Implications of conditional expectation in portfolio theory

Sergio Ortobelli ¹, Noureddine Kouaissah ²

Abstract
This paper examines and explores the implications of using conditional expectation estimators in portfolio theory. In particular, we focus on two financial applications – (i) approximation of the conditional expectation within large-scale portfolio selection problems, and (ii) performance valuation considering the heavy tails of returns. The aim is to examine the extent to which bandwidth selection and kernel functions impact portfolio returns estimation. Thus, we compare the ex-post wealth obtained from applying the portfolio strategies, which use alternative performance measures based on a conditional expectation.

Keywords
Conditional expectation, large-scale portfolio selection, performance measures, bandwidth selection.

JEL Classification: G11, G12.

1. Introduction
In financial literature, it is well-known that asset returns are not normally distributed and show heavy tails, for a survey of criticism see Mandelbrot [9], Fama [5] and Rachev and Mittnik [11]. Empirical research has established few stylized facts about asset returns: (a) clustering of volatility, (b) skewness, and (c) fat tails (see for instance Kim et al. [8]). Over the years, a significant number of studies on portfolio selection problems have been published. Most of these have suggested different formulations based on operations research models that try to overcome the mean-variance shortcomings.

In this paper, we assess the impact of non-parametric techniques that use conditional expectation estimators in portfolio theory. Conditional expectation is a fundamental concept in probability and statistics and is extremely useful in financial modeling. It plays a significant role in portfolio theory and various pricing and risk management problems. In particular, we examine the impact of bandwidth selection and kernel function choices in two financial applications: (i) approximation problems within large-scale portfolio selection problems, and (ii) performance valuation considering the heavy tails of returns.

The first contribution of this paper is to investigate the impact of an alternative return approximation method that depends on $k$-factors in large-scale portfolio problems (e.g., the $k$-fund separation model of Ross [13]). In particular, we examine a non-parametric approximation of returns based on factors obtained using a principal components analysis (PCA). The most common approach to estimate the relationship between returns and $k$-

¹ Sergio Ortobelli Lozza, University of Bergamo, Italy – VSB – TU of Ostrava, Czech Republic, sol@unibg.it
² Noureddine Kouaissah, University of Bergamo, Italy, noureddine.kouaissah@unibg.it

The research was supported through the Czech Science Foundation (GACR) under project 17-19981S and by VSB-TU Ostrava under the SGS Project SP2018/34.
factors is the linear approximation based on the ordinary least squares (OLS) estimator (see Ross [13]). This approximation appears to be good enough when the returns are normally distributed. However, Ortobelli et al. [15] found that the non-parametric regression outperforms its parametric counterpart. Moreover, we believe there is substantial evidence of non-linearity in the financial returns (see, e.g., Rachev et al. [12]). For this reason, we use a non-parametric regression analysis to approximate the returns. This approach relaxes the assumptions of linearity and is suitable even for non-Gaussian distributions. The empirical analysis is provided using a portfolio consisting of S&P 500 components assuming that no short sales are allowed since in general the markets admit only limited short sales and during a crisis period the short sales are strictly limited or even not allowed.

The second contribution of this paper is to properly evaluate portfolio choices that take into account the tails of a portfolio distribution. Portfolio distributional tails play a crucial role in finance because they represent the probabilities to obtain losses or gains. In this regard, we consider the behavior of the portfolio returns when some particular events have an impact on the portfolio or on the market (positive or negative trends, crisis, etc.). In particular, we forecast the conditional expected portfolio returns with respect to a given σ-algebra of events (generated either by possible profits or by possible losses) using a new alternative conditional expectation estimator proposed by Ortobelli et al. [14]. Thus, we empirically test sound portfolio performance measures recently proposed by Ortobelli et al. [15] when the most significant events (for investors) happen. Moreover, we illustrate how the new performance measures are impacted by the choice of the kernel function and bandwidth selection. Here, we use an ex-post empirical analysis to show their greater capacity to produce wealth in the US market considering some different rules.

The rest of the paper is organized as follows. In Section 2, we discuss and examine the impact of the approximation method within large-scale portfolio selection problems. In Section 3, we introduce and empirically test two performance measures based on a conditional expectation. Finally, our conclusions are summarized in Section 4.

