risk-free asset weight. All three optimization methods suggest that a conservative investor (annualized required return of 5%) will allocate more than 70% of the wealth to the risk-free asset. For an aggressive investment strategy (annualized required return of 20%), all three methods recommend a significant exposure (more than 16%) to the Information Technology sector. The proposed optimization and MaxVaR optimizations also choose Health Care (34.7% and 23.5%, respectively), while the mean-variance optimization prefer Consumer Discretionary stocks (16.2%) and continues to allocate a significant portion of the wealth to the risk free asset (30.7%). For higher annualized required returns, the portfolios pointed out by the proposed methodology are composed by a larger number of assets. Finally, it is also noteworthy that our proposed approach selects portfolios with lower expected time to achieve the required return than the alternative methods for low IH risk levels.
First passage times in portfolio optimization: a novel nonparametric approach

Sample: daily data from January 2005 to December 2022

Optimization in (11) and (12), \( N = 20 \), Mean-variance

\( r^+ = 5\% \), \( r^- = -5\% \)

<table>
<thead>
<tr>
<th>Allocation</th>
<th>ann. target return</th>
<th>Mean-variance</th>
<th>MaxVaR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5%</td>
<td>6.8%</td>
<td>10%</td>
</tr>
<tr>
<td>Risk-free asset</td>
<td>0.740</td>
<td>0.500</td>
<td>0.302</td>
</tr>
<tr>
<td>Communication Services</td>
<td>0.029</td>
<td>0.000</td>
<td>0.025</td>
</tr>
<tr>
<td>Consumer Discretionary</td>
<td>0.131</td>
<td>0.025</td>
<td>0.160</td>
</tr>
<tr>
<td>Consumer Staples</td>
<td>0.000</td>
<td>0.156</td>
<td>0.066</td>
</tr>
<tr>
<td>Energy</td>
<td>0.007</td>
<td>0.000</td>
<td>0.033</td>
</tr>
<tr>
<td>Financials</td>
<td>0.041</td>
<td>0.066</td>
<td>0.006</td>
</tr>
<tr>
<td>Health Care</td>
<td>0.033</td>
<td>0.067</td>
<td>0.165</td>
</tr>
<tr>
<td>Industrials</td>
<td>0.000</td>
<td>0.081</td>
<td>0.036</td>
</tr>
<tr>
<td>Information Technology</td>
<td>0.010</td>
<td>0.027</td>
<td>0.057</td>
</tr>
<tr>
<td>Materials</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Real Estate</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Utilities</td>
<td>0.000</td>
<td>0.077</td>
<td>0.150</td>
</tr>
<tr>
<td>Stock Indices</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Number of assets</td>
<td>11</td>
<td>12</td>
<td>19</td>
</tr>
<tr>
<td>( E(T^+(\omega)) )</td>
<td>188</td>
<td>173</td>
<td>125</td>
</tr>
<tr>
<td>( P(T^-(-\omega) \leq N^*) )</td>
<td>0.005</td>
<td>0.010</td>
<td>0.021</td>
</tr>
<tr>
<td>Portfolio beta</td>
<td>0.290</td>
<td>0.392</td>
<td>0.567</td>
</tr>
</tbody>
</table>

