Università degli Studi di Napoli "Federico II"

Dipartimento di Scienze Economiche e Statistiche

Corso di Dottorato in Economia

XXXIII ciclo

Two P-Splines applications to portfolio selection problems

Tutor
Prof. Massimo Aria

Candidato
Marco Scaglione

Coordinatore
Prof. Marco Pagano

Anno Accademico 2020-2021

To my F. the sole constant among all the variables.

Contents

List of Figures

List of Tables

Incipit

In statistics to smooth a dataset means to create an approximating function that attempts to capture important patterns in the data, leaving out noise or other scale-structures pheonomena. The data points are modified so individual points higher than the adjacent points are reduced and points that are lower than the adjacent points are increased leading to a smoother data, under the smoothing label falls a set of tools which objective is to remove errors and noise from data.
Noise has negative effects on predictions and descriptions of data, and one of the main task of the statistician is to produce information that minimizes the amount of errors in the data, this is because in empirical applications one usually believe that data are the best available description of the world and at the same time has no interest in explaining and identifying all the sources of variation.
Also in economics, at least under the reductionist paradigm, smoothness has a huge role in the formulation of theories and models trying to describe the rational behaviour, and noise is identified as one of the main source for the theories and models failures. The main difference between the two approaches is that the statistician produces an analysis that attempts to minimize the unexplained variation in the data, while the econometrician models the error to isolate the desired effect that its analysis aims to inspect.
In this manuscript I provide two applications of an established smoothing technique, the Penalized Spline smoother, to solve the roughness issue in econometric analysis of the portfolio selection problem. In Chapter 1 I face the problem of statistical hedge ratio estimation through quantile regression, here the result of the P-spline application is to avoid model specification and to smooth the quantile regression objective function. In Chapter 2 the smoother is applied in a time series filtering framework in order to achieve smoother Principal Component Analysis and provide smoother principal component for regression analysis, this allows robust feature selection from principal component loadings in order to perform portfolio selection with index tracking purpose.

Chapter 1

P-spline Quantile Regression Hedge Ratio

Modern portfolio theory, as defined by Markowitz (1952), suffers of two major flaws, either one assumes Gaussian asset returns or the agent utility function is assumed to be quadratic, these allow portfolio weights estimation by ordinary least squares regression method. In this chapter I try to loose these assumptions applying P-spline quantile regression method to the task of Choquet risk minimization in an exercise of statistical portfolio estimation.

1.1 Introduction

Portfolio optimization is the selection process of the best asset allocation according to some criterion. It is one of the most controversial and long-lived discussion topics since the dawn of civilization and the invention of writing, its references can be found in the Old Testament[1] as in the Gospel [2]. However, aside from these exotic references, the topic has been developing as the financial market itself, drawing the attention of several scholars belonging to different fields that produced a plethora of theories, strategies and methods to be applied in order to achieve a scientific and sound investment process.
In this chapter I am going to focus on a specific subject within the broader topic of portfolio optimization, discussing the estimation issues concerning the most famous hedging procedures and attempting to tackle them through the application of an already well studied method in the domain of statistics, the P(enalized)-Spline estimator.
The cornerstone of modern portfolio theory is expected utility maximization as elaborated by Bernoulli(1737), Ramsey(1931), de Finetti(1937), von Neumann and Morgenstein (1944) and many other authors[46]. Many are its pro-

[1] Genesis 41:34-36 "Let Pharaoh appoint commissioners over the land to tackle a fifth of the harvest of the Egypt during the seven years of abundance...".
[2] Matthew 25:14-30, better known as the "Parable of the Talents".

posed variations, one of the most successful has been the family of non-additive, or rank-dependent, formulations of Quiggin (1982)[121], Yaari(1987)[140] and Schmeidler(1989)[127] , that replaces the Lebesgue integral with the Choquet integral[33] , thus accentuating the probability of the least favorable outcomes and yielding a pessimistic decision criterion.
In this kind of framework the selection problem between two random variables, X and Y, characterized by their distribution function F_x and F_y, is solved through their quantile functions $F^{-1}(t)$ and $G^{-1}(t)$, and Choquet utility introduces a distortion of the original probability assessment that allows to integrate $dv(t)$ with respect to some other probability measures defined on the interval $[0,1]$.
The distortion function v inflates or deflates the probabilities according to the rank ordering of the outcomes, and can be seen as a reflection of the optimism or pessimism of the decisor agents, leading to the quite schizophrenic situation of a decision maker that accepts probabilities represented by the distribution functions F and G and then distorts these probability before making decisions.
To be short, while in the traditional expected utility framework (supposing that the initial wealth is embodied in the two random variables), one prefers X to Y if

$$\mathbb{E}_F\, u(X) = \int_{-\infty}^{\infty} u(x)\, dF(x) > \int_{-\infty}^{\infty} u(y)\, dG(y) = \mathbb{E}_G\, u(Y), \tag{1.1}$$

in the Choquet framework, X is preferred to Y if

$$\mathbb{E}_{v,F}\, u(X) = \int_0^1 u(F^{-1}(\tau))\, dv(\tau) > \int_0^1 u(G^{-1}(\tau))\, dv(\tau) = \mathbb{E}_{v,G}\, u(Y). \tag{1.2}$$

Because in a risk mitigation exercise the flaws of mean-variance optimization emerge more critically (indeed non-Gaussian tail behaviour of empirical return distribution leads to undesired risk taking, thus sub-optimal behaviour), in this paper the comparison between this two approach will be conducted in the statistical estimation of the optimal hedge ratio within twenty-three stock indeces and their respective front month future contract.
This chapter links to a wide branch of literature trying to stress the problem of determining the optimal hedging whenever none of the standard conditions occur, the next section will provide an insightful review of this branch, from a theoretical and empirical point of view. Section number three will briefly review the quantile regression model, its estimation through P-spline and the inference procedure (however the P-spline estimator has been more comprehensively treated in appendix A) and section four presents its empirical application to the estimation of the optimal hedge ratio for twenty-three major stock indices. The fifth, and last part, will be devoted to a discussion on the results and on the subsequent conclusions.

1.2 Literature Review

1.2.1 Mean-Variance and Pessimistic risk hedging

Mean-Variance and its less fortunate variations

The basic concept of hedging is to combine positions in different assets to make a portfolio that reduces fluctuations in its value, i.e. a portfolio whose return dispersion is less than the sum of the dispersions of its component returns. In this section I will always consider a portfolio consisting of w_S shares of wealth on a long position in the spot market and w_F shares of wealth on a short position in the futures market. The return of the portfolio, R_P, will thus be given by:

$$R_P = \frac{w_S S_t R_S - w_F F_t R_F}{w_S S_t} = R_S - hR_F,$$

where F_t, S_t are the futures and spot prices in time t, R_F, R_S are respectively futures and spot returns related to $t-1$ and thus $h = \frac{w_F F_t}{w_S S_t}$ is the hedge ratio. This latter quantity depends on a particular objective function to be optimized, and leads to the first distinction within the strategies to be discussed next.
I consider the static case, where h doesn't change over time. In this case the most famous hedge ratio comes from Harry Markowitz' *"Modern Portfolio Theory"*[106] in his seminal papers from the fifties, adapted to the hedging problem by Jhonson(1960)[85, 44].
In this framework the hedge ratio is the one minimizing the portfolio risk as resulting by the variance of the changes in its value, stated as

$$Var(R_P) = w_S^2 Var(R_S) + w_F^2 Var(R_F) - 2w_S w_F Cov(R_S, R_F),$$

thus the MV hedge ratio is given by

$$h^*_{MV} = w_F/w_S = \frac{Cov(R_F, R_S)}{Var(R_F)} = \rho\frac{\sigma_S}{\sigma_F}. \tag{1.3}$$

On the same theoretical foundations lays another estimation strategy, now incorporating the portfolio return in the hedging and based on the risk-return trade-off as formulated by William F. Sharpe [132, 133, 134] and developed in the hedging framework by Howard and D'Antonio [71, 72, 73], that considers the optimal level of contracts that maximizes the portfolio's excess return to its volatility:

$$\underset{w_F}{Max}\frac{\mathbb{E}(R_P) - R_f}{\sigma_P},$$

where R_f is the risk-free interest rate. In this case one has that the optimal shares of future position is given by

$$w_F^* = -w_S\frac{\left(\frac{S}{F}\right)\left(\frac{\sigma_S}{\sigma_F}\right)\left[\frac{\sigma_S}{\sigma_F}\left(\frac{\mathbb{E}(R_F)}{\mathbb{E}(R_S)-R_f}\right) - \rho\right]}{1 - \frac{\sigma_S}{\sigma_F}\left(\frac{\mathbb{E}(R_F)\rho}{\mathbb{E}(R_S)-R_f}\right)},$$

that allows to determine the optimal hedge ratio as

$$h_S^* = -\frac{\left(\frac{\sigma_S}{\sigma_F}\right)\left[\frac{\sigma_S}{\sigma_F}\left(\frac{\mathbb{E}(R_F)}{\mathbb{E}(R_S)-R_f}\right)-\rho\right]}{1-\frac{\sigma_S}{\sigma_F}\left(\frac{\mathbb{E}(R_F)\rho}{\mathbb{E}(R_S)-R_f}\right)}, \tag{1.4}$$

if one makes the standard assumption that $E(R_F) = 0$ then

$$h_S^* = h_{MV}^*.$$

As pointed out by Chen et al. (2001)[30], the Sharpe ratio is highly non-linear function of the hedge ratio, and this can lead to solutions that may minimize rather than maximize the Sharpe ratio.

A first deviation from the MPT framework (which assume either quadratic utility function or normally distributed assets returns) is the stochastic dominance approach, in which the analyst knows that the agents on the market maximize their utility from returns but ignores their utility functions. One example, pertinent to the application to be developed, is the strand of literature applying the Extended Mean Gini coefficient[130, 31, 95][3] defined as:

$$\Gamma_v(R_P) = -vCov(R_P, (1 - F(R_P)^{v-1})), \tag{1.5}$$

where $F(\cdot)$ is the cumulative distribution and v is the risk aversion parameter. To define the optimal hedge ratio in this framework, one differentiates the portfolio equation with respect to h and obtains:

$$h_{MEG}^* = \frac{-Cov(R_F, (1 - F(R_P))^{v-1})}{\frac{\partial Cov(R_F,(1-F(R_P))^{v-1})}{\partial h}}. \tag{1.6}$$

This latter expression is not easy to compute due to the partial derivative at the denominator, thus scholars rely on search grid optimization methods to approximate its value.

Aside from the purely theoretical point of view, also the knowledge about the empirical distribution of returns led to some modification in the criterion for the determination of the optimal hedge ratio, indeed during the seventies has been pointed out that returns distributions are skewned and leptokurtic, proving that co-skewness and co-kurtosis of asset returns (which measure the contribution that an asset makes to the skewness and kurtosis of a portfolio) are priced by the markets, thus to not optimality of minimum variance hedging due to ambiguous effects on portfolio returns distribution third and fourth moments. Intensive studies [24, 55, 68] have been conducted on portfolio optimization on

[3]The Gini Index is a measure of statistical dispersion, developed by Corrado Gini to measure how far a country's wealth deviates from a totally equal distribution, that is:

$$G = \frac{1}{\mu}\int_0^1\int_0^1 |Q(X_1) - Q(X_2)| dX_1 dX_2$$

where $Q(\cdot)$ is the quantile function.

the mean-VaR (Value-at-Risk) space, leading to criteria that have the following form over a given time period τ:

$$\alpha\text{-VaR}(R_P) = Z_\alpha \sigma_P \sqrt{\tau} - \mathbb{E}[R_P]\tau$$

which results in the *"zero-VaR"*[75] hedge ratio, given by

$$h^*_{VaR} = \rho\frac{\sigma_S}{\sigma_F} - \mathbb{E}[R_F]\frac{\sigma_S}{\sigma_F}\sqrt{\frac{1-\rho^2}{Z_\alpha^2\sigma_F^2 - \mathbb{E}[R_F]^2}}. \tag{1.7}$$

Choquet expected utility and pessimistic portfolio allocation

The story of the development of the pessimistic framework for portfolio allocation is not clear and linear as the one of the mean-variance portfolio because the former has been developed almost autonomously in two different fields with different motivations.
The pure economics theoretical roots for the pessimistic portfolio allocation has been developed in the field of the Behavioral Economics, the field under which one could group all the studies highlighting (either experimentally or rethorically) the paradoxes induced by classical expected utility formulation and providing alternative framework overcoming them. Trying to overcome the puzzling observation made by Friedman and Savage (1948)[58], that many people are going to buy insurance and gamble at the same time, Quigging (1982)[121] and Schmeidler (1989)[127] noted that some distortion functions initially concave and then convex may explain this behavior, so if one defines the simplest distortion $v_\alpha(t) = \min\{t/\alpha, 1\}$ then has $\mathbb{E}_{v_\alpha} u(X) = \alpha^{-1}\int_0^{-\alpha} u(F^{-1}(t))dt$, meaning that the α least-favorable outcome has an inflated probability while the $1-\alpha$ proportion of the most-favorable outcomes are entirely discounted. Following Schmeidler's article comonotonicity definitions[4], one can loose the independence axiom and achieve the monotone invariance of the quantile function in the formulation of the pessimistic portfolio theory as stated by Bassett, Koenker and Kordas in their 2004 article[12].
The other pillar upon which is based pessimistic portfolio allocation is a branch of literature emerged in the late nineties in the field of quantitative finance concerning the portfolio risks measures. An influential article in this branch is the one by Artzner et al. (1999)[8] which defines the axiomatic foundation for "coherent" risk measures.

Definition. For real-valued random variables $x \in \chi$ on (Ω, Y^S) a mapping $\varrho : \chi \to R$ is called a coherent risk measure if it is:

---

[4]In Schmeidler own word:

Definition. Two acts, f and g, in Y^S are comonotonic if for no s and c in S, $f(s) \succ f(t)$ and $g(t) \succ g(s)$.

(*Comonotonic independence axiom*) For all pairwise comonotonic acts f, g and h in L and for all α in $]0,1[$: $f \succ g$ implies $\alpha f + (\alpha)h \succ \alpha g + (1-\alpha)h$.

1. Monotone: $x, y \in \chi$, with $x \leq\ y \Rightarrow \varrho(x) \geq \varrho(y)$.
2. Subadditive: $x, y, x+y \in \chi, \Rightarrow \varrho(x+y) \leq \varrho(x) + varrho(y)$.
3. Linearly homogeneous: For all $\lambda \geq 0$ and $x \in \chi$, $\varrho(\lambda, x) = \lambda\varrho(x)$.
4. Translation invariant: For all $\lambda \in \mathbb{R}$ and $x \in \chi$, $\varrho(\lambda + x) = \varrho(x) - \lambda$.

This definition eliminated many conventional risk measures traditionally used in finance, ruling out all those based on second moments by monotonicity requirment and those quantile-based (including the α-VaR) by subadditivity.
This drove a boost in the research of robust risk measures and subsequently in strategies for portfolio optimization in the subsequents plethora of new spaces developed from these new metrics. Conditional Value-at-Risk by Rockafellar and Uryasev(2000) (from now on C-VaR)[122] is a measure of downside risk that overcomes the shortfalls of traditional α-VaR in terms of coherence,

$$\text{C-VaR} = (1-c)^{-1} \int_{-1}^{VaR} R_P p(R_P) dR_P,$$

meaning that is something like the mean of the loss exceeding α-Var and leading to an optimization problem that can be stated in the following way

$$h^*_{C-VaR} = \underset{h \in \mathbb{R}}{argmin} \quad u(\alpha, h) = \underset{(h,v) \in \mathbb{R} \times \mathbb{R}}{argmin} F_\alpha(h, v), \tag{1.8}$$

with $F_\alpha(h, v) = v + \alpha^{-1}\,\mathbb{E}[(-r_p - v)^+]$ and $(x)^+ = max(x, 0)$. Other prominent example of coherent risk measure are the expected shortfall proposed by Acerbi and Tasche in (2002)[1] and tail conditional expectation[8].
In the framework of pessimistic portfolio allocation a considerable coherent risk measure is

$$\varrho_{v_\alpha}(x) = -\int_0^1 F^{-1}(t) dv(t) = -\alpha^{-1} \int_0^\alpha F^{-1}(t) dt,$$

that leads straightforward to the following optimization problem

$$\min_h \varrho_{v_\alpha}(R_P) - \lambda\mu(R_P), \tag{1.9}$$

with the intuitive meaning of minimizing the α-risk measure subject to a contraint on mean return (another $alpha$-risk too, since $\mu(R_P) = -\varrho_{v_1}(R_P)$).

1.2.2 Estimation Methods of the Optimal Hedge Ratio

OLS Method

The conventional[44, 85] approach to the MV hedge ratio estimation is a straight forward application of the classical regression model on the difference of the prices, written as

$$\Delta S_t = \beta_0 + \beta_1 \Delta F_t + \varepsilon_t, \tag{1.10}$$

whence is immediate to understand that the empirical counterpart of the Minimum Variance hedge ratio (1.3) is β_1, this approach has the downside that the hedge ratio is estimated using unconditional sample moments, thus giving the same weight to all past informations instead of giving more reliance to the newer ones. Another approach[112], also relying on the OLS method, suggest the use of conditional covariance and variances in order to obtain a conditional version of the optimal hedge ratio with the following form:

$$h^*_{MV} = \frac{w_S}{w_F} = \frac{Cov(\Delta S, \Delta F)|\Omega_{t-1}}{Var(\Delta F)|\Omega_{t-1}},$$

where Ω_{t-1} is the current information, including a vector of variables $\mathbf{X_{t-1}}$ and the spot and futures price changes as generated by the following equilibrium model

$$\Delta S_t = X_{t-1}\alpha + u_t$$
$$\Delta F_t = X_{t-1}\beta + v_t$$

that allows an agile computation through a straightforward application of the Frisch-Waugh theorem, leading to

$$\hat{h}|X_{t-1} = \frac{\hat{\sigma}_{uv}}{\hat{\sigma}^2_v}, \tag{1.11}$$

where $\hat{\sigma}_{uv}$ is the sample covariance between the residuals u_t and v_t, and $\hat{\sigma}^2_v$ is the sample variance of the residual v_t. Here, again, one can see that if the spot and futures prices follow a random walk, with or without drift, the two estimation strategies produce the same results.

ARCH and GARCH Method

Since the development of ARCH and GARCH models onward, the OLS hedge ratio estimation method has been generalized to take into account the heteroskedastic nature of the error term, so the unconditional sample variance and covariance in (1.3) have been substituted by the conditional variance and covariance from the GARCH model, allowing an update of the hedge ratio over the hedging period, as in the following bivariate GARCH model[25, 9]

$$\begin{bmatrix}\Delta S_t \\ \Delta F_f\end{bmatrix} = \begin{bmatrix}\mu_1 \\ \mu_2\end{bmatrix} + \begin{bmatrix}\varepsilon_{1,t} \\ \varepsilon_{2,t}\end{bmatrix} \Leftrightarrow \Delta Y_t = \mu + \varepsilon_t,$$

$$\varepsilon_t|\Omega_{t-1} \sim N(0, \Sigma_t), \quad \Sigma_t = \begin{bmatrix}\Sigma_{1,1,t} & \Sigma_{1,2,t} \\ \Sigma_{1,2,t} & \Sigma_{2,2,t}\end{bmatrix},$$

$$vec(\Sigma_t) = C + A \cdot vec(\varepsilon_{t-1}\varepsilon'_{t-1}) + B \cdot vec(\Sigma_{t-1}).$$

Then, the conditional MV hedge ratio at time t is given by

$$\hat{h}_{MV,t} = \frac{\Sigma_{1,2,t-1}}{\Sigma_{2,2,t-1}}. \tag{1.12}$$

There are extensions[128, 101] of this model to allow positions in more than two contracts according to the same logic, also some authors proposed regime-switching GARCH models introducing a state variable $s_t = \{1, 2\}$ in the data generating process, assumed to follow a first-order Markov process, with the state transition probabilities assumed to follow a logistic distribution, affecting the expression of the conditional covariance matrix and the time varying conditional MV hedge ratio between the spot and futures returns.

Cointegration and Error Correction Method

The estimation methods as far discussed do not allow spot and futures returns to be non-stationary, leading to misspecification of the model (1.10). Engle and Granger (1987)[50] proposed the inclusion of an error correction term in the equation. Indeed if an arbitrage condition ties the two returns they can't drift far apart in the long run thus, if both series obey a random walk, one can expect them to be cointegrated, leading to the need of cointegration analysis.
This latter requires the fulfillment of a two steps procedure, first is necessary to test each series for a unit root, through some standard test as those by Dicky and Fuller or Phillips and Perron [41, 119], then to perform a cointegration test as those proposed by Engle and Granger themself or Johansen and Juselius[84]. In the context of hedge ratios estimation, if the spot and futures price are found to be cointegrated, then the h^*_{MV} can be obtained attending to the following procedure, first one has to estimate the following cointegrating regression

$$S_t = a + bF_t + u_t.$$

Then one estimates the error correction model

$$\Delta S_t = \rho u_{t-1} + \beta \Delta F_{t-1} + \sum_{i}^{m} \delta_i \Delta F_{t-i} + \sum_{j=1}^{n} \theta_i \Delta S_{t-j} + \varepsilon_j,$$

with u_t the residuals of the cointegrating regression, the optimal hedge ratio is thus given by β. Some scholars[102] assume that the long-run cointegrating relationship is $(S_t - F_t)$, so they find more appropriate to estimate an error correction model of this form

$$\Delta S_t = \rho(S_{t-1} - F_{t-1}) + \beta \Delta F_t + \sum_{i=1}^{m} \delta_i \Delta F_{t-i} + \sum_{j=1}^{n} \theta_i \Delta S_{t-j} + \varepsilon_j.$$

An alternative model[34] combines the error correction with the biavariate GARCH in the following way

$$\begin{bmatrix} \Delta \ln(S_t) \\ \Delta \ln(F_t) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \alpha_S(\ln(S_{t-1}) - \ln(F_{t-1})) \\ \alpha_S(\ln(S_{t-1}) - \ln(F_{t-1})) \end{bmatrix} + \begin{bmatrix} e_{1,t} \\ e_{2,t} \end{bmatrix},$$

where the error processes follow a GARCH process, again the hedge ratio at time (t) is given by

$$\hat{h}_{t-1} = \frac{H_{1,2,t-1}}{H_{2,2,t-1}}.$$

Mean Extended Gini coefficient estimation

So far I have discussed the statistical methods linked to portfolio choices of agent with quadratic utility over the mean-variance space, or assuming the normality of the return distribution, now I will discuss a family of statistical methods which applications to the portfolio problem are based on the second order stochastic dominance, thus involving the cumulative distribution function estimation in order to rank alternative prospects, and their linear combinations, in a *coherent* way. Contrary to the methods listed above, those to be discussed in this section will generally not provide an analytical solution of the model, and closed form expression for the hedge ratio, instead their solution will require an iterative research of the minimum over a grid of parameters. One of the methods [95] discussed in this section is the estimation of the MEG hedge ratio requiring the minimization, over a grid of portfolio weights, of the portfolio MEG coefficient (1.5), through an estimate of the cumulative distribution function $F(R_P)$, achieved by the empirical distribution method as follows:

$$\hat{F}(R_{P,t}) = \frac{Rank(R_{P,t})}{T},$$

where T is the sample size. Once one has obtained a set of probability distribution functions, the MEG is obtained by substituting the sample covariance to the theoretical one

$$\hat{\Gamma}_v(R_P) = -\frac{v}{T}\sum_{t=1}^{N}(R_{P,t} - \bar{R}_P)((1 - \hat{F}(R_{P,t}))^{v-1} - \bar{\Theta}) \tag{1.13}$$

where $\bar{R}_P$ is the mean return of the portfolio, $\bar{\Theta} = \frac{1}{n}\sum_{t=1}^{T}(1 - \hat{F}(R_{P,t}))^{v-1}$ and v is a parameter representing the risk adversion of the agent.
Shalit(1995)[129] proposal is to find the MEG hedge ratio through the instrumental variable method, allowing for the derivation of an analytical solution to the hedging problem, indeed assuming the equivalence between the cumulative distribution of the terminal wealth and that of the future price he avoided the partial derivative at the denominator of (1.6) achieving this expression[5]

$$\hat{h}^{IV}_{MEG} = \frac{Cov(S_{t+1}, (1 - Q(F_{t+1}))^{v-1})}{Cov(F_{t+1}, (1 - Q(F_{t+1}))^{v-1})} \tag{1.14}$$

Some authors[103] pointed out that this method produces unsatisfactory results in the smoothness of the estimator (if it has any relevance), also it has been showed that this estimation method is asymptotically deficient in comparison to a properly chosen kernel estimator, thus requiring a bigger sample to achieve the convergence. Those motivations lead some researchers to investigate the properties of Nadaraya-Watson estimators for the MEG coefficient, leading to

[5] To avoid confusion, since this expression involves futures prices that I have previously addressed as F_t, I have used $Q(\cdot)$ to represent the cumulative distribution function, that before has been declared as $F(\cdot)$.

the same expression has (1.13), but with the cumulative distribution function now estimated by

$$\hat{F}(R_P) = \frac{1}{N}\sum_{t=1}^{T} K((R_P - R_{P,t})/\imath),$$

where $\imath$ is the bandwidth and $K(x) = \frac{1}{\sqrt{2\pi}} exp(-x^2/2)$ is the Gaussian kernel. This procedure produces more robust results than the empirical distribution method, but it implies the choices of the right kernel and of the appropriate bandwidth, nonetheless the results in the literature shows no improvement in the hedging performance and the insensitivity of the hedge ratio to the bandwidth selection.

C-Var estimation methods

Traditionally, VaR and CVaR are computed as the negative of the 1% quantile of the historical distribution of returns over a prespecified period, relying on the assumption of the normality of the distribution for consistent results. However when the assumption of normality doesn't hold this estimation method becomes less accurate. The "standard" method for the VaR-optimal hedge ratio estimation relies on numerical procedures. One starts with an arbitrary hedge ratio, computes the portfolio returns and the historical distribution approach is used to estimate the VaR of the resulting portfolio. Then a numerical optimization procedure (usually a grid search approach) is used to find the value of the hedge ratio that minimizes the portfolio VaR and that is the minimum-VaR hedge ratio. The minimum-CVaR hedge ratio is calculated with the same numerical procedure.
A drawback of this approach is its reliance on a large historical sample of data of returns for both the assets included in the optimization procedure, unavoidable because by construction it is only possible to measure the empirical frequency of a relatively rare event by using a sample in which there are sufficient occurrences of such events, leading the researchers to focus on methods much less dependent on historical data. One of these is focused on mathematical expansion applied to the approximation of the quantiles of the probability distribution.
If the returns are supposed to be drawn from a location-scale family of distributions and, for the sake of notation, one assumes that its mean is zero, then the $(1-\tau)$ percent VaR of a portfolio can be written as

$$VaR_P(1-\tau) = -\sigma_P q_P(\tau)$$

where $q_P(\tau)$ is the τ percent quantile of the standardized distribution of hedge portfolio returns and σ_P it's his standard deviation. An analytical expression for the minimum-VaR hedge ratio can be derived from the Cornish-Fisher expansion, that approximates $q_P(\tau)$ using the higher moments of the distribution of hedge portfolio returns, thus

$$\tilde{q}_P(\tau; s_P, k_P) = q(\tau)+\frac{1}{6}[q(\tau)^2-1]s_P+\frac{1}{24}[q(\tau)^3-3q(\tau)](k_P-3)-\frac{1}{36}[2q(\tau)^3-5q(\tau)]s_P^2,$$

where $q(\tau)$ is the τ percent quantile of the standard normal distribution and s_P and k_P are respectively the skewness coefficient and the kurtosis coefficient of the hedge portfolio, substituting this corrected quantile in the VaR expression one obtaines the Cornish-Fisher VaR. The optimal h in this case can be obtained differentiating this last objective function with respect to the hedge ratio and setting its first derivative equal to zero, yielding the following first order condition

$$\frac{\partial \sigma_P}{\partial h}(A_1 + A_2 s_P + A_3 k_P + A_4 s_P^2) + \sigma_P \left(A_2 \frac{\partial s_P}{\partial h} + A_3 \frac{\partial k_P}{\partial h} + 2A_4 s_P \frac{\partial s_P}{\partial h} \right) = 0,$$

where $A_1 = q(\tau) - \frac{1}{8}[q(\tau)^3 - 3c(\tau)]$, $A_2 = \frac{1}{6}[q(\tau)^2 - 1]$, $A_3 = \frac{1}{24}[q(\tau)^3 - 3q(\tau)]$, and $A_4 = -\frac{1}{36}[2q(\tau)^2 - 5q(\tau)]$. One then replaces the population moments with their sample estimates and solve for the minimum-VaR hedge ratio, h^{CF}_{VaR}.
The C-VaR of a portfolio can be approximated in the same way as

$$\begin{aligned}\text{C-VaR}_P(1-\tau) = -\sigma_P \bigg(& M_1 + \frac{1}{6}(M_2 - 1)s_P + \frac{1}{24}(M_3 - 3M_1)k_P + \\ & -\frac{1}{36}(2M_3 - 5M_1)s_P^2 \bigg),\end{aligned}$$

with $M_i = \frac{1}{\tau}\int_{-\infty}^{c(\tau)} x^i f(x)dx$ and $f(\cdot)$ is the standard normal probability density function. Differentiating this latter expression with respect to h and then setting the first derivative equal to zero yields the following first order condition

$$\frac{\partial \sigma_P}{\partial h}(B_1 + B_2 s_P + B_3 k_P + B_4 s_P^2) + \sigma_P \left(B_2 \frac{\partial s_P}{\partial h} + B_3 \frac{\partial k_P}{\partial h} + 2B_4 s_P \frac{\delta s_P}{\delta h} \right) = 0,$$

where $B_1 = M_1 - \frac{1}{8}[M_3 - 3M_1]$, $B_2 = \frac{1}{6}[M_2 - 1]$, $B_3 = \frac{1}{24}[M_3 - 3M_1]$, and $B_4 = -\frac{1}{36}[2M_3 - 5M_1]$, and again the solution is found by substituting the population moments with thier sample estimates and solve for the minimum-CVaR hedge ratio, h^{CF}_{CVar} numerically.
One issue in the application of kernel estimation in this context, that doesn't arise in the MEG framework, is the boundary effect, alias the inconsistency of the kernel estimation at finite points at the end of the support. This issue has been sometimes addressed with the technique of weighted double kernel local linear (WDKLL) estimator[22, 74]. Given a symmetric kernel $K(\cdot)$, notice that

$$\mathbb{E}[K_{\imath_0}(y - Y_t)|X_t = x] = f(y|x) + \frac{\imath_0^2}{2}\mu_2(K)f^{2,0}(y|x) + o(\imath_0^2),$$

where $f(y|x)$ is the conditional probability density function of Y_t, $X_t = x$, $K_{\imath_0}(u) = K(u/\imath_0)/\imath_0$, $\mu_2(K) = \int_{-\infty}^{\infty} u^2 K(u)du$ and $f^{2,0} = \frac{\partial^2 f(y|x)}{\partial y^2}$. So if one considers $K_{\imath_0}(y - Y_y)$ as a first estimation of $f(y|x)$, then one could express the left hand side of the former equation as a nonparametric regression of the observed variable versus X_t and apply the local linear fitting scheme, leading to the locally weighted least squares regression problem

$$\sum_{t=1}^{n}[K_{\imath_0}(y-Y_y)-a-b(X_t-x)]^2 W_\imath(x-X_t).$$

Minimizing this expression with respect to a and b leads to the following estimators

$$\hat{f}(y|x)=\sum_{t=1}^{T}W_t(x,\imath)K_{\imath_0}(y-Y_y).$$

The double kernel local linear estimator of the cumulative distribution function is then obtained by integration of this latter formula, in the following way

$$\hat{F}(y|x)=\int_{-\infty}^{\infty}\hat{f}(y|x)dy=\sum_{t=1}^{T}W_t(x,\imath)K_{\imath_0}(y-Y_t).$$

This procedure can be made more reliable with a better choice of both the kernel involved, and with ad hoc procedure for the bandwidth selection.
However the estimator is composed, one substitute it in the model (1.8), that has already been proven to be a convex problem irregardless of the tuning parameters and thus can be solved by search grid optimization[141, 74].
Quantile regression also provide a valid technnique to estimate C-VaR [51].

Pessimistic portfolio estimation and quantile hedge ratio

Empirical strategies for minimizing τ-risk lead immediately to the methods of quantile regression[92, 21]. Let

$$\rho_\tau(u)=u(\tau-I(u<0)) \tag{1.15}$$

denote the piecewise linear (quantile) loss function (also known as "*check*" function), and consider the problem,

$$\min_{\xi\in\mathbb{R}}\mathbb{E}[\rho_\tau(x-\xi)],$$

any minimizer of this problem is a τ-quantile of the random variable x, and minimizing the τ-quantile objective function is equivalent to the evaluation of the sum of expected return and the τ-risk of x, then multiplyied by τ[6]

$$\min_{\xi\in\mathbb{R}}\mathbb{E}[\rho_\tau(x-\xi)]=\tau(\mu+\varrho_{v_\tau}(x)). \tag{1.16}$$

This allowed Bassett et al.(2004) [12] to formulate the problem as

$$\hat{\varrho}_{v_\tau}(x)=(n\tau)^{-1}\min_{\xi\in\mathbb{R}}\sum_{i=1}^{n}\rho_\tau(x_i-\xi)-\hat{\mu}_{n'},$$

[6]Theorem 2 in Bassett et al.(2004)

thus formulating the problem as a quantile regression one and, instead of solving it for a quantity representing τ-th sample quantile, they solve for p coefficients of a linear function estimating the τ-th conditional quantile function.
Finally, a recent trend in the literature[104, 135, 10, 105] applied this methodology to the study of hedging effectiveness, estimating the parameter vector $[\alpha(\tau), \beta(\tau)]$ obtained as the minimizers of the sum of the check functions calculated over a sample of returns, namely

$$[\hat{\alpha}(\tau), \hat{\beta}(\tau)] = \underset{\alpha(\tau),\beta(\tau)}{argmin} \sum_{t=1}^{T} [\rho_\tau (R_{S,t} - \alpha(\tau) - \beta(\tau) R_{F,t}]. \tag{1.17}$$

such that $\hat{\beta}(\tau)$ is the quantile hedge ratio at quantile τ.

1.3 P-Spline Estimation

1.3.1 Linear Spline in Regression Analysis

Before the spread of CAD technologies spline was the name given to thin strips of wood widely used (mostly in naval design) to draw smooth curve through a set of given knots, they were very flexible (thus allowing them to curve enough to pass for each knots) and their curvature may be increased applying weights within each knot. Those tools inspired mathematicians (mostly in the field of numerical analysis) to name *splines* a family of piecewise continuous functions joining multiple polynomials to generate smooth curve through a set of points. Thus a linear spline, mathemathically speaking, can be defined as

$$f(x) = \beta_0 + \beta_1 x + \sum_{k=1}^{K} b_k (x - \omega_k)_+$$

where b_k is the weight of each linear function and $(x - \omega_k)_+$ refers to the $k - th$ function on the knot ω_k and the notation indicates that below that knot the function value is defined to be zero

$$(x - \omega_k)_+ = \begin{cases} x - \omega_k, & \text{if } x - \omega_k > 0, \\ 0, & \text{if } x - \omega_k \leq 0. \end{cases} \tag{1.18}$$

Referring to the regression model framework it means that the basis of the model would be

$$[\mathbf{1} \quad \mathbf{x} \quad (\mathbf{x} - \omega_1)_+ \quad \dots \quad (\mathbf{x} - \omega_K)_+]$$

allowing a wide variety of shapes to be fit. A more comprehensive treatment on the splines can be found in appendix A.