2. Practical and theoretical aspects of return approximation in portfolio selection problems

In this section, we focus on approximation methods within large-scale portfolio selection problems. In practice, there are many different ways to reduce the dimensionality dependence of a large-scale portfolio selection problem see Ortobelli et al. [16]. In this section, we widely discuss a non-parametric regression model to reduce the dimensionality of a large-scale portfolio problem. In particular, we reduce the dimensionality of the problem and approximate the return series using a multifactor model that depends on a proper number (i.e., not too large) of factors. Following Ortobelli et al. [15], we determine the principal components by applying the PCA to a proper Pearson linear correlation matrix of returns.

We consider n risky assets defined on a probability space \((\Omega, \mathcal{F}, \mathbb{P})\), where \(\mathcal{F}\) is a σ-algebra of events on \(\Omega\) (i.e., a collection of subsets of closed under all countable set operations: union, intersection, and complement). We point out the portfolio gross returns \(x'z\), where \(x = [x_1, ..., x_n]'\) is the vector of nonnegative allocations among \(n\) risky limited liability investments, with gross returns \(z = [z_1, ..., z_n]'\). According to the portfolio literature, a portfolio gross return admits a finite mean and belongs to \(L^p(\Omega, \mathcal{F}, \mathbb{P})\), (the set of all \(\mathcal{F}\) measurable real random variables \(X : \Omega \to \mathbb{R}\) such that the \(p\)-th absolute moment is finite i.e., \(E(|X|^p) < 1\) for some \(p \geq 1\).
With parametric regression models, we can replace the original \( n \) correlated time series \( \{z_i\}_{i=1}^n \) with the \( n \) uncorrelated time series \( \{R_i\}_{i=1}^n \) (obtained by the PCA), assuming that each \( z_i \) is a linear combination of \( R_i \). In particular, the dimensionality reduction is obtained by choosing only the first factors that sufficiently summarize a large part of the variability.

In this setting, we call portfolio factors \( f_i \) the first \( s \) principal components \( R_i \) with significant variability, while the remaining \( n-s \) principal components with smaller variability are summarized by an error, \( \varepsilon \). Typically, the OLS estimator is widely used to approximate the returns using the following linear relation:

\[
z_i = b_{i,0} + \sum_{j=1}^{s} b_{i,j} f_j + \sum_{j=s+1}^{n} b_{i,j} R_j = b_{i,0} + \sum_{j=1}^{s} b_{i,j} f_j + \varepsilon_i, \quad i = 1, \ldots, n
\]  

(1)

where \( z_i \) is the gross return for the asset \( i \), \( b_{i,0} \) is the fixed intercept for the asset \( i \), \( b_{i,j} \) is the coefficient for factor \( f_j \), \( s \) is the number of factors, \( \varepsilon_i \) is the error term for asset \( i \) and \( n \) is the number of assets.

In general, the OLS estimator is a well-established and very useful procedure for solving regression problems when the returns are normally distributed. However, the returns are often characterized by a heavy-tailed distribution (see, e.g., Rachev and Mittnik [11]). Therefore, we cannot assume that the dependence between the returns and the principal components is linear. For this reason, we use a non-parametric regression analysis as an alternative to the classic parametric approach (1). In several financial models (APT, CAPM, etc.) the returns are assumed to be elliptically distributed (e.g., Chamberlain [4] and Ingersoll [7]), and the large-scale portfolio problem is solved by approximating the returns using a regression model on some uncorrelated market factors. Differently, we reduce the complexity of the large portfolio model using a non-parametric regression model, where \( s \) factors are determined by applying a PCA to a linear correlation measure and the \((s + 1)\)th factor \( M_{s+1} \) is a market index, that is,

\[
z_i = E(z_i|f_1, f_2, \ldots, f_s, M_{s+1}) + \varepsilon_i = m(f) + \varepsilon
\]  

(2)