Annualized log-return statistics

<table>
<thead>
<tr>
<th></th>
<th>mean (%)</th>
<th>s.d. (%)</th>
<th>mean/s.d.</th>
<th>skewness</th>
<th>kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6.923</td>
<td>7.868</td>
<td>11.007</td>
<td>18.073</td>
<td>4.899</td>
</tr>
<tr>
<td></td>
<td>6.703</td>
<td>8.629</td>
<td>12.681</td>
<td>18.078</td>
<td>2.990</td>
</tr>
<tr>
<td></td>
<td>4.956</td>
<td>6.709</td>
<td>9.658</td>
<td>18.553</td>
<td>5.000</td>
</tr>
<tr>
<td></td>
<td>1.072</td>
<td>0.902</td>
<td>0.868</td>
<td>1.052</td>
<td>1.623</td>
</tr>
<tr>
<td></td>
<td>1.623</td>
<td>1.409</td>
<td>1.998</td>
<td>1.522</td>
<td>1.067</td>
</tr>
<tr>
<td></td>
<td>1.522</td>
<td>1.496</td>
<td>1.759</td>
<td>1.255</td>
<td>1.505</td>
</tr>
<tr>
<td></td>
<td>1.505</td>
<td>1.484</td>
<td>1.662</td>
<td>1.155</td>
<td>1.242</td>
</tr>
<tr>
<td></td>
<td>7.242</td>
<td>17.302</td>
<td>13.850</td>
<td>9.465</td>
<td>5.955</td>
</tr>
<tr>
<td></td>
<td>16.067</td>
<td>15.213</td>
<td>13.278</td>
<td>10.93</td>
<td></td>
</tr>
</tbody>
</table>
| Notes: | (i) "ann. mean / ann. s.d." refers to the ratio of the annualized log return’s mean over the annualized log return’s standard deviation; (ii) for the proposed portfolio optimization, the annualized target return is assumed to be equal to \((1 + r^+)(252/N^*)\) \( N^* \) comes from the return restriction \( E(T^+(\omega)) \) \( \leq N^* \) \( \leq N^* \), with \( N^* = 252, 186, 126, 68 \); (iii) in order to compute the MaxVaR, we use, respectively, the sample mean and standard deviation of the simple return time series as estimates for the drift and volatility parameters of the GBM; (iv) for mean-variance and MaxVaR optimization, the reward restriction requires that the daily return’s mean is equal or higher than \([(\text{ann. target return})^{1/252} - 1]\); (v) portfolio beta is computed relative to the S&P 500 index.

Table 1. Optimal portfolios suggested by the considered methodologies
3.2. Out-of-sample analysis

As stated by Gilli and Schumann (2011), in order to evaluate the quality of a specific portfolio optimization, we need to look at its out-of-sample performance and check whether there is a meaningful relationship between in-sample and out-of-sample results. To this end, we use a rolling window approach with weekly and monthly rebalancing intervals. More specifically, using daily data, we consider 13 years of rolling-window\(^{11}\) data to obtain the portfolios that will be held in the following 5 or 20 days (out-of-sample). Since we have 18 years of daily data available, from 2005 to 2022, the out-of-sample exercise will generate a time series with 5 years of forecasts rebalanced every 5 or 20 days, which requires 252 and 63 portfolio optimizations, respectively. Every time the portfolio is rebalanced, the previous portfolio composition is considered as starting values for the threshold accepting optimization algorithm, in order to foster solutions with lower turnover values (average percentage of wealth traded per rebalancing period).

Table 2 presents the out-of-sample results. The proposed optimization exhibits more consistency between in-sample and out-of-sample results, since the out-of-sample dynamics of the resulting portfolios seem to satisfy, with an exception (the annualized target return of 20% case), the implicit reward restrictions associated with the optimization problems. More specifically, unlike mean-variance and MaxVaR optimizations, the annualized means of the out-of-sample returns obtained with our approach are very close to the required annualized target returns. On the other hand, the two alternative optimization methods considered suggest portfolios with considerably lower out-of-sample return’s volatility. However, this result seems to be caused by the fact that mean-variance and MaxVaR optimization heavily rely on the risk-free asset. For instance, even for the annualized target return of 20%, about 38% (on average) of the portfolio value is assigned to the risk-free asset. Table 2 also provides the Herfindahl-Hirschman index (HHI) as a proxy for concentration and the portfolio turnover (turnover) as a proxy of trading activity (lower values for this indicator mean that the composition of the portfolio remained stable over time). The portfolios obtained using the approach we introduce have lower concentration, but substantially higher turnover, which increases the transaction costs. Overall, the main findings are similar for weekly and monthly rebalancing intervals. Therefore, a possible strategy to reduce the transaction costs associated to the portfolio optimization we propose is to consider 20 days instead of 5 days rebalancing intervals.