1.3.2 P-Spline in a nutshell

The decision about the optimal number of knots is crucial because it affects the number of parameters to be estimate in the process furthermore, since its optimization can be time intensive and memory consuming, in the literature scholars have developed an alternative method, *penalized spline* (abbreviated P-Spline) where the $\mathbf{b}$ in (1.18) is constrained by a penalty function, in order to optimize the fit and avoiding overfitting the data, thus leading to a modification of the minimizing criterion that now can be formally stated as

$$\begin{aligned} \min_{\beta} \quad & |\mathbf{y} - \mathbf{X}\beta|^2 \\ \text{s.t.} \quad & \beta^{\mathbf{T}}\mathbf{D}\beta \leq C \end{aligned}$$

or through Lagrange multipliers

$$\min_{\beta} |\mathbf{y} - \mathbf{X}\beta|^2 + \lambda^2 \beta^T \mathbf{D}\beta$$

where the D is a symmetric penalty matrix. Thus, given a smoothing parameter λ, the least-square spline estimator of y is given by

$$\hat{y} = \sum_{p=1}^{P} b_{\lambda,p} x_p, \tag{1.19}$$

where $\mathbf{b}_\lambda = (b_{\lambda,1}, \ldots, b_{\lambda,p})$ is the estimator of the vector of parameter β, specifically if one defines $\mathbf{X}_\lambda = \{x_i(\omega_k)\}_{i \in [1,n], k \in [1,K]}$ then $\mathbf{b}_\lambda$ is the solution of the normal equations

$$\mathbf{X}_\lambda^{\mathbf{T}} \mathbf{X}_\lambda \mathbf{b}_\lambda = \mathbf{X}_\lambda^{\mathbf{T}} \mathbf{y},$$

and if $\mathbf{X}_\lambda$ has rank K then

$$\mathbf{b}_\lambda = (\mathbf{X}_\lambda^{\mathbf{T}} \mathbf{X}_\lambda + \lambda^{\mathbf{2}} \mathbf{D})^{-\mathbf{1}} \mathbf{X}_\lambda^{\mathbf{T}} \mathbf{y},$$

so the fit can be expressed in the following way

$$\hat{\mathbf{y}} = \mathbf{X}(\mathbf{X}^{\mathbf{T}} \mathbf{X} + \lambda^{\mathbf{2}} \mathbf{D})^{-\mathbf{1}} \mathbf{X}^{\mathbf{T}} \mathbf{y}.$$

1.3.3 P-Spline Quantile Regression

It is well known that the minimization of $S = \sum_{i=1}^{n} (y_i - g)^2$ brings to the solution of $g = \sum y_i / n$, the arithmetic mean. It is not the case if one moves to the L-1 norm, such that $S = \sum_i^n |y_i - g|$, bringing the median as solution, but only after one has sorted the data. If there are covariates, the solution is found through linear programming technique, and this leads to the case of quantile regression. Koenker et al. (1994)[93, 94] solved the median smoothing problem with unpenalized B-splines straightforward, minimizing $S_1 = \|y - \hat{y}\|^1$, where $\hat{y}$ is defined as in (1.19), this approach can be generalized to any quantile, with the application of the quantile loss function (also know as *check* function) (1.15), Bollaerts et al. (2006)[16] introduced monotonicity restrictions to avoid quantile cross in isotropic and anisotropic P-spline regression quantile.
Penalization in this framework requires care and extra work, because the sum of differences does not easily combine with the sum of absolute residuals. Let

me start with a clear statement of the problem, that is the minimization in the vector of spline coefficients α of the objective function,

$$S_1 = \|y - \hat{y}\|^1 + \lambda \|D^d \mathbf{b}_\lambda\|^2. \tag{1.20}$$

Several approaches have been proposed in the literature for solving this problem with standard linear programming technique, considering that the optimization software takes a response vector and a design matrix as inputs. The first one [47] is to proceed by data augmentation, one defines a design matrix including the penalization term as $(B^+)' = [B'|\lambda D']'$ and extends y as $y^+ = [y'|\mathbf{0}']'$, so the rows of B are extended by λD and y is extended by a vector with $(n-d)$ zeros, with d being the difference order in the penalty term, feeding this augmented problem to a standard linear programming software yields the desired result. One alternative is to drop linear programming and switch to iterative algorithms[126] , so one should combine the sum of absolute values of the residuals with the sum of squares in the penalty, the key viewpoint to understand this approach is to notice that for any scalar u, $\|u\|^1 = u^2/\|u\|^1 = wu^2$, with $w = 1/\|u\|^1$. This identity allows to write a sum of absolute values as a weighted sum of squares, whether u is a vector then the identity extends to $\|u\|^1 = Wu^2$ with $W = diag(w)$ (i.e. a diagonal matrix with $w_i = 1/\|w_i\|^1$), this leads to a chicken-and-egg problem, since one needs u to compute W and viceversa, solved performing standard P-spline fitting, and using its $\tilde{\mathbf{b}}_\lambda$ as starting value for the residuals computation, then the objective function can be written as $(y - \hat{y})'\tilde{W}(y - B\hat{y}) + \lambda\|D\mathbf{b}_\lambda\|^2$ where $\tilde{w}_i = 1/\|\tilde{u}_i\|$ and $\tilde{u}_i = y - \hat{y}$. These estimation strategy relies on the knowledge of the smoothing parameter λ, and since it is not usual to have this information in advance, the estimation strategy is usually loaded with many other computations and loops through coefficient and smoothing parameter optimizations. For these reasons in the empirical analysis I opted for a more general and stronger methodology, based on the intuition that the function (1.20) is convex since it is the sum of convex functions, allowing me to rely on Disciplined Convex Programming[7] as in [6, 7] and streamline the optimization process.

1.3.4 Optimal smoothing parameter selection

The second step of the estimation strategy concerns about the research of the optimal smoothing parameter λ, this is a crucial step since all the qualities ascribed to the P-spline estimators compared to its more famous competitors (such as local regression method, smoothing splines and kernel smoothers) depend on that. The most used criteria are Akaike's Information Criterion, Schwartz Information's Criterion and Cross Validation, however there is a growing branch of literature highlighting the flaws of these methods, because they all require

[7]The reader that is curious about Disciplined Convex Programming may read Stephen Boyd PhD dissertation thesis [18], while the impatient reader may find more insightful his article for the presentation of the R package CVXR [59]

the estimation of the model over a vector of different parameters (or over different samples as in CV) that is not desirable in L_1 optimization, they can be very sensitive to outlying observations and this is not adequate in a quantile regression task, and they can go astray in presence of serial correlation that is a contingency that I can't exclude since, due to hardware limitations, I was not able to regress over lagged returns.
L-Curve (and V-Curve)[57] criterion seemed a promising alternative due to their robustness to serial correlation, but in my application it brought ambiguous results in the location of a suitable λ[8].
Recent advancement in the quantile regression framework from the bayesian perspective [61, 62] allows the application of the Harville-Fellner-Schall (HFS) algorithm to the selection of the smoothing parameter as in [138, 111].
The HFS algorithm, in L_2-norm is based on the intepretation of the P-spline as a mixed model,

$$\begin{aligned}
&y = \underset{n\times p}{\mathbf{X}}\,\beta + \underset{n\times(p-d)}{\mathbf{Z}}\,u + \varepsilon \text{ with } \varepsilon \sim \mathbb{N}(0,\Sigma), u \sim \mathbb{N}(0,\Omega),\\
&cov(\varepsilon) = \Sigma = \sigma_\varepsilon^2 \mathbf{I}_p,\\
&cov(u) = \Omega = \psi^2 \mathbf{I}_{p-d},
\end{aligned}$$

thus, considering $\hat{y} = X\beta + Zu$ with $u = D\mathbf{b}_\lambda$ the smoothing parameters can be expressed as the ratio between the two estimated variances

$$\hat{\lambda} = \frac{\hat{\sigma}_\varepsilon^2}{\hat{\psi}^2}.$$

The algorithm proceeds as follows:

1. Fixes a starting value for the smoothing parameter $\lambda^{(0)}$,
2. Fits the model minimising the objective function
$$\sum_{i=1}^{n} \rho_\tau\left(y_i - \hat{y}\right) + \lambda \sum_{j=1}^{p-d} \|D^d \mathbf{b}_\lambda\|^2,$$
3. Computes the variances $\hat{\sigma_\varepsilon}^2$ and $\hat{\psi}^2$,
4. Puts $\hat{\lambda} = \frac{\hat{\sigma}_\varepsilon^2}{\hat{\psi}^2}$,
5. Sets $\hat{\lambda} \to \lambda^{(0)}$ and iterates steps 2 to 4 until convergence is achieved.

Problems arise because the variance is a concept based on mean measure, so it is not simple to establish a measure of a quantile-based variance, also my quantile regression framework is distribution free, meaning that I have to choose a reliable distribution either for error and random effect to estimate the variance

[8]The resulting L and V curves can be provided at the request of the reader

components. To solve these issues I rely on the strategy shown in [111], thus I estimated the standard errors instead of variances, and also I assumed that the errors distribute as an asymmetric Laplace variable and used the Maximum Likelihood estimator [61, 142]

$$\hat{\sigma}_\varepsilon = n^{-1}\sum_i^n \rho_\tau(y_i - \mu_{\tau,i}) = n^{-1}\sum_i^n \rho_\tau(y_i - \hat{y}),$$

with $\mu_{\tau,i}$ being the τ quantile of the conditional distribution.
The estimation of the random effect standard error goes in the same direction, assuming $u_i \sim ALD(\mu_i, \psi, \tau)$, so the join density of (y_i, u_i) becomes

$$f(y_i, u_i) = \frac{1}{(4\psi)^{n+1}\lambda^n}\exp\left\{-\frac{1}{2\sigma_\varepsilon}\left[\sum_i^n(\|y_i - \mu_i\|^1) + \lambda\|u_i\|^1\right]\right\},$$

that for similarity to the penalized quantile regression in [91] seems related to a penalized model by [111, 138], then the random effect variance ψ estimator becomes

$$\hat{\psi} = \sum_i^n \|\hat{u}_i\|^1.$$

1.3.5 Confidence Interval and hypothesis testing

So far I descrived the procedure used to estimate the parameter

$$\beta(\tau) = \sum_{p=1}^P b_{\lambda,p}. \tag{1.21}$$

To retrieve its confidence interval and provide hypothesis testing I have to estimate the variance of $\hat{\beta}(\tau)$ in order to derive its asymptotic distribution that is given by:

$$\sqrt{n}[\hat{\beta}(\tau) - \beta(\tau)] \to N(0, \omega^2(\tau)\mathbf{V})$$
$$\text{with } \mathbf{V} = \lim_{n\to\infty} n^{-1}\mathbf{X}^T\mathbf{X}$$

where $\omega^2(\tau)$, the scale parameter at the selected quantile is defined as

$$\omega^2(\tau) = \frac{\tau(1-\tau)}{f(F^{-1}(\tau))^2}$$

The density at the selected quantile is unkown and has to be estimated by Siddiqui(1960)[136] estimator [9]:

$$s(\tau) = \frac{1}{f(F^{-1}(\tau))} = \frac{F^{-1}(t+h) - F^{-1}(t-h)}{2h}$$
$$h = n^{-1/3}\left[\frac{4.5\phi^4(\Phi^{-1}(\tau)}{(2\Phi^{-1}(\tau)^2+1)^2}\right]^2.$$

[9] According to the procedure discussed in the book by Davino, Furno and Vistocco [39] at Chapter 5.

Thus, one wanting to compute the confidence intervals and test the hypothesis $H_0 : \beta(\tau) = 0$ can rely on $\sqrt{\hat{\omega}^2(\tau)}$ as standard error for $\hat{\beta}(\tau)$, such that the confidence interval is given by

$$P\left(\hat{\beta}(\tau) - \sqrt{\hat{\omega}^2(\tau)} \times z_{1-\alpha/2} \leq \beta(\tau) \leq \hat{\beta}(\tau) + \sqrt{\hat{\omega}^2(\tau)} \times z_{1-\alpha/2}\right) = 1 - \alpha,$$

and the Student-t test with $n-p$ degrees of freedom to verify the null hypothesis is $t = \hat{\beta}(\tau)/\hat{\omega}(\tau)$.

1.4 Results

I have performed statistical pessimistic hedge ratio estimation on twenty-three major stock indeces and their front month future (i.e. their corresponding future contract with the nearest expiration date), conducting the experiment on a six year time window starting on 02/01/2014 and ending on 02/07/2020 (2-nd January 2014 - 2-nd July 2020), such that most of my series have more than 1600 observations after cleaning operations. Each series has been splitted in two parts, the first (*training set*) is used to "train" the methods and obtain the coefficients, which statistical validity is verified on the second period (*testing set*), also used to verify the performance of the hedged portfolio.
For the sake of exposure I will provide in the body of the chapter only the analysis and the results for the Amsterdam Exchange Index because it's the first in alphabetical order, however analysis and result for the rest of the sample are showed in the Appendix 1.

1.4.1 Preliminary Analysis

To verify ex-ante the opportunity to hedge risk in a pessimistic way, one needs to asses first if and how the correlation between the assets taken in consideration changes over time. Indeed if the correlation is proved to be stable (moreover after random shocks in the assets volatilities) there is no practical reason to avoid standard Mean-Variance (OLS) estimation, however if one has a clue about the sensitivity of the correlation to shocks in the volatilities of the assets included in the portfolio then there should be space for the application of more sophisticated hedging strategies.
To assess the possible usefulness of a pessimistic hedging estimation I analyze the relationship between the assets taken in consideration through the application of the famous DCC-Garch model [17, 48, 49], that consist in a two step procedure to analyse the time conditional correlation defined as

$$\rho_{S,F,t} = \frac{\mathbb{E}_{t-1}[r_{S,t} r_{F,t}]}{\sqrt{\mathbb{E}_{t-1}[r^2_{S,t}]\,\mathbb{E}_{t-1}[r^2_{F,t}]}}.$$

The first step consists in the estimation for each series of the return $r_{i,t}$ and its conditional volatility $\sigma_{i,t}$ using a GARCH model, then denoting the diagonal matrix of the conditional volatilities D_t and the standardized residuals as

$$\nu_t \equiv D_t^{-1}(r_t - \mu),$$

and defining the Bollerlev's Constant Conditional Correlation (CCC) estimator as

$$\bar{R} \equiv \frac{1}{T}\sum_{t=1}^{T} \nu_t \nu_t',$$

then the Dynamic Conditional Correlations are

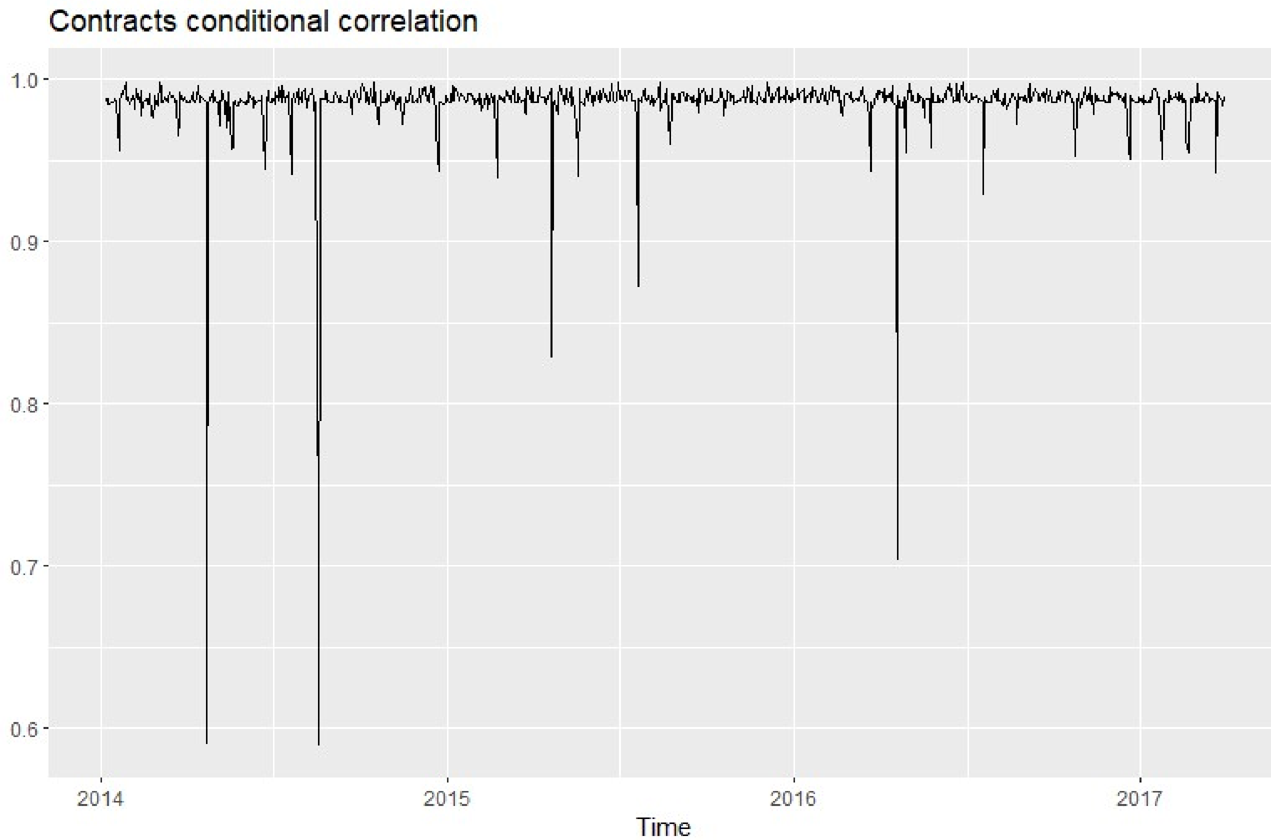
$$Q_t = \bar{R} + \alpha(\nu_{t-1}\nu_{t-1}' - \bar{R}) + \beta(Q_{t-1} - \bar{R}), \tag{1.22}$$

which parameters can be estimated simultaneously through maximum log-likelihood estimation.
In the next page I show the plot of the conditional correlation and the table of the parameters estimated with the DCC-Garch(1,1) model. Before looking at the graph and reading the table I need to make some clarifications, this analysis has been conducted just to investigate the behaviour of the correlation between the assets in the time frame covered in the training set from a qualitative point of view, I don't intend to apply the results in the prediction of the correlations between the considered assets. This is crucial because changes the interpretation that I give of the parameters and their statistical significance.
As the plot shows, the correlation is always positive and nearly always close to the unity, even though one can see there is a slight noise, few outliers are present and this can be the first hint for the occurrence of under-hedging with the traditional approach, indeed a perfect positive correlation implies that one can fully hedge the risk bore by the Index position through the short selling of an equal amount of Future assets, however a lower correlation implies that if a fall occurs in the Index value the subsequent losses won't be matched by the gains in the short Future position, making the position over hedged such that some part of the hedging is useless.
A little bit more informative is the parameter interpretation, for what matters in this context I am not uncomfortable by the small t-statistics for θ^i and ω^i because I have selected DCC due to its generality and wideness, also these parameters should not affects the estimation of the α^i and β^i parameters that describes most of the behaviour I am interested in. $\alpha^{Idx}, \beta^{Idx}, \alpha$ and β^{Fut} in the parameters table suggest that the variance models are not misspecified, while the statistical significance of the α^{Cor} shows the absorbing pattern of the co-movement to shocks in the variance, β^{Cor} value and significance suggest that a CCC model maybe a more appropriate alternative to model the correlation behaviour.

1.4.2 P-spline quantile regression results

Each future's log-return training series has been expanded on a collection of Cubic B-spline spanning $\min\{40, n/4\}$ knots[10] uniformly distributed on a line

[10] According to an empirical rule stated in [123].



Parameters

	Estimate	Std. Error	t value	Pr(>|t|)
μ^{Idx}	0.000755	0.000347	2.173237	0.029762
θ^{Idx}	0.045852	0.035624	1.287119	0.198053
ω^{Idx}	3e-06	4e-06	0.884951	0.376183
α^{Idx}	0.129531	0.028161	4.599677	4e-06
β^{Idx}	0.850358	0.022851	37.212602	0
μ^{Fut}	0.000779	0.000326	2.390025	0.016847
θ^{Fut}	0.041467	0.044282	0.936419	0.349058
ω^{Fut}	4e-06	9e-06	0.422874	0.672387
α^{Fut}	0.134186	0.030443	4.407699	1e-05
β^{Fut}	0.843639	0.064163	13.148311	0
α^{Cor}	0.328601	0.149933	2.191657	0.028404
β^{Cor}	0	0.041587	1e-06	1

Figure 1.1: GARCH-DCC Results for AEX Index and Future correlation.

ranging from $min(\mathbf{x}_{Rf}) - k_\varepsilon$ to $max(\mathbf{x}_{Rf}) + k_\varepsilon$, where k_ε is the distance between each knot, defined as $\frac{max(_Rf)-min(\mathbf{x}_Rf)}{min\{40,n/4\}}$, and $\mathbf{x}_{Rf}$ is the vector containing the training set of the future's return series,[11] obtaining from each vector a matrix $\mathbf{B}_x$. Due to the implemented knots placement the order of the difference operator (that in my analysis is set to five) in (1.20) has no straightforward interpretation and has been chosen to be just greater than the number of non-zero cells per row in matrix $\mathbf{B}_x$.
Following the strategy showed in [104] I estimated quantile hedge ratio using the a narrow set of quantiles[12]. Results, according to the procedure exposed above are statistically significant both on the training than on the testing set over the whole sample, the Total Absolute Quantile Loss (that is a measure of the goodness of fit in this context) doesn't show any substantial variation within the different quantiles and between the training and testing sets, so does the coefficient standard error.

	$\hat{\beta}$	$\hat{\sigma}_\beta$	5%	95%	t statistic	Total Loss
1%	-0.61904	0.00002	-0.61908	-0.61899	-32073.73011	5.23704
2%	-0.58942	0.00002	-0.58947	-0.58937	-28886.05496	5.14451
5%	-0.57685	0.00005	-0.57697	-0.57672	-10854.49206	5.11532
10%	-0.54330	0.00003	-0.54336	-0.54324	-21539.80771	5.02519
20%	-0.55236	0.00002	-0.55241	-0.55231	-26519.58831	5.09347
30%	-0.55492	0.00002	-0.55495	-0.55488	-36002.10569	5.14060
40%	-0.54086	0.00001	-0.54089	-0.54083	-45852.02112	5.13233
50%	-0.55523	0.00001	-0.55525	-0.55520	-48370.20417	5.21908
60%	-0.52411	0.00001	-0.52414	-0.52407	-35955.09319	5.15220
70%	-0.49404	0.00001	-0.49407	-0.49402	-53178.04247	5.08741
80%	-0.49264	0.00002	-0.49268	-0.49260	-28132.01081	5.11981
90%	-0.48555	0.00003	-0.48561	-0.48549	-18161.74439	5.13240
95%	-0.45797	0.00005	-0.45808	-0.45785	-9247.80434	5.05511
98%	-0.42902	0.00002	-0.42907	-0.42897	-19658.60937	4.96536
99%	-0.41046	0.00002	-0.41050	-0.41042	-24584.02704	4.90433

Table 1.1: Shows $\hat{\beta}$, its standard error, its confidence interval, Student-t statistics at 95% against the null $H_0 : \beta(\tau) = 0$ and the Total Loss, over the training set for the model for the Amsterdam Exchange Index against its front month future, $n = 829$.

The main difference that I notice in comparison with Lien et al. [104] results, is that there is no inverted "U-shape" pattern through the different quantiles,

[11] Here, my approach differs from other applications of P-spline smoother in the literature [26, 78, 79], because my splines range through the value assumed by the variable in the training period, instead of ranging through the time domain of the series, so my basis can be thought as the probability that each observation falls in four adjacent bins in an evenly spaced histogram, rather than the coefficients of a moving average smoother.

[12] i.e. 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 98%, 99%

	$\hat{\beta}$	$\hat{\sigma}_\beta$	5%	95%	t statistic	Total Loss
1%	-0.61904	0.00004	-0.61914	-0.61894	-14253.71683	4.58752
2%	-0.58942	0.00004	-0.58951	-0.58933	-15132.23199	4.50578
5%	-0.57685	0.00006	-0.57700	-0.57670	-8972.47060	4.47539
10%	-0.54330	0.00002	-0.54336	-0.54325	-22852.33096	4.38897
20%	-0.55236	0.00001	-0.55239	-0.55233	-38773.12122	4.43110
30%	-0.55492	0.00001	-0.55495	-0.55489	-44000.82141	4.45489
40%	-0.54086	0.00001	-0.54088	-0.54084	-64630.06242	4.43126
50%	-0.55523	0.00001	-0.55524	-0.55521	-84947.45650	4.48888
60%	-0.52411	0.00001	-0.52412	-0.52409	-70116.94669	4.41579
70%	-0.49404	0.00001	-0.49406	-0.49403	-65271.42141	4.34508
80%	-0.49264	0.00001	-0.49268	-0.49261	-36725.89378	4.35691
90%	-0.48555	0.00002	-0.48560	-0.48551	-25821.86456	4.35213
95%	-0.45797	0.00003	-0.45804	-0.45789	-14447.42849	4.27956
98%	-0.42902	0.00001	-0.42906	-0.42899	-29017.13066	4.19967
99%	-0.41046	0.00002	-0.41051	-0.41041	-19392.60950	4.14695

Table 1.2: Shows $\hat{\beta}$, its standard error, its confidence interval, Student-t statistics at 95% against the null $H_0 : \beta(\tau) = 0$ and the Total Loss, over the testing set for the model for the Amsterdam Exchange Index against its front month future, $n = 829$.

neither I observe the almost identical parameters estimated in the stock index subsets. Instead what I observe is an increasing monotonic pattern through the different quantiles, never approaching the OLS $\hat{\beta}_F$ at any level.
To test the economic significance of the result I have derived the portoflio weights by combination the definitions of hedge ratio and portfolio weights in the following way:

$$h = \frac{w_F}{w_S},\ w_S + w_F = 1$$

$$\hat{w}_F = \frac{1}{1+\hat{h}},\ \hat{w}_S = \frac{\hat{h}}{1+\hat{h}}.$$

After obtaining $\hat{h}$ from the statistical procedure, I have computed the portfolio returns over the testing period. What one can immediately see looking at figure Figure 1.1, is that there is no substantial difference between the different weights configurations, due to the equivalence between the Index and Future distribution[13]. This is a pattern that is persistent over the whole dataset and is consistent with the results of Bassett et al.(2004)[12].

[13] Kolmogorov-Smirnov and Mann-Whitney tests confirms this interpretation, the results are not shown due to redundancy and will be provided to request of the reader.

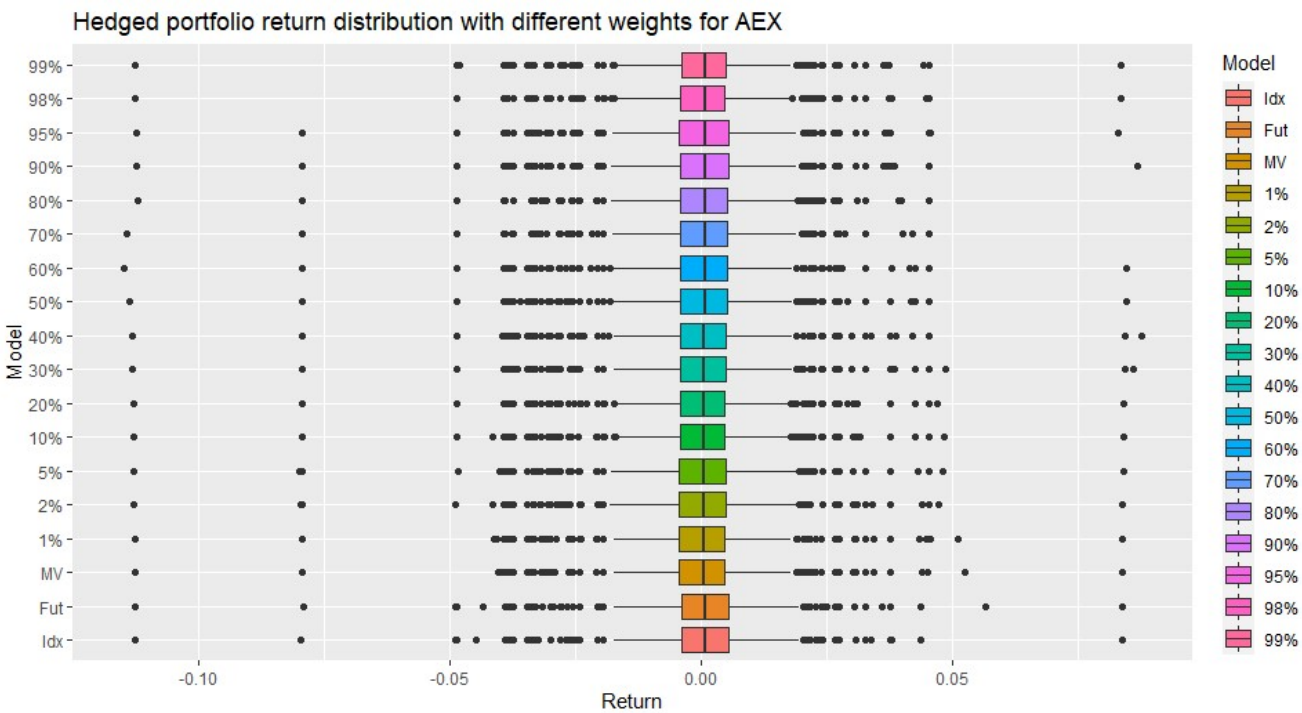


Figure 1.2: Box-plot of the return distribution of Amsterdam Exchange Index, its front month future, OLS Mean-Variance hedged portfolio, and the conditional quantile regression hedged portfolio.

1.4.3 Extreme Analysis results

Figure 2 shows the box-plot of the return distribution of portfolios built with different weight configurations over the testing period, as one can see every configuration shows an high number of outliers, thus it may be useful to confront juts the tails of these distributions. To perform this task I rely on the comparison of several estimator of the *Tail Index*, the α parameter of a Paretian distribution, that means the positive constant for which

$$1 - F(x) = x^{-\alpha} l(x),$$

where $l(x)$ is the slowly varying function at infinity. For convenience the literature on the subject focused on the estimation of the quantity $\gamma = 1/\alpha$, as emerged due to maximum likelihood considerations in Hill(1975)[69], the subsequent "Hill Estimator" is an estimate of the mean excess function of the log-transformed data replacing the expected value by the empirical average of those sample values larger than a given threshold, thus given a sample $\{X_1, ..., X_n\}$ of which $\{X_{1,n}, ..., X_{n,n}\}$ is the ordered sample, the estimator for the right tail is given by

$$\hat{\gamma}_H = \frac{1}{k} \sum_{i=1}^{k} (\log X_{n-i+1,n} - \log X_{n-k,n}) = \frac{1}{k} \sum_{i=1}^{k} log \frac{X_{n-i+1,n}}{X_{n-k,n}}.$$

The Hill estimator can't be negative, this implies the need to perform some dirty tricks to compute it when dealing with distributions with $\gamma < 0$, to avoid the potentially induced bias I have implemented also the Moment Estimator [40], defining the log-moments of the sample as

$$M_n^j = \frac{1}{k} \sum_{i=1}^{k} \left(\log \frac{X_{n-i+1,n}}{X_{n-k,n}} \right)^j,$$

then one can correct the $\hat{\gamma}_H$ to obtain

$$\hat{\gamma}_M = \hat{\gamma}_H + 1 - \frac{1}{2} \left(1 - \frac{(\hat{\gamma}_H)^2}{M_n^2} \right)^{-1}.$$

The last alternative implemented in the analysis for the tail index is the Adjusted Hill estimator[64]:

$$\hat{\gamma}_{Adj} = \hat{\gamma}^H \left(1 - \frac{\hat{\beta}}{1 - \hat{\rho}} \left(\frac{n}{k} \right)^{\hat{\rho}} \right),$$

$$\hat{\beta}(k) = \left(\frac{k}{n} \right)^{\hat{\rho}} \frac{\left(\frac{1}{k} \sum_{i=1}^{k} \left(\frac{i}{k} \right)^{\hat{\rho}} \right) \left(\frac{1}{k} \sum_{i=1}^{k} U_i \right) - \left(\frac{1}{k} \sum_{i=1}^{k} \left(\frac{i}{k} \right)^{-\hat{\rho}} U_i \right)}{\left(\frac{1}{k} \sum_{i=1}^{k} \left(\frac{i}{k} \right)^{-\hat{\rho}} \right) \left(\frac{1}{k} \sum_{i=1}^{k} \left(\frac{i}{k} \right)^{-\hat{\rho}} U_i \right) - \left(\frac{1}{k} \sum_{i=1}^{k} \left(\frac{i}{k} \right)^{-2\hat{\rho}} \right)},$$

$$\hat{\rho}_{\tau}(k) = - \left| \frac{3(T_n^{\tau}(k) - 1)}{T_n^{\tau}(k) - 3} \right|,$$

here U_i are the $X_{i,n}$ log-spacings, a little discussion has to be done about the second order parameter estimator $\hat{\rho}$, the tuning parameter τ depends by the value of ρ, if $\rho \in (-\inf, -1)$ $\tau = 0$, instead if $\rho \in [-1, 0)$ $\tau = 1$, I set $\tau = 0$ by default, if the resulting $\rho < -1$ then I repeat the estimation for $tau = 1$[14]. The strategy adopted to determine the threshold k is based on the minimization of the maximal Kolmogorov-Smirnov distance between different log-spaced tail sequences as explained in [37].
Pareto model is a very common choice in financial risk managment [27] and to assess the goodness of my strategy I confront the tail index estimators so far described to compare the tail behaviour of the return distribution for different pessimistic hedge ratios and the scale parameter of the fitted Student t distribution.
While there is no appreciable difference between the different strategies in the mean/normal domain, one can clearly see that the pessimistic quantile hedging always produces better results at the tails of the distributions, indeed Table 3 shows clearly that the pessimistic portfolios always have thinner left tails than the Mean-Variance one, implying that extreme losses for the pessimistic portfolios are less intense than those of the Mean-Variance. This is the first time that this result appears in the literatue.
Anothe, more interesting, fact is that pessimistic hedged portfolio return distributions' right tails are always fatter than that of the Mean-Variance, meaning that extreme gains of the pessimistic portfolios are higher than those of the Mean-Variance one.

[14] τ affects the estimation through the value $T_n^\tau(k)$

$$T_n^\tau(k) \equiv \begin{cases} \dfrac{\left(M_n^{(1)}(k)\right)^\tau - \left(\frac{1}{2}M_n^{(2)}(k)\right)^{\frac{\tau}{2}}}{\left(\frac{1}{2}M_n^{(2)}(k)\right)^{\frac{\tau}{2}} - \left(\frac{1}{6}M_n^{(3)}(k)\right)^{\frac{\tau}{3}}}, & \text{if } \tau > 0, \\ \dfrac{\log\left(M_n^{(1)}(k)\right) - \frac{1}{2}\log\left(\frac{1}{2}M_n^{(2)}(k)\right)}{\frac{1}{2}\log\left(\frac{1}{2}M_n^{(2)}(k)\right) - \frac{1}{3}\log\left(\frac{1}{6}M_n^{(3)}(k)\right)}, & \text{if } \tau = 0. \end{cases}$$

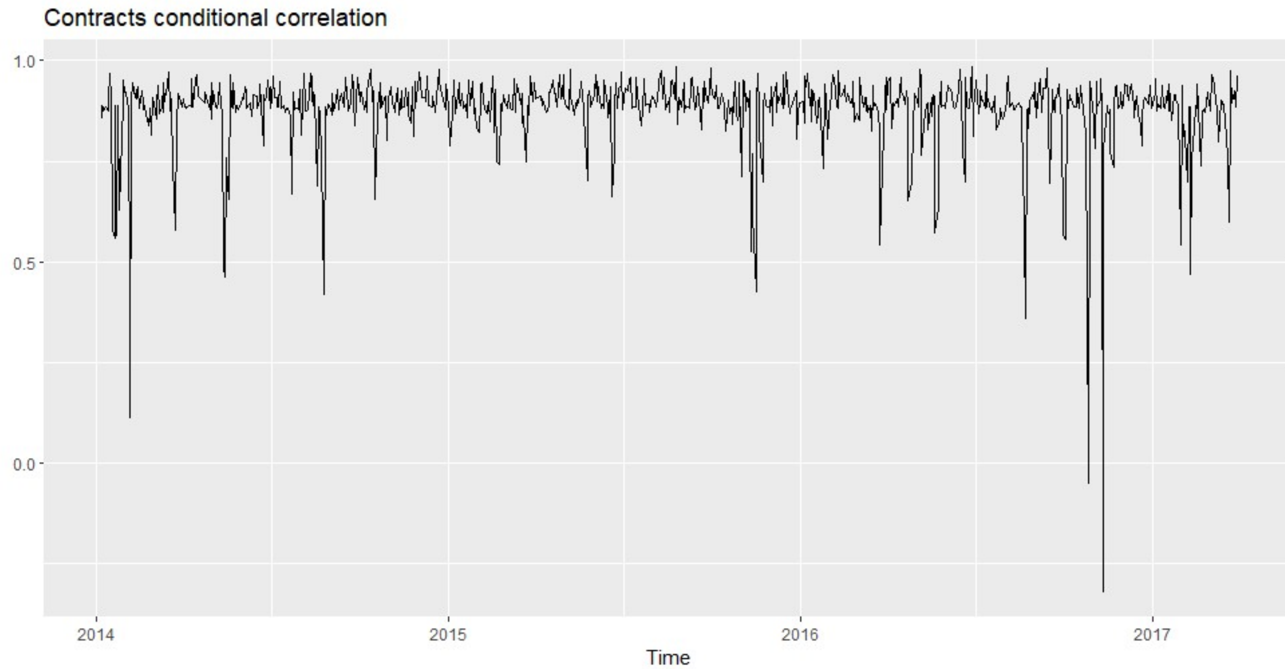
	Left Tail					Right Tail					Student-t Scale	
	k	γ_H	γ_M	γ_{Adj}	ρ_H	k	γ_H	γ_M	γ_{Adj}	ρ_H	σ	σ_{Err}
MV	81	0.51750	0.46170	0.51750	-0.58333	83	0.43049	0.30390	0.43048	-0.86589	0.00585	0.00026
1%	81	0.51525	0.43968	0.51513	-0.51580	83	0.46456	0.34337	0.46244	-0.11517	0.00605	0.00028
2%	81	0.51006	0.44699	0.50999	-0.59267	83	0.45180	0.35303	0.44943	-0.04561	0.00598	0.00027
5%	81	0.50799	0.45014	0.50795	-0.62537	83	0.44969	0.35334	0.44769	-0.06859	0.00599	0.00027
10%	81	0.51792	0.44045	0.51785	-0.50775	83	0.44244	0.35562	0.44135	-0.14370	0.00593	0.00027
20%	81	0.50775	0.45204	0.50773	-0.63980	83	0.44112	0.36005	0.44043	-0.19623	0.00599	0.00027
30%	81	0.50483	0.45531	0.50482	-0.67834	83	0.44206	0.35921	0.44127	-0.18060	0.00588	0.00027
40%	81	0.52061	0.43733	0.52051	-0.47348	83	0.43718	0.36324	0.43681	-0.25793	0.00590	0.00027
50%	81	0.50461	0.45555	0.50460	-0.68122	83	0.44217	0.35911	0.44137	-0.17875	0.00589	0.00027
60%	81	0.52353	0.43447	0.52337	-0.43945	83	0.43305	0.36308	0.43275	-0.27565	0.00596	0.00027
70%	81	0.52773	0.42886	0.52739	-0.37757	83	0.45807	0.30950	0.45777	-0.43762	0.00600	0.00027
80%	81	0.52760	0.42900	0.52727	-0.37878	83	0.45777	0.30961	0.45746	-0.43615	0.00600	0.00027
90%	81	0.52835	0.42786	0.52797	-0.36624	83	0.45626	0.31034	0.45594	-0.42625	0.00600	0.00027
95%	81	0.52495	0.43111	0.52462	-0.39727	83	0.46435	0.28894	0.46428	-0.65476	0.00598	0.00027
98%	81	0.52293	0.43294	0.52265	-0.41293	83	0.47358	0.26471	0.47357	-0.88564	0.00597	0.00027
99%	81	0.52170	0.43421	0.52146	-0.42377	83	0.47105	0.26657	0.47104	-0.86742	0.00596	0.00027

Table 1.3: Shows different Tail Index Values and the Student-t scale parameter over the testing set for the model of the Amsterdam Exchange Index against its front month future, $n = 829$.