One of the most used estimators for \( m(\cdot) \) is the multivariate version of the Nadaraya-Watson kernel estimator. However, this estimator presents certain disadvantages. In particular, it corresponds to the local constant fit and may be biased, depending on the marginal density of the design. To overcome these drawbacks, a general class of non-parametric regression estimator based on locally weighted least squares has been proposed (see, e.g., Ruppert and Wand [17]). In this setting, an estimate of the regression function \( m(f) \) is readily obtained by estimating the parameter \( a \) as the argument in the following minimization problem:

\[
\min_{a} \sum_{i=1}^{T} \left( z_i - a - b'(f_{(i)} - f) \right)^2 K_h(f_{(i)} - f)
\]  

(3)

where \( H \) is an \( s \times s \) symmetric positive definite matrix that depends on the number of observations \( T \), \( f_{(i)} \) is the \( i \)-th observation of vector \( f \), and \( K_h(\cdot) \) is a multivariate kernel estimator. Several kernel functions exist in the literature, see among others Scott [18] for their review. We provide some possible choices of the kernel function in the univariate setting (multivariate kernel function could be seen as the product of the univariate kernel):

- Uniform kernel \( K(u) = \frac{1}{2} \mathbf{1}(|u| \leq 1) \);
- Epanechnikov kernel \( K(u) = \frac{3}{4} (1 - u^2) \mathbf{1}(|u| \leq 1) \);
- Quadratic kernel \( K(u) = \frac{15}{16} (1 - u^2)^2 \mathbf{1}(|u| \leq 1) \);
Gaussian kernel $K(u) = (2\pi)^{-1/2} \exp\left(-\frac{u^2}{2}\right)$.

Several researchers have shown that the choice of kernel is not critical, while the performance of the smoothed regression function is more a question of bandwidth choice. Fan and Gijbels [6] give a survey on bandwidth selection for the univariate local polynomial smoothing technique, which contains the Nadaraya–Watson estimator as a special case. However, there is little guidance in the literature on bandwidth selection for multivariate kernel density estimation, which certainly remains an important issue in empirical studies. The most widely used bandwidth selection methods are the rule-of-thumb and the plug-in bandwidth selections. In particular, the former is the normal reference rule for kernel density estimation presented in Bowman and Azzalini [2]. For general multivariate kernel estimators, Scott [18] suggests the following bandwidth selectors:

$$\hat{h}_i = \hat{\sigma}_i n^{-1/(d+4)},$$

where $\hat{\sigma}_i$ is the usual estimate of the standard deviation of each variable $x_i$, and $n$ is the sample size.

To emphasize the implications of non-parametric approach, we examine recent performance measures, proposed by Ortobelli et al. [15], based on the conditional expectation that takes into account the portfolio distributional behavior on the tails. More specifically, the first suggested performance measure is based on two different $\sigma$-algebras (the $\sigma$-algebra generated by the portfolio losses, and the $\sigma$-algebra generated by the portfolio profits). The second performance measure considers $\sigma$-algebras generated by the joint losses and gains of all assets in the market. These $\sigma$-algebras are approximated using $\sigma$-algebras generated by proper partitions (of losses or of gains). Due to space limitation, for theoretical and practical discussion on the new performance measures, we refer to Ortobelli et al. [15]. The TOK ratio (see Ortobelli et al. [15]) is a highly flexible performance measure that considers the expected portfolio returns, given the $\sigma$-algebras generated by the portfolio profits and the $\sigma$-algebra generated by the portfolio losses. Moreover, in view of this approach, the second performance measure JTOK considers an important feature of the market, namely, the joint losses and gains among the assets. This performance measure takes into account the heavy tails of all returns jointly (i.e., the joint losses and gains among the assets) that should be considered by portfolio managers.

The aim of this paper is not to discuss or to define the new performance measures, which are well presented in Ortobelli et al. [15], but to examine the extent to which these performance measures are impacted by estimating the conditional expectation. In particular, we investigate the implications of bandwidth selection and kernel function choice. The following section as illustration presents the differences among two well-known reference rules and distinguishes two defined kernel functions.