\(^{11}\) For instance, DeMiguel et al. (2013) and Georgantas et al. (2021) also consider rolling-window approaches to assess the portfolio selection methodologies’ out-of-sample performance.
First passage times in portfolio optimization: a novel nonparametric approach

### Weekly rebalancing

<table>
<thead>
<tr>
<th>Optimization in (11) and (12), $N = 20$, $r^* = 5%$, $r = -5%$</th>
<th>Mean-variance</th>
<th>MaxVaR$_{0.95}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Composition</strong></td>
<td><strong>ann. target return</strong></td>
<td><strong>ann. target return</strong></td>
</tr>
<tr>
<td>5%</td>
<td>6.9%</td>
<td>10%</td>
</tr>
<tr>
<td>average no. of assets</td>
<td>8.6</td>
<td>9.9</td>
</tr>
<tr>
<td>average risk-free assets</td>
<td>0.687</td>
<td>0.564</td>
</tr>
<tr>
<td><strong>MaxTurnover</strong></td>
<td>0.360</td>
<td>0.471</td>
</tr>
</tbody>
</table>

#### Annualized statistics for the out-of-sample log-return time series

| | | | |
|---|---|---|
| **mean (%)** | 5.006 | 6.921 | 15.565 |
| **s.d. (%)** | 6.752 | 11.013 | 16.033 |
| **mean/s.d.** | 0.741 | 0.628 | 0.908 |
| **skewness** | -0.553 | -0.201 | -0.731 |
| **kurtosis** | 15.336 | 25.160 | 18.906 |

### Monthly rebalancing

<table>
<thead>
<tr>
<th>Optimization in (11) and (12), $N = 20$, $r^* = 5%$, $r = -5%$</th>
<th>Mean-variance</th>
<th>MaxVaR$_{0.95}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Composition</strong></td>
<td><strong>ann. target return</strong></td>
<td><strong>ann. target return</strong></td>
</tr>
<tr>
<td>5%</td>
<td>6.9%</td>
<td>10%</td>
</tr>
<tr>
<td>average no. of assets</td>
<td>10.7</td>
<td>10.8</td>
</tr>
<tr>
<td>average risk-free assets</td>
<td>0.616</td>
<td>0.514</td>
</tr>
<tr>
<td><strong>MaxTurnover</strong></td>
<td>0.430</td>
<td>0.612</td>
</tr>
</tbody>
</table>

#### Annualized statistics for the out-of-sample log-return time series

| | | | |
|---|---|---|
| **mean (%)** | 6.867 | 7.911 | 12.281 |
| **s.d. (%)** | 8.716 | 10.332 | 17.332 |
| **mean/s.d.** | 0.788 | 0.766 | 0.709 |
| **skewness** | -0.275 | -0.169 | -0.463 |
| **kurtosis** | 16.396 | 21.183 | 18.102 |

### Stock indices performance out-of-sample

<table>
<thead>
<tr>
<th></th>
<th>GSPC</th>
<th>DJI</th>
<th>IXIC</th>
<th>NDX</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Composition</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>mean (%)</strong></td>
<td>7.244</td>
<td>5.872</td>
<td>8.330</td>
<td>10.742</td>
</tr>
<tr>
<td><strong>mean/s.d.</strong></td>
<td>0.331</td>
<td>0.269</td>
<td>0.325</td>
<td>0.406</td>
</tr>
<tr>
<td><strong>skewness</strong></td>
<td>-0.791</td>
<td>-0.925</td>
<td>-0.619</td>
<td>-0.516</td>
</tr>
<tr>
<td><strong>kurtosis</strong></td>
<td>7.244</td>
<td>5.872</td>
<td>8.330</td>
<td>10.742</td>
</tr>
</tbody>
</table>

**Notes:**
1. "ann. mean / ann. s.d." refers to the ratio of the annualized log return's mean over the annualized log return's standard deviation;
2. for the proposed portfolio optimization, the annualized target return is assumed to be equal to $(1 + r^*)^{(252/N^*)}$, where $N^*$ comes from the return restriction $E(T^+(\omega)) \leq N^*$, with $N^* = 252, 186, 126, 58$;
3. in order to compute the MaxVaR, we use, respectively, the sample mean and standard deviation of the simple return time series as estimates for the drift and volatility parameters of the GBM;
4. for mean-variance and MaxVaR optimization, the reward restriction requires that the daily return's mean is equal or higher than $[\text{ann. target return}]^{1/252} - 1$;
5. $HHI_{j} = \sum_{k=1}^{m} w_{j,k}^2$ is the Herfindahl-Hirschman index and $\text{Turnover} = \sum_{j=1}^{m} |w_{j,k+1} - w_{j,k}|$, where $w_{j,k}$ and $w_{j,k+1}$ denote, respectively, the portfolio weights after and before the rebalancing $k$.