1.5 Conclusion

In this chapter I focused on the application of P-Spline method to the estimation of a quantile regression model for the solution of a Pessimistic static hedging portfolio allocation.
The results shown refute the evidences of Lien et al.(2014)[104] that the two approaches provide the same hedge ratios estimation for stock indeces. I don't have a strong interpretation of the reasons behind this difference, but I have two suspects. First, the standard linear quantile regression performed in Lien work may lack the ability to capture non-linearities in the relationship between the two variables, while the P-Spline estimator doesn't need a precise specification of the functional relation to estimate (indeed it only needs a large enough basis and penalty order). At the same time, the procedure that I used to build the confidence interval which I based my hypothesis testing upon, strongly depends on the error density estimation at the selected quantile, which in turn depends on the sample size, so I can't exclude that a large enough sample size may shrink the estimated coefficient's standard errors to zero, thus leading to too narrows confidence interval. Further investigations are needed to clarify this question and will be the subject of future researches.
From the economic side of the problem, my results support the evidence present in the literature that in mean terms there is no economically significant difference between the Pessimistic and the Mean-Variance hedging. Looking at my result one may even argue that stock index future hedging is useless at all, while this could be a tempting statement to make it is not true for several reason. First of all, my analysis has been conducted only on the daily hedging horizon, that is not a very realistic one, conducting the same analysis on several different horizons may give more insight about the time relationship between the spot and future index return quantile. Second, while comparing the distribution obtained from the different configurations, I haven't considered the time domain, that in any hedging exercise is crucial to determine rebalancing gains and losses, this exercise would have needed a richer dataset, requiring the registration and modeling of transaction costs, it should also required the definition of an optimal rebalancing timing, that is still a controversial topic in the field of quantitative finance, and which solution would have been out of the scope of this paper. Third, preliminary results shows that while static hedging maybe obsolete on stock index due to market efficiency, dynamic hedging is still a viable risk mitigation technique. All these aspects requires further investigations and will be the subject of future researches.
However, the situation changes drastically in the extreme domain, indeed extreme analysis results are fresh and encouraging, showing for the first time that Pessimistic quantile hedge ratio is able to achieve results in flattening the tail of the loss distribution while, at the same time, making the gains distribution tail heavier.

1.6 Appendix I: Graphs and Tables of Chapter 1



Parameters

	Estimate	Std. Error	t value	Pr(>|t|)
μ^{Idx}	0.000202	0.000275	0.732984	0.463568
θ^{Idx}	0.020121	0.03566	0.564257	0.572579
ω^{Idx}	2e-06	4e-06	0.470398	0.638071
α^{Idx}	0.076113	0.06705	1.135176	0.256302
β^{Idx}	0.897909	0.081179	11.060915	0
μ^{Fut}	0.000171	0.000281	0.60892	0.542577
θ^{Fut}	0.003114	0.041624	0.074815	0.940362
ω^{Fut}	1e-06	2e-06	0.524512	0.599922
α^{Fut}	0.047896	0.015404	3.109289	0.001875
β^{Fut}	0.942422	0.018201	51.779926	0
α^{Cor}	0.362771	0.07526	4.820232	1e-06
β^{Cor}	0	0.148867	0	1

Figure 1.3: GARCH-DCC Results for ASX Index and Future correlation.

	$\hat{\beta}$	$\hat{\sigma}_\beta$	5%	95%	t statistic	Total Loss
1%	-0.34476	0.00001	-0.34479	-0.34473	-28945.43273	3.59835
2%	-0.29364	0.00001	-0.29367	-0.29361	-21314.71032	3.46400
5%	-0.23870	0.00003	-0.23876	-0.23863	-8532.20210	3.32250
10%	-0.20308	0.00002	-0.20313	-0.20304	-10736.33226	3.23436
20%	-0.16838	0.00001	-0.16841	-0.16836	-17096.71691	3.15362
30%	-0.16255	0.00001	-0.16257	-0.16252	-17057.83992	3.14866
40%	-0.12372	0.00001	-0.12373	-0.12370	-17861.81007	3.05598
50%	-0.11911	0.00001	-0.11913	-0.11909	-15678.41122	3.05381
60%	-0.12296	0.00001	-0.12298	-0.12295	-17560.23750	3.07417
70%	-0.10558	0.00001	-0.10560	-0.10555	-10798.80441	3.03762
80%	-0.07142	0.00001	-0.07144	-0.07140	-8942.22785	2.95599
90%	0.07417	0.00001	0.07414	0.07419	7117.86526	2.58280
95%	0.08200	0.00002	0.08195	0.08204	4282.17402	2.56669
98%	0.14058	0.00001	0.14056	0.14060	18135.87813	2.41821
99%	0.15325	0.00001	0.15323	0.15326	25372.49045	2.38664

	$\hat{\beta}$	$\hat{\sigma}_\beta$	5%	95%	t statistic	Total Loss
1%	-0.34476	0.00001	-0.34479	-0.34473	-28945.43273	3.59835
2%	-0.29364	0.00001	-0.29367	-0.29361	-21314.71032	3.46400
5%	-0.23870	0.00003	-0.23876	-0.23863	-8532.20210	3.32250
10%	-0.20308	0.00002	-0.20313	-0.20304	-10736.33226	3.23436
20%	-0.16838	0.00001	-0.16841	-0.16836	-17096.71691	3.15362
30%	-0.16255	0.00001	-0.16257	-0.16252	-17057.83992	3.14866
40%	-0.12372	0.00001	-0.12373	-0.12370	-17861.81007	3.05598
50%	-0.11911	0.00001	-0.11913	-0.11909	-15678.41122	3.05381
60%	-0.12296	0.00001	-0.12298	-0.12295	-17560.23750	3.07417
70%	-0.10558	0.00001	-0.10560	-0.10555	-10798.80441	3.03762
80%	-0.07142	0.00001	-0.07144	-0.07140	-8942.22785	2.95599
90%	0.07417	0.00001	0.07414	0.07419	7117.86526	2.58280
95%	0.08200	0.00002	0.08195	0.08204	4282.17402	2.56669
98%	0.14058	0.00001	0.14056	0.14060	18135.87813	2.41821
99%	0.15325	0.00001	0.15323	0.15326	25372.49045	2.38664

Table 1.4: Shows $\hat{\beta}$ inference results over the testing and training set for the model for the Australian Securities Exchange against its front month future, $n = 825$.

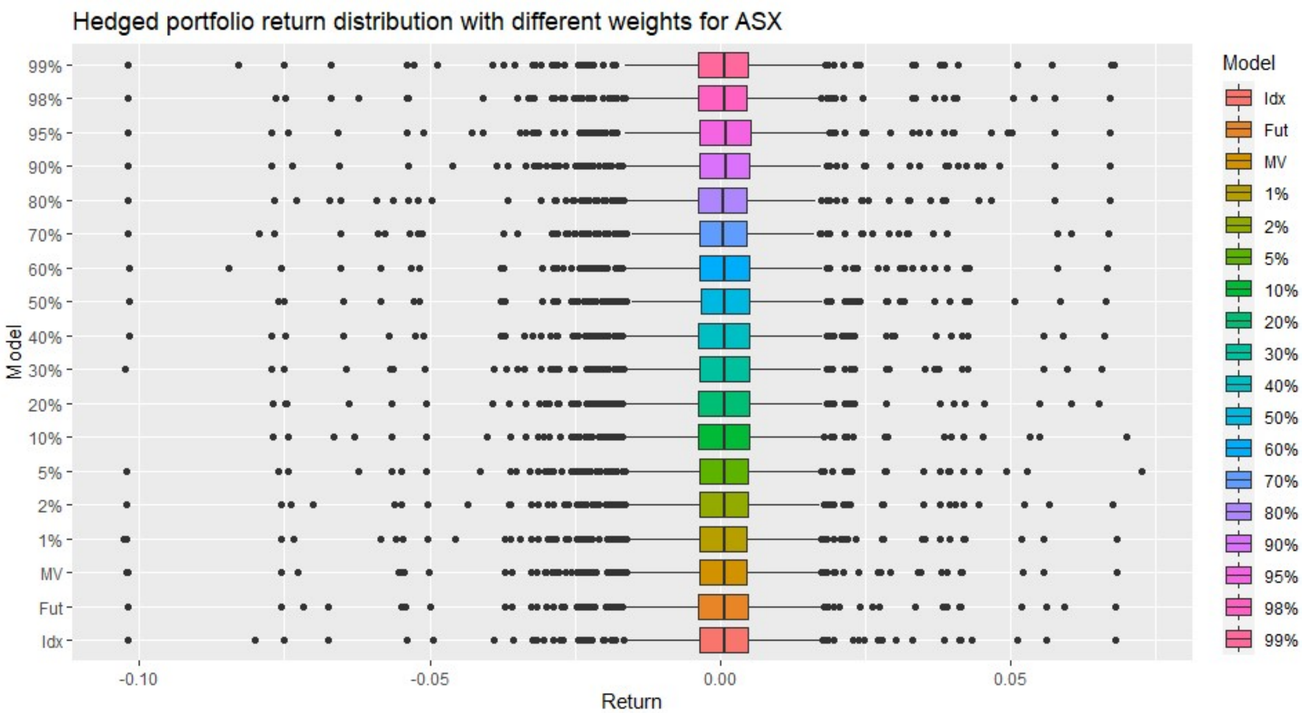
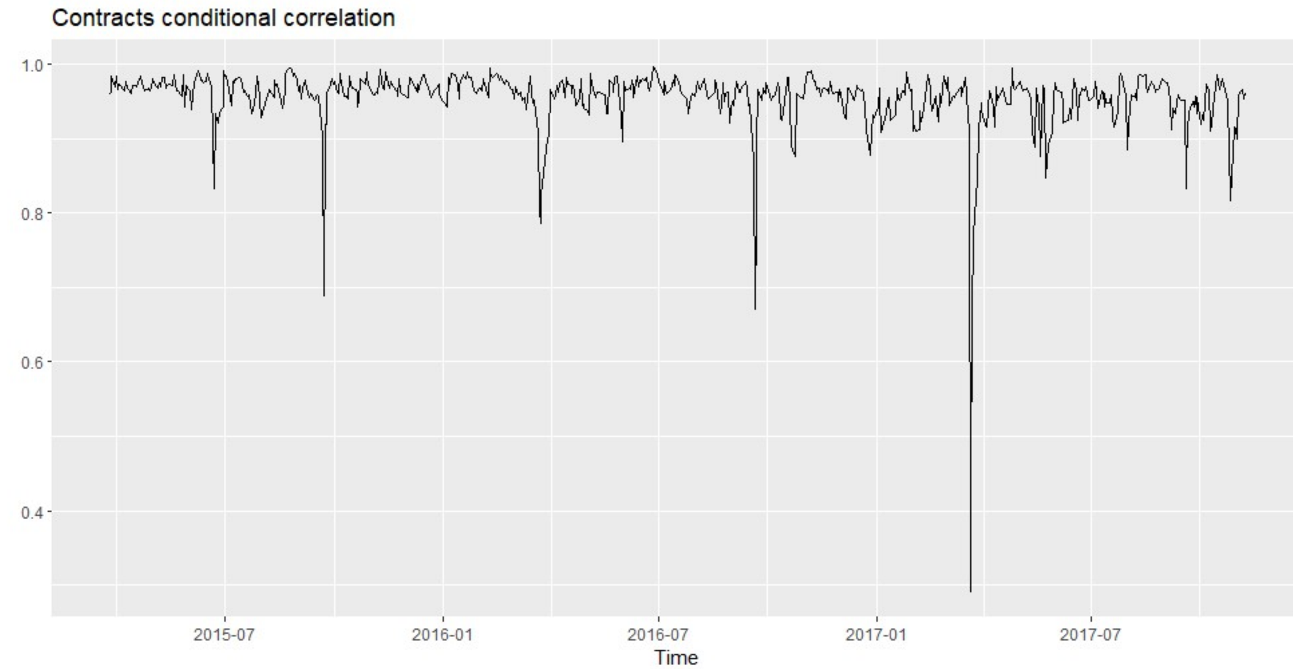


Figure 1.4: Box-plot of the return distribution of Australian Securities Exchange its front month future, OLS Mean-Variance hedged portfolio, and the conditional quantile regression hedged portfolio.

	Left Tail					Right Tail					Student-t Scale	
	k	γ_H	γ_M	γ_{Adj}	ρ_H	k	γ_h	γ_M	γ_{Adj}	ρ_H	σ	σ_{Err}
MV	80	0.68384	0.41444	0.68276	-0.39587	82	0.52007	0.46016	0.52002	-0.45090	0.00523	0.00022
1%	80	0.62310	0.46819	0.61961	-0.13906	82	0.49151	0.40483	0.49002	-0.18681	0.00536	0.00023
2%	80	0.61378	0.48357	0.61178	-0.26313	82	0.48682	0.42870	0.48674	-0.44779	0.00537	0.00023
5%	80	0.63788	0.46628	0.63165	-0.08447	82	0.48123	0.45618	0.48124	-0.77062	0.00524	0.00022
10%	80	0.64919	0.45738	0.63932	-0.00333	82	0.48875	0.45705	0.48875	-0.71974	0.00593	0.00031
20%	80	0.63292	0.47795	0.62988	-0.16733	82	0.49354	0.45662	0.49355	-0.67756	0.00540	0.00023
30%	80	0.63067	0.48085	0.62829	-0.19162	82	0.49402	0.45690	0.49403	-0.67636	0.00517	0.00022
40%	80	0.63500	0.47462	0.63104	-0.13429	82	0.49711	0.45874	0.49711	-0.66994	0.00499	0.00021
50%	80	0.63646	0.47261	0.63205	-0.11709	82	0.49754	0.45871	0.49754	-0.66633	0.00502	0.00021
60%	80	0.63524	0.47429	0.63120	-0.13143	82	0.49718	0.45873	0.49719	-0.66934	0.00499	0.00021
70%	80	0.63942	0.46835	0.63388	-0.08123	82	0.49876	0.45869	0.49877	-0.65670	0.00504	0.00021
80%	80	0.63920	0.46696	0.63282	-0.07136	82	0.50441	0.45519	0.50440	-0.58285	0.00524	0.00022
90%	80	0.65966	0.44272	0.65413	-0.12090	82	0.53556	0.42416	0.53104	-0.09858	0.00516	0.00022
95%	80	0.65500	0.44801	0.64815	-0.08190	82	0.53553	0.42452	0.53127	-0.10052	0.00516	0.00022
98%	80	0.64853	0.45427	0.63873	-0.03695	82	0.53538	0.42684	0.53220	-0.11171	0.00517	0.00022
99%	80	0.64971	0.45283	0.64050	-0.04935	82	0.53538	0.42722	0.53218	-0.11279	0.00521	0.00022

Table 1.5: tab:Table 3Shows different Tail Index Values and the Student-t scale parameter over the testing set for the model of the Australian Security Exchange against its front month future



Parameters

	Estimate	Std. Error	t value	Pr(>|t|)
μ^{Idx}	0.000849	0.00051	1.662685	0.096375
θ^{Idx}	0.014107	0.045556	0.309673	0.756809
ω^{Idx}	4e-06	5.5e-05	0.081838	0.934775
α^{Idx}	0.112318	0.101151	1.110407	0.266824
β^{Idx}	0.858948	0.244878	3.50766	0.000452
μ^{Fut}	0.000777	5e-04	1.554019	0.12018
θ^{Fut}	0.037815	0.044828	0.843558	0.398916
ω^{Fut}	4e-06	7e-06	0.531494	0.595077
α^{Fut}	0.099585	0.03068	3.245973	0.001171
β^{Fut}	0.877989	0.042034	20.887462	0
α^{Cor}	0.321778	0.110124	2.921959	0.003478
β^{Cor}	0.414446	0.349613	1.185443	0.235842

Figure 1.5: GARCH-DCC Results for ATX Index and Future correlation.

	$\hat{\beta}$	$\hat{\sigma}_\beta$	5%	95%	t statistic	Total Loss
1%	-0.55713	0.00003	-0.55718	-0.55707	-22267.12207	4.23509
2%	-0.54390	0.00002	-0.54394	-0.54385	-28322.01661	4.20394
5%	-0.51654	0.00004	-0.51664	-0.51645	-12439.60449	4.14356
10%	-0.48980	0.00004	-0.48989	-0.48971	-12513.99684	4.09335
20%	-0.44435	0.00002	-0.44439	-0.44431	-23654.93884	4.01270
30%	-0.39018	0.00001	-0.39021	-0.39015	-27765.23556	3.90478
40%	-0.24325	0.00002	-0.24329	-0.24321	-15414.26905	3.53112
50%	-0.15822	0.00001	-0.15825	-0.15820	-12814.24765	3.32563
60%	-0.22767	0.00001	-0.22769	-0.22765	-24563.13447	3.56098
70%	-0.27763	0.00001	-0.27766	-0.27760	-21046.39522	3.74394
80%	-0.32891	0.00001	-0.32894	-0.32888	-24063.56890	3.93393
90%	-0.36125	0.00002	-0.36130	-0.36119	-15472.78957	4.07061
95%	-0.35387	0.00004	-0.35396	-0.35378	-9023.64736	4.06903
98%	-0.28714	0.00002	-0.28718	-0.28711	-18506.99951	3.88043
99%	-0.28134	0.00002	-0.28138	-0.28130	-15830.29838	3.86682
	$\hat{\beta}$	$\hat{\sigma}_\beta$	5%	95%	t statistic	Total Loss
1%	-0.55713	0.00004	-0.55722	-0.55703	-14064.32089	5.24959
2%	-0.54390	0.00004	-0.54398	-0.54381	-15369.94622	5.19688
5%	-0.51654	0.00008	-0.51673	-0.51636	-6469.15870	5.08255
10%	-0.48980	0.00004	-0.48989	-0.48970	-12158.41092	4.95907
20%	-0.44435	0.00002	-0.44439	-0.44430	-22696.77550	4.74319
30%	-0.39018	0.00002	-0.39022	-0.39014	-23523.52264	4.50148
40%	-0.24325	0.00001	-0.24328	-0.24322	-16553.34524	3.95451
50%	-0.15822	0.00001	-0.15825	-0.15820	-15145.52929	3.62787
60%	-0.22767	0.00001	-0.22770	-0.22764	-17663.35413	3.80460
70%	-0.27763	0.00001	-0.27766	-0.27760	-20593.30632	3.91557
80%	-0.32891	0.00001	-0.32894	-0.32888	-26842.25704	4.02759
90%	-0.36125	0.00003	-0.36131	-0.36119	-14338.51014	4.07599
95%	-0.35387	0.00007	-0.35402	-0.35372	-5395.28518	4.02567
98%	-0.28714	0.00003	-0.28722	-0.28707	-9371.53699	3.80183
99%	-0.28134	0.00003	-0.28141	-0.28127	-9046.82122	3.77865

Table 1.6: Shows $\hat{\beta}$ inference results over the testing and training set for the model for the Austrian Traded Index against its front month future, $n = 644$.

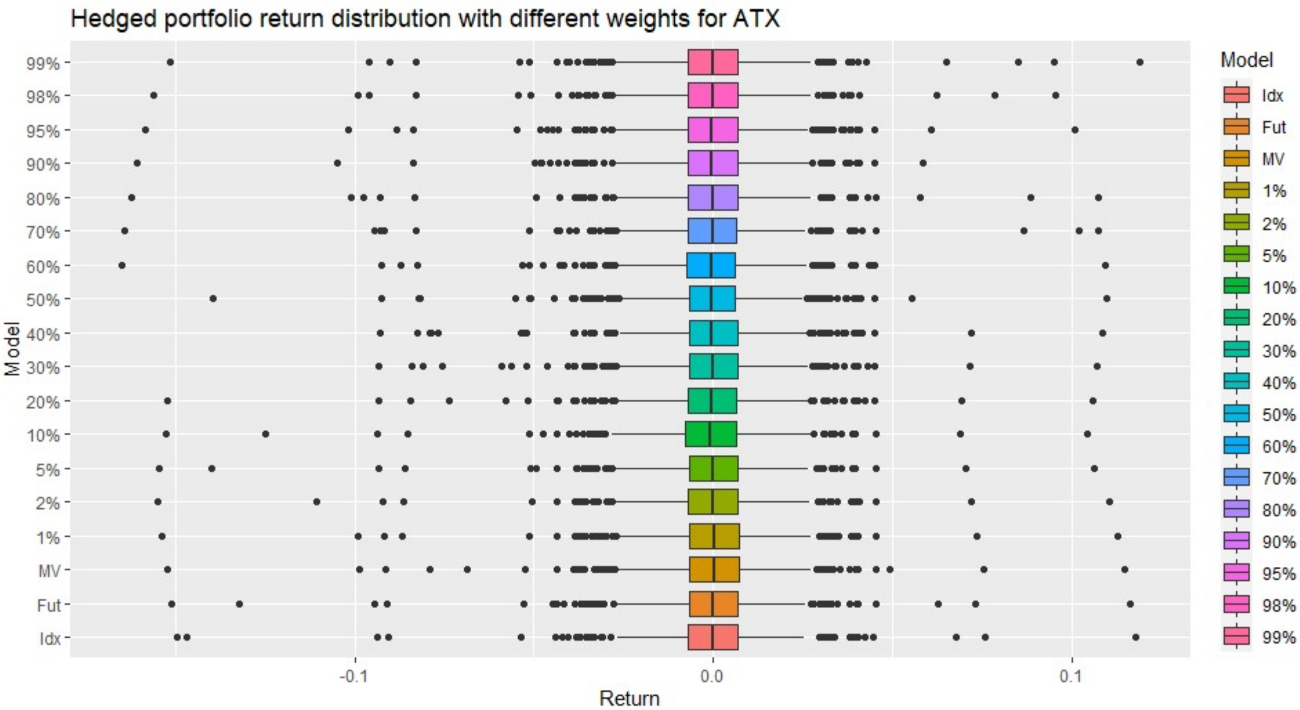
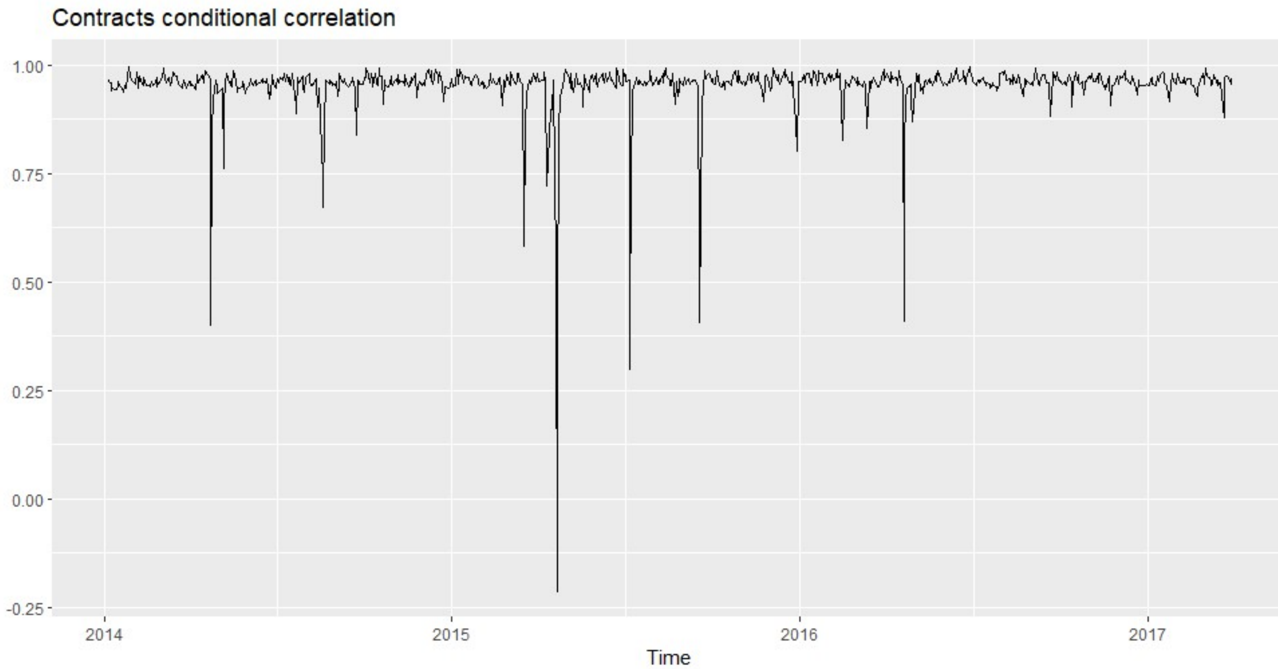


Figure 1.6: Box-plot of the return distribution of Austrian Traded Index its front month future, OLS Mean-Variance hedged portfolio, and the conditional quantile regression hedged portfolio.

	Left Tail					Right Tail					Student-t Scale	
	k	γ_H	γ_M	γ_{Adj}	ρ_H	k	γ_h	γ_M	γ_{Adj}	ρ_H	σ	σ_{Err}
MV	62	0.59583	0.42841	0.59520	-0.43046	64	0.58880	0.25174	0.58851	-0.59947	0.00846	0.00040
1%	62	0.62468	0.28771	0.62465	-0.93837	64	0.56408	0.30537	0.56390	-0.63712	0.00852	0.00042
2%	62	0.62399	0.29356	0.62396	-0.91140	64	0.56596	0.29380	0.56576	-0.62760	0.00851	0.00042
5%	62	0.63442	0.28456	0.63440	-0.96925	64	0.57431	0.25904	0.57418	-0.69928	0.00840	0.00042
10%	62	0.58803	0.36244	0.58755	-0.54908	64	0.57615	0.23765	0.57604	-0.72582	0.00836	0.00041
20%	62	0.61336	0.33952	0.61327	-0.74814	64	0.57021	0.22789	0.57010	-0.69999	0.00832	0.00041
30%	62	0.61576	0.34848	0.61566	-0.74810	64	0.52206	0.29520	0.52103	-0.34831	0.00829	0.00040
40%	62	0.63146	0.33831	0.63146	-1.44673	64	0.51681	0.33035	0.50936	-0.08874	0.00829	0.00040
50%	62	0.59640	0.39655	0.59625	-0.69023	64	0.52196	0.34199	0.50785	-0.03675	0.00828	0.00040
60%	62	0.63533	0.33463	0.63533	-1.50398	64	0.51465	0.33756	0.50246	-0.04434	0.00832	0.00040
70%	62	0.62223	0.34847	0.62221	-0.93158	64	0.51939	0.31555	0.51590	-0.18651	0.00833	0.00040
80%	62	0.62737	0.33864	0.62734	-0.89707	64	0.50591	0.32090	0.50177	-0.15354	0.00833	0.00040
90%	62	0.61160	0.35799	0.61148	-0.72580	64	0.50284	0.32403	0.49819	-0.13253	0.00833	0.00041
95%	62	0.60802	0.36396	0.60786	-0.69344	64	0.50347	0.32335	0.49925	-0.13567	0.00833	0.00041
98%	62	0.61949	0.35181	0.61946	-0.89331	64	0.50838	0.33028	0.50067	-0.07817	0.00833	0.00040
99%	62	0.62118	0.34974	0.62115	-0.91706	64	0.51886	0.31493	0.51547	-0.19137	0.00833	0.00040

Table 1.7: Shows different Tail Index estimators value and the Student-t scale parameter over the testing set for the model of the Austrian Traded Index against its front month future



Parameters

	Estimate	Std. Error	t value	Pr(>|t|)
μ^{Idx}	0.000653	0.000365	1.790435	0.073384
θ^{Idx}	0.035036	0.035354	0.991003	0.321684
ω^{Idx}	6e-06	2e-06	2.636009	0.008389
α^{Idx}	0.149581	0.024552	6.09251	0
β^{Idx}	0.79472	0.042466	18.714296	0
μ^{Fut}	0.000682	0.000325	2.099168	0.035802
θ^{Fut}	-0.002351	0.038838	-0.060542	0.951724
ω^{Fut}	6e-06	3e-06	1.686709	0.091659
α^{Fut}	0.146619	0.034181	4.289509	1.8e-05
β^{Fut}	0.809346	0.054532	14.841681	0
α^{Cor}	0.347315	0.115515	3.006669	0.002641
β^{Cor}	0.177089	0.437171	0.405079	0.685419

Figure 1.7: GARCH-DCC Results for BEL Index and Future correlation.

	$\hat{\beta}$	$\hat{\sigma}_\beta$	5%	95%	t statistic	Total Loss
1%	-0.60592	0.00002	-0.60596	-0.60587	-30845.09949	4.70062
2%	-0.59042	0.00002	-0.59047	-0.59038	-32277.33657	4.65953
5%	-0.65658	0.00004	-0.65669	-0.65648	-14720.90463	4.86734
10%	-0.64246	0.00002	-0.64252	-0.64240	-26020.05907	4.84825
20%	-0.62173	0.00002	-0.62177	-0.62169	-36580.30651	4.83132
30%	-0.61286	0.00001	-0.61289	-0.61283	-42999.30935	4.84899
40%	-0.59628	0.00001	-0.59630	-0.59625	-55586.60350	4.84273
50%	-0.57352	0.00001	-0.57354	-0.57350	-54221.56464	4.81662
60%	-0.59373	0.00001	-0.59374	-0.59371	-77004.89069	4.92221
70%	-0.58719	0.00002	-0.58722	-0.58715	-37426.69182	4.94542
80%	-0.58466	0.00001	-0.58469	-0.58463	-41302.30925	4.98090
90%	-0.46152	0.00002	-0.46157	-0.46147	-21627.17222	4.63348
95%	-0.31353	0.00003	-0.31360	-0.31346	-10562.87930	4.18553
98%	-0.23321	0.00001	-0.23324	-0.23318	-18596.45551	3.94297
99%	-0.18816	0.00001	-0.18819	-0.18813	-14568.84363	3.80441
	$\hat{\beta}$	$\hat{\sigma}_\beta$	5%	95%	t statistic	Total Loss
1%	-0.60592	0.00006	-0.60605	-0.60579	-10739.01166	5.54857
2%	-0.59042	0.00004	-0.59051	-0.59034	-15717.81702	5.48824
5%	-0.65658	0.00006	-0.65673	-0.65643	-10332.48453	5.73317
10%	-0.64246	0.00003	-0.64252	-0.64240	-24640.47383	5.67100
20%	-0.62173	0.00002	-0.62177	-0.62169	-37673.63802	5.57537
30%	-0.61286	0.00001	-0.61289	-0.61283	-53578.58502	5.52470
40%	-0.59628	0.00001	-0.59631	-0.59625	-42273.12535	5.44536
50%	-0.57352	0.00001	-0.57354	-0.57350	-74017.75178	5.34332
60%	-0.59373	0.00001	-0.59375	-0.59370	-55053.39572	5.40163
70%	-0.58719	0.00001	-0.58721	-0.58716	-59991.99958	5.36028
80%	-0.58466	0.00001	-0.58469	-0.58463	-47614.05074	5.33392
90%	-0.46152	0.00002	-0.46157	-0.46148	-23101.08796	4.86275
95%	-0.31353	0.00004	-0.31362	-0.31344	-7898.64522	4.31034
98%	-0.23321	0.00002	-0.23326	-0.23316	-11181.35926	4.01218
99%	-0.18816	0.00002	-0.18820	-0.18811	-9139.97340	3.84639

Table 1.8: Shows $\hat{\beta}$ inference results over the testing and training set for the model for the Brussels Stock Exchange against its front month future, $n = 829$.

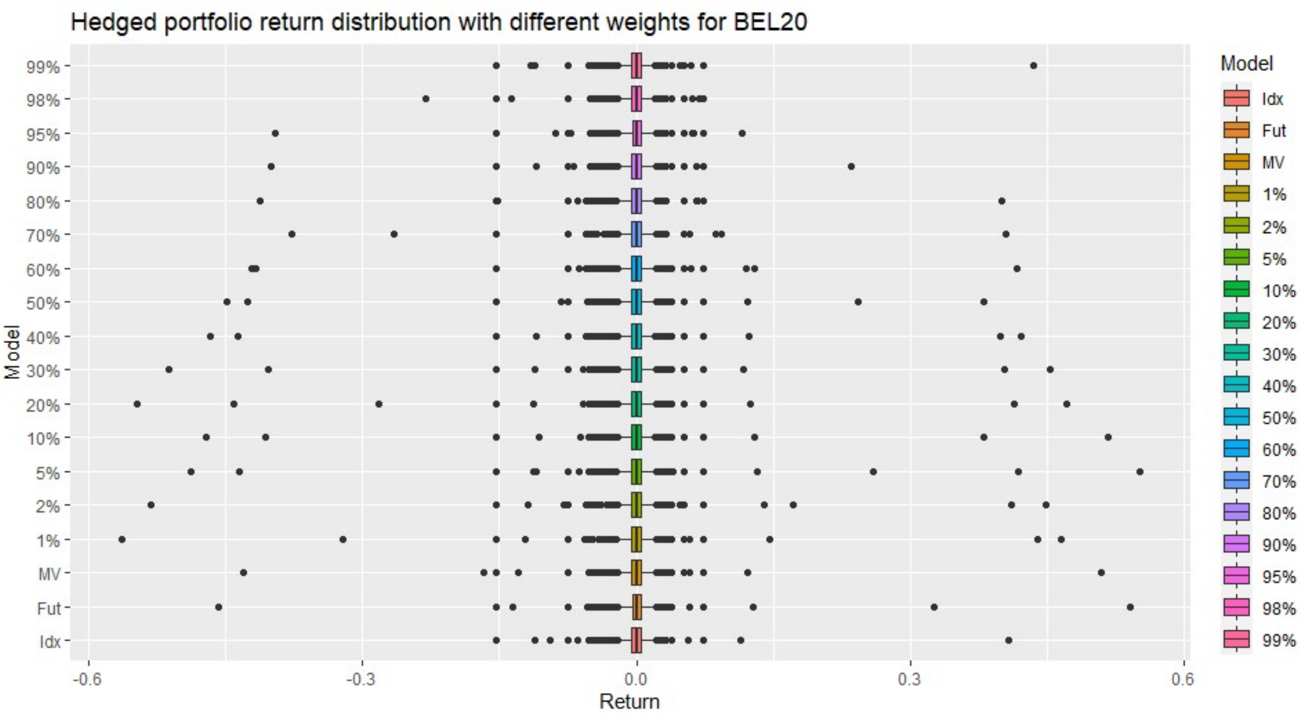
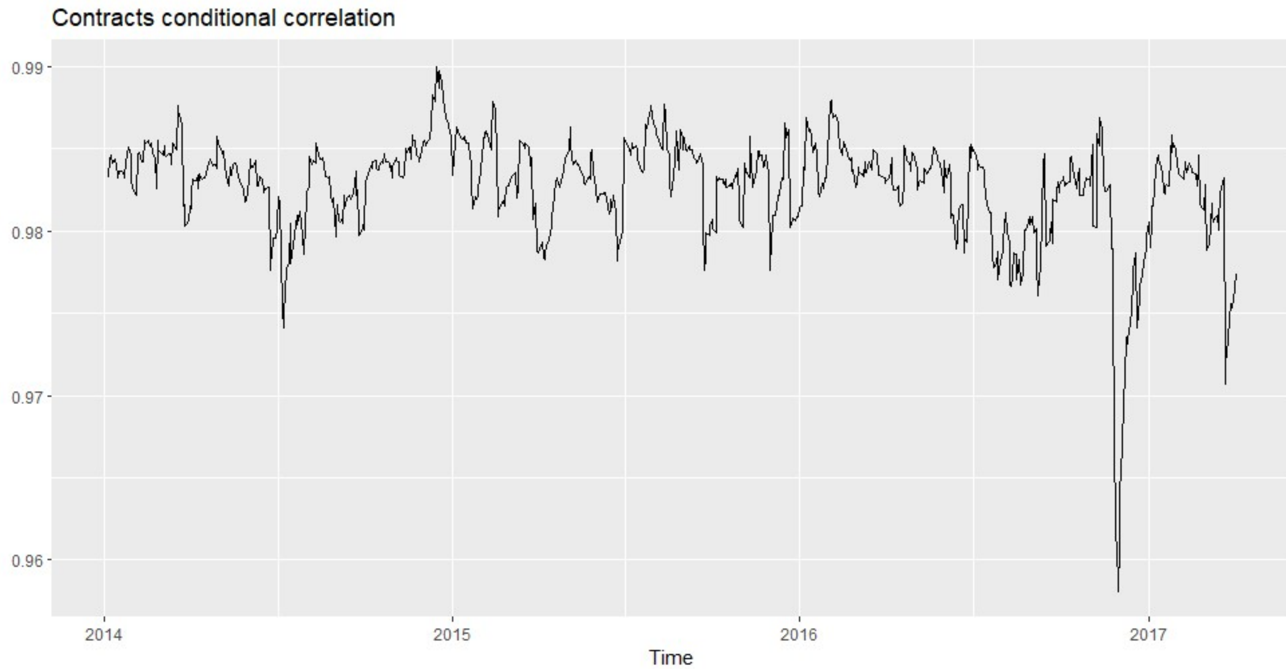


Figure 1.8: Box-plot of the return distribution of Brussel Stock Exchange index, its front month future, OLS Mean-Variance hedged portfolio, and the conditional quantile regression hedged portfolios.