3. Empirical analysis

In this section, we compare the optimal portfolio approaches solved according to the performance measures (i.e., TOK and JTOK ratios) with different parameter choices, see Ortobelli et al. [15] for more details. In other words, the ex-post wealth of the optimal portfolios is evaluated considering the new performance measures and different approximation methods. We use all active stocks on the S&P 500 index from October 28, 2003, to June 10, 2016. The data set is obtained from the Thomson Reuters DataStream database and consists of 3177 daily observations. In particular, we employ the proposed techniques to reduce the dimensionality of large-scale portfolio problems (see Section 2).
Specifically, we perform a PCA on a Pearson correlation matrix of the returns. Then, we approximate the portfolio returns using the RW regression model. Moreover, in order to evaluate the conditional expected value with respect to a given $\sigma$-algebra, we use the OLP estimator as suggested by Ortobelli et al. [14].

We recalibrate the portfolio every month (20 trading days), respectively. To guarantee sufficient diversification, which is known to lead to greater long-run portfolio performance (see, e.g., Statman [19]), we set the upper bound of investment to each asset to 20% (i.e., $x_i \leq 0.2$). We use a moving average window of 500 trading days to compute each optimal portfolio. Following Biglova et al. [3], we use a larger number of observations (namely 500) to compute and estimate the JTOK ratio, since we have to account the joint risk of several assets. Starting with an initial wealth $W_0 = 1$, which we invest on October 28, 2003, we evaluate the ex-post wealth sample paths obtained by maximizing the new performance measures. Therefore, at the $k$-th optimization, three steps are performed to compute the ex-post final wealth:

**Step 1** Apply the PCA to the Pearson correlation matrix. Then, approximate the portfolio returns as suggested in Section 2 (via the RW estimator). Here, we investigate the implications of bandwidth choices and kernel functions. In particular, due to space limitations, we use two methods for bandwidth selection (namely Scott’s rule and Bowman and Azzalini rule (for more details on this rule see Bowman and Azzalini [2]) and two different kernel functions (namely, Gaussian and Epanechnikov). An extended version of this paper considers various rules and different kernel functions to shed some light on the impact of their choices.

**Step 2** Determine the market portfolio that maximizes either TOK or JTOK performance measures applied to the approximated returns. Observe that when we apply the new performance measures, the computational complexity must be analyzed since the conditional expected portfolio is not a linear function of portfolio weights. Consequently, we may have many local maxima and an increased computational complexity. Clearly, standard optimization algorithms may be not adequately suited to optimize portfolio problems. Thus, in order to optimize performance measures in an acceptable computational time, we use as a starting point the optimal solution obtained using a heuristic algorithm for overall optimization [the ones proposed by Angelelli and Ortobelli [1] for portfolio optimization]. Then, we improve this solution by applying the Matlab heuristic function pattern search implemented in Matlab 2015 to solve global optimization problems.

**Step 3** Compute the ex-post final wealth. In this step, we consider proportional transaction costs of 20 basis points.

We apply the algorithm until the observations are available. The results of this analysis are summarized in Tables 1-2. In both Tables, we present summary statistics (mean, standard deviation, VaR5%, CVaR5%, final wealth) for the ex-post log returns obtained by maximizing the TOK and JTOK ratios with Pearson correlation matrix and using two estimators of the conditional expectation (i.e., OLP and RW). Table 1 illustrates the results of the Gaussian kernel function and two bandwidth rules. Table 2 represents the case of the Epanechnikov kernel function and the two reference rules.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>St.dev.</th>
<th>VaR5%</th>
<th>CVaR5%</th>
<th>Final W</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scott’s rule</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOK(5)</td>
<td>0.134</td>
<td>2.839</td>
<td>4.008</td>
<td>6.485</td>
<td>18.2577</td>
</tr>
<tr>
<td>TOK(7)</td>
<td>0.112</td>
<td>2.721</td>
<td>3.838</td>
<td>6.433</td>
<td>10.9141</td>
</tr>
</tbody>
</table>

Table 1: Statistics of the ex-post returns obtained by maximizing the TOK and JTOK ratios using the Gaussian kernel function and two reference rules.
Table 2: Statistics of the ex-post returns obtained by maximizing the TOK and JTOK ratios using the Epanechnikov kernel function and two reference rules