### Table 2. Out-of-sample analysis from January 2018 to December 2022
4. Conclusions

In this paper, we introduce a simple nonparametric framework that incorporates IH risk in portfolio optimization. The proposed approach is based on first-hitting probabilities, taking into account the entire asset price path, which allows us to manage the probability of breaching the maximum accepted drawdown during the investment horizon. Given the reported difficulties of the GBM approach in describing some financial data’s observed features (see, for instance, Bakshi and Panayotov 2010 and Lassance et al. 2022, Section EC.2), several alternative processes implying more flexible distributions for the returns have been considered. However, most of them do not have closed-form analytical expressions for the first passage probability, which makes it difficult to find a tractable way to estimate this function. In order to overcome these limitations, we introduce a novel nonparametric method to estimate the first-hitting time probabilities.

The proposed optimization process incorporates the path that leads to the final investment outcome and aims to minimize the probability of crossing a drawdown threshold subject to a return-based constraint: the expected time to achieve the target cumulative return should be less or equal than a predefined number of days. The investors have four parameters ($r^+, r^-, n^-$ and $n^+$ in (18) and (19)) to incorporate their IH risk-return preferences through the trade-off between the probability of crossing a drawdown threshold and the expected time to attain a target return. We provide an empirical application considering 87 companies from the S&P 100 index, a risk-free asset and four stock indices, in which we compare the in-sample and out-of-sample performance of the portfolio optimization framework we introduce with the widely used mean-variance model and an alternative optimization problem that minimizes the MaxVaR measure of Boudoukh et al. (2004) subject to a return constraint. The proposed optimization exhibits more consistency between in-sample and out-of-sample results. Except for the most ambitious case (annualized target return of 20%), the reward restriction in (19) is always satisfied. The alternative approaches result in considerably lower out-of-sample returns and volatility, mostly explained by the relatively greater weight they assign to the risk-free asset.

Summing up, the flexible and easy to implement optimization framework we introduce exhibits interesting out-of-sample performance and, therefore, could be a useful tool for portfolio managers. A possible avenue for future work would be to compare our methodology with prospect theory optimization problems (as in, for instance, Best and Grauer 2016). However, it is not an easy task since prospect theory does not explicitly incorporate return objectives in its framework.
References


Appendix A - Auxiliary Algorithms

Algorithm 1: Neighborhood Algorithm
1. Set \( \epsilon \), which is determined by a draw of a uniformly distributed, over \([0, 0.005]\), random variable;
2. Randomly select \( j_1 \in \{\text{assets with weight} \geq w_{\text{min}}\} \);
3. Randomly select \( j_2 \in \{\text{assets with weight} < w_{\text{max}}\} \);
4. \( \omega_{j_1} = \omega_{j_1} - \epsilon \);
5. \( \omega_{j_2} = \omega_{j_2} + \epsilon \);

where \( w_{\text{min}} \) and \( w_{\text{max}} \) are such that \( w_{\text{min}} < w_j < w_{\text{max}} \) for \( j = 1, \ldots, m \).

Algorithm 2: Threshold Sequence Algorithm
1. Set the number of thresholds \( n_{\text{rounds}} \) and the number of random steps \( n_{\text{deltas}} \);
2. for \( i = 1 : n_{\text{deltas}} \) do
   a. Randomly generate a feasible current solution, say, \( x^c \);
   b. Generate \( x^n \in \mathcal{N}(x^c) \) and compute \( \Delta_i = |\Psi(x^n) - \Psi(x^c)| \), where \( \mathcal{N}(x^c) \) is a neighbor of the current solution defined using the neighborhood algorithm above;
   c. Set \( x^c = x^n \);
end for;
3. Compute the empirical distribution CDF of \( \Delta_i \), \( i = 1, \ldots, n_{\text{deltas}} \);
4. Compute the threshold sequence \( \tau_k = \text{CDF}^{-1}(\frac{n_{\text{rounds}} - k}{n_{\text{rounds}}}), k = 1, \ldots, n_{\text{rounds}}, \) where \( n_{\text{rounds}} \) are equidistant quantiles.