	Left Tail					Right Tail					Student-t Scale	
	k	γ_H	γ_M	γ_{Adj}	ρ_H	k	γ_h	γ_M	γ_{Adj}	ρ_H	σ	σ_{Err}
MV	80	0.64356	0.59389	0.64349	-0.45519	83	0.48911	0.48896	0.48911	-1.00509	0.00864	0.00013
1%	76	0.66903	0.70412	0.67187	-0.06455	83	0.52317	0.69502	0.52365	-0.48708	0.00654	0.00029
2%	77	0.65479	0.69997	0.65483	-0.57218	83	0.52814	0.68514	0.52920	-0.36622	0.00633	0.00028
5%	75	0.69428	0.73161	0.69448	-0.38596	83	0.54734	0.71828	0.54915	-0.30754	0.00658	0.00029
10%	75	0.69005	0.72182	0.69016	-0.42835	83	0.54376	0.71099	0.54553	-0.30692	0.00654	0.00029
20%	76	0.67523	0.71193	0.67628	-0.19349	83	0.53062	0.70196	0.53142	-0.41889	0.00637	0.00028
30%	76	0.67176	0.70744	0.67349	-0.13184	83	0.52537	0.69826	0.52591	-0.47135	0.00659	0.00029
40%	76	0.66526	0.69973	0.66813	-0.06334	83	0.53009	0.68779	0.53130	-0.34891	0.00637	0.00028
50%	77	0.64905	0.69291	0.64905	-1.15754	83	0.52085	0.67814	0.52140	-0.45439	0.00642	0.00028
60%	76	0.66427	0.69861	0.66645	-0.10640	83	0.52875	0.68678	0.52984	-0.36400	0.00642	0.00028
70%	77	0.65369	0.69856	0.65371	-0.66179	83	0.52674	0.68374	0.52770	-0.38037	0.00635	0.00028
80%	77	0.65283	0.69749	0.65284	-0.74158	83	0.52565	0.68267	0.52652	-0.39225	0.00648	0.00029
90%	79	0.65175	0.63043	0.65171	-0.47885	83	0.51507	0.61568	0.51518	-0.59188	0.00632	0.00028
95%	80	0.62435	0.56809	0.62425	-0.52874	83	0.48660	0.51834	0.48660	-2.15471	0.00631	0.00028
98%	80	0.62508	0.52160	0.62438	-0.34707	83	0.48145	0.43725	0.48145	-0.53448	0.00637	0.00028
99%	80	0.63252	0.48694	0.62980	-0.16896	83	0.48837	0.36456	0.48572	-0.09648	0.00638	0.00028

Table 1.9: Shows different Tail Index estimators value and the Student-t scale parameter over the testing set for the model of the Bruxelles Stock Exchange against its front month future



Parameters

	Estimate	Std. Error	t value	Pr(>|t|)
μ^{Idx}	0.000347	0.000294	1.181367	0.237457
θ^{Idx}	0.075267	0.038455	1.957266	0.050316
ω^{Idx}	4e-06	1e-06	3.654027	0.000258
α^{Idx}	0.078373	0.008573	9.141724	0
β^{Idx}	0.866127	0.016459	52.622827	0
μ^{Fut}	0.000354	0.000301	1.178218	0.23871
θ^{Fut}	0.070773	0.036115	1.959647	0.050037
ω^{Fut}	4e-06	1e-06	3.620847	0.000294
α^{Fut}	0.076456	0.007834	9.758979	0
β^{Fut}	0.87104	0.014606	59.636343	0
α^{Cor}	0.031117	0.02371	1.312374	0.189394
β^{Cor}	0.862111	0.121248	7.110336	0

Figure 1.9: GARCH-DCC Results for BMV Index and Future correlation.

	$\hat{\beta}$	$\hat{\sigma}_\beta$	5%	95%	t statistic	Total Loss
1%	-0.25709	0.00001	-0.25712	-0.25706	-19231.92468	3.13395
2%	-0.25265	0.00001	-0.25268	-0.25262	-20577.01752	3.12435
5%	-0.26026	0.00002	-0.26032	-0.26021	-11010.61673	3.14939
10%	-0.26345	0.00001	-0.26349	-0.26342	-17839.64750	3.16674
20%	-0.24345	0.00001	-0.24348	-0.24342	-21213.16860	3.13312
30%	-0.22244	0.00001	-0.22246	-0.22242	-24614.82493	3.09632
40%	-0.19337	0.00001	-0.19339	-0.19335	-22847.86025	3.03785
50%	-0.15451	0.00001	-0.15452	-0.15449	-21586.59834	2.95286
60%	-0.18040	0.00001	-0.18042	-0.18038	-22130.39013	3.03811
70%	-0.20444	0.00001	-0.20447	-0.20442	-19830.70121	3.11924
80%	-0.19714	0.00001	-0.19716	-0.19712	-19087.57410	3.11714
90%	-0.18531	0.00001	-0.18534	-0.18528	-15168.78117	3.10267
95%	-0.17884	0.00002	-0.17889	-0.17879	-8635.87923	3.09382
98%	-0.11896	0.00001	-0.11898	-0.11893	-11531.58025	2.93737
99%	-0.08835	0.00001	-0.08837	-0.08832	-8116.56260	2.85641

	$\hat{\beta}$	$\hat{\sigma}_\beta$	5%	95%	t statistic	Total Loss
1%	-0.25709	0.00003	-0.25715	-0.25703	-10154.86503	3.97970
2%	-0.25265	0.00002	-0.25269	-0.25261	-15410.14813	3.96218
5%	-0.26026	0.00004	-0.26036	-0.26017	-6193.54216	3.97724
10%	-0.26345	0.00002	-0.26350	-0.26341	-14511.65413	3.97158
20%	-0.24345	0.00001	-0.24347	-0.24343	-24503.89161	3.87608
30%	-0.22244	0.00001	-0.22246	-0.22241	-20719.16912	3.77853
40%	-0.19337	0.00001	-0.19338	-0.19335	-26940.43919	3.65702
50%	-0.15451	0.00001	-0.15452	-0.15449	-20449.54781	3.50672
60%	-0.18040	0.00001	-0.18042	-0.18038	-21141.36040	3.55699
70%	-0.20444	0.00001	-0.20447	-0.20442	-22515.75977	3.60040
80%	-0.19714	0.00001	-0.19716	-0.19711	-18414.63394	3.54796
90%	-0.18531	0.00002	-0.18535	-0.18527	-11444.29356	3.48232
95%	-0.17884	0.00004	-0.17892	-0.17875	-5023.77006	3.44806
98%	-0.11896	0.00002	-0.11899	-0.11892	-7621.17539	3.26125
99%	-0.08835	0.00001	-0.08838	-0.08832	-6518.82625	3.16784

Table 1.10: Shows $\hat{\beta}$ inference results over the testing and training set for the model for the Bolsa Mexicana de Valores against its front month future, $n = 815$.

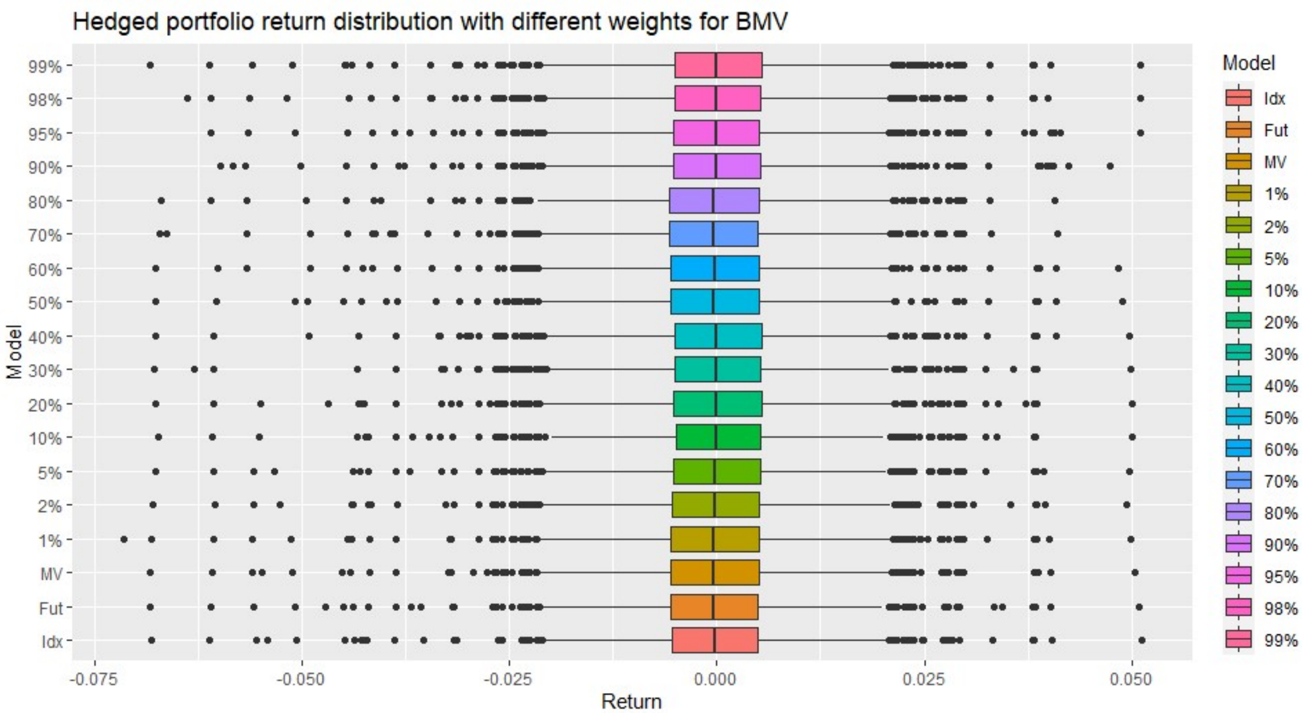
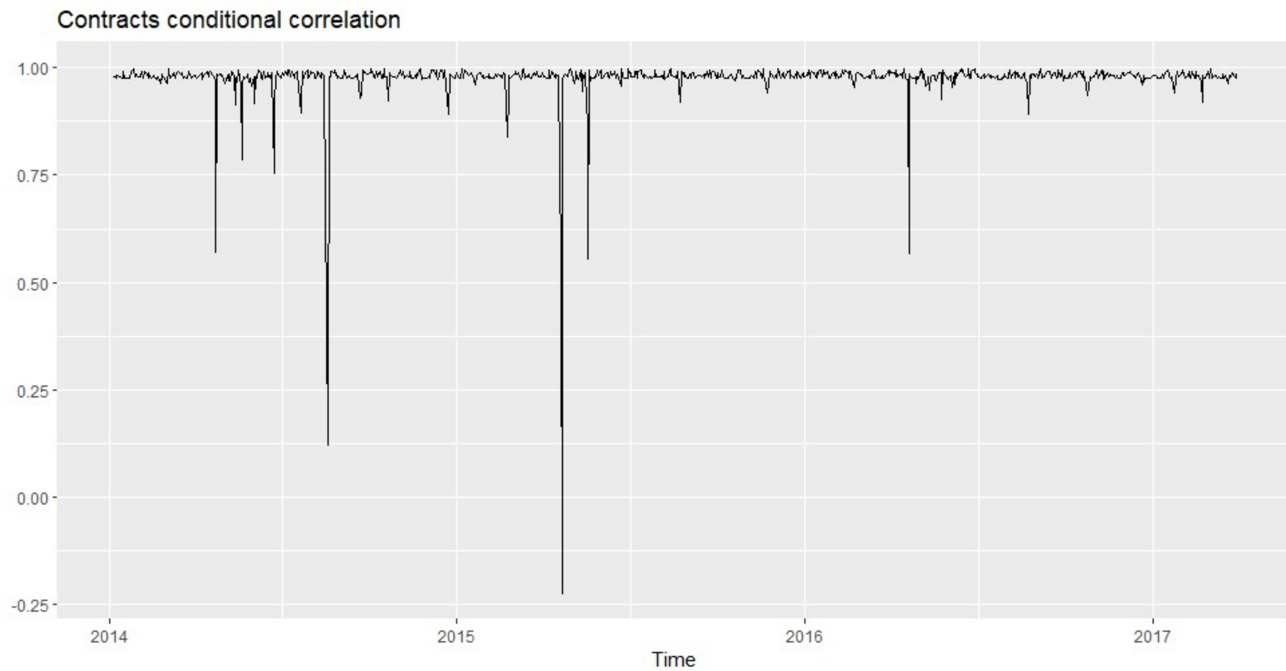


Figure 1.10: Box-plot of the return distribution of Bolsa Mexicana de Valores index, its front month future, OLS Mean-Variance hedged portfolio, and the conditional quantile regression hedged portfolios.

	Left Tail					Right Tail					Student-t Scale	
	k	γ_H	γ_M	γ_{Adj}	ρ_H	k	γ_h	γ_M	γ_{Adj}	ρ_H	σ	σ_{Err}
MV	80	0.53481	0.29358	0.52603	-0.03231	82	0.48708	-0.01650	0.48514	-0.33997	0.00710	0.00031
1%	80	0.50686	0.35831	0.50581	-0.28336	82	0.49735	0.05463	0.49563	-0.32286	0.00807	
2%	80	0.50700	0.35965	0.50604	-0.29213	82	0.49870	0.05072	0.49704	-0.33094	0.00808	
5%	80	0.50852	0.35460	0.50724	-0.25817	82	0.49639	0.05744	0.49462	-0.31693	0.00808	
10%	80	0.51020	0.35084	0.50865	-0.23281	82	0.49540	0.06031	0.49360	-0.31083	0.00808	
20%	80	0.50950	0.35811	0.50832	-0.28126	82	0.50143	0.04281	0.49988	-0.34680	0.00809	
30%	80	0.51335	0.35622	0.51198	-0.26916	82	0.49970	0.04721	0.49779	-0.33104	0.00808	
40%	80	0.51998	0.35019	0.51823	-0.23235	82	0.50243	0.03983	0.50058	-0.33984	0.00806	
50%	80	0.52243	0.35003	0.52058	-0.23544	82	0.49392	0.06466	0.49098	-0.27076	0.00807	
60%	80	0.52553	0.34326	0.52306	-0.19082	82	0.49951	0.04829	0.49734	-0.31693	0.00806	
70%	80	0.51516	0.35605	0.51381	-0.26915	82	0.50197	0.04094	0.50011	-0.34041	0.00807	
80%	80	0.51835	0.35217	0.51675	-0.24460	82	0.50330	0.03734	0.50153	-0.34647	0.00806	
90%	80	0.52211	0.34804	0.52017	-0.21963	82	0.50060	0.04511	0.49856	-0.32563	0.00806	
95%	80	0.52602	0.34262	0.52349	-0.18712	82	0.49916	0.04930	0.49696	-0.31417	0.00806	
98%	80	0.52589	0.34645	0.52379	-0.21749	82	0.48677	0.08600	0.48228	-0.20691	0.00806	
99%	80	0.52775	0.34412	0.52541	-0.20699	82	0.50035	0.04955	0.49783	-0.29099	0.00806	

Table 1.11: Shows different Tail Index estimators value and the Student-t scale parameter over the testing set for the model of the Bolsa Mexicana de Valores against its front month future



Parameters

	Estimate	Std. Error	t value	Pr(>|t|)
μ^{Idx}	0.000588	0.000384	1.52977	0.126074
θ^{Idx}	-0.016572	0.019274	-0.859814	0.389891
ω^{Idx}	3e-06	1e-06	2.07141	0.038321
α^{Idx}	0.113092	0.034174	3.309342	0.000935
β^{Idx}	0.873789	0.032457	26.92142	0
μ^{Fut}	0.000593	0.000374	1.58703	0.112506
θ^{Fut}	-0.0344	0.041059	-0.837836	0.402123
ω^{Fut}	2e-06	6e-06	0.406982	0.684021
α^{Fut}	0.109873	0.077479	1.4181	0.156161
β^{Fut}	0.882011	0.076684	11.501961	0
α^{Cor}	0.472259	0.14626	3.228903	0.001243
β^{Cor}	0	0.007146	0	1

Figure 1.11: GARCH-DCC Results for CAC40 Index and Future correlation.

	$\hat{\beta}$	$\hat{\sigma}_{\beta}$	5%	95%	t statistic	Total Loss
1%	-0.85273	0.00003	-0.85279	-0.85266	-30155.54376	6.62427
2%	-0.80702	0.00003	-0.80708	-0.80695	-28146.84295	6.46339
5%	-0.76255	0.00006	-0.76268	-0.76242	-13518.67368	6.31320
10%	-0.75152	0.00004	-0.75162	-0.75142	-17927.92589	6.28936
20%	-0.74499	0.00002	-0.74504	-0.74494	-35316.55612	6.29740
30%	-0.74250	0.00002	-0.74254	-0.74246	-41378.94339	6.31993
40%	-0.73971	0.00002	-0.73975	-0.73967	-41896.17146	6.34126
50%	-0.73983	0.00001	-0.73985	-0.73980	-62346.66048	6.37320
60%	-0.73795	0.00002	-0.73798	-0.73791	-46072.48269	6.39775
70%	-0.73130	0.00001	-0.73133	-0.73128	-61194.45467	6.40454
80%	-0.70085	0.00002	-0.70090	-0.70081	-34762.16297	6.32243
90%	-0.68954	0.00003	-0.68961	-0.68946	-21505.19012	6.31089
95%	-0.67354	0.00005	-0.67365	-0.67342	-13677.12852	6.26619
98%	-0.61014	0.00001	-0.61018	-0.61011	-42962.82820	6.03721
99%	-0.58665	0.00002	-0.58670	-0.58661	-29695.40200	5.95191

	$\hat{\beta}$	$\hat{\sigma}_{\beta}$	5%	95%	t statistic	Total Loss
1%	-0.85273	0.00006	-0.85286	-0.85260	-15151.25595	5.90197
2%	-0.80702	0.00005	-0.80712	-0.80691	-17362.78731	5.75699
5%	-0.76255	0.00007	-0.76271	-0.76239	-10906.90873	5.61578
10%	-0.75152	0.00004	-0.75161	-0.75143	-19555.35618	5.58037
20%	-0.74499	0.00002	-0.74502	-0.74495	-48107.28692	5.55879
30%	-0.74250	0.00001	-0.74253	-0.74247	-60527.40392	5.55001
40%	-0.73971	0.00001	-0.73973	-0.73968	-69016.87486	5.54028
50%	-0.73983	0.00001	-0.73986	-0.73979	-51199.94096	5.53977
60%	-0.73795	0.00001	-0.73797	-0.73792	-73529.90150	5.53293
70%	-0.73130	0.00001	-0.73133	-0.73128	-61509.43811	5.51103
80%	-0.70085	0.00001	-0.70089	-0.70082	-47242.22904	5.41379
90%	-0.68954	0.00003	-0.68960	-0.68948	-26835.24822	5.37712
95%	-0.67354	0.00005	-0.67366	-0.67342	-13413.59417	5.32609
98%	-0.61014	0.00003	-0.61020	-0.61008	-23156.60665	5.12532
99%	-0.58665	0.00003	-0.58672	-0.58659	-21577.41029	5.05096

Table 1.12: tab:Table 7 Shows $\hat{\beta}$ inference results over the testing and training set for the model for the Cotation Assistée en Continu against its front month future, $n = 829$.

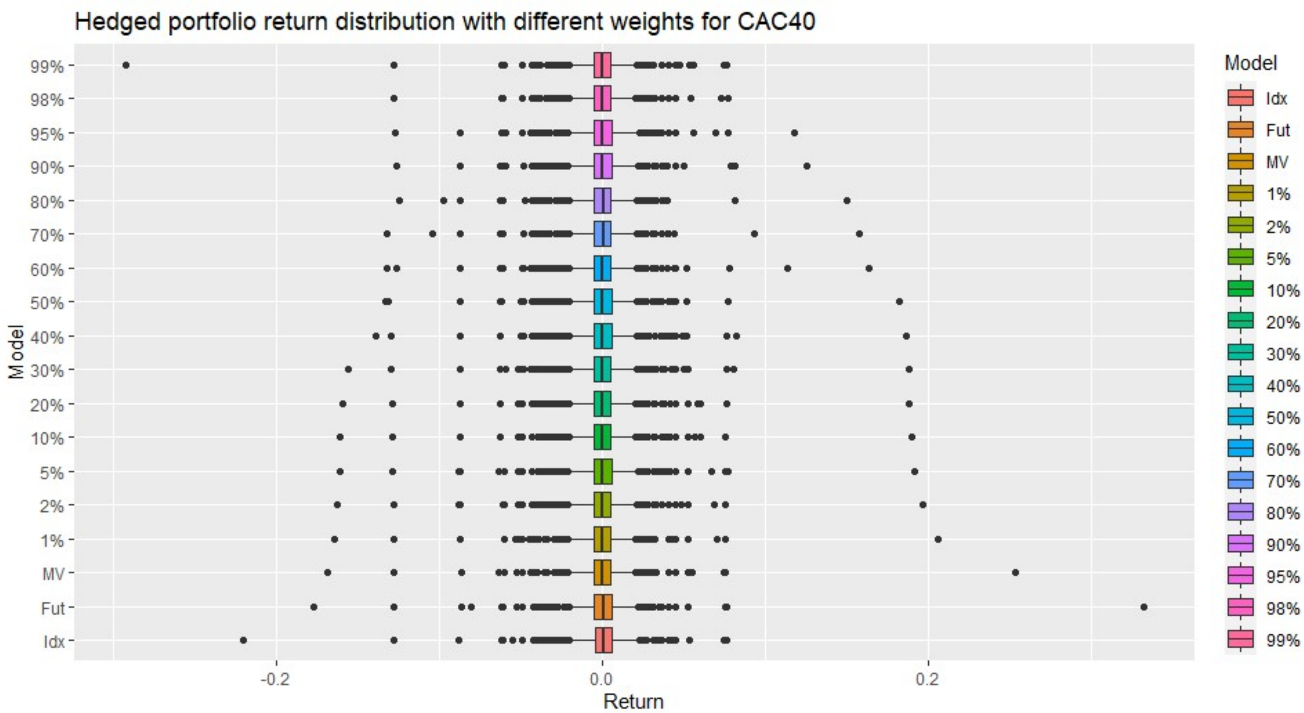
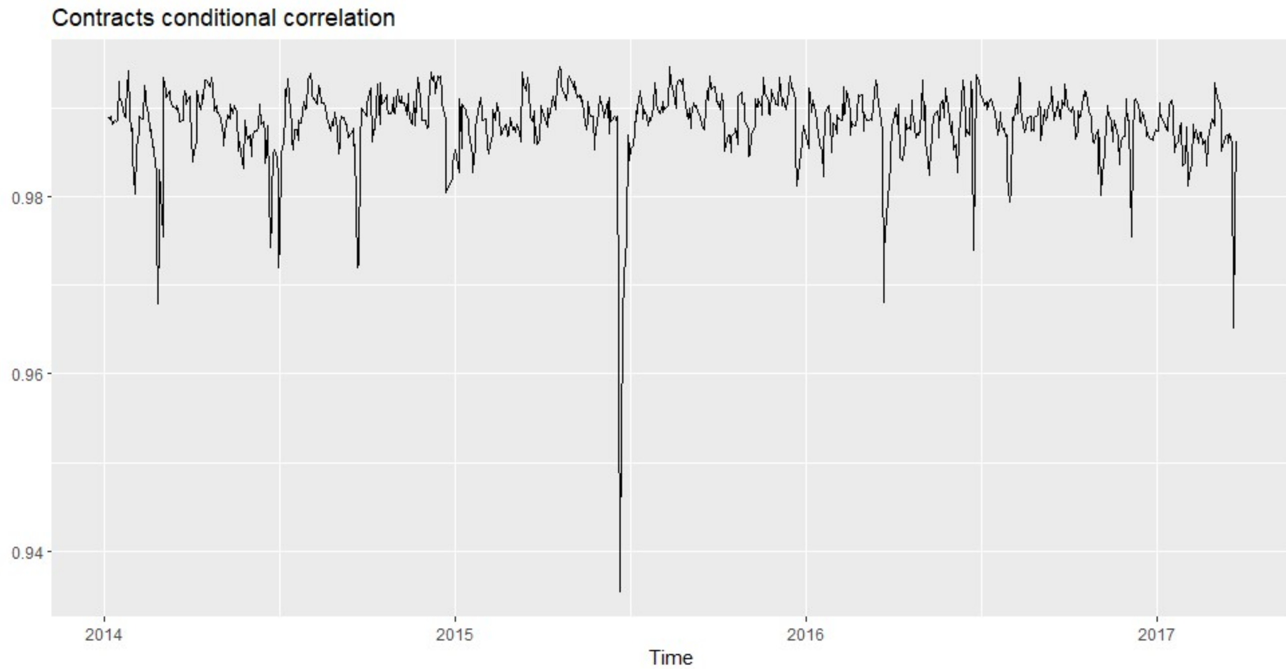


Figure 1.12: Box-plot of the return distribution of Cotation Assistée en Continu index, its front month future, OLS Mean-Variance hedged portfolio, and the conditional quantile regression hedged portfolios.

	Left Tail					Right Tail					Student-t Scale	
	k	γ_H	γ_M	γ_{Adj}	ρ_H	k	γ_h	γ_M	γ_{Adj}	ρ_H	σ	σ_{Err}
MV	81	0.54668	0.42967	0.54346	-0.07544	83	0.48064	0.24855	0.48045	-0.55460	0.00624	0.00028
1%	79	0.54327	0.52675	0.54327	-0.42661	83	0.61457	0.53629	0.61457	-1.52337	0.00690	0.00032
2%	80	0.54520	0.50415	0.54489	-0.14679	83	0.56036	0.48227	0.56036	-3.74410	0.00938	
5%	80	0.51788	0.52093	0.51788	-1.04851	83	0.52227	0.46214	0.52227	-2.03992	0.00671	0.00031
10%	80	0.54440	0.49476	0.54415	-0.30082	83	0.52727	0.44441	0.52727	-2.51031	0.00674	0.00031
20%	81	0.54479	0.49474	0.54452	-0.31815	83	0.52373	0.44135	0.52373	-2.32429	0.00679	0.00032
30%	81	0.54226	0.49716	0.54212	-0.39064	83	0.52734	0.43377	0.52734	-2.66663	0.00681	0.00032
40%	81	0.53946	0.49985	0.53940	-0.47011	83	0.52690	0.43060	0.52690	-2.75863	0.00682	0.00032
50%	81	0.53958	0.49974	0.53951	-0.46674	83	0.52631	0.43153	0.52631	-2.70903	0.00682	0.00032
60%	81	0.53772	0.50154	0.53768	-0.51909	83	0.53526	0.41723	0.53526	-3.42036	0.00683	0.00032
70%	81	0.54169	0.49721	0.54163	-0.43214	83	0.53586	0.40668	0.53586	-3.85631	0.00684	0.00032
80%	81	0.53356	0.50420	0.53356	-0.66718	83	0.50899	0.40706	0.50899	-3.38236	0.00644	0.00029
90%	81	0.52826	0.50929	0.52826	-0.79285	83	0.49985	0.40958	0.49985	-2.89315	0.00643	0.00029
95%	81	0.53970	0.49848	0.53967	-0.58396	83	0.51516	0.37470	0.51516	-4.12206	0.00670	0.00032
98%	81	0.57137	0.45982	0.56782	-0.09398	83	0.50067	0.35772	0.50067	-3.28878	0.00636	0.00029
99%	81	0.57298	0.45448	0.56902	-0.07217	83	0.48386	0.37257	0.48386	-2.41669	0.00633	0.00029

Table 1.13: Shows different Tail Index estimators value and the Student-t scale parameter over the testing set for the model of the CAC40 against its front month future



Parameters

	Estimate	Std. Error	t value	Pr(>|t|)
μ^{Idx}	0.000657	0.000389	1.687365	0.091533
θ^{Idx}	0.006945	0.040401	0.171898	0.863517
ω^{Idx}	3e-06	4e-06	0.673577	0.50058
α^{Idx}	0.105953	0.051351	2.063319	0.039082
β^{Idx}	0.880375	0.055437	15.880748	0
μ^{Fut}	0.00063	0.00037	1.701674	0.088817
θ^{Fut}	0.024748	0.039087	0.633136	0.526645
ω^{Fut}	3e-06	4e-06	0.642156	0.520772
α^{Fut}	0.099655	0.047096	2.115992	0.034346
β^{Fut}	0.886736	0.050884	17.426678	0
α^{Cor}	0.109248	0.029271	3.732349	0.00019
β^{Cor}	0.647111	0.125103	5.172623	0

Figure 1.13: GARCH-DCC Results for DAX Index and Future correlation.

	$\hat{\beta}$	$\hat{\sigma}_\beta$	5%	95%	t statistic	Total Loss
1%	-0.39556	0.00001	-0.39559	-0.39553	-26712.60170	5.11946
2%	-0.38160	0.00002	-0.38164	-0.38156	-21816.86574	5.07206
5%	-0.36587	0.00004	-0.36596	-0.36577	-8691.43394	5.02496
10%	-0.35097	0.00003	-0.35105	-0.35090	-11242.53268	4.98744
20%	-0.33355	0.00002	-0.33360	-0.33350	-15924.51840	4.95689
30%	-0.32455	0.00002	-0.32459	-0.32451	-17840.92617	4.95675
40%	-0.31874	0.00001	-0.31877	-0.31870	-22607.81519	4.96807
50%	-0.30048	0.00001	-0.30051	-0.30045	-27562.94726	4.93228
60%	-0.28881	0.00001	-0.28884	-0.28879	-27392.23455	4.92053
70%	-0.27154	0.00001	-0.27157	-0.27152	-23524.74697	4.88684
80%	-0.25035	0.00002	-0.25040	-0.25031	-12700.95429	4.83725
90%	-0.20200	0.00002	-0.20205	-0.20194	-8301.92781	4.68180
95%	-0.18501	0.00003	-0.18507	-0.18494	-6977.81216	4.63103
98%	-0.17863	0.00001	-0.17865	-0.17861	-17480.21391	4.61517
99%	-0.17127	0.00001	-0.17129	-0.17125	-19174.05039	4.58954

	$\hat{\beta}$	$\hat{\sigma}_\beta$	5%	95%	t statistic	Total Loss
1%	-0.39556	0.00003	-0.39564	-0.39548	-11431.92532	4.69094
2%	-0.38160	0.00003	-0.38167	-0.38153	-12445.56086	4.64419
5%	-0.36587	0.00005	-0.36598	-0.36575	-7334.36744	4.59260
10%	-0.35097	0.00003	-0.35103	-0.35091	-13357.87175	4.54500
20%	-0.33355	0.00002	-0.33359	-0.33351	-18418.33537	4.49162
30%	-0.32455	0.00001	-0.32458	-0.32452	-29864.80101	4.46675
40%	-0.31874	0.00001	-0.31876	-0.31871	-35288.50347	4.45267
50%	-0.30048	0.00001	-0.30050	-0.30046	-35130.81258	4.39612
60%	-0.28881	0.00001	-0.28884	-0.28879	-27201.05536	4.36194
70%	-0.27154	0.00001	-0.27157	-0.27152	-30046.80159	4.30853
80%	-0.25035	0.00001	-0.25038	-0.25032	-18675.81309	4.24156
90%	-0.20200	0.00002	-0.20204	-0.20196	-12428.05824	4.08142
95%	-0.18501	0.00004	-0.18510	-0.18491	-4488.32679	4.02582
98%	-0.17863	0.00003	-0.17869	-0.17857	-6954.25699	4.00552
99%	-0.17127	0.00002	-0.17132	-0.17121	-7200.48656	3.98079

Table 1.14: Shows $\hat{\beta}$ inference results over the testing and training set for the model for the Deutscher Aktienindex against its front month future, $n = 820$.

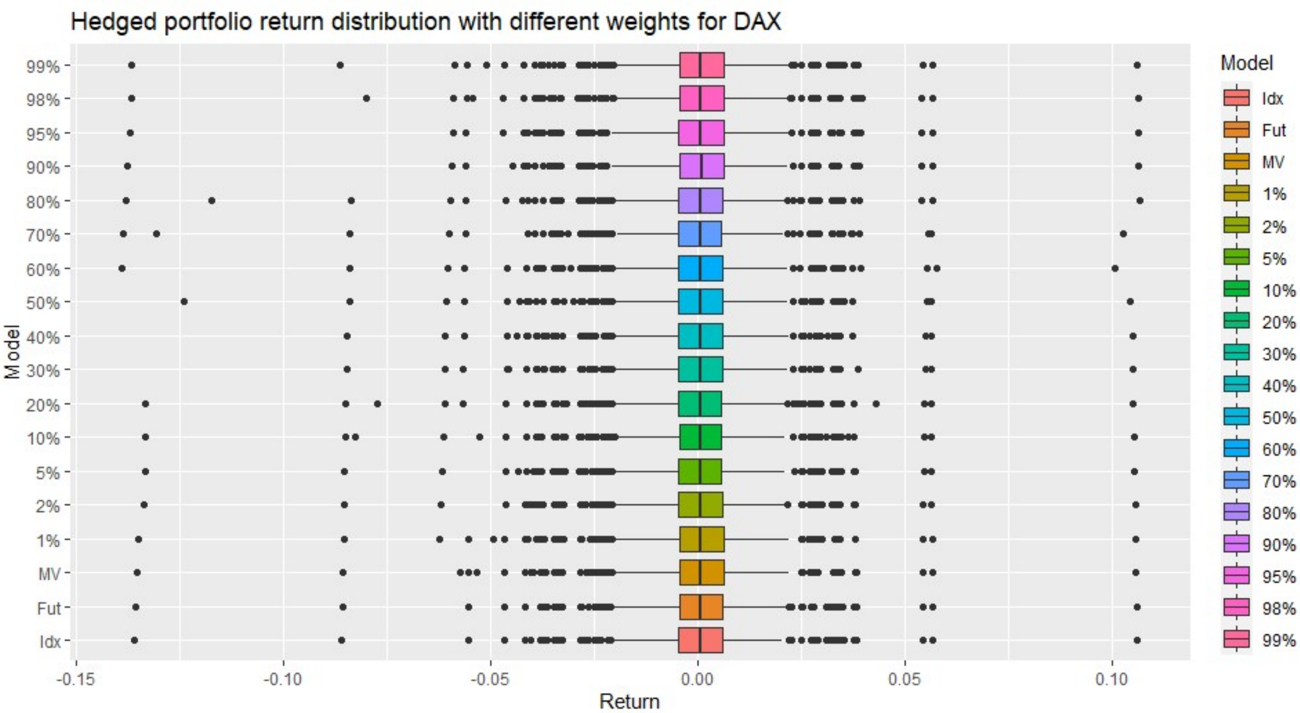
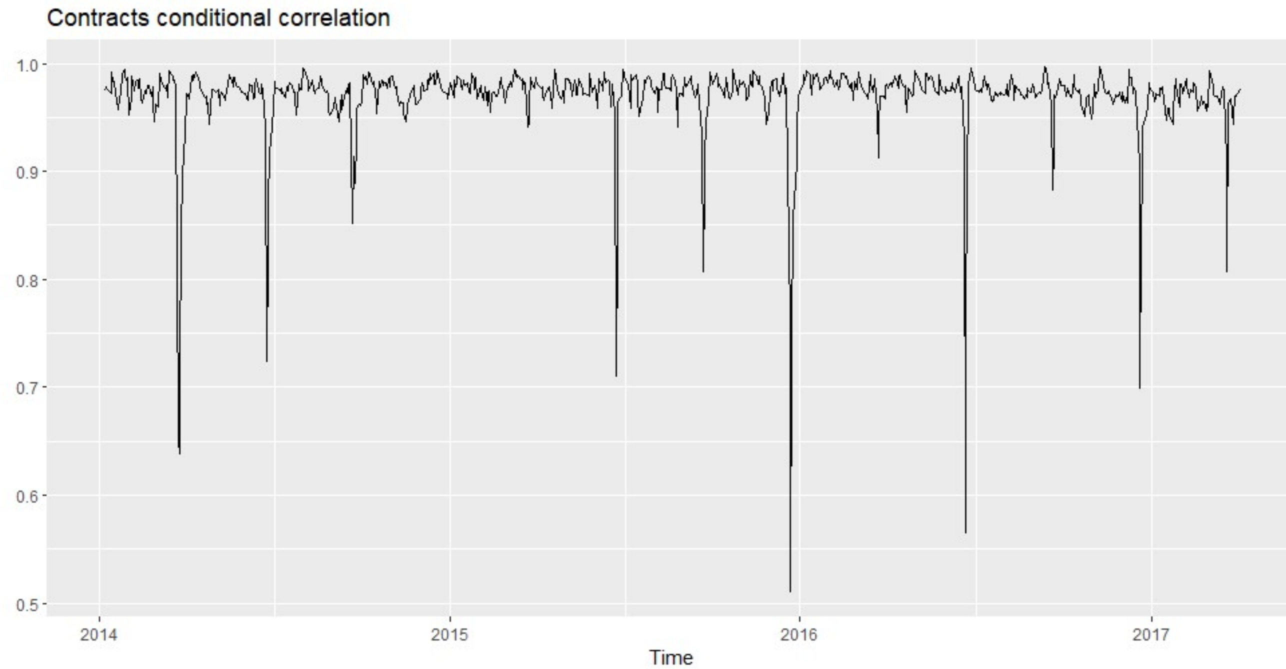


Figure 1.14: Box-plot of the return distribution of Deutscher Aktienindex, its front month future, OLS Mean-Variance hedged portfolio, and the conditional quantile regression hedged portfolios.

	Left Tail					Right Tail					Student-t Scale	
	k	γ_H	γ_M	γ_{Adj}	ρ_H	k	γ_h	γ_M	γ_{Adj}	ρ_H	σ	σ_{Err}
MV	80	0.47151	0.38433	0.46952	-0.07998	82	0.40786	0.38737	0.40786	-0.69342	0.00707	0.00033
1%	80	0.46143	0.46712	0.46143	-1.07237	82	0.45067	0.39446	0.45061	-0.35525	0.00697	0.00031
2%	80	0.45556	0.47244	0.45556	-1.21374	82	0.44997	0.39319	0.44988	-0.33853	0.00697	0.00031
5%	80	0.45626	0.47032	0.45626	-1.17746	82	0.44920	0.39210	0.44906	-0.32490	0.00697	0.00031
10%	80	0.45483	0.47070	0.45483	-1.19966	82	0.44846	0.39138	0.44829	-0.31697	0.00698	0.00031
20%	80	0.44935	0.47510	0.44935	-1.32347	82	0.44456	0.39538	0.44454	-0.40037	0.00698	0.00031
30%	80	0.44782	0.47547	0.44782	-1.34883	82	0.44486	0.39384	0.44481	-0.37443	0.00699	0.00032
40%	80	0.44801	0.47445	0.44801	-1.33446	82	0.44404	0.39438	0.44399	-0.38767	0.00699	0.00032
50%	80	0.44856	0.47149	0.44856	-1.29179	82	0.44074	0.39743	0.44070	-0.45695	0.00699	0.00032
60%	80	0.44887	0.46978	0.44887	-1.26646	82	0.44185	0.39486	0.44182	-0.41020	0.00699	0.00032
70%	80	0.44175	0.47578	0.44175	-1.43389	82	0.43577	0.40228	0.43578	-0.57157	0.00691	0.00031
80%	80	0.44604	0.46864	0.44604	-1.28938	82	0.43741	0.39825	0.43745	-0.49853	0.00696	0.00031
90%	80	0.44178	0.46882	0.44178	-1.34607	82	0.42735	0.40997	0.42736	-0.77425	0.00694	0.00031
95%	80	0.44267	0.46643	0.44267	-1.30354	82	0.42644	0.40931	0.42645	-0.77560	0.00694	0.00031
98%	80	0.44316	0.46536	0.44316	-1.28328	82	0.43130	0.40156	0.43133	-0.61257	0.00697	0.00032
99%	80	0.44374	0.46407	0.44374	-1.25956	82	0.43686	0.39248	0.43695	-0.42699	0.00697	0.00032

Table 1.15: Shows different Tail Index estimators value and the Student-t scale parameter over the testing set for the model of the DAX against its front month future



Parameters

	Estimate	Std. Error	t value	Pr(>|t|)
μ^{Idx}	0.00055	0.00022	2.497441	0.012509
θ^{Idx}	-0.06567	0.038927	-1.686996	0.091604
ω^{Idx}	6e-06	1e-06	6.807698	0
α^{Idx}	0.213277	0.033354	6.394272	0
β^{Idx}	0.686995	0.045709	15.029753	0
μ^{Fut}	0.000595	0.000235	2.53712	0.011177
θ^{Fut}	-0.02658	0.042387	-0.62709	0.5306
ω^{Fut}	8e-06	1e-06	11.272982	0
α^{Fut}	0.263417	0.045549	5.783134	0
β^{Fut}	0.624026	0.051791	12.048819	0
α^{Cor}	0.287834	0.044406	6.481801	0
β^{Cor}	0.41892	0.060531	6.920721	0

Figure 1.15: GARCH-DCC Results for DJA Index and Future correlation.