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>St.dev.</th>
<th>VaR5%</th>
<th>CVaR5%</th>
<th>Final W</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Scott’s rule</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOK(5)</td>
<td>0.137</td>
<td>2.828</td>
<td>4.011</td>
<td>6.315</td>
<td>18.2497</td>
</tr>
<tr>
<td>TOK(7)</td>
<td>0.114</td>
<td>2.732</td>
<td>3.829</td>
<td>6.458</td>
<td>10.8438</td>
</tr>
<tr>
<td>TOK(10)</td>
<td>0.142</td>
<td>3.424</td>
<td>4.911</td>
<td>7.837</td>
<td>13.8911</td>
</tr>
<tr>
<td>JTok(3)</td>
<td>0.102</td>
<td>2.201</td>
<td>3.404</td>
<td>5.381</td>
<td>10.7215</td>
</tr>
<tr>
<td>JTok(5)</td>
<td>0.077</td>
<td>2.211</td>
<td>3.442</td>
<td>5.461</td>
<td>5.0864</td>
</tr>
<tr>
<td>JTok(10)</td>
<td>0.074</td>
<td>2.169</td>
<td>3.339</td>
<td>5.371</td>
<td>4.9124</td>
</tr>
<tr>
<td><strong>Bowman and Azzalini rule</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOK(5)</td>
<td>0.129</td>
<td>2.739</td>
<td>3.401</td>
<td>6.317</td>
<td>17.4242</td>
</tr>
<tr>
<td>TOK(7)</td>
<td>0.110</td>
<td>2.531</td>
<td>3.239</td>
<td>6.250</td>
<td>10.1245</td>
</tr>
<tr>
<td>TOK(10)</td>
<td>0.133</td>
<td>3.129</td>
<td>4.528</td>
<td>7.260</td>
<td>12.9015</td>
</tr>
<tr>
<td>JTok(3)</td>
<td>0.099</td>
<td>2.127</td>
<td>3.381</td>
<td>5.250</td>
<td>10.3321</td>
</tr>
<tr>
<td>JTok(5)</td>
<td>0.074</td>
<td>2.109</td>
<td>3.238</td>
<td>5.129</td>
<td>4.9891</td>
</tr>
<tr>
<td>JTok(10)</td>
<td>0.071</td>
<td>2.037</td>
<td>3.201</td>
<td>5.217</td>
<td>4.4369</td>
</tr>
</tbody>
</table>

From Tables 1-2, we observe that:

- The impact of bandwidth choice is slightly greater than the choice of the kernel function. The results are slightly different from Scott’s rule to the Bowman and Azzalini rule even maintaining the same evolution. The differences impacted by kernel function is minimal and this in line with previous research that has shown that the choice of kernel is not critical, while the performance of the smoothed regression function is more a question of bandwidth choice.

- The TOK strategies present the greatest average, final wealth, Sharpe ratio (mean/St.dev.). However, they show the highest risk (standard deviation, VaR5%, CVaR5%).

- The strategies based on TOK and JTOK performance measures perform much better than the S&P 500 benchmark, which shows the worst results in terms of mean, final wealth, and Sharpe ratio. However, it achieves the lowest risk (standard deviation, VaR5%, CV aR5%) among the strategies.

Overall, these results confirm and provide strong support for the proposed performance measures, and the proposed methodology to reduce the dimensionality of the problem, taking into consideration the bandwidth selection problem.
4. Conclusion

Portfolio selection problems often involve unknown parameters that have to be properly approximated from the data. Therefore, in this paper, we consider the implications of conditional expectation estimators, especially the bandwidth selection and kernel function choice, within the portfolio theory. In particular, we focus on two financial problems, namely, approximation problems within large-scale portfolio selection problems and optimal portfolio choices with consistent estimations of the conditional expected returns. Thus, we test new performance measures that account for the heavy-tailed distribution of the returns and their joint risk. In this context, we stressed the important role of bandwidth selection with respect to the kernel function choice. In addition, we confirm that the proposed portfolio selection models yield the best performance and outperform the market benchmark. Therefore, the proposed empirical analysis supports the significance of the conditional expectation estimators within the portfolio theory.

Further research could involve theoretical and empirical studies. On the one hand, a natural extension of this research would take into account several rules and parameters. On the other hand, the use of semi-nonparametric approaches could be considered. Future research will investigate these aspects.

References