Algorithm 3: Threshold Accepting (TA) Algorithm
1. Use the threshold sequence algorithm to construct the threshold sequence \( \tau \);
2. Randomly generate a feasible current solution, say, \( x^c \);
3. Set \( x^* = x^c \);
4. for \( r = 1 : n_{\text{rounds}} \) do
   for \( i = 1 : n_{\text{steps}} \), where \( n_{\text{steps}} \) is the number of steps per threshold, do
   generate \( x^n \in \mathcal{N}(x^c) \) and compute \( \Delta = \Psi(x^n) - \Psi(x^c) \)
   if \( \Delta < \tau_r \) then \( x^c = x^n \);
   if \( \Psi(x^c) \leq \Psi(x^*) \) then \( x^c = x^* \);
end for;
end for;
5. Return \( x^* \).
Appendix B - Technical Proofs

Proof of the result in (16)

Let \( w_t' = \left( S_{t-1}...S_{t-(k-1)} \right) \) and consider the OLS estimator,

\[
\hat{\theta} = \left( \frac{1}{n} \sum_t w_t w_t' \right)^{-1} \frac{1}{n} \sum_t w_t S_t.
\]

Thus, we can establish that,

\[
\frac{1}{n} \sum_t w_t w_t' = \frac{1}{n} \sum_t \left( \begin{array}{cc} S_{t-1}...S_{t-(k-1)} & S_{t-1}...S_{t-k} \end{array} \right) \xrightarrow{p} \left( \begin{array}{cc} \pi_{k-1} & \pi_k \\ \pi_k & \pi_k \end{array} \right),
\]

\[
\frac{1}{n} \sum_t w_t S_t = \frac{1}{n} \sum_t \left( \begin{array}{c} S_{t}S_{t-1}...S_{t-(k-1)} \\ S_{t}S_{t-1}...S_{t-k} \end{array} \right) \xrightarrow{p} \left( \begin{array}{c} \pi_k \\ \pi_{k+1} \end{array} \right).
\]

Since the Markov chain \( \{S_t\} \) is stationary, \( \{S_t\} \) is also weakly dependent, all moments are finite, and the law of large numbers holds. Consequently,

\[
\hat{\theta} \xrightarrow{p} \left( \frac{\pi_{k-1}}{\pi_k} \right)^{-1} \left( \begin{array}{c} \pi_k \\ \pi_k \end{array} \right) = \left( \begin{array}{c} \frac{\pi_{k+1}-\pi_k}{\pi_{k+1} \pi_{k-1}} \\ \frac{\pi_{k+1}-\pi_k}{\pi_{k+1} \pi_{k-1}} \end{array} \right). \tag{B.1}
\]

The vector in (B.1) simplifies using the formulas \( \pi_{k+1} = p_k \pi_k \) and \( \pi_k = p_{k-1} \pi_{k-1} \). Hence, after some calculations we get,

\[
\hat{\theta} \xrightarrow{p} \left( \frac{p_k(1-p_k)}{p_k-p_{k-1}} \right).
\]

Proof for single regressor parameter estimator in (17)

The OLS estimator of \( \theta \) is, \( \hat{\theta} = \sum_t x_t k S_t \sum_t x_t k \). Under certain regularity conditions the numerator converges to,

\[
\sum_t x_t k S_t = \sum_t \mathcal{I} \left( S_{t-1}=1, S_{t-2}=1, ..., S_{t-k}=1 \right) \xrightarrow{p} P \left( S_t = 1, S_{t-1} = 1, ..., S_{t-k} = 1 \right)
\]

and the denominator to,

\[
\sum_t x_t^2 k = \sum_t \mathcal{I} \left( S_{t-1}=1, ..., S_{t-k}=1 \right) \xrightarrow{p} P \left( S_{t-1} = 1, ..., S_{t-k} = 1 \right).
\]

where \( \mathcal{I} \) is the indicator function.

Therefore, it follows that,

\[
\hat{\theta} \xrightarrow{p} \frac{P \left( S_t = 1, S_{t-1} = 1, ..., S_{t-k} = 1 \right)}{P \left( S_{t-1} = 1, ..., S_{t-k} = 1 \right)} = P \left( S_t = 1 | S_{t-1} = 1, ..., S_{t-k} = 1 \right).
\]