	$\hat{\beta}$	$\hat{\sigma}_\beta$	5%	95%	t statistic	Total Loss
1%	-0.30588	0.00001	-0.30590	-0.30586	-29716.78685	2.88616
2%	-0.28982	0.00001	-0.28984	-0.28980	-33132.80693	2.85510
5%	-0.21725	0.00003	-0.21731	-0.21719	-8391.36079	2.70987
10%	-0.15470	0.00002	-0.15474	-0.15466	-9341.47123	2.59075
20%	-0.14271	0.00001	-0.14273	-0.14268	-12939.53985	2.59144
30%	-0.13292	0.00001	-0.13293	-0.13290	-19699.17821	2.59636
40%	-0.14844	0.00001	-0.14845	-0.14843	-28726.10205	2.65605
50%	-0.12363	0.00001	-0.12365	-0.12362	-19657.91610	2.62754
60%	-0.14870	0.00001	-0.14872	-0.14869	-27734.60923	2.70901
70%	-0.15220	0.00001	-0.15222	-0.15218	-16873.70686	2.74306
80%	-0.11274	0.00001	-0.11276	-0.11272	-13976.29424	2.67962
90%	-0.08867	0.00002	-0.08871	-0.08863	-5390.11176	2.64978
95%	-0.08043	0.00001	-0.08047	-0.08040	-5869.84669	2.64321
98%	-0.05361	0.00001	-0.05363	-0.05360	-8786.80298	2.58861
99%	-0.04797	0.00001	-0.04799	-0.04795	-5983.37945	2.57796
	$\hat{\beta}$	$\hat{\sigma}_\beta$	5%	95%	t statistic	Total Loss
1%	-0.30588	0.00004	-0.30597	-0.30579	-8170.89590	4.11451
2%	-0.28982	0.00003	-0.28989	-0.28975	-9907.62499	4.06758
5%	-0.21725	0.00006	-0.21739	-0.21710	-3512.36542	3.85187
10%	-0.15470	0.00003	-0.15478	-0.15462	-4502.36354	3.67151
20%	-0.14271	0.00001	-0.14273	-0.14268	-13364.31425	3.65990
30%	-0.13292	0.00001	-0.13294	-0.13289	-14319.12315	3.65469
40%	-0.14844	0.00001	-0.14845	-0.14843	-28110.77827	3.72919
50%	-0.12363	0.00001	-0.12365	-0.12362	-19972.68367	3.67577
60%	-0.14870	0.00001	-0.14872	-0.14869	-23664.82308	3.78143
70%	-0.15220	0.00001	-0.15222	-0.15218	-22690.97485	3.81844
80%	-0.11274	0.00001	-0.11276	-0.11271	-10999.16952	3.71572
90%	-0.08867	0.00002	-0.08871	-0.08862	-4692.61356	3.66171
95%	-0.08043	0.00003	-0.08051	-0.08035	-2368.34415	3.64683
98%	-0.05361	0.00001	-0.05365	-0.05358	-3577.16305	3.56587
99%	-0.04797	0.00003	-0.04805	-0.04789	-1415.29410	3.54967

Table 1.16: Shows $\hat{\beta}$ inference results over the testing and training set for the model for the Dow Jones Industrial Average against its front month future, $n = 817$.

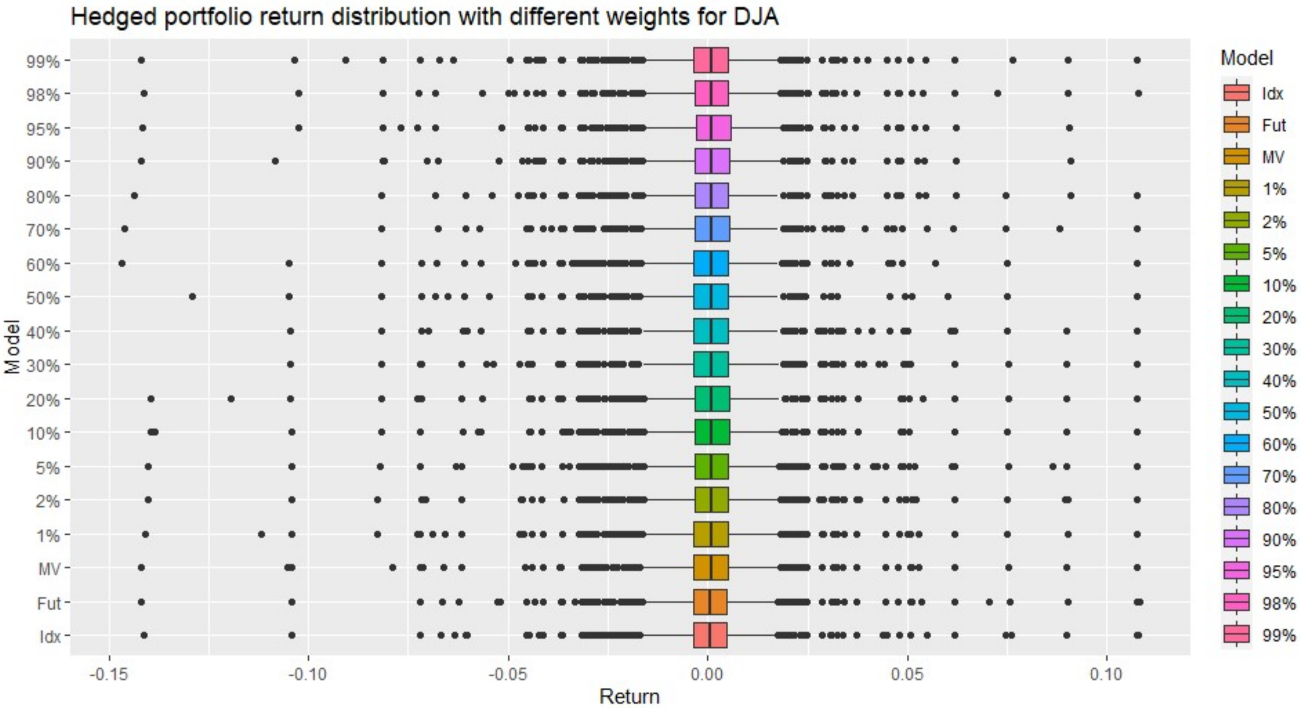
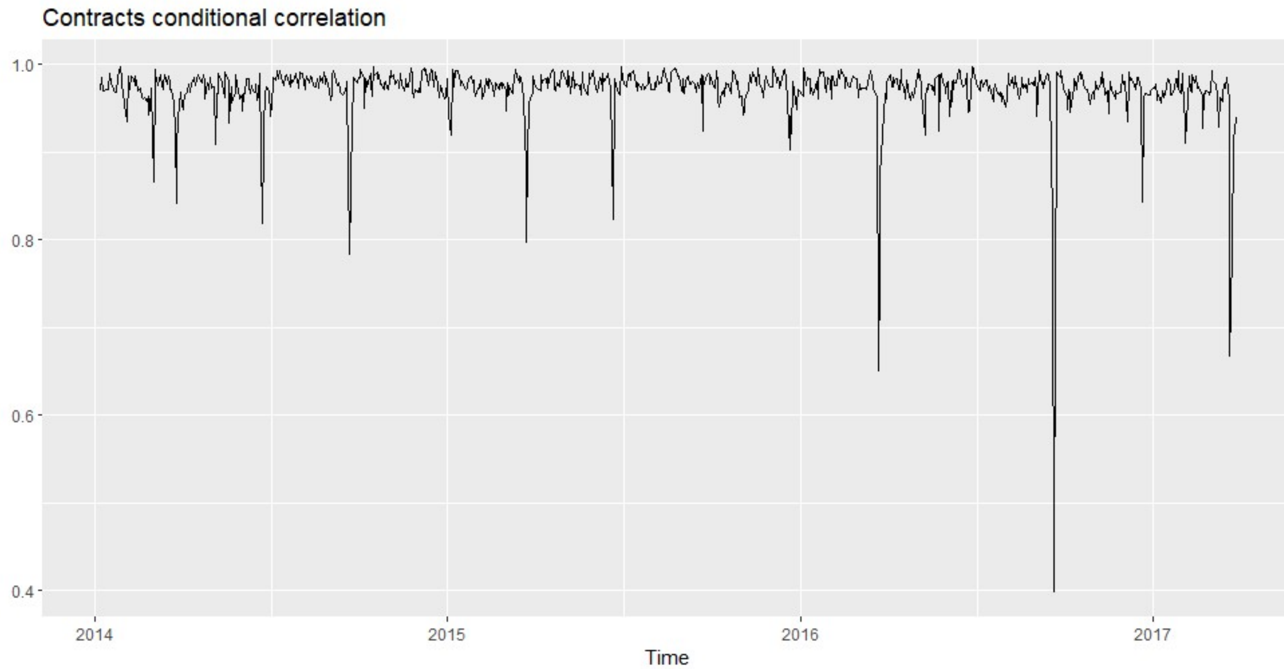


Figure 1.16: Box-plot of the return distribution of Dow Jones Industrial Average index, its front month future, OLS Mean-Variance hedged portfolio, and the conditional quantile regression hedged portfolios.

	Left Tail					Right Tail					Student-t Scale	
	k	γ_H	γ_M	γ_{Adj}	ρ_H	k	γ_h	γ_M	γ_{Adj}	ρ_H	σ	σ_{Err}
MV	78	0.68545	0.36537	0.68427	-0.41472	82	0.53311	0.46517	0.53274	-0.35324	0.00512	0.00025
1%	78	0.72135	0.29127	0.72132	-0.90046	82	0.54877	0.46126	0.54320	-0.09509	0.00798	
2%	78	0.72344	0.28578	0.72341	-0.90850	82	0.55009	0.45979	0.54229	-0.07087	0.00797	
5%	78	0.67048	0.35105	0.67023	-0.63650	82	0.53339	0.48046	0.53322	-0.44881	0.00803	0.00002
10%	78	0.67234	0.34167	0.67208	-0.62637	82	0.54071	0.47467	0.54034	-0.34520	0.00523	0.00026
20%	78	0.67246	0.34193	0.67215	-0.61004	82	0.54163	0.47319	0.54118	-0.32341	0.00514	0.00025
30%	78	0.67684	0.33609	0.67658	-0.62459	82	0.54237	0.47202	0.54184	-0.30650	0.00518	0.00026
40%	78	0.67241	0.34174	0.67213	-0.61820	82	0.54119	0.47389	0.54079	-0.33367	0.00523	0.00026
50%	78	0.70285	0.29713	0.70277	-0.76743	82	0.54305	0.47096	0.54243	-0.29115	0.00840	0.00012
60%	78	0.67241	0.34174	0.67212	-0.61856	82	0.54117	0.47392	0.54077	-0.33415	0.00523	0.00026
70%	78	0.67237	0.34168	0.67210	-0.62320	82	0.54090	0.47436	0.54052	-0.34055	0.00523	0.00026
80%	78	0.70916	0.28689	0.70909	-0.79172	82	0.54383	0.46976	0.54310	-0.27396	0.00840	0.00012
90%	78	0.70579	0.29133	0.70569	-0.75930	82	0.54547	0.46726	0.54444	-0.23893	0.00512	0.00025
95%	78	0.70465	0.29297	0.70455	-0.74800	82	0.54592	0.46655	0.54479	-0.22939	0.00512	0.00025
98%	78	0.70020	0.29983	0.70007	-0.70683	82	0.54323	0.46889	0.54250	-0.27679	0.00516	0.00026
99%	78	0.69900	0.30175	0.69885	-0.69665	82	0.54268	0.46937	0.54202	-0.28665	0.00515	0.00025

Table 1.17: Shows different Tail Index estimators value and the Student-t scale parameter over the testing set for the model of the DJA against its front month future



Parameters

	Estimate	Std. Error	t value	Pr(>|t|)
μ^{Idx}	0.000535	0.000368	1.454447	0.145822
θ^{Idx}	-0.031624	0.040581	-0.779284	0.435813
ω^{Idx}	3e-06	3e-06	0.815119	0.415004
α^{Idx}	0.09779	0.048593	2.012443	0.044173
β^{Idx}	0.891212	0.047521	18.75411	0
μ^{Fut}	0.000363	0.000476	0.762226	0.445925
θ^{Fut}	-0.025216	0.043422	-0.580727	0.561425
ω^{Fut}	4e-06	5e-06	0.652471	0.514098
α^{Fut}	0.089374	0.023965	3.72943	0.000192
β^{Fut}	0.893328	0.020037	44.584575	0
α^{Cor}	0.393304	0.065571	5.998186	0
β^{Cor}	0.258541	0.232639	1.11134	0.266422

Figure 1.17: GARCH-DCC Results for EXX Index and Future correlation.

	$\hat{\beta}$	$\hat{\sigma}_{\beta}$	5%	95%	t statistic	Total Loss
1%	-0.79062	0.00003	-0.79068	-0.79055	-29414.78867	6.70820
2%	-0.78014	0.00003	-0.78020	-0.78007	-28561.92558	6.67069
5%	-0.76303	0.00006	-0.76317	-0.76290	-13104.91266	6.61200
10%	-0.74405	0.00003	-0.74414	-0.74397	-21343.43708	6.54988
20%	-0.72862	0.00002	-0.72866	-0.72857	-35251.81535	6.51021
30%	-0.71158	0.00003	-0.71164	-0.71153	-27733.93042	6.46421
40%	-0.77157	0.00002	-0.77160	-0.77153	-48463.26012	6.71041
50%	-0.75708	0.00001	-0.75711	-0.75704	-57076.17875	6.67416
60%	-0.72193	0.00002	-0.72197	-0.72189	-42133.85431	6.55882
70%	-0.72251	0.00001	-0.72254	-0.72248	-59003.94966	6.57949
80%	0.29581	0.00001	0.29578	0.29583	27864.29879	2.72858
90%	-0.11573	0.00002	-0.11577	-0.11569	-6506.86690	4.28742
95%	-0.30077	0.00004	-0.30086	-0.30068	-7603.99123	5.00348
98%	-0.35168	0.00002	-0.35172	-0.35164	-21082.17701	5.20345
99%	-0.35106	0.00002	-0.35110	-0.35102	-20548.63709	5.20255

	$\hat{\beta}$	$\hat{\sigma}_{\beta}$	5%	95%	t statistic	Total Loss
1%	-0.79062	0.00005	-0.79073	-0.79050	-15643.63513	5.78717
2%	-0.78014	0.00005	-0.78024	-0.78003	-17001.36553	5.75195
5%	-0.76303	0.00007	-0.76320	-0.76287	-10554.01001	5.69381
10%	-0.74405	0.00003	-0.74413	-0.74398	-22027.51859	5.62851
20%	-0.72862	0.00002	-0.72865	-0.72858	-46089.60845	5.57258
30%	-0.71158	0.00002	-0.71162	-0.71155	-44653.13570	5.51143
40%	-0.77157	0.00001	-0.77160	-0.77153	-56527.64514	5.70501
50%	-0.75708	0.00001	-0.75710	-0.75705	-70890.57011	5.65223
60%	-0.72193	0.00001	-0.72196	-0.72190	-56311.64784	5.53130
70%	-0.72251	0.00001	-0.72253	-0.72249	-73921.99028	5.52842
80%	0.29581	0.00001	0.29579	0.29582	46903.71790	2.20234
90%	-0.11573	0.00002	-0.11577	-0.11569	-6731.31607	3.52591
95%	-0.30077	0.00003	-0.30085	-0.30069	-8748.24553	4.13016
98%	-0.35168	0.00002	-0.35173	-0.35163	-16843.07784	4.29591
99%	-0.35106	0.00003	-0.35113	-0.35099	-11499.03882	4.29346

Table 1.18: Shows $\hat{\beta}$ inference results over the testing and training set for the model for the EURO STOXX 50 against its front month future, $n = 827$.

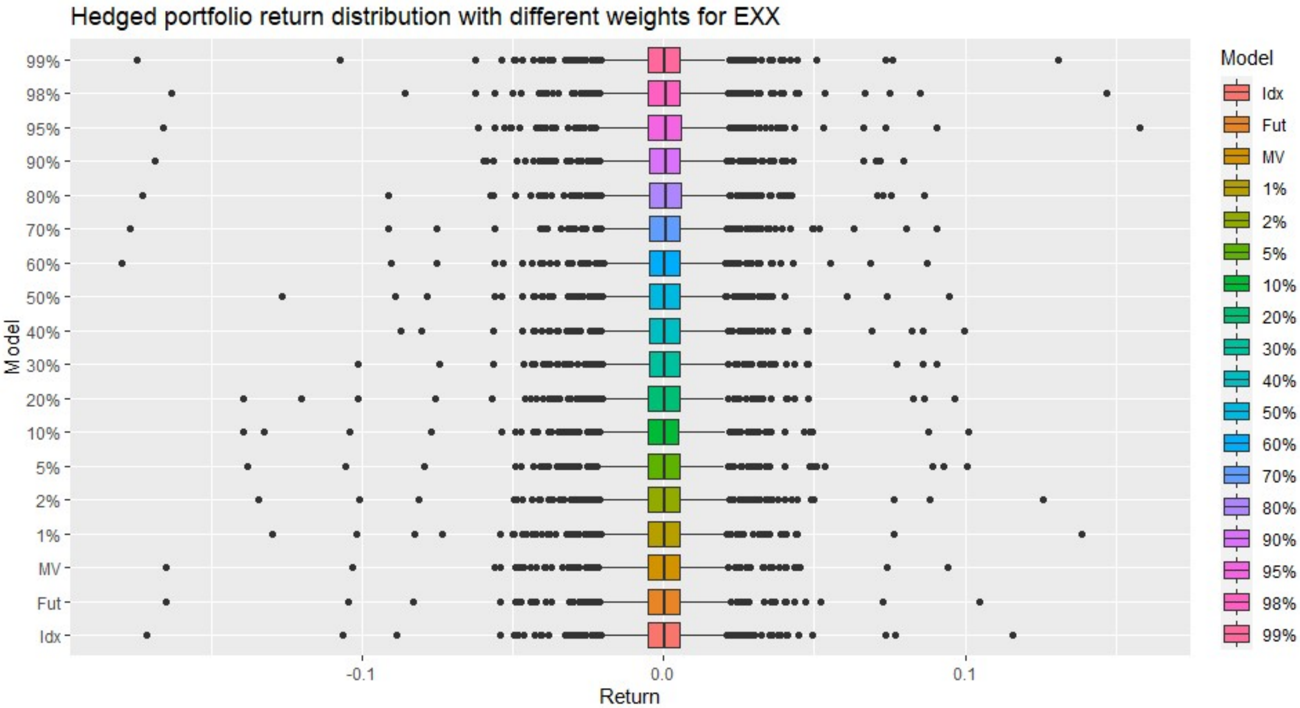
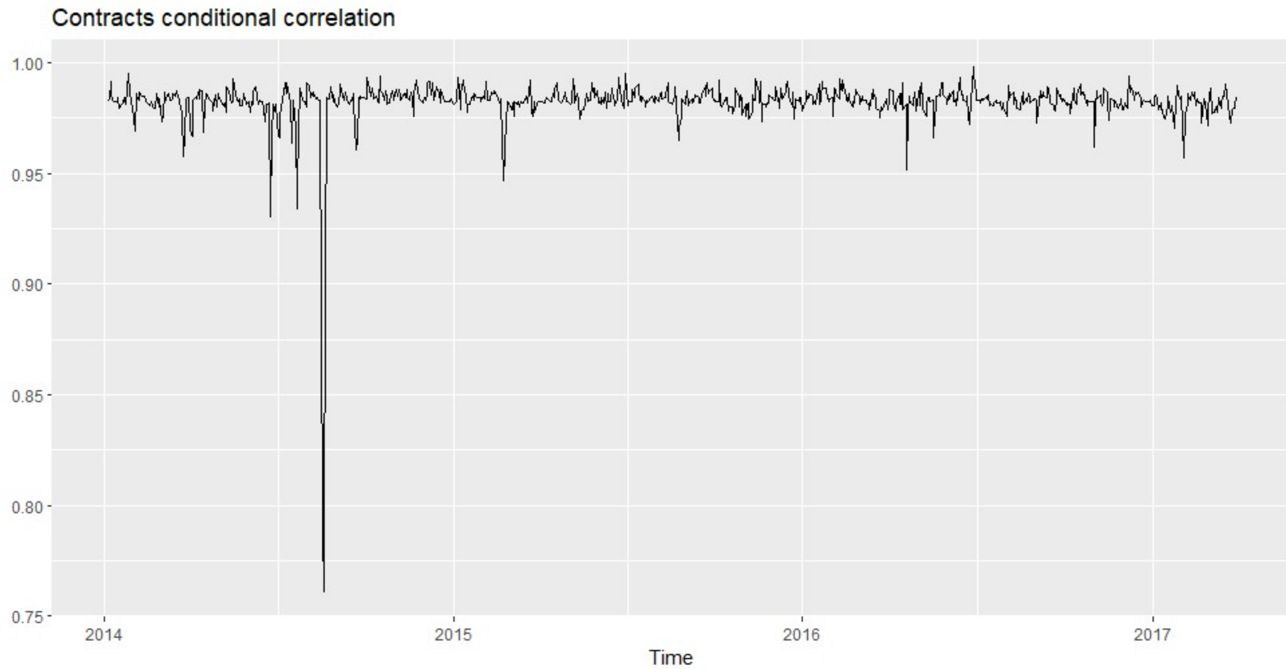


Figure 1.18: Box-plot of the return distribution of EURO STOXX 50 index, its front month future, OLS Mean-Variance hedged portfolio, and the conditional quantile regression hedged portfolios.

	Left Tail					Right Tail					Student-t Scale	
	k	γ_H	γ_M	γ_{Adj}	ρ_H	k	γ_h	γ_M	γ_{Adj}	ρ_H	σ	σ_{Err}
MV	80	0.56235	0.39482	0.55854	-0.12086	83	0.44076	0.39275	0.44074	-0.53058	0.00637	0.00029
1%	80	0.47550	0.44327	0.47524	-0.01853	83	0.52500	0.51204	0.52501	-0.83871	0.00756	0.00035
2%	80	0.49342	0.42984	0.49338	-0.70464	83	0.52774	0.49672	0.52775	-0.64913	0.00745	0.00034
5%	80	0.51374	0.41742	0.51374	-1.20904	83	0.54942	0.45401	0.54564	-0.10622	0.00724	0.00033
10%	80	0.51399	0.43121	0.51389	-0.54796	83	0.54753	0.43637	0.54121	-0.08954	0.00712	0.00032
20%	80	0.52353	0.42804	0.52345	-0.58953	83	0.53322	0.43780	0.53234	-0.26070	0.00679	0.00031
30%	80	0.50794	0.45019	0.50602	-0.00696	83	0.52300	0.43321	0.52249	-0.34579	0.00678	0.00031
40%	80	0.50825	0.41750	0.50825	-1.27513	83	0.54085	0.47311	0.54029	-0.30521	0.00735	0.00034
50%	80	0.50700	0.43134	0.50693	-0.58029	83	0.55292	0.44335	0.53511	-0.02869	0.00709	0.00032
60%	80	0.52101	0.43328	0.52082	-0.45653	83	0.53555	0.42867	0.53393	-0.20622	0.00682	0.00031
70%	80	0.52300	0.43073	0.52288	-0.51930	83	0.53482	0.43009	0.53334	-0.21836	0.00681	0.00031
80%	80	0.55995	0.40172	0.55606	-0.15336	83	0.45117	0.37879	0.45076	-0.33652	0.00628	0.00028
90%	80	0.54884	0.42866	0.54468	-0.03494	83	0.42783	0.41585	0.42784	-0.89768	0.00637	0.00029
95%	80	0.54704	0.43548	0.54398	-0.06843	83	0.46891	0.36429	0.46803	-0.27106	0.00619	0.00027
98%	80	0.55506	0.42139	0.55401	-0.31365	83	0.48130	0.34253	0.47470	-0.08891	0.00617	0.00027
99%	80	0.55505	0.42145	0.55400	-0.31227	83	0.48123	0.34262	0.47468	-0.08927	0.00617	0.00027

Table 1.19: Shows different Tail Index estimators value and the Student-t scale parameter over the testing set for the model of the EXX50 against its front month future



Parameters

	Estimate	Std. Error	t value	Pr(>|t|)
μ^{Idx}	0.000284	0.000468	0.606423	0.544234
θ^{Idx}	0.026068	0.050752	0.513638	0.607505
ω^{Idx}	7e-06	6e-06	1.181181	0.237531
α^{Idx}	0.1146	0.019092	6.002566	0
β^{Idx}	0.855707	0.036611	23.372871	0
μ^{Fut}	0.000253	0.000483	0.523312	0.600757
θ^{Fut}	-0.001937	0.057439	-0.033722	0.973099
ω^{Fut}	6e-06	5e-06	1.283708	0.199244
α^{Fut}	0.098825	0.01402	7.048909	0
β^{Fut}	0.873616	0.025167	34.712718	0
α^{Cor}	0.177888	0.04265	4.170865	3e-05
β^{Cor}	0.235414	0.196513	1.197954	0.230935

Figure 1.19: GARCH-DCC Results for IBEX35 Index and Future correlation.

	$\hat{\beta}$	$\hat{\sigma}_\beta$	5%	95%	t statistic	Total Loss
1%	-1.34829	0.00004	-1.34838	-1.34820	-33220.17901	9.37232
2%	-1.29819	0.00003	-1.29827	-1.29811	-37349.55544	9.17359
5%	-0.98964	0.00006	-0.98979	-0.98950	-15892.99629	7.94541
10%	-1.01908	0.00004	-1.01917	-1.01900	-27475.22934	8.06914
20%	-1.00875	0.00003	-1.00883	-1.00867	-30936.63795	8.04010
30%	-0.98903	0.00003	-0.98910	-0.98897	-35087.96845	7.97332
40%	-0.98449	0.00002	-0.98452	-0.98445	-62297.18940	7.96721
50%	-0.98441	0.00002	-0.98444	-0.98437	-64538.44865	7.97899
60%	-0.89389	0.00002	-0.89392	-0.89385	-56708.86507	7.62655
70%	-0.92887	0.00002	-0.92891	-0.92882	-46150.42311	7.77919
80%	-0.93379	0.00003	-0.93385	-0.93372	-33665.05800	7.81084
90%	-0.97481	0.00004	-0.97490	-0.97473	-26096.30237	7.98862
95%	-0.85949	0.00005	-0.85961	-0.85937	-16802.84438	7.52794
98%	-1.02571	0.00003	-1.02577	-1.02565	-40292.19287	8.20443
99%	-1.07693	0.00002	-1.07697	-1.07688	-57545.18428	8.41315

	$\hat{\beta}$	$\hat{\sigma}_\beta$	5%	95%	t statistic	Total Loss
1%	-1.34829	0.00005	-1.34840	-1.34817	-27290.13304	8.23814
2%	-1.29819	0.00005	-1.29831	-1.29808	-26845.73882	8.05417
5%	-0.98964	0.00008	-0.98982	-0.98946	-12765.84039	6.94927
10%	-1.01908	0.00003	-1.01916	-1.01900	-30129.80439	7.01981
20%	-1.00875	0.00002	-1.00880	-1.00870	-47973.06902	6.91859
30%	-0.98903	0.00002	-0.98907	-0.98899	-53172.72933	6.78588
40%	-0.98449	0.00001	-0.98452	-0.98446	-72620.21411	6.70597
50%	-0.98441	0.00001	-0.98444	-0.98437	-65880.03441	6.64134
60%	-0.89389	0.00001	-0.89392	-0.89385	-64676.84323	6.27570
70%	-0.92887	0.00001	-0.92890	-0.92884	-72241.53563	6.32955
80%	-0.93379	0.00002	-0.93384	-0.93374	-40411.66956	6.28308
90%	-0.97481	0.00003	-0.97488	-0.97475	-33400.58402	6.35295
95%	-0.85949	0.00005	-0.85961	-0.85937	-16456.56837	5.95021
98%	-1.02571	0.00003	-1.02578	-1.02565	-36399.13272	6.46491
99%	-1.07693	0.00004	-1.07702	-1.07683	-27209.71407	6.62237

Table 1.20: Shows $\hat{\beta}$ inference results over the testing and training set for the model for the Indice Bursatil Espanol against its front month future, $n = 829$.

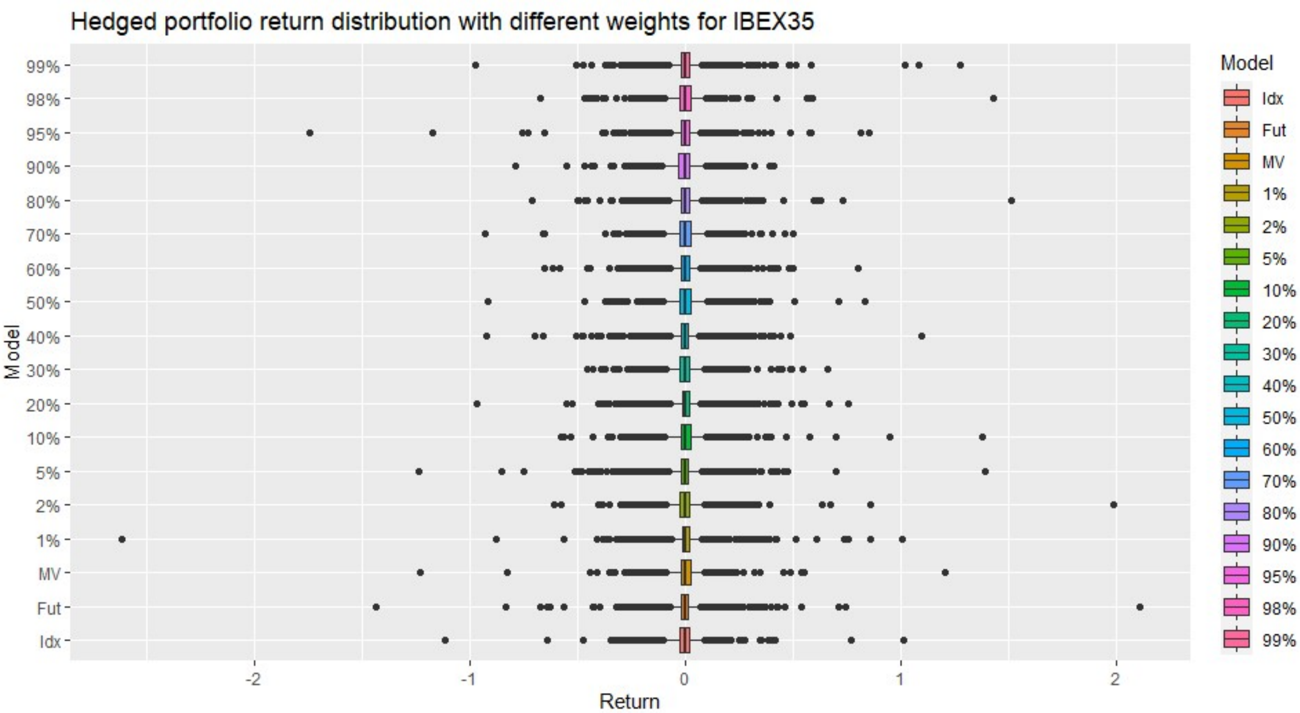
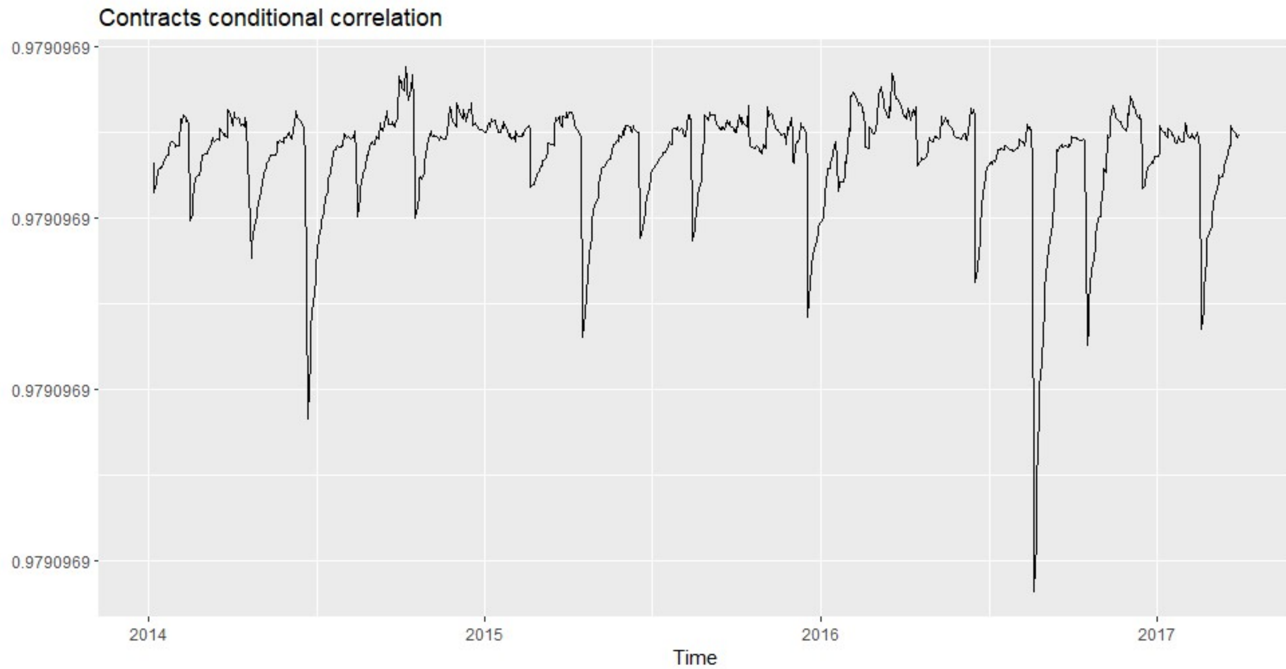


Figure 1.20: Box-plot of the return distribution of Indice Bursatil Espanol, its front month future, OLS Mean-Variance hedged portfolio, and the conditional quantile regression hedged portfolios.

	Left Tail					Right Tail					Student-t Scale	height
k	γ_H	γ_M	γ_{Adj}	ρ_H	k	γ_h	γ_M	γ_{Adj}	ρ_H	σ	σ_{Err}	
MV	81	0.51391	0.42039	0.51284	-0.22023	83	0.43353	0.33825	0.43243	-0.16017	0.00814	0.00060
1%	81	0.48337	0.31083	0.48320	-0.55369	83	0.36643	0.32715	0.36636	-0.29534	0.00952	0.00040
2%	81	0.48694	0.29060	0.48679	-0.59028	83	0.38721	0.26292	0.38720	-0.80073	0.01002	0.00042
5%	68	0.33052	0.25497	0.33032	-0.39529	83	0.50401	0.41037	0.50279	-0.16673	0.14688	0.00606
10%	65	0.45973	0.43369	0.45973	-0.64913	83	0.30966	0.30113	0.30966	-0.91837	0.08252	0.00333
20%	49	0.43714	0.45323	0.43714	-1.19088	83	0.32432	0.27260	0.32427	-0.42174	0.17778	0.00720
30%	69	0.32710	0.26598	0.32631	-0.16984	83	0.50390	0.41004	0.50274	-0.17496	0.13861	0.00572
40%	72	0.32933	0.27647	0.32870	-0.13283	83	0.48752	0.42972	0.48717	-0.23007	0.09753	0.00405
50%	72	0.32962	0.27606	0.32905	-0.14828	83	0.48716	0.43014	0.48684	-0.23962	0.09701	0.00403
60%	80	0.45400	0.36978	0.45400	-1.39308	83	0.43451	0.41270	0.43452	-0.65927	0.01500	0.00064
70%	79	0.41503	0.35085	0.41473	-0.30374	83	0.48763	0.37331	0.48750	-0.52961	0.02124	0.00091
80%	79	0.42508	0.31439	0.42508	-1.20518	83	0.49483	0.36832	0.49478	-0.65726	0.02261	0.00097
90%	76	0.34599	0.27921	0.34595	-0.59708	83	0.48463	0.42334	0.48384	-0.12235	0.05951	0.00250
95%	80	0.49556	0.34374	0.49556	-4.99902	83	0.42159	0.37878	0.42147	-0.29288	0.01189	0.00050
98%	69	0.48543	0.40035	0.48359	-0.07955	83	0.31192	0.28769	0.31193	-0.78779	0.06178	0.00250
99%	78	0.43285	0.39930	0.43245	-0.29289	83	0.39538	0.05664	0.39531	-0.68375	0.02282	0.00093

Table 1.21: Shows different Tail Index estimators value and the Student-t scale parameter over the testing set for the model of the IBEX35 against its front month future



Parameters

	Estimate	Std. Error	t value	Pr(>|t|)
μ^{Idx}	0.00053	0.000505	1.05057	0.293456
θ^{Idx}	-0.018625	0.033752	-0.551822	0.58107
ω^{Idx}	5e-06	5e-06	0.949049	0.342595
α^{Idx}	0.050839	0.010199	4.984524	1e-06
β^{Idx}	0.928463	0.011459	81.028066	0
μ^{Fut}	0.000533	0.000516	1.031835	0.30215
θ^{Fut}	-0.035627	0.033402	-1.066602	0.286151
ω^{Fut}	5e-06	3e-06	1.881205	0.059944
α^{Fut}	0.049039	0.006103	8.034975	0
β^{Fut}	0.929951	0.008384	110.921165	0
α^{Cor}	0	0	5.4e-05	0.999957
β^{Cor}	0.888575	0.028712	30.947567	0

Figure 1.21: GARCH-DCC Results for iBOV Index and Future correlation.

	$\hat{\beta}$	$\hat{\sigma}_\beta$	5%	95%	t statistic	Total Loss
1%	-0.04253	0.00001	-0.04255	-0.04250	-3542.63180	4.86100
2%	0.00675	0.00001	0.00672	0.00678	502.39325	4.63030
5%	0.06927	0.00003	0.06920	0.06933	2491.40995	4.34262
10%	0.13533	0.00002	0.13528	0.13538	6797.08166	4.04247
20%	0.18708	0.00001	0.18705	0.18711	15260.68074	3.81921
30%	0.17104	0.00001	0.17101	0.17107	14214.77688	3.91682
40%	0.15370	0.00001	0.15367	0.15373	13197.39234	4.02164
50%	0.18616	0.00001	0.18613	0.18618	15969.03928	3.88706
60%	0.22783	0.00001	0.22781	0.22786	19299.98444	3.70663
70%	0.24594	0.00001	0.24592	0.24597	21376.59129	3.63896
80%	0.25363	0.00001	0.25360	0.25367	16974.14609	3.62113
90%	0.27138	0.00001	0.27135	0.27141	19034.19576	3.55380
95%	0.28600	0.00003	0.28594	0.28607	10340.45681	3.49188
98%	0.32435	0.00001	0.32432	0.32439	22439.73432	3.30999
99%	0.37586	0.00001	0.37583	0.37588	35881.68915	3.06260
	$\hat{\beta}$	$\hat{\sigma}_\beta$	5%	95%	t statistic	Total Loss
1%	-0.04253	0.00004	-0.04262	-0.04243	-1002.49610	4.72437
2%	0.00675	0.00002	0.00670	0.00680	331.59403	4.50208
5%	0.06927	0.00004	0.06918	0.06936	1821.14623	4.22584
10%	0.13533	0.00002	0.13528	0.13538	6270.49430	3.93880
20%	0.18708	0.00001	0.18705	0.18711	16255.63836	3.73160
30%	0.17104	0.00001	0.17101	0.17106	16108.42608	3.83846
40%	0.15370	0.00001	0.15368	0.15372	16818.29211	3.95272
50%	0.18616	0.00001	0.18613	0.18618	16233.29153	3.83086
60%	0.22783	0.00001	0.22781	0.22785	26911.05825	3.66219
70%	0.24594	0.00001	0.24592	0.24597	23980.86719	3.60454
80%	0.25363	0.00001	0.25361	0.25366	22882.77559	3.59636
90%	0.27138	0.00001	0.27135	0.27141	22323.64394	3.53812
95%	0.28600	0.00003	0.28594	0.28606	11287.66841	3.48008
98%	0.32435	0.00001	0.32432	0.32439	22517.45246	3.29890
99%	0.37586	0.00001	0.37583	0.37588	33089.92081	3.04763

Table 1.22: Shows $\hat{\beta}$ inference results over the testing and training set for the model for the Indice Bovespa against its front month future, $n = 801$.

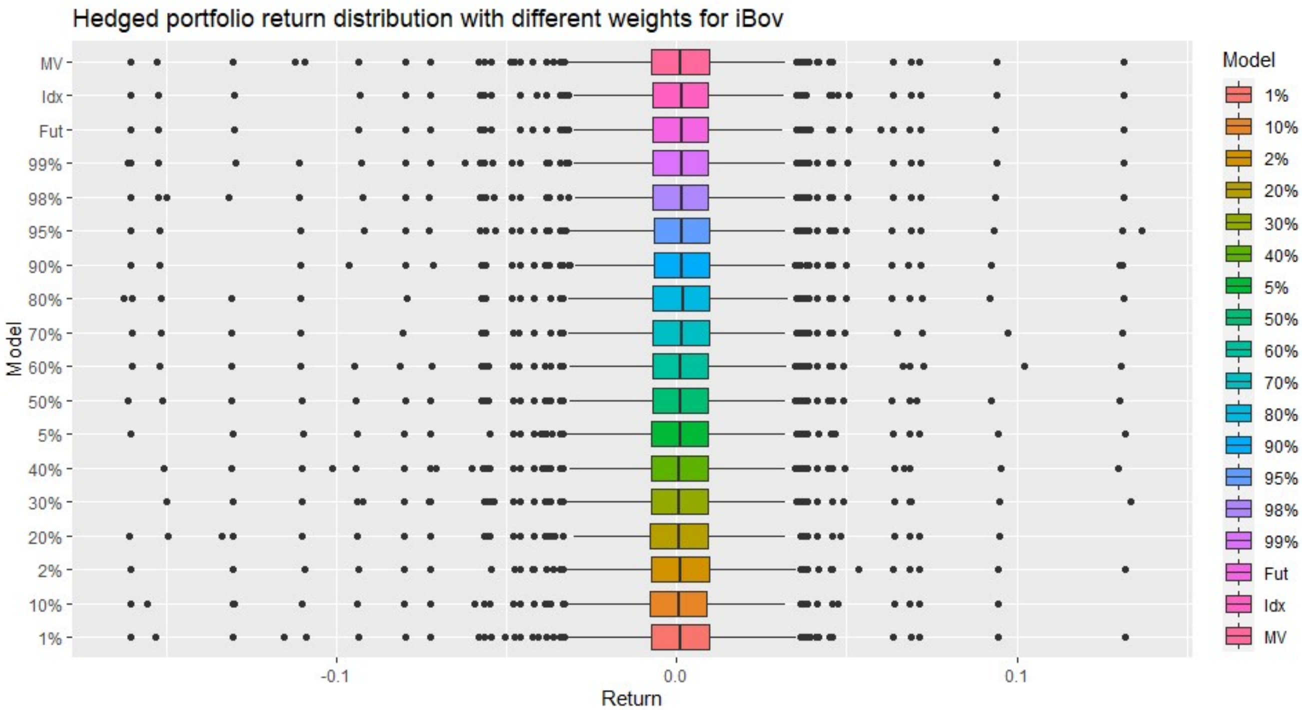
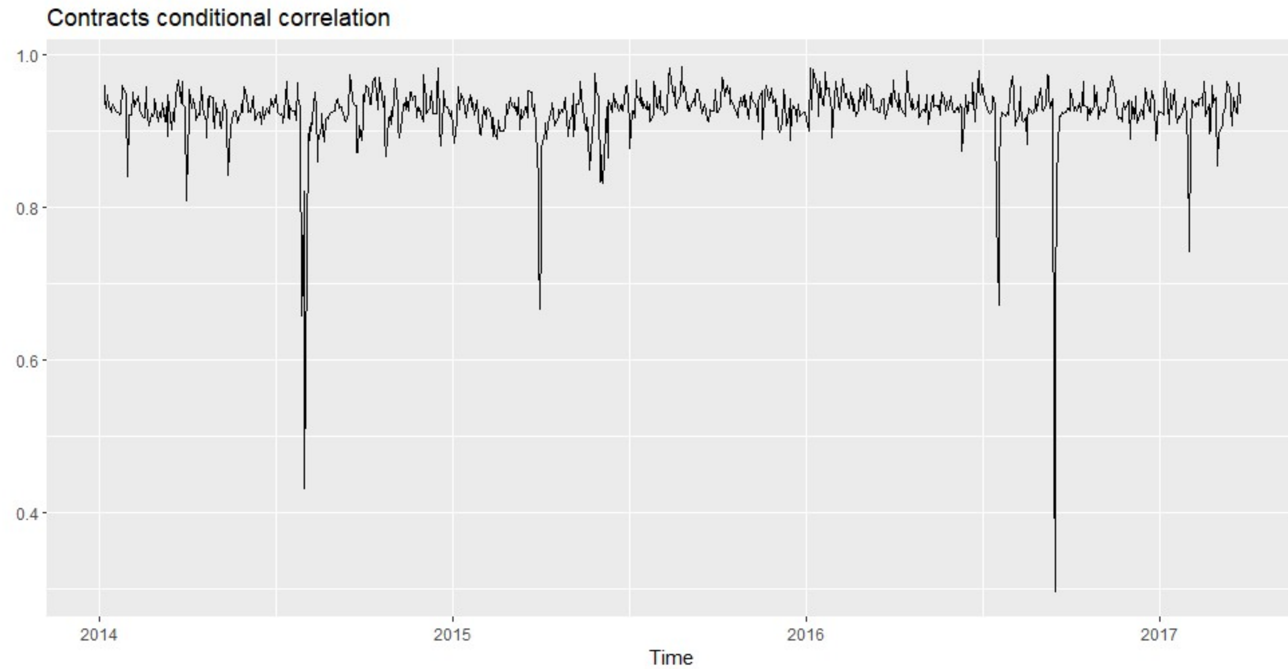


Figure 1.22: Box-plot of the return distribution of Indice Bovespa, its front month future, OLS Mean-Variance hedged portfolio, and the conditional quantile regression hedged portfolios.

	Left Tail					Right Tail					Student-t Scale	
	k	γ_H	γ_M	γ_{Adj}	ρ_H	k	γ_h	γ_M	γ_{Adj}	ρ_H	σ	σ_{Err}
MV	77	0.53040	0.54717	0.53040	-1.27498	80	0.42100	0.37286	0.39449	-0.00630	0.01039	0.00044
1%	77	0.57531	0.48933	0.57478	-0.34994	80	0.42886	0.34907	0.42880	-0.57525	0.01040	0.00043
2%	77	0.57338	0.49397	0.57244	-0.24409	80	0.43004	0.34884	0.42997	-0.59002	0.01042	0.00044
5%	77	0.56589	0.50455	0.56354	-0.03153	80	0.42847	0.35267	0.42837	-0.50930	0.01043	0.00044
10%	77	0.56345	0.50924	0.56230	-0.14607	80	0.42882	0.35314	0.42873	-0.51631	0.01042	0.00044
20%	77	0.55919	0.51459	0.55876	-0.29251	80	0.42160	0.36594	0.42055	-0.15409	0.01042	0.00044
30%	77	0.56047	0.51298	0.55989	-0.24805	80	0.42320	0.36306	0.42260	-0.23849	0.01042	0.00044
40%	77	0.56189	0.51120	0.56108	-0.19904	80	0.42657	0.35720	0.42639	-0.40650	0.01042	0.00044
50%	77	0.55926	0.51450	0.55883	-0.28998	80	0.42164	0.36586	0.42060	-0.15643	0.01042	0.00044
60%	77	0.55609	0.51848	0.55590	-0.40162	80	0.41987	0.36947	0.41696	-0.05165	0.01041	0.00044
70%	77	0.55478	0.52013	0.55464	-0.44830	80	0.41915	0.37092	0.40049	-0.00879	0.01041	0.00044
80%	77	0.55423	0.52081	0.55412	-0.46778	80	0.41887	0.37147	0.39564	-0.00704	0.01041	0.00044
90%	77	0.55300	0.52236	0.55291	-0.51196	80	0.41822	0.37271	0.41481	-0.04342	0.01042	0.00044
95%	77	0.55201	0.52360	0.55194	-0.54756	80	0.41770	0.37371	0.41587	-0.07322	0.01042	0.00044
98%	77	0.56212	0.51490	0.56171	-0.26399	80	0.41639	0.37626	0.41569	-0.15036	0.01042	0.00044
99%	77	0.55681	0.52028	0.55670	-0.42286	80	0.41506	0.37901	0.41471	-0.23429	0.01042	0.00044

Table 1.23: Shows different Tail Index estimators value and the Student-t scale parameter over the testing set for the model of the iBov against its front month future



Parameters

	Estimate	Std. Error	t value	Pr(>|t|)
μ^{Idx}	0.000105	0.000229	0.456987	0.64768
θ^{Idx}	0.065943	0.039908	1.652359	0.098461
ω^{Idx}	1e-06	6e-06	0.132493	0.894595
α^{Idx}	0.0726	0.109865	0.660809	0.508735
β^{Idx}	0.91412	0.112558	8.121341	0
μ^{Fut}	5.8e-05	0.000312	0.185954	0.852481
θ^{Fut}	-0.00932	0.040508	-0.230082	0.818028
ω^{Fut}	1e-06	7e-06	0.099455	0.920777
α^{Fut}	0.05404	0.086947	0.621535	0.534247
β^{Fut}	0.937616	0.089498	10.476444	0
α^{Cor}	0.20414	0.037279	5.475972	0
β^{Cor}	0.249858	0.073427	3.402819	0.000667

Figure 1.23: GARCH-DCC Results for MASCI Index and Future correlation.

	$\hat{\beta}$	$\hat{\sigma}_\beta$	5%	95%	t statistic	Total Loss
1%	-0.20333	0.00001	-0.20335	-0.20330	-19539.54260	2.84496
2%	-0.18859	0.00001	-0.18860	-0.18857	-23392.15571	2.80665
5%	-0.15500	0.00002	-0.15504	-0.15496	-8216.75621	2.71911
10%	-0.14237	0.00001	-0.14240	-0.14234	-10452.61790	2.68414
20%	-0.12909	0.00001	-0.12912	-0.12907	-11899.92615	2.64489
30%	-0.17779	0.00001	-0.17780	-0.17777	-27481.16446	2.76260
40%	-0.34351	0.00001	-0.34353	-0.34350	-46637.05747	3.17807
50%	-0.31303	0.00001	-0.31304	-0.31302	-55890.82336	3.09388
60%	-0.26768	0.00001	-0.26769	-0.26766	-39841.13839	2.97248
70%	-0.15388	0.00001	-0.15390	-0.15386	-20294.13060	2.67887
80%	-0.06320	0.00001	-0.06321	-0.06318	-7625.82006	2.44699
90%	0.01074	0.00001	0.01072	0.01076	1052.57770	2.25965
95%	0.04566	0.00002	0.04561	0.04571	2137.45908	2.17426
98%	0.07756	0.00001	0.07754	0.07759	7315.12162	2.09763
99%	0.08442	0.00001	0.08440	0.08444	9080.51841	2.08104

	$\hat{\beta}$	$\hat{\sigma}_\beta$	5%	95%	t statistic	Total Loss
1%	-0.20333	0.00003	-0.20339	-0.20327	-7541.58156	3.45243
2%	-0.18859	0.00002	-0.18862	-0.18855	-12129.95646	3.40720
5%	-0.15500	0.00003	-0.15508	-0.15492	-4680.68033	3.30332
10%	-0.14237	0.00002	-0.14241	-0.14233	-8307.64953	3.25845
20%	-0.12909	0.00001	-0.12911	-0.12907	-15561.81367	3.20434
30%	-0.17779	0.00001	-0.17781	-0.17777	-17970.09060	3.33115
40%	-0.34351	0.00001	-0.34353	-0.34350	-44047.10417	3.79832
50%	-0.31303	0.00001	-0.31305	-0.31301	-43588.46377	3.69159
60%	-0.26768	0.00001	-0.26770	-0.26766	-30469.84575	3.54289
70%	-0.15388	0.00001	-0.15390	-0.15386	-15702.19392	3.19973
80%	-0.06320	0.00001	-0.06322	-0.06317	-6356.86523	2.92693
90%	0.01074	0.00001	0.01071	0.01077	852.32287	2.70584
95%	0.04566	0.00002	0.04562	0.04570	2800.99069	2.60499
98%	0.07756	0.00001	0.07755	0.07758	9333.03635	2.51540
99%	0.08442	0.00001	0.08439	0.08444	7622.72586	2.49575

Table 1.24: Shows $\hat{\beta}$ inference results over the testing and training set for the model for the MSCI Singapore against its front month future, $n = 831$.

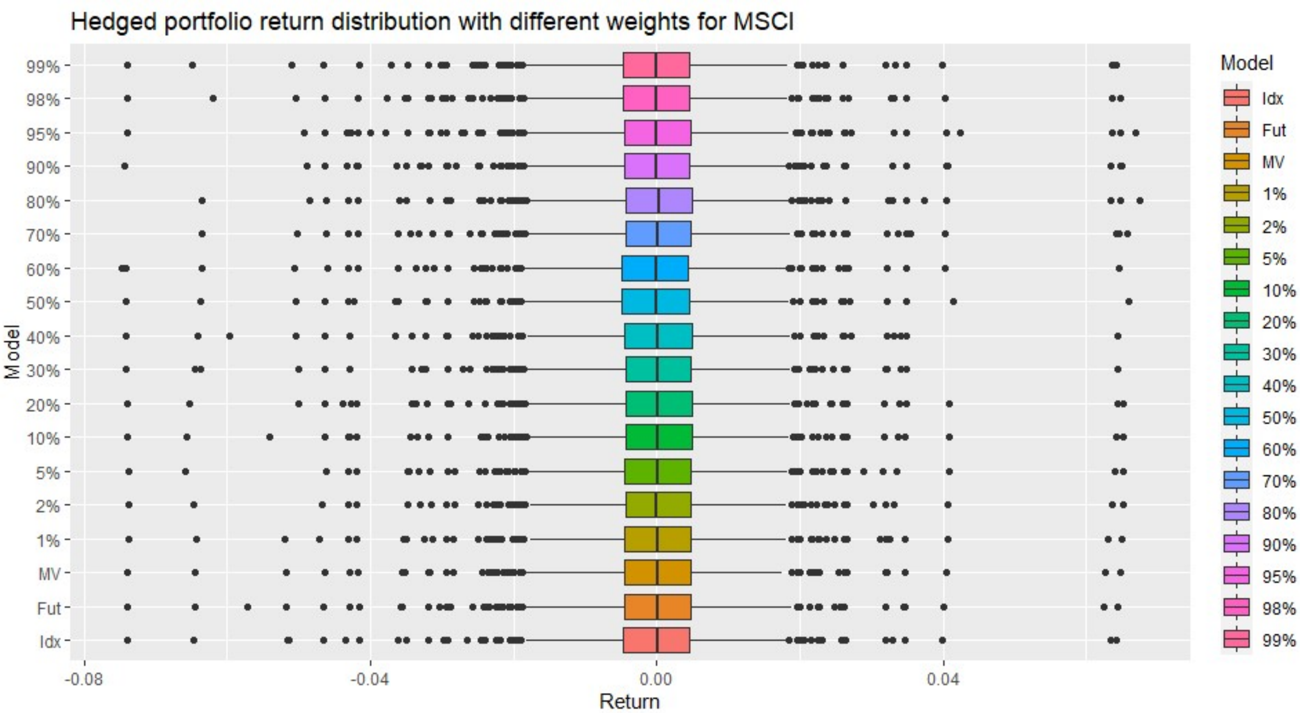
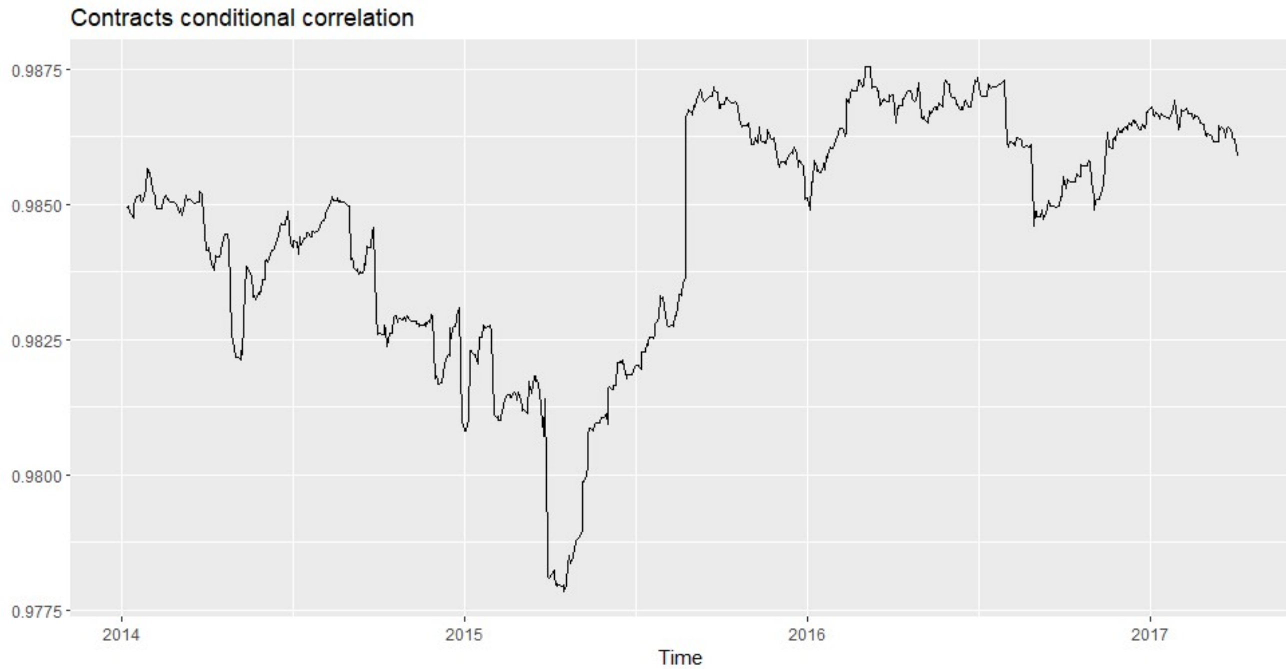


Figure 1.24: Box-plot of the return distribution of MASCI Singapore, its front month future, OLS Mean-Variance hedged portfolio, and the conditional quantile regression hedged portfolios.

k	Left Tail					Right Tail					Student-t Scale	height
	γ_H	γ_M	γ_{Adj}	ρ_H	k	γ_h	γ_M	γ_{Adj}	ρ_H	σ	σ_{Err}	
MV	82	0.53984	0.28779	0.53967	-0.59376	83	0.41346	0.31213	0.41346	-2.53456	0.00637	0.00027
1%	82	0.51989	0.32170	0.51908	-0.36383	83	0.43044	0.29136	0.43044	-1.21172	0.00677	0.00049
2%	82	0.51219	0.33485	0.51047	-0.25418	83	0.43281	0.28477	0.43281	-1.40464	0.00677	0.00049
5%	82	0.52412	0.31583	0.52350	-0.41370	83	0.43788	0.27068	0.43788	-1.83424	0.00676	0.00049
10%	82	0.52427	0.31593	0.52366	-0.41322	83	0.43967	0.26577	0.43967	-1.98822	0.00589	0.00025
20%	82	0.52441	0.31618	0.52382	-0.41110	83	0.43967	0.26462	0.43967	-2.07529	0.00674	0.00046
30%	82	0.51850	0.32470	0.51756	-0.34140	83	0.43449	0.28007	0.43449	-1.54556	0.00676	0.00048
40%	82	0.50780	0.33670	0.50579	-0.21425	83	0.43228	0.30737	0.43193	-0.39068	0.00611	0.00026
50%	82	0.51518	0.32508	0.51416	-0.32309	83	0.42864	0.30892	0.42842	-0.43809	0.00609	0.00026
60%	82	0.52365	0.31247	0.52313	-0.43149	83	0.41980	0.31938	0.41947	-0.36356	0.00606	0.00026
70%	82	0.52414	0.31583	0.52351	-0.41372	83	0.43804	0.27024	0.43804	-1.84808	0.00676	0.00048
80%	82	0.52555	0.31534	0.52499	-0.41930	83	0.42133	0.29752	0.42133	-1.62783	0.00605	0.00026
90%	82	0.52009	0.32487	0.51914	-0.34635	83	0.41655	0.30424	0.41655	-1.69766	0.00614	0.00026
95%	82	0.52783	0.31219	0.52733	-0.44571	83	0.41231	0.31182	0.41231	-1.55636	0.00611	0.00027
98%	82	0.52854	0.31108	0.52799	-0.45464	83	0.41158	0.31298	0.41158	-1.58959	0.00610	0.00026
99%	82	0.52868	0.31089	0.52815	-0.45607	83	0.41069	0.31464	0.41069	-1.54671	0.00609	0.00026

Table 1.25: Shows different Tail Index estimators value and the Student-t scale parameter over the testing set for the model of the MSCI against its front month future



Parameters

	Estimate	Std. Error	t value	Pr(>|t|)
μ^{Idx}	0.000663	0.000333	1.989589	0.046636
θ^{Idx}	0.092195	0.032901	2.802191	0.005076
ω^{Idx}	5e-06	0	24.766278	0
α^{Idx}	0.043809	0.001961	22.344498	0
β^{Idx}	0.895202	0.007723	115.912393	0
μ^{Fut}	0.000684	0.000323	2.116904	0.034268
θ^{Fut}	0.064409	0.03603	1.787646	0.073833
ω^{Fut}	5e-06	0	21.685935	0
α^{Fut}	0.045277	0.002068	21.888864	0
β^{Fut}	0.894623	0.008011	111.678296	0
α^{Cor}	0.005074	0.003442	1.47413	0.140447
β^{Cor}	0.993915	0.007257	136.9585	0

Figure 1.25: GARCH-DCC Results for NFTY Index and Future correlation.

	$\hat{\beta}$	$\hat{\sigma}_{\beta}$	5%	95%	t statistic	Total Loss
1%	-0.62070	0.00001	-0.62074	-0.62067	-46658.54633	4.04904
2%	-0.59306	0.00002	-0.59310	-0.59301	-30930.04945	3.98576
5%	-0.58466	0.00004	-0.58476	-0.58457	-14392.75660	3.98338
10%	-0.57036	0.00002	-0.57042	-0.57031	-25304.23040	3.97818
20%	-0.55035	0.00002	-0.55039	-0.55031	-30053.79227	3.98831
30%	-0.54127	0.00001	-0.54129	-0.54125	-55584.39485	4.02559
40%	-0.54735	0.00001	-0.54737	-0.54733	-59919.15944	4.10256
50%	-0.56892	0.00001	-0.56895	-0.56889	-47043.22400	4.22191
60%	-0.54225	0.00001	-0.54228	-0.54223	-45622.93413	4.21058
70%	-0.52599	0.00001	-0.52601	-0.52596	-48342.66315	4.22611
80%	-0.51590	0.00002	-0.51594	-0.51586	-31946.46515	4.25781
90%	-0.49422	0.00002	-0.49426	-0.49417	-24656.24020	4.25554
95%	-0.49409	0.00002	-0.49415	-0.49404	-22269.39682	4.28464
98%	-0.48785	0.00001	-0.48787	-0.48783	-53216.40729	4.28425
99%	-0.48665	0.00001	-0.48666	-0.48663	-74424.15025	4.28664

	$\hat{\beta}$	$\hat{\sigma}_{\beta}$	5%	95%	t statistic	Total Loss
1%	-0.62070	0.00005	-0.62082	-0.62059	-12508.33226	4.79917
2%	-0.59306	0.00003	-0.59312	-0.59299	-22106.75691	4.71877
5%	-0.58466	0.00005	-0.58478	-0.58454	-11263.66579	4.69999
10%	-0.57036	0.00002	-0.57041	-0.57031	-27288.20587	4.66760
20%	-0.55035	0.00002	-0.55039	-0.55031	-33653.93982	4.62808
30%	-0.54127	0.00001	-0.54130	-0.54124	-42644.09876	4.62107
40%	-0.54735	0.00001	-0.54738	-0.54732	-45474.91928	4.65979
50%	-0.56892	0.00001	-0.56894	-0.56890	-61062.27262	4.74585
60%	-0.54225	0.00001	-0.54228	-0.54223	-44360.48600	4.68493
70%	-0.52599	0.00001	-0.52601	-0.52596	-53678.17858	4.65527
80%	-0.51590	0.00001	-0.51593	-0.51587	-43264.24610	4.64426
90%	-0.49422	0.00002	-0.49426	-0.49417	-26619.71266	4.59708
95%	-0.49409	0.00005	-0.49420	-0.49399	-10704.19693	4.60654
98%	-0.48785	0.00002	-0.48789	-0.48780	-23617.02199	4.59303
99%	-0.48665	0.00004	-0.48674	-0.48655	-12291.71326	4.59126

Table 1.26: Shows $\hat{\beta}$ inference results over the testing and training set for the model of the NIFTY 50 Index against its front month future, $n = 799$.

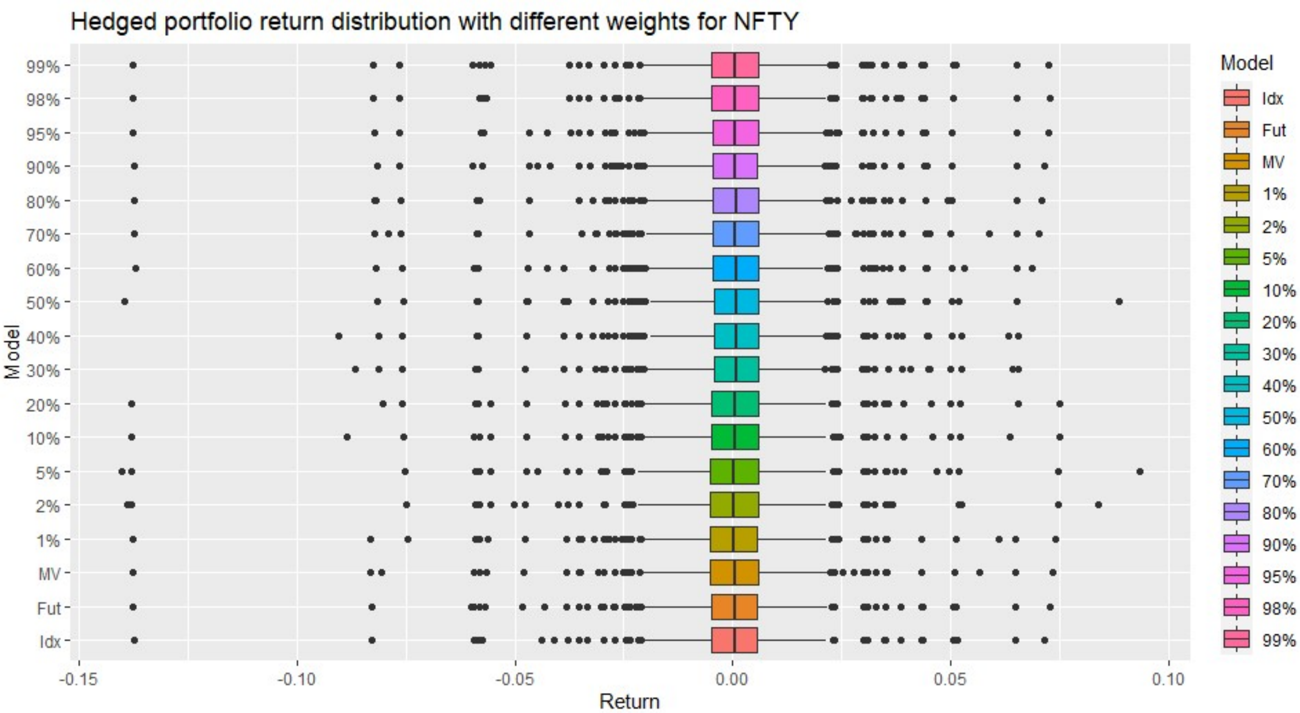
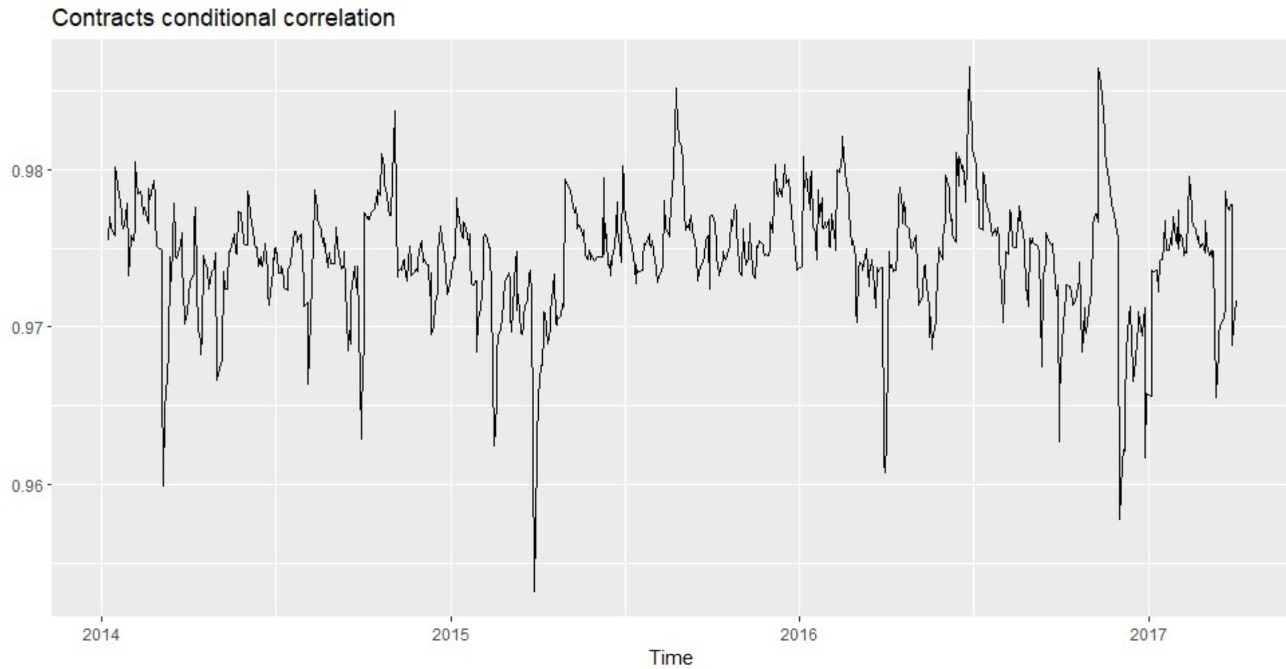


Figure 1.26: Box-plot of the return distribution of NIFTY 50 index, its front month future, OLS Mean-Variance hedged portfolio, and the conditional quantile regression hedged portfolios.

	Left Tail					Right Tail					Student-t Scale	
	k	γ_H	γ_M	γ_{Adj}	ρ_H	k	γ_H	γ_M	γ_{Adj}	ρ_H	σ	σ_{Err}
MV	77	0.56317	0.53732	0.56316	-0.65776	80	0.51792	0.41306	0.51520	-0.18893	0.00643	0.00027
1%	77	0.48863	0.53129	0.48863	-2.25378	80	0.51937	0.40772	0.51834	-0.29817	0.00672	0.00028
2%	77	0.49773	0.53254	0.49773	-1.90091	80	0.50772	0.42420	0.50751	-0.44532	0.00666	0.00028
5%	77	0.49832	0.53448	0.49832	-1.90506	80	0.50261	0.43002	0.50251	-0.50714	0.00665	0.00028
10%	77	0.50446	0.53342	0.50446	-1.68611	80	0.49651	0.43667	0.49647	-0.58161	0.00659	0.00027
20%	77	0.51472	0.53054	0.51473	-1.34628	80	0.49657	0.43565	0.49652	-0.56517	0.00668	0.00028
30%	77	0.51780	0.53028	0.51780	-1.26518	80	0.49548	0.43688	0.49544	-0.57777	0.00666	0.00028
40%	77	0.51569	0.53061	0.51569	-1.32252	80	0.49584	0.43652	0.49580	-0.57501	0.00667	0.00028
50%	77	0.50531	0.53312	0.50531	-1.65524	80	0.49845	0.43416	0.49839	-0.55135	0.00659	0.00027
60%	77	0.51746	0.53033	0.51746	-1.27418	80	0.49518	0.43725	0.49514	-0.58265	0.00666	0.00028
70%	77	0.51874	0.53321	0.51874	-1.29401	80	0.49008	0.44358	0.49007	-0.65838	0.00664	0.00028
80%	77	0.51443	0.53926	0.51443	-1.48943	80	0.48902	0.44459	0.48901	-0.67037	0.00663	0.00028
90%	77	0.51039	0.54722	0.51039	-1.68391	80	0.48681	0.44695	0.48681	-0.69887	0.00661	0.00027
95%	77	0.51047	0.54718	0.51047	-1.68140	80	0.48680	0.44697	0.48680	-0.69903	0.00661	0.00027
98%	77	0.51274	0.54634	0.51274	-1.61759	80	0.48622	0.44764	0.48622	-0.70721	0.00661	0.00027
99%	77	0.51247	0.54675	0.51247	-1.62869	80	0.48611	0.44778	0.48611	-0.70885	0.00661	0.00027

Table 1.27: Shows different Tail Index estimators value and the Student-t scale parameter over the testing set for the model of the NFTY against its front month future



Parameters

	Estimate	Std. Error	t value	Pr(>|t|)
μ^{Idx}	0.000729	0.000382	1.908899	0.056275
θ^{Idx}	-0.06217	0.040383	-1.539524	0.123676
ω^{Idx}	9e-06	4e-06	2.546807	0.010871
α^{Idx}	0.157488	0.029678	5.306485	0
β^{Idx}	0.804119	0.034651	23.206207	0
μ^{Fut}	0.000816	0.000379	2.151526	0.031435
θ^{Fut}	-0.071395	0.039688	-1.798902	0.072034
ω^{Fut}	1e-05	4e-06	2.282169	0.022479
α^{Fut}	0.173232	0.033622	5.152295	0
β^{Fut}	0.788088	0.041814	18.847307	0
α^{Cor}	0.0366	0.017876	2.047467	0.040612
β^{Cor}	0.793627	0.098709	8.040106	0

Figure 1.27: GARCH-DCC Results for NKK Index and Future correlation.

	$\hat{\beta}$	$\hat{\sigma}_\beta$	5%	95%	t statistic	Total Loss
1%	-0.17044	0.00002	-0.17049	-0.17040	-8472.63986	4.51483
2%	-0.15077	0.00002	-0.15081	-0.15073	-8340.29428	4.43950
5%	-0.09726	0.00004	-0.09735	-0.09716	-2363.36349	4.23504
10%	-0.00902	0.00003	-0.00908	-0.00895	-333.49859	3.89761
20%	0.10705	0.00001	0.10702	0.10709	7214.81467	3.46016
30%	0.29871	0.00001	0.29869	0.29873	35589.31927	2.73016
40%	0.48246	0.00001	0.48245	0.48247	93498.13550	2.05112
50%	0.48791	0.00000	0.48790	0.48792	127068.78252	2.04001
60%	0.27680	0.00001	0.27678	0.27681	52027.38915	2.85270
70%	0.22603	0.00001	0.22600	0.22605	24995.22300	3.06576
80%	0.17377	0.00001	0.17375	0.17379	17624.52382	3.28742
90%	0.17497	0.00002	0.17493	0.17501	10303.65715	3.29714
95%	0.13649	0.00002	0.13643	0.13655	5507.94681	3.45942
98%	0.11692	0.00001	0.11690	0.11695	10522.53461	3.54300
99%	0.12495	0.00002	0.12491	0.12498	8042.21725	3.51213
	$\hat{\beta}$	$\hat{\sigma}_\beta$	5%	95%	t statistic	Total Loss
1%	-0.17044	0.00003	-0.17051	-0.17037	-5685.50308	3.69642
2%	-0.15077	0.00002	-0.15081	-0.15073	-8986.08963	3.63567
5%	-0.09726	0.00004	-0.09736	-0.09716	-2308.04724	3.47083
10%	-0.00902	0.00002	-0.00907	-0.00896	-374.37206	3.19910
20%	0.10705	0.00001	0.10704	0.10707	12605.37360	2.85150
30%	0.29871	0.00001	0.29870	0.29873	49099.69142	2.27187
40%	0.48246	0.00000	0.48245	0.48247	103490.73894	1.73334
50%	0.48791	0.00000	0.48790	0.48792	139282.18967	1.72588
60%	0.27680	0.00001	0.27678	0.27681	45978.16502	2.37276
70%	0.22603	0.00001	0.22601	0.22604	43932.02142	2.54341
80%	0.17377	0.00001	0.17374	0.17379	16139.58210	2.72108
90%	0.17497	0.00001	0.17494	0.17499	15460.29736	2.73002
95%	0.13649	0.00002	0.13643	0.13655	5497.06774	2.85995
98%	0.11692	0.00001	0.11690	0.11695	10218.07314	2.92703
99%	0.12495	0.00001	0.12492	0.12498	9553.21597	2.90249

Table 1.28: Shows $\hat{\beta}$ inference results over the testing and training set for the model of the NIKKEI 225 Index against its front month future, $n = 793$.

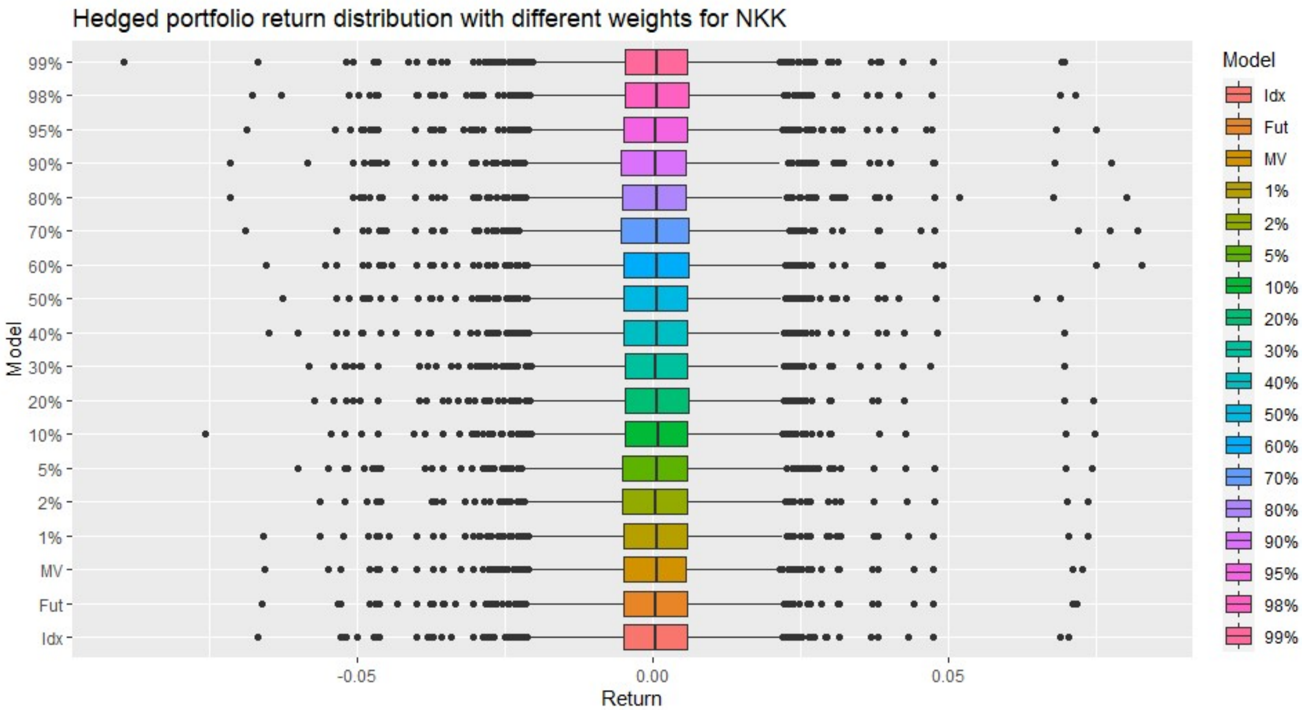
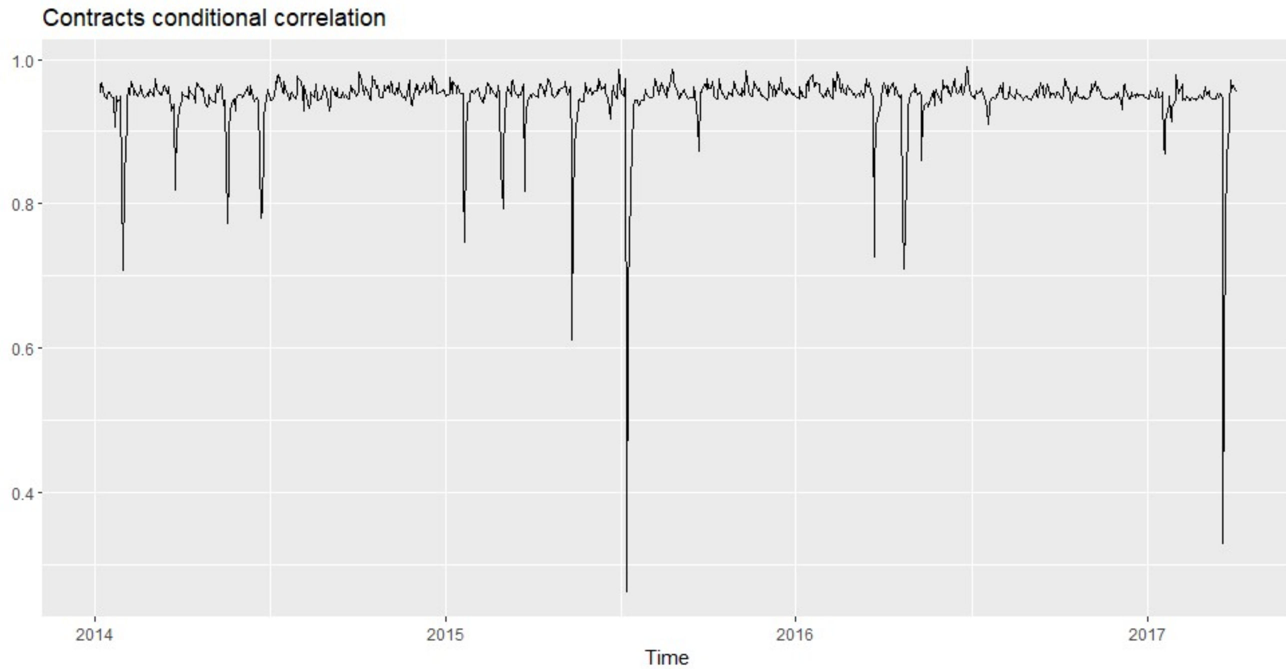


Figure 1.28: Box-plot of the return distribution of NIKKEI 225 index, its front month future, OLS Mean-Variance hedged portfolio, and the conditional quantile regression hedged portfolios.

	Left Tail					Right Tail					Student-t Scale	
	k	γ_H	γ_M	γ_{Adj}	ρ_H	k	γ_h	γ_M	γ_{Adj}	ρ_H	σ	σ_{Err}
MV	78	0.54290	0.23067	0.53917	-0.20815	79	0.40675	0.29209	0.40028	-0.04888	0.00878	
1%	78	0.52886	0.10908	0.52676	-0.32025	79	0.45786	0.26171	0.45759	-0.50241	0.00713	0.00033
2%	78	0.52716	0.11896	0.52494	-0.30901	79	0.45902	0.25539	0.45883	-0.54497	0.00837	
5%	78	0.52205	0.14842	0.51892	-0.25993	79	0.46032	0.24212	0.46022	-0.63895	0.00870	
10%	78	0.55093	0.11281	0.54955	-0.38784	79	0.46774	0.20912	0.46772	-0.89975	0.00867	
20%	78	0.56281	0.11314	0.56166	-0.41605	79	0.45612	0.22285	0.45610	-0.84885	0.00865	
30%	78	0.57663	0.10906	0.57588	-0.47750	79	0.43919	0.26237	0.43883	-0.45184	0.00865	
40%	78	0.55517	0.17571	0.55336	-0.33809	79	0.42411	0.28664	0.42168	-0.15885	0.00868	
50%	78	0.55505	0.17649	0.55323	-0.33672	79	0.42389	0.28670	0.42153	-0.15751	0.00868	
60%	78	0.57835	0.10243	0.57764	-0.48721	79	0.44185	0.25529	0.44164	-0.53257	0.00866	
70%	78	0.57781	0.09714	0.57707	-0.48569	79	0.44518	0.24582	0.44508	-0.64620	0.00865	
80%	78	0.56920	0.11007	0.56820	-0.44066	79	0.44930	0.23623	0.44925	-0.74311	0.00865	
90%	78	0.56931	0.11002	0.56831	-0.44109	79	0.44915	0.23655	0.44911	-0.74023	0.00865	
95%	78	0.56566	0.11217	0.56453	-0.42568	79	0.45325	0.22805	0.45323	-0.81171	0.00865	
98%	78	0.56565	0.10891	0.56460	-0.42819	79	0.45514	0.22459	0.45512	-0.83676	0.00865	
99%	78	0.56578	0.10995	0.56470	-0.42790	79	0.45436	0.22598	0.45434	-0.82706	0.00865	

Table 1.29: Shows different Tail Index estimators value and the Student-t scale parameter over the testing set for the model of the NKK against its front month future



Parameters

	Estimate	Std. Error	t value	Pr(>|t|)
μ^{Idx}	-0.000258	0.000475	-0.543697	0.58665
θ^{Idx}	0.146775	0.036692	4.000189	6.3e-05
ω^{Idx}	1.2e-05	2e-06	7.390161	0
α^{Idx}	0.152055	0.020273	7.500428	0
β^{Idx}	0.788538	0.021693	36.350378	0
μ^{Fut}	-0.000305	0.000478	-0.638877	0.522903
θ^{Fut}	0.103558	0.015595	6.640423	0
ω^{Fut}	1.4e-05	3e-06	5.019861	1e-06
α^{Fut}	0.136252	0.016327	8.345063	0
β^{Fut}	0.795343	0.026343	30.192042	0
α^{Cor}	0.162387	0.150511	1.078909	0.280628
β^{Cor}	0.377837	0.119909	3.151034	0.001627

Figure 1.29: GARCH-DCC Results for PSI Index and Future correlation.

	$\hat{\beta}$	$\hat{\sigma}_{\beta}$	5%	95%	t statistic	Total Loss
1%	-0.76104	0.00003	-0.76111	-0.76096	-24230.30881	7.58944
2%	-0.70901	0.00002	-0.70907	-0.70896	-29350.34915	7.35493
5%	-0.64551	0.00005	-0.64564	-0.64539	-12078.71612	7.06090
10%	-0.60107	0.00004	-0.60115	-0.60098	-16525.23311	6.84206
20%	-0.56993	0.00002	-0.56998	-0.56987	-22996.65475	6.65913
30%	-0.56484	0.00002	-0.56487	-0.56480	-36543.65259	6.59037
40%	-0.56104	0.00002	-0.56108	-0.56100	-31131.26633	6.52746
50%	-0.54539	0.00001	-0.54542	-0.54537	-43375.48016	6.41451
60%	-0.52301	0.00001	-0.52304	-0.52298	-35344.61807	6.27419
70%	-0.51329	0.00002	-0.51333	-0.51324	-26082.19344	6.18822
80%	-0.38348	0.00002	-0.38352	-0.38344	-22429.08745	5.61052
90%	-0.37859	0.00002	-0.37864	-0.37853	-17132.10142	5.54979
95%	-0.36479	0.00003	-0.36486	-0.36472	-12015.47150	5.47377
98%	-0.29740	0.00001	-0.29743	-0.29737	-20986.07360	5.19080
99%	-0.25357	0.00001	-0.25360	-0.25354	-20070.02793	5.01096

	$\hat{\beta}$	$\hat{\sigma}_{\beta}$	5%	95%	t statistic	Total Loss
1%	-0.76104	0.00002	-0.76110	-0.76098	-30459.01458	5.20142
2%	-0.70901	0.00002	-0.70906	-0.70897	-35389.81052	5.04053
5%	-0.64551	0.00004	-0.64560	-0.64542	-16186.67351	4.84156
10%	-0.60107	0.00003	-0.60113	-0.60100	-20907.52486	4.69800
20%	-0.56993	0.00001	-0.56995	-0.56990	-47214.53310	4.58690
30%	-0.56484	0.00001	-0.56486	-0.56481	-55089.93891	4.55501
40%	-0.56104	0.00001	-0.56106	-0.56102	-58061.52326	4.52711
50%	-0.54539	0.00001	-0.54541	-0.54538	-67226.37526	4.46370
60%	-0.52301	0.00001	-0.52303	-0.52299	-50583.47206	4.38040
70%	-0.51329	0.00001	-0.51332	-0.51326	-41066.67961	4.33529
80%	-0.38348	0.00001	-0.38350	-0.38345	-37249.52524	3.93341
90%	-0.37859	0.00002	-0.37862	-0.37855	-23466.55771	3.90407
95%	-0.36479	0.00003	-0.36486	-0.36472	-12689.23984	3.85593
98%	-0.29740	0.00001	-0.29742	-0.29738	-31141.29477	3.65376
99%	-0.25357	0.00001	-0.25360	-0.25354	-19907.81497	3.52468

Table 1.30: Shows $\hat{\beta}$ inference results over the testing and training set for the model of the Portuguese Stock Index against its front month future, $n = 827$.

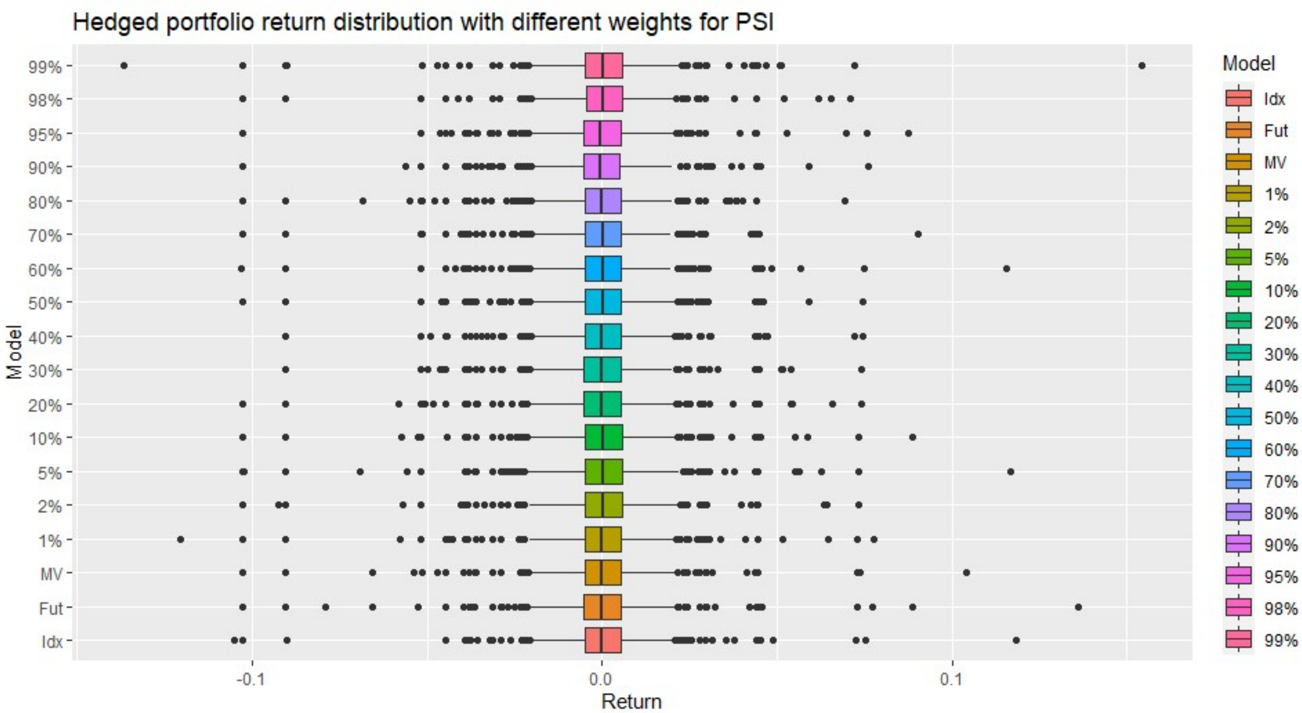
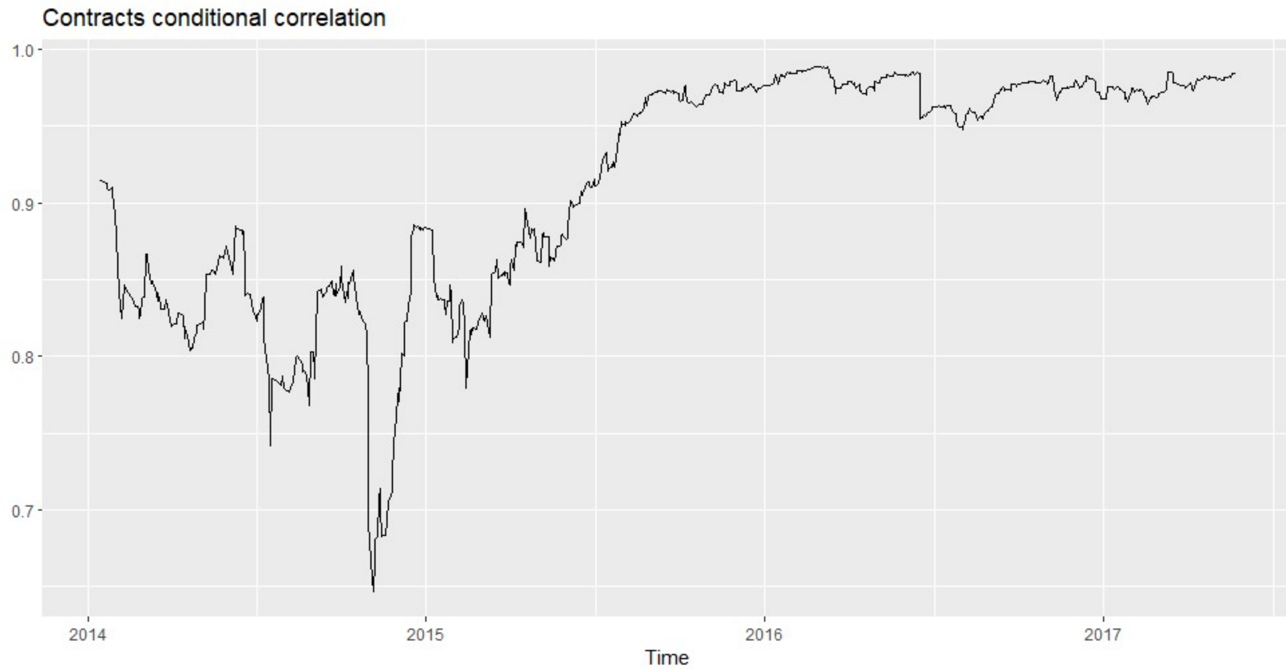


Figure 1.30: Box-plot of the return distribution of Portuguese Stock index, its front month future, OLS Mean-Variance hedged portfolio, and the conditional quantile regression hedged portfolios.

	Left Tail					Right Tail					Student-t Scale	
	k	γ_H	γ_M	γ_{Adj}	ρ_H	k	γ_h	γ_M	γ_{Adj}	ρ_H	σ	σ_{Err}
MV	81	0.49262	0.34159	0.49262	-2.99056	83	0.44838	0.22756	0.44837	-0.90185	0.00654	0.00027
1%	80	0.56812	0.49611	0.56781	-0.38726	83	0.69689	0.57435	0.68635	-0.06786	0.00608	0.00028
2%	80	0.56172	0.42210	0.56024	-0.27308	83	0.60410	0.53272	0.60349	-0.33385	0.00667	0.00030
5%	80	0.53176	0.39889	0.53110	-0.36395	83	0.58953	0.42093	0.58757	-0.25492	0.00677	0.00030
10%	80	0.50711	0.41214	0.50372	-0.04304	83	0.55941	0.37920	0.55563	-0.18275	0.00676	0.00030
20%	81	0.48989	0.42754	0.48958	-0.28254	83	0.54417	0.34845	0.54044	-0.17663	0.00673	0.00029
30%	81	0.48630	0.43171	0.48616	-0.37078	83	0.54088	0.34577	0.53664	-0.15876	0.00664	0.00029
40%	81	0.48368	0.43470	0.48359	-0.43506	83	0.53850	0.34361	0.53363	-0.14756	0.00664	0.00029
50%	81	0.47923	0.44018	0.47922	-0.56081	83	0.53108	0.33295	0.52484	-0.11699	0.00668	0.00029
60%	81	0.49188	0.41739	0.49082	-0.19147	83	0.53091	0.30146	0.52723	-0.18698	0.00664	0.00028
70%	81	0.49413	0.40922	0.49200	-0.07005	83	0.55192	0.24541	0.55120	-0.43736	0.00666	0.00029
80%	81	0.47898	0.38183	0.47891	-0.65784	83	0.49333	0.16962	0.49321	-0.65007	0.00665	0.00028
90%	81	0.47784	0.38251	0.47778	-0.66862	83	0.48302	0.18920	0.48279	-0.56870	0.00827	0.00012
95%	81	0.47464	0.38492	0.47459	-0.68745	83	0.48796	0.16650	0.48788	-0.69018	0.00657	0.00028
98%	81	0.48157	0.37196	0.48157	-1.73011	83	0.47016	0.15060	0.47016	-1.60601	0.00812	0.00010
99%	81	0.47934	0.37845	0.47934	-1.84822	83	0.45986	0.16979	0.45986	-1.78789	0.00659	0.00027

Table 1.31: Shows different Tail Index estimators value and the Student-t scale parameter over the testing set for the model of the PSI against its front month future



Parameters

	Estimate	Std. Error	t value	Pr(>|t|)
μ^{Idx}	0.000191	0.000627	0.30476	0.760549
θ^{Idx}	0.041029	0.037744	1.087029	0.277024
ω^{Idx}	1e-05	7e-06	1.496812	0.134442
α^{Idx}	0.079444	0.016174	4.911946	1e-06
β^{Idx}	0.900267	0.022081	40.770864	0
μ^{Fut}	0.000239	0.000797	0.299917	0.76424
θ^{Fut}	-0.009867	0.037511	-0.26304	0.79252
ω^{Fut}	9e-06	0.000117	0.077191	0.938472
α^{Fut}	0.100677	0.193572	0.520101	0.602993
β^{Fut}	0.885836	0.174506	5.076257	0
α^{Cor}	0.040999	0.008162	5.023387	1e-06
β^{Cor}	0.958056	0.011951	80.165215	0

Figure 1.31: GARCH-DCC Results for RTS Index and Future correlation.

	$\hat{\beta}$	$\hat{\sigma}_\beta$	5%	95%	t statistic	Total Loss
1%	-1.60642	0.00006	-1.60656	-1.60629	-28379.34269	16.68986
2%	-1.52434	0.00005	-1.52446	-1.52422	-29159.07301	16.14636
5%	-1.46895	0.00011	-1.46921	-1.46868	-12899.08081	15.76726
10%	-1.52410	0.00008	-1.52429	-1.52390	-18253.35747	16.09963
20%	-1.41087	0.00004	-1.41098	-1.41077	-31489.72662	15.30659
30%	-1.30683	0.00004	-1.30693	-1.30673	-30966.28927	14.57800
40%	-1.27467	0.00004	-1.27477	-1.27456	-28582.24862	14.31838
50%	-1.25960	0.00004	-1.25970	-1.25950	-29071.89727	14.17026
60%	-1.17224	0.00004	-1.17232	-1.17215	-32439.70615	13.55868
70%	-1.16296	0.00004	-1.16304	-1.16287	-31238.16468	13.45042
80%	-1.14853	0.00005	-1.14865	-1.14842	-22552.66765	13.30979
90%	-1.11995	0.00006	-1.12008	-1.11982	-19575.61574	13.07976
95%	-1.02573	0.00012	-1.02600	-1.02547	-8875.63891	12.45891
98%	-0.91197	0.00005	-0.91210	-0.91185	-17026.05444	11.72545
99%	-0.81428	0.00004	-0.81437	-0.81419	-21301.97644	11.10527
	$\hat{\beta}$	$\hat{\sigma}_\beta$	5%	95%	t statistic	Total Loss
1%	-1.60642	0.00010	-1.60667	-1.60618	-15384.84495	10.96509
2%	-1.52434	0.00006	-1.52447	-1.52421	-27537.40358	10.62098
5%	-1.46895	0.00010	-1.46918	-1.46871	-14457.95064	10.39535
10%	-1.52410	0.00005	-1.52422	-1.52398	-29045.00351	10.64342
20%	-1.41087	0.00003	-1.41094	-1.41081	-49593.45019	10.19173
30%	-1.30683	0.00003	-1.30689	-1.30677	-47621.71439	9.77676
40%	-1.27467	0.00002	-1.27472	-1.27461	-54138.88790	9.66674
50%	-1.25960	0.00003	-1.25966	-1.25954	-48625.18166	9.62904
60%	-1.17224	0.00003	-1.17230	-1.17217	-44548.28408	9.28160
70%	-1.16296	0.00002	-1.16301	-1.16290	-49451.18267	9.26731
80%	-1.14853	0.00002	-1.14859	-1.14848	-48466.34519	9.23073
90%	-1.11995	0.00004	-1.12004	-1.11986	-29957.94216	9.13280
95%	-1.02573	0.00007	-1.02591	-1.02556	-13993.26689	8.73875
98%	-0.91197	0.00003	-0.91205	-0.91189	-27791.75142	8.25482
99%	-0.81428	0.00003	-0.81436	-0.81420	-23626.15768	7.83575

Table 1.32: Shows $\hat{\beta}$ inference results over the testing and training set for the model of the Russian Trading System Index against its front month future, $n = 782$.

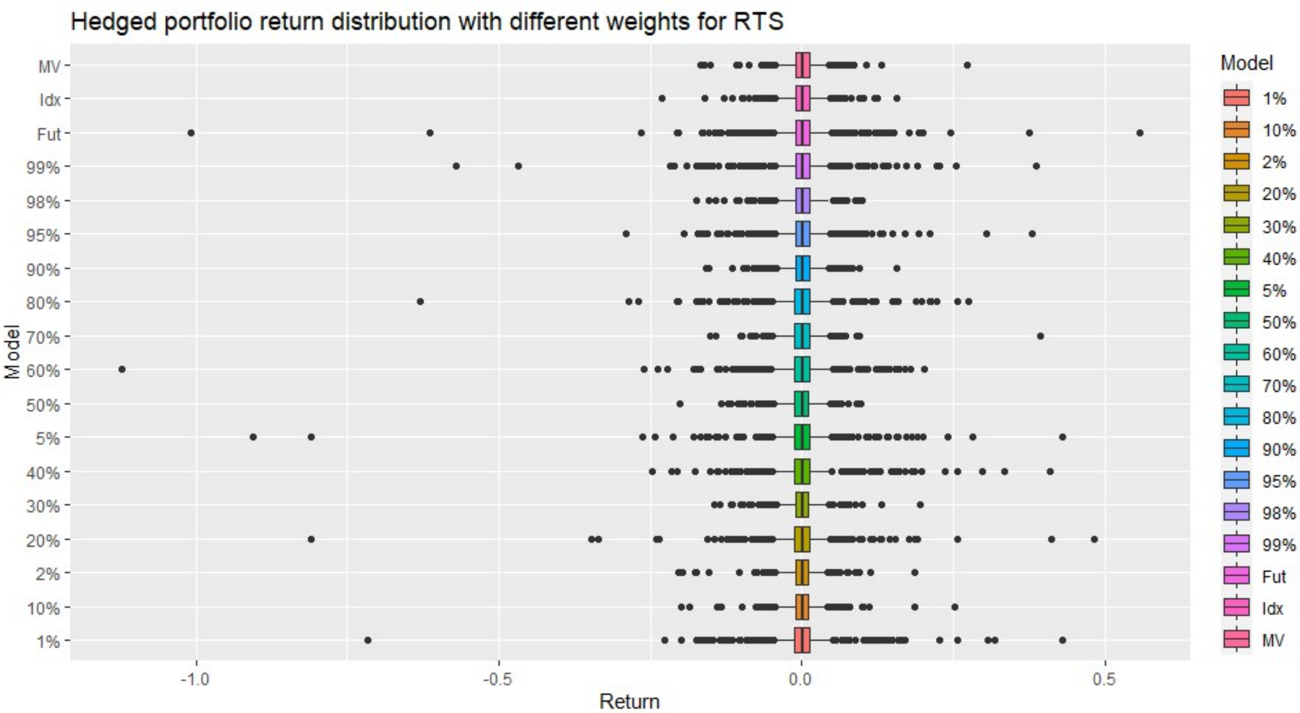
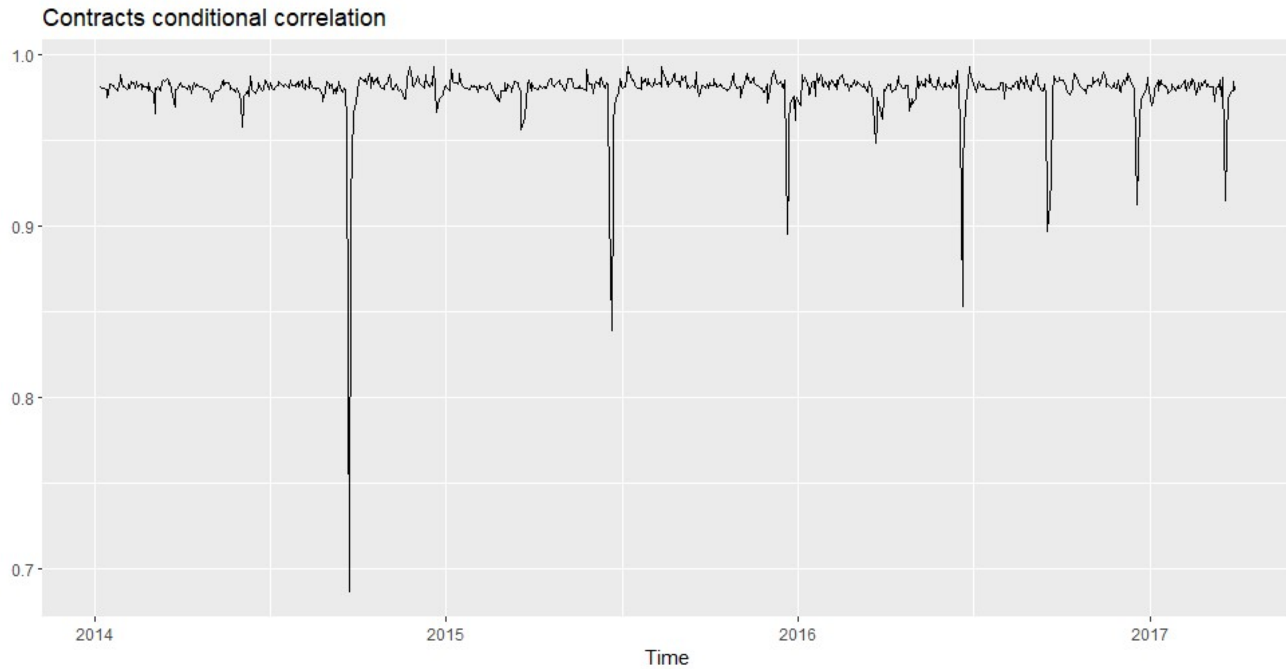


Figure 1.32: Box-plot of the return distribution of Russian Trading System index, its front month future, OLS Mean-Variance hedged portfolio, and the conditional quantile regression hedged portfolios.

	Left Tail					Right Tail					Student-t Scale	
	k	γ_H	γ_M	γ_{Adj}	ρ_H	k	γ_h	γ_M	γ_{Adj}	ρ_H	σ	σ_{Err}
MV	76	0.50481	0.52939	0.50481	-1.34166	78	0.40438	0.27801	0.39429	-0.02261	0.00982	0.00041
1%	75	0.54685	0.45905	0.54671	-0.46923	78	0.34563	0.20258	0.34562	-0.84047	0.01141	0.00049
2%	75	0.51664	0.49212	0.51664	-0.85089	78	0.33452	0.20757	0.33448	-0.65284	0.01174	0.00051
5%	75	0.52828	0.47951	0.52830	-0.71356	78	0.33422	0.19019	0.33419	-0.71479	0.01199	0.00052
10%	75	0.51668	0.49207	0.51669	-0.85032	78	0.33451	0.20751	0.33447	-0.65299	0.01174	0.00051
20%	75	0.53205	0.47113	0.53204	-0.64814	78	0.33667	0.15628	0.33666	-0.90065	0.01241	0.00054
30%	75	0.54039	0.45323	0.54032	-0.52644	78	0.35116	0.05218	0.35116	-1.54085	0.01364	0.00060
40%	75	0.52378	0.46716	0.52379	-0.68045	78	0.35818	0.01682	0.35818	-1.54612	0.01430	0.00063
50%	75	0.52261	0.46567	0.52260	-0.67375	78	0.36158	0.00622	0.36158	-1.47312	0.01671	
60%	75	0.49673	0.47817	0.49673	-0.85555	78	0.39709	-0.04578	0.39701	-0.70989	0.01831	0.00079
70%	75	0.49229	0.48593	0.49230	-0.94793	78	0.40008	-0.04675	0.39999	-0.70433	0.01922	0.00084
80%	74	0.50377	0.48123	0.50377	-0.80694	78	0.39388	-0.00182	0.39367	-0.57863	0.02013	0.00086
90%	73	0.49901	0.50063	0.49901	-1.01732	78	0.36993	0.14055	0.36055	-0.03387	0.02341	0.00099
95%	60	0.44301	0.58813	0.44301	-2.59685	78	0.38144	0.21514	0.37866	-0.13861	0.09165	0.00385
98%	75	0.43020	0.31476	0.42990	-0.38177	78	0.49685	0.45490	0.49684	-0.45034	0.02804	0.00118
99%	76	0.46455	0.41520	0.46449	-0.52435	78	0.46297	0.37190	0.46162	-0.18127	0.01595	0.00067

Table 1.33: Shows different Tail Index estimators value and the Student-t scale parameter over the testing set for the model of the RTS against its front month future



Parameters

	Estimate	Std. Error	t value	Pr(>|t|)
μ^{Idx}	0.000286	0.000322	0.88649	0.375354
θ^{Idx}	0.000769	0.040166	0.019148	0.984723
ω^{Idx}	3e-06	2e-06	1.127564	0.259504
α^{Idx}	0.096541	0.027323	3.533307	0.00041
β^{Idx}	0.881032	0.03236	27.225949	0
μ^{Fut}	0.000297	0.000318	0.933872	0.35037
θ^{Fut}	0.002225	0.036661	0.0607	0.951598
ω^{Fut}	3e-06	7e-06	0.467593	0.640076
α^{Fut}	0.10455	0.040847	2.559535	0.010481
β^{Fut}	0.871021	0.065442	13.309916	0
α^{Cor}	0.129507	0.06236	2.076761	0.037824
β^{Cor}	0.418793	0.170412	2.457529	0.01399

Figure 1.33: GARCH-DCC Results for SA Index and Future correlation.

	$\hat{\beta}$	$\hat{\sigma}_\beta$	5%	95%	t statistic	Total Loss
1%	-0.15844	0.00002	-0.15848	-0.15840	-9277.87030	3.69110
2%	-0.12246	0.00001	-0.12249	-0.12243	-9762.82958	3.57667
5%	-0.06752	0.00003	-0.06759	-0.06746	-2530.80741	3.40382
10%	-0.06040	0.00001	-0.06043	-0.06036	-4177.53954	3.38532
20%	-0.06256	0.00002	-0.06260	-0.06252	-3731.30645	3.40050
30%	-0.07732	0.00001	-0.07735	-0.07729	-6897.31244	3.45605
40%	-0.12229	0.00001	-0.12231	-0.12227	-14266.19541	3.60962
50%	-0.19994	0.00001	-0.19997	-0.19992	-18997.30028	3.87062
60%	-0.11215	0.00001	-0.11217	-0.11213	-13183.02733	3.59426
70%	-0.05482	0.00001	-0.05484	-0.05480	-6263.20261	3.41703
80%	0.00050	0.00001	0.00048	0.00052	63.47577	3.24614
90%	0.05696	0.00001	0.05693	0.05699	4165.30089	3.07085
95%	0.08076	0.00002	0.08071	0.08080	4044.15477	2.99737
98%	0.09881	0.00001	0.09879	0.09883	12479.74020	2.94113
99%	0.11317	0.00001	0.11315	0.11319	12795.06053	2.89556
	$\hat{\beta}$	$\hat{\sigma}_\beta$	5%	95%	t statistic	Total Loss
1%	-0.15844	0.00003	-0.15850	-0.15838	-6140.54601	4.03550
2%	-0.12246	0.00002	-0.12250	-0.12242	-6617.73423	3.90701
5%	-0.06752	0.00003	-0.06760	-0.06744	-1967.57883	3.71549
10%	-0.06040	0.00002	-0.06044	-0.06036	-3438.24455	3.69611
20%	-0.06256	0.00002	-0.06259	-0.06252	-4151.15917	3.71511
30%	-0.07732	0.00001	-0.07734	-0.07730	-10507.82970	3.77876
40%	-0.12229	0.00001	-0.12231	-0.12228	-16187.66195	3.95185
50%	-0.19994	0.00001	-0.19997	-0.19992	-19212.08219	4.24823
60%	-0.11215	0.00001	-0.11218	-0.11213	-10250.96313	3.93891
70%	-0.05482	0.00001	-0.05483	-0.05480	-7899.99765	3.74431
80%	0.00050	0.00001	0.00048	0.00053	43.71964	3.55974
90%	0.05696	0.00002	0.05692	0.05700	3626.64996	3.37494
95%	0.08076	0.00003	0.08068	0.08083	2473.79486	3.29955
98%	0.09881	0.00002	0.09877	0.09885	5539.47870	3.24209
99%	0.11317	0.00002	0.11312	0.11322	5324.01029	3.19539

Table 1.34: Shows $\hat{\beta}$ inference results over the testing and training set for the model of the Johannesburg Stock Exchange Index against its front month future, $n = 782$.

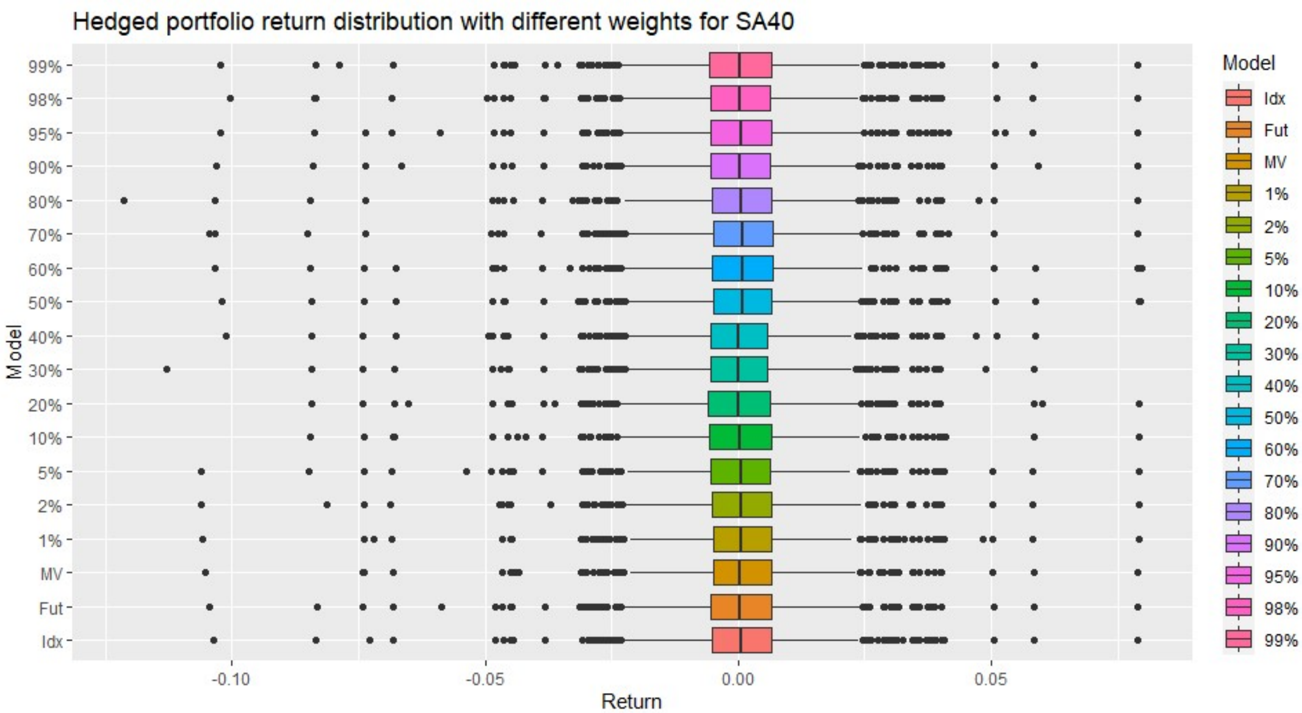
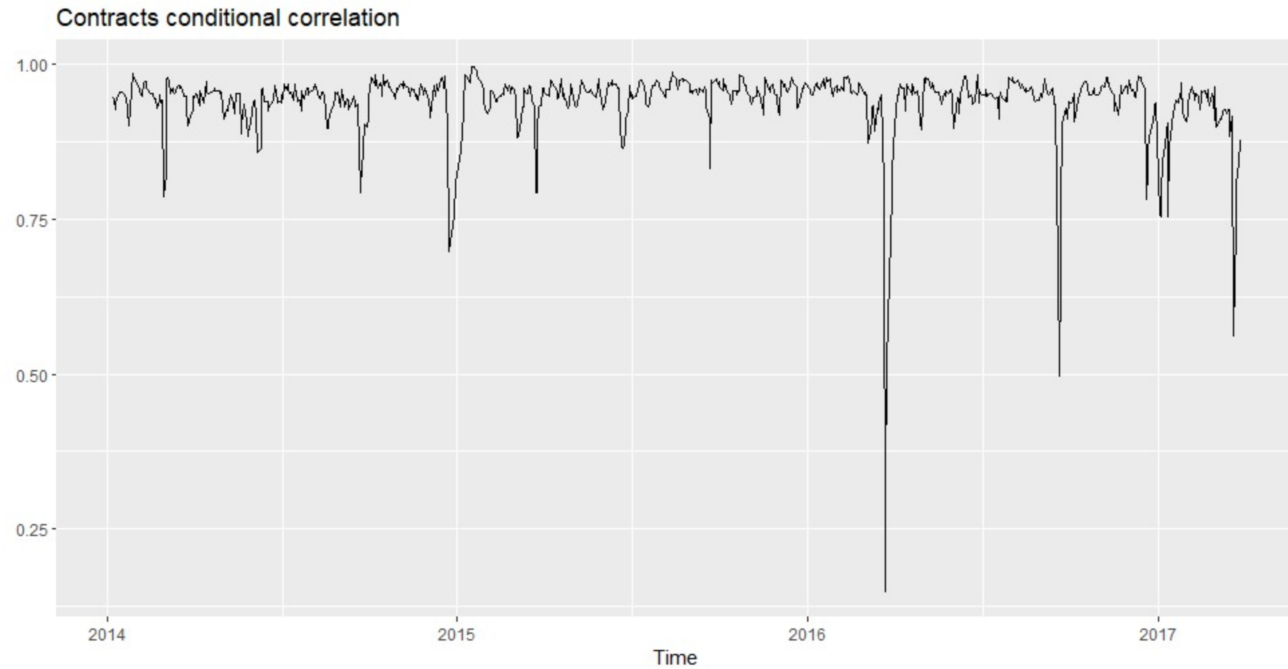


Figure 1.34: Box-plot of the return distribution of Johannesburg Stock Exchange index, its front month future, OLS Mean-Variance hedged portfolio, and the conditional quantile regression hedged portfolios.

	Left Tail					Right Tail					Student-t Scale	
	k	γ_H	γ_M	γ_{Adj}	ρ_H	k	γ_h	γ_M	γ_{Adj}	ρ_H	σ	σ_{Err}
MV	79	0.49831	0.36546	0.49811	-0.51179	81	0.47398	0.22425	0.46849	-0.09162	0.00847	0.00039
1%	80	0.46368	0.37997	0.46307	-0.29309	81	0.45474	0.29461	0.45076	-0.07738	0.00806	0.00036
2%	80	0.46169	0.38369	0.46085	-0.24808	81	0.45967	0.29147	0.45419	-0.04743	0.00792	0.00035
5%	80	0.46423	0.38402	0.46333	-0.27588	81	0.45916	0.29736	0.45575	-0.08814	0.00778	0.00035
10%	80	0.46378	0.38546	0.46272	-0.24450	81	0.45958	0.29763	0.45596	-0.09169	0.00778	0.00035
20%	80	0.46372	0.38531	0.46267	-0.24735	81	0.45941	0.29777	0.45613	-0.09231	0.00778	0.00035
30%	80	0.46529	0.38147	0.46469	-0.33088	81	0.45864	0.29682	0.45506	-0.08254	0.00779	0.00035
40%	80	0.46170	0.38369	0.46085	-0.24826	81	0.45912	0.29254	0.45402	-0.05505	0.00793	0.00035
50%	80	0.45860	0.38709	0.45682	-0.08843	81	0.44703	0.29581	0.44369	-0.10094	0.00927	
60%	80	0.46207	0.38372	0.46130	-0.25594	81	0.45645	0.29809	0.45312	-0.09314	0.00787	0.00035
70%	80	0.46393	0.38587	0.46285	-0.23604	81	0.47036	0.27721	0.46458	-0.04730	0.00778	0.00035
80%	80	0.47055	0.38364	0.46965	-0.27535	81	0.47139	0.27690	0.46415	-0.03735	0.00783	0.00035
90%	80	0.46798	0.39144	0.46631	-0.11509	81	0.47039	0.27951	0.45826	-0.01062	0.00772	0.00034
95%	80	0.46898	0.39186	0.46756	-0.10937	81	0.47209	0.27274	0.46617	-0.05423	0.00781	0.00035
98%	80	0.46751	0.39543	0.46551	-0.03464	81	0.46890	0.27685	0.46105	-0.02733	0.00782	0.00035
99%	80	0.46517	0.39992	0.46345	-0.06180	81	0.46641	0.28028	0.44381	-0.00376	0.00771	0.00034

Table 1.35: Shows different Tail Index estimators value and the Student-t scale parameter over the testing set for the model of the SA40 against its front month future



Parameters

	Estimate	Std. Error	t value	Pr(>|t|)
μ^{Idx}	0.000509	0.000323	1.574003	0.115487
θ^{Idx}	0.077935	0.051256	1.520497	0.128386
ω^{Idx}	4e-06	1.8e-05	0.2135	0.830937
α^{Idx}	0.192487	0.055328	3.479038	0.000503
β^{Idx}	0.788511	0.164978	4.779501	2e-06
μ^{Fut}	0.000325	0.000387	0.839751	0.401048
θ^{Fut}	0.030115	0.056696	0.531156	0.59531
ω^{Fut}	5e-06	4e-06	1.116343	0.264275
α^{Fut}	0.140113	0.047803	2.931054	0.003378
β^{Fut}	0.823848	0.062885	13.10078	0
α^{Cor}	0.214519	0.065571	3.27155	0.00107
β^{Cor}	0.568485	0.226955	2.504834	0.012251

Figure 1.35: GARCH-DCC Results for SMI Index and Future correlation.

	$\hat{\beta}$	$\hat{\sigma}_\beta$	5%	95%	t statistic	Total Loss
1%	-0.96283	0.00003	-0.96289	-0.96277	-38309.29701	5.74267
2%	-0.82860	0.00002	-0.82865	-0.82855	-36442.48932	5.34232
5%	-0.47240	0.00003	-0.47248	-0.47232	-13550.82701	4.28255
10%	-0.45761	0.00003	-0.45768	-0.45755	-16331.01899	4.24133
20%	-0.56293	0.00002	-0.56297	-0.56290	-37447.71799	4.56055
30%	-0.49635	0.00001	-0.49638	-0.49633	-39969.68934	4.36736
40%	-0.61678	0.00001	-0.61680	-0.61675	-53451.63685	4.73288
50%	-0.62870	0.00001	-0.62872	-0.62867	-53454.07083	4.77438
60%	-0.53188	0.00001	-0.53190	-0.53185	-52608.41528	4.49008
70%	-0.58473	0.00001	-0.58476	-0.58471	-62543.79746	4.65420
80%	-0.99898	0.00002	-0.99902	-0.99894	-54933.85635	5.90622
90%	-0.98171	0.00003	-0.98177	-0.98165	-36353.48353	5.86113
95%	-0.96228	0.00005	-0.96239	-0.96217	-20458.20011	5.80595
98%	-0.86120	0.00002	-0.86125	-0.86115	-39627.40978	5.50289
99%	-0.80128	0.00002	-0.80133	-0.80123	-37254.48564	5.32289

	$\hat{\beta}$	$\hat{\sigma}_\beta$	5%	95%	t statistic	Total Loss
1%	-0.96283	0.00004	-0.96292	-0.96275	-26543.43151	5.25288
2%	-0.82860	0.00003	-0.82866	-0.82854	-31999.61005	4.88446
5%	-0.47240	0.00005	-0.47251	-0.47229	-9889.45902	3.91059
10%	-0.45761	0.00002	-0.45766	-0.45757	-24050.02571	3.88272
20%	-0.56293	0.00001	-0.56296	-0.56290	-41353.32896	4.20038
30%	-0.49635	0.00001	-0.49638	-0.49633	-53808.71500	4.04106
40%	-0.61678	0.00001	-0.61680	-0.61676	-70543.62409	4.40659
50%	-0.62870	0.00001	-0.62872	-0.62867	-61761.37274	4.46858
60%	-0.53188	0.00001	-0.53190	-0.53185	-56518.50524	4.22026
70%	-0.58473	0.00001	-0.58476	-0.58471	-60886.60825	4.39877
80%	-0.99898	0.00002	-0.99903	-0.99893	-49219.08277	5.62975
90%	-0.98171	0.00003	-0.98178	-0.98164	-32909.00052	5.61413
95%	-0.96228	0.00004	-0.96236	-0.96220	-27242.61967	5.57433
98%	-0.86120	0.00001	-0.86124	-0.86117	-57760.02768	5.28695
99%	-0.80128	0.00001	-0.80131	-0.80124	-53884.77691	5.11405

Table 1.36: Shows $\hat{\beta}$ inference results over the testing and training set for the model of the Swiss Market Index against its front month future, $n = 814$.

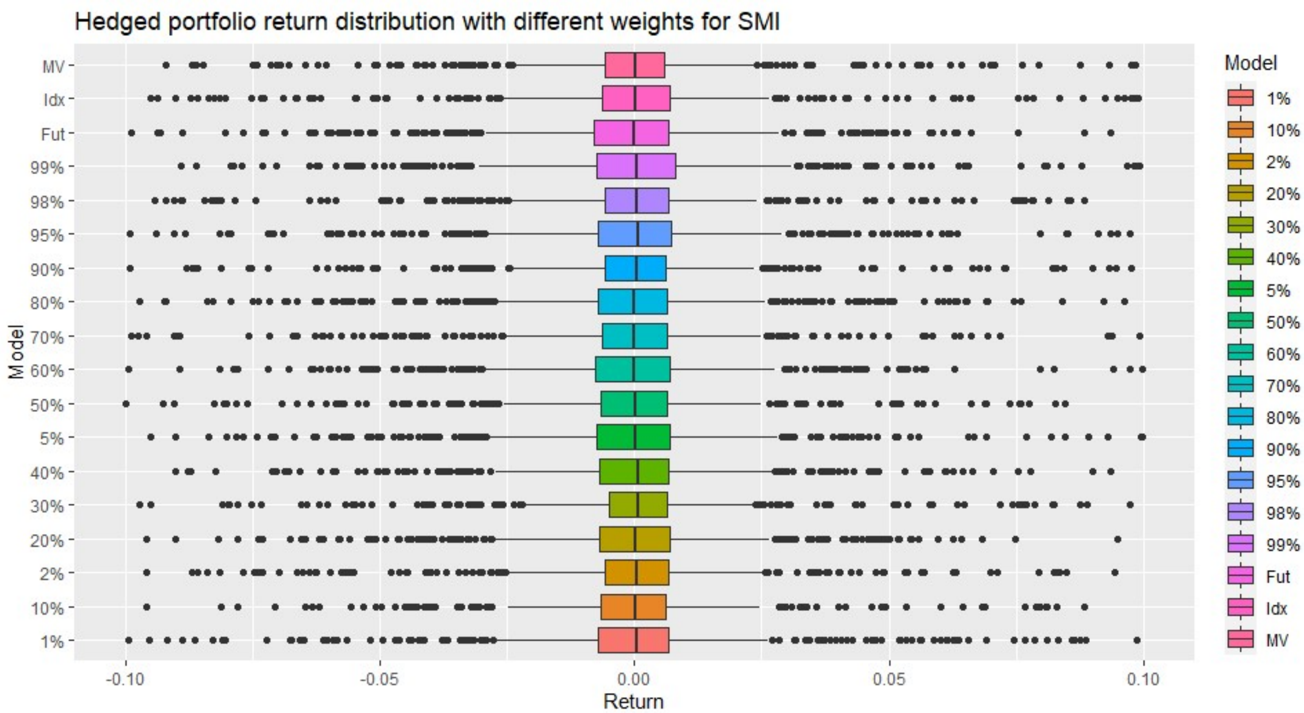
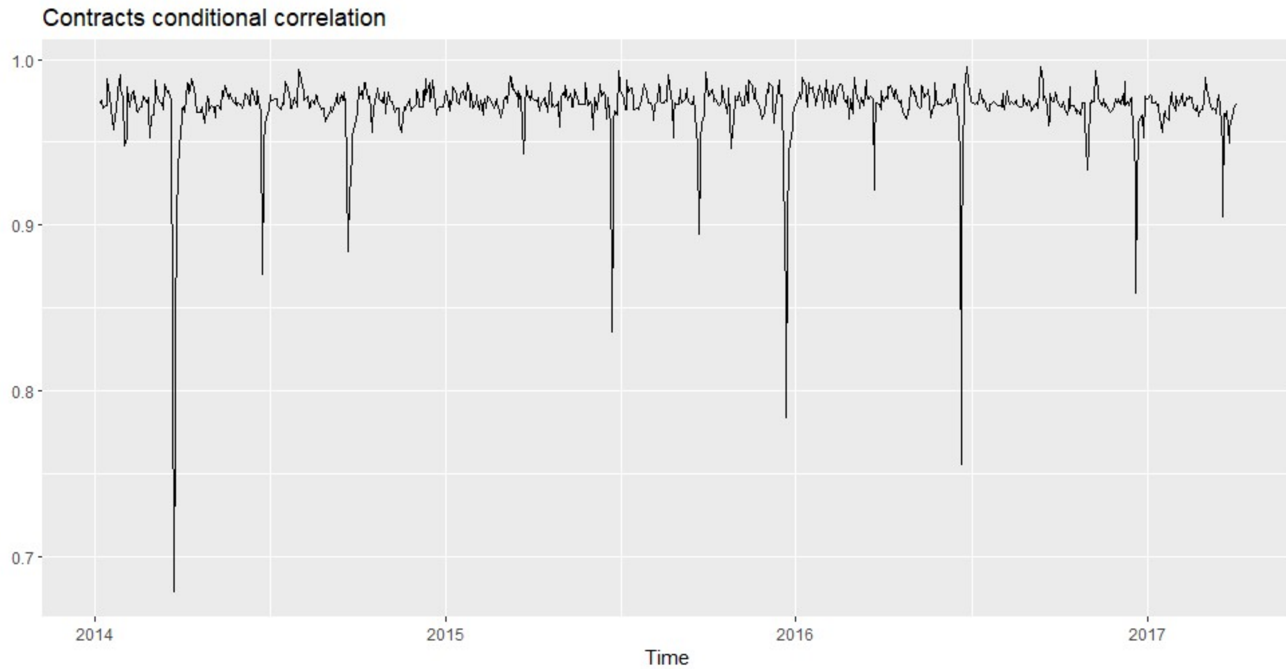


Figure 1.36: Box-plot of the return distribution of Swiss Market index, its front month future, OLS Mean-Variance hedged portfolio, and the conditional quantile regression hedged portfolios.

	Left Tail					Right Tail					Student-t Scale	
	k	γ_H	γ_M	γ_{Adj}	ρ_H	k	γ_h	γ_M	γ_{Adj}	ρ_H	σ	σ_{Err}
MV	80	0.53188	0.41885	0.53129	-0.35124	81	0.42706	0.11093	0.42706	-3.53606	0.00622	0.00027
1%	62	0.53358	0.54352	0.53357	-0.85081	81	0.69279	0.65322	0.69267	-0.39432	0.03734	0.00163
2%	77	0.55193	0.48846	0.55193	-4.62718	81	0.54715	0.51723	0.54715	-3.07405	0.00971	0.00042
5%	80	0.48395	0.46989	0.48395	-0.87767	81	0.45779	0.13140	0.45779	-1.59258	0.00717	0.00052
10%	80	0.47138	0.48002	0.47138	-1.07611	81	0.44839	0.14941	0.44839	-1.52323	0.00715	0.00055
20%	80	0.49459	0.48885	0.49459	-0.94682	81	0.47297	0.15866	0.47295	-0.92556	0.00660	0.00029
30%	80	0.48571	0.47547	0.48571	-0.90925	81	0.47537	0.08060	0.47537	-1.91493	0.00641	0.00028
40%	79	0.49902	0.49549	0.49902	-0.96336	81	0.45753	0.27210	0.45751	-0.81877	0.00682	0.00029
50%	79	0.49280	0.50431	0.49280	-1.12553	81	0.44239	0.31425	0.44233	-0.60479	0.00677	0.00029
60%	80	0.49590	0.47784	0.49590	-0.83905	81	0.49279	0.06347	0.49279	-1.78329	0.00657	0.00028
70%	80	0.49651	0.49212	0.49651	-0.95759	81	0.46615	0.20808	0.46613	-0.83056	0.00674	0.00029
80%	81	0.49971	0.55811	0.50037	-0.21116	81	0.67572	0.69186	0.67572	-1.23465	1.36379	0.05958
90%	43	0.51515	0.57435	0.51603	-0.27042	81	0.67915	0.68037	0.67915	-1.01791	0.07579	0.00330
95%	62	0.53230	0.54427	0.53229	-0.82020	81	0.69309	0.65228	0.69299	-0.37356	0.03680	0.00160
98%	76	0.53775	0.52126	0.53775	-2.77915	81	0.56905	0.55725	0.56905	-4.16885	0.01112	0.00048
99%	78	0.55477	0.48552	0.55476	-0.83801	81	0.49149	0.51728	0.47718	-0.01067	0.00888	0.00038

Table 1.37: Shows different Tail Index estimators value and the Student-t scale parameter over the testing set for the model of the SMI against its front month future



Parameters

	Estimate	Std. Error	t value	Pr(>|t|)
μ^{Idx}	0.000569	0.000214	2.65565	0.007916
θ^{Idx}	-0.073686	0.036482	-2.019825	0.043402
ω^{Idx}	6e-06	1e-06	5.023991	1e-06
α^{Idx}	0.200028	0.035614	5.616552	0
β^{Idx}	0.705482	0.046844	15.060225	0
μ^{Fut}	0.000628	0.000224	2.809407	0.004963
θ^{Fut}	-0.067315	0.044814	-1.50209	0.133074
ω^{Fut}	7e-06	1e-06	7.294025	0
α^{Fut}	0.255701	0.039007	6.555267	0
β^{Fut}	0.655077	0.044127	14.845279	0
α^{Cor}	0.181363	0.061236	2.96171	0.003059
β^{Cor}	0.382796	0.106314	3.600617	0.000317

Figure 1.37: GARCH-DCC Results for SP500 Index and Future correlation.

	$\hat{\beta}$	$\hat{\sigma}_\beta$	5%	95%	t statistic	Total Loss
1%	-0.42689	0.00002	-0.42693	-0.42686	-26112.10246	3.20930
2%	-0.42194	0.00001	-0.42197	-0.42190	-28848.14694	3.20188
5%	-0.38685	0.00003	-0.38692	-0.38678	-13281.70444	3.13434
10%	-0.37334	0.00002	-0.37339	-0.37330	-19747.69077	3.12157
20%	-0.32121	0.00001	-0.32125	-0.32118	-22098.82594	3.03801
30%	-0.31307	0.00001	-0.31309	-0.31305	-39158.14194	3.05279
40%	-0.30381	0.00001	-0.30383	-0.30380	-50372.15995	3.06455
50%	-0.28804	0.00001	-0.28805	-0.28802	-47638.76601	3.06057
60%	-0.27523	0.00001	-0.27525	-0.27521	-33054.30820	3.06291
70%	-0.21994	0.00001	-0.21996	-0.21992	-25232.08250	2.96347
80%	-0.10278	0.00001	-0.10280	-0.10277	-12278.72877	2.71266
90%	-0.08500	0.00002	-0.08504	-0.08497	-5653.53527	2.69743
95%	-0.07006	0.00002	-0.07010	-0.07002	-4124.63205	2.67479
98%	-0.05168	0.00001	-0.05170	-0.05166	-6135.54616	2.63805
99%	-0.03645	0.00001	-0.03647	-0.03643	-4570.98543	2.60350

	$\hat{\beta}$	$\hat{\sigma}_\beta$	5%	95%	t statistic	Total Loss
1%	-0.42689	0.00004	-0.42700	-0.42679	-9761.59417	4.22947
2%	-0.42194	0.00003	-0.42202	-0.42186	-12206.20182	4.21889
5%	-0.38685	0.00007	-0.38701	-0.38669	-5733.11648	4.12736
10%	-0.37334	0.00004	-0.37344	-0.37325	-9181.19005	4.10703
20%	-0.32121	0.00002	-0.32125	-0.32118	-20861.23135	3.99057
30%	-0.31307	0.00001	-0.31309	-0.31306	-46176.94046	4.00356
40%	-0.30381	0.00001	-0.30382	-0.30380	-57528.59031	4.01271
50%	-0.28804	0.00001	-0.28805	-0.28802	-42611.05961	4.00144
60%	-0.27523	0.00001	-0.27525	-0.27522	-38433.08497	3.99844
70%	-0.21994	0.00001	-0.21996	-0.21992	-23899.17389	3.86253
80%	-0.10278	0.00001	-0.10281	-0.10276	-9798.02923	3.52975
90%	-0.08500	0.00001	-0.08504	-0.08497	-5670.22262	3.50473
95%	-0.07006	0.00003	-0.07014	-0.06998	-2065.24497	3.47266
98%	-0.05168	0.00002	-0.05173	-0.05163	-2267.82137	3.42325
99%	-0.03645	0.00003	-0.03653	-0.03637	-1104.11523	3.37772

Table 1.38: Shows $\hat{\beta}$ inference results over the testing and training set for the model of the S&P500 against its front month future, $n = 818$.

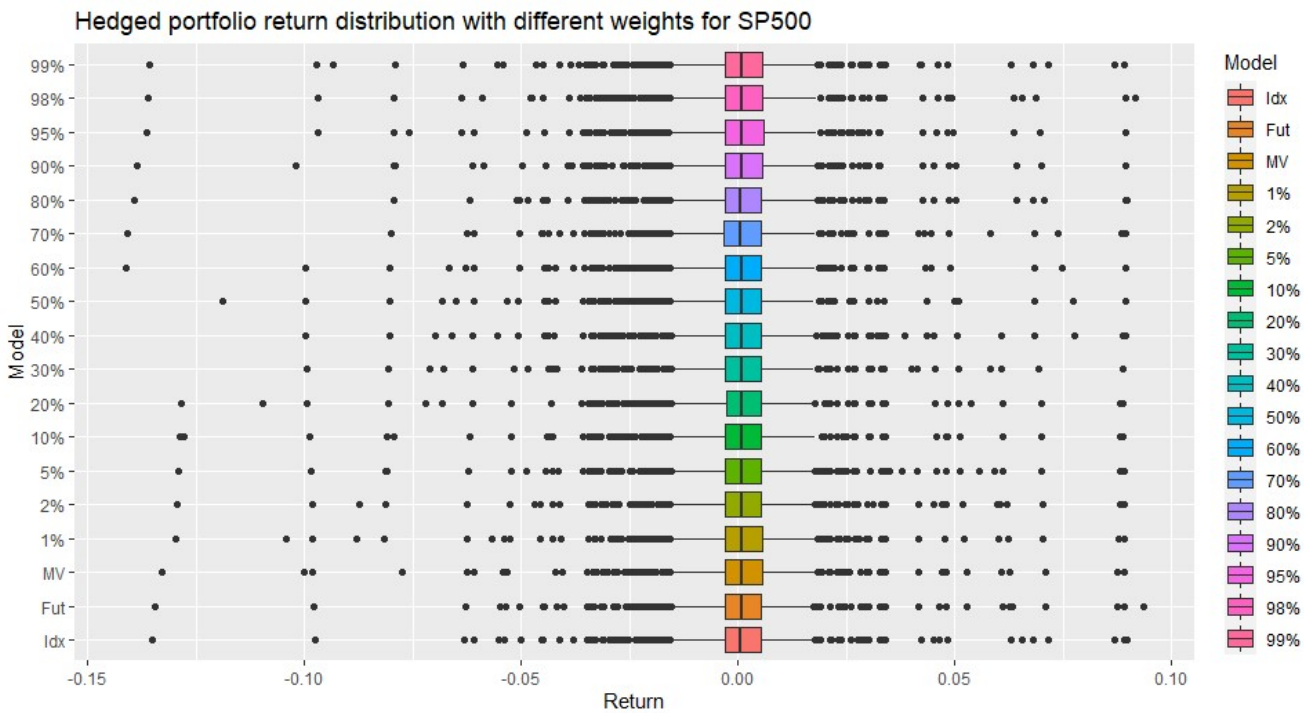
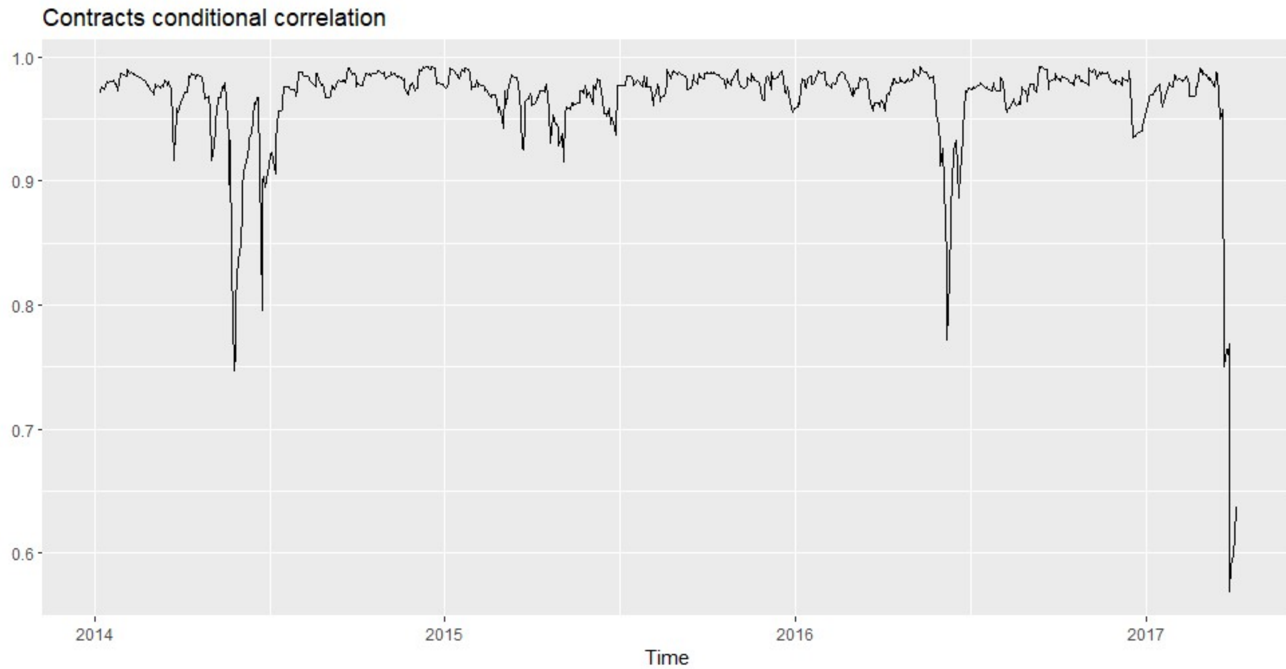


Figure 1.38: Box-plot of the return distribution of S&P 500, its front month future, OLS Mean-Variance hedged portfolio, and the conditional quantile regression hedged portfolios.

	Left Tail					Right Tail					Student-t Scale	
	k	γ_H	γ_M	γ_{Adj}	ρ_H	k	γ_h	γ_M	γ_{Adj}	ρ_H	σ	σ_{Err}
MV	78	0.73797	0.23292	0.73784	-0.74224	82	0.52099	0.50051	0.52099	-0.80223	0.00506	0.00025
1%	78	0.71951	0.30847	0.71937	-0.71847	82	0.52310	0.53762	0.52310	-1.13802	0.00773	
2%	78	0.71709	0.30896	0.71694	-0.71684	82	0.51859	0.54124	0.51859	-1.21653	0.00773	
5%	78	0.70065	0.31635	0.70046	-0.67692	82	0.51300	0.54785	0.51301	-1.33554	0.00501	0.00024
10%	78	0.70110	0.31069	0.70093	-0.68907	82	0.51362	0.54568	0.51362	-1.31213	0.00499	0.00024
20%	78	0.70304	0.29457	0.70290	-0.70421	82	0.50763	0.54541	0.50763	-1.38020	0.00508	0.00025
30%	78	0.70276	0.29309	0.70262	-0.70388	82	0.50481	0.54707	0.50482	-1.42668	0.00505	0.00025
40%	78	0.70359	0.28961	0.70345	-0.70986	82	0.50168	0.54894	0.50168	-1.47880	0.00799	0.00011
50%	78	0.70436	0.28507	0.70423	-0.71498	82	0.49654	0.55203	0.49654	-1.56472	0.00532	0.00027
60%	78	0.70828	0.27651	0.70816	-0.73383	82	0.49292	0.55359	0.49292	-1.62197	0.00514	0.00026
70%	78	0.70690	0.26909	0.70677	-0.72684	82	0.49389	0.54925	0.49389	-1.57917	0.00498	0.00024
80%	78	0.71937	0.21079	0.71931	-0.83257	82	0.50509	0.53628	0.50509	-1.32281	0.00507	0.00025
90%	78	0.72074	0.20595	0.72068	-0.83406	82	0.50324	0.53697	0.50324	-1.34894	0.00506	0.00025
95%	78	0.72180	0.20245	0.72174	-0.83438	82	0.50173	0.53758	0.50173	-1.37054	0.00506	0.00025
98%	78	0.72327	0.19824	0.72321	-0.83506	82	0.50213	0.53648	0.50213	-1.35484	0.00506	0.00025
99%	78	0.72449	0.19504	0.72443	-0.83546	82	0.50259	0.53550	0.50259	-1.33971	0.00506	0.00025

Table 1.39: Shows different Tail Index estimators value and the Student-t scale parameter over the testing set for the model of the S&P500 against its front month future



Parameters

	Estimate	Std. Error	t value	Pr(>|t|)
μ^{Idx}	0.000341	0.000281	1.212351	0.225378
θ^{Idx}	0.073084	0.035665	2.04921	0.040442
ω^{Idx}	2e-06	5e-06	0.521147	0.602265
α^{Idx}	0.138274	0.034019	4.064656	4.8e-05
β^{Idx}	0.820055	0.069989	11.716864	0
μ^{Fut}	0.000359	0.000295	1.218162	0.223162
θ^{Fut}	0.056113	0.035771	1.568671	0.116725
ω^{Fut}	2e-06	4e-06	0.654795	0.5126
α^{Fut}	0.115504	0.023879	4.837074	1e-06
β^{Fut}	0.844413	0.031117	27.136901	0
α^{Cor}	0.127307	0.029079	4.377996	1.2e-05
β^{Cor}	0.82109	0.045585	18.01213	0

Figure 1.39: GARCH-DCC Results for TSX Index and Future correlation.

	$\hat{\beta}$	$\hat{\sigma}_{\beta}$	5%	95%	t statistic	Total Loss
1%	-0.16677	0.00001	-0.16680	-0.16675	-15722.25176	2.59890
2%	-0.16187	0.00001	-0.16189	-0.16184	-16622.06592	2.58957
5%	-0.15153	0.00002	-0.15159	-0.15148	-6226.22217	2.57134
10%	-0.12798	0.00002	-0.12802	-0.12794	-8105.22372	2.52654
20%	-0.12500	0.00001	-0.12502	-0.12498	-13442.03265	2.53563
30%	-0.10044	0.00001	-0.10047	-0.10042	-10818.35785	2.49561
40%	-0.09362	0.00001	-0.09364	-0.09361	-17147.43782	2.49548
50%	-0.08796	0.00000	-0.08797	-0.08794	-17947.39288	2.49780
60%	-0.07437	0.00000	-0.07438	-0.07436	-15390.84303	2.48167
70%	-0.06317	0.00001	-0.06319	-0.06316	-9451.42630	2.47071
80%	-0.07705	0.00001	-0.07706	-0.07703	-12559.91447	2.51809
90%	-0.06417	0.00001	-0.06420	-0.06414	-4948.53986	2.50290
95%	-0.02229	0.00002	-0.02233	-0.02225	-1237.67634	2.41171
98%	0.05574	0.00001	0.05572	0.05575	7562.77815	2.23226
99%	0.07979	0.00001	0.07977	0.07981	10588.93081	2.17718
	$\hat{\beta}$	$\hat{\sigma}_{\beta}$	5%	95%	t statistic	Total Loss
1%	-0.16677	0.00004	-0.16685	-0.16669	-4743.73131	2.81606
2%	-0.16187	0.00002	-0.16191	-0.16182	-8269.89014	2.80434
5%	-0.15153	0.00004	-0.15162	-0.15144	-3912.85978	2.77967
10%	-0.12798	0.00002	-0.12802	-0.12794	-7625.62590	2.72347
20%	-0.12500	0.00001	-0.12502	-0.12498	-16284.61840	2.71620
30%	-0.10044	0.00001	-0.10046	-0.10043	-18026.86141	2.65755
40%	-0.09362	0.00000	-0.09364	-0.09361	-19720.22309	2.64115
50%	-0.08796	0.00000	-0.08796	-0.08795	-23403.12949	2.62751
60%	-0.07437	0.00000	-0.07438	-0.07436	-16683.14842	2.59521
70%	-0.06317	0.00001	-0.06319	-0.06316	-11920.31484	2.56856
80%	-0.07705	0.00001	-0.07706	-0.07704	-14407.28086	2.60124
90%	-0.06417	0.00001	-0.06419	-0.06415	-7324.79929	2.57062
95%	-0.02229	0.00003	-0.02236	-0.02223	-817.08156	2.47147
98%	0.05574	0.00001	0.05571	0.05577	4581.50388	2.28734
99%	0.07979	0.00002	0.07975	0.07983	4343.60153	2.23147

Table 1.40: Shows $\hat{\beta}$ inference results over the testing and training set for the model of the S&P Toronto Stock Exchange Index against its front month future, $n = 812$.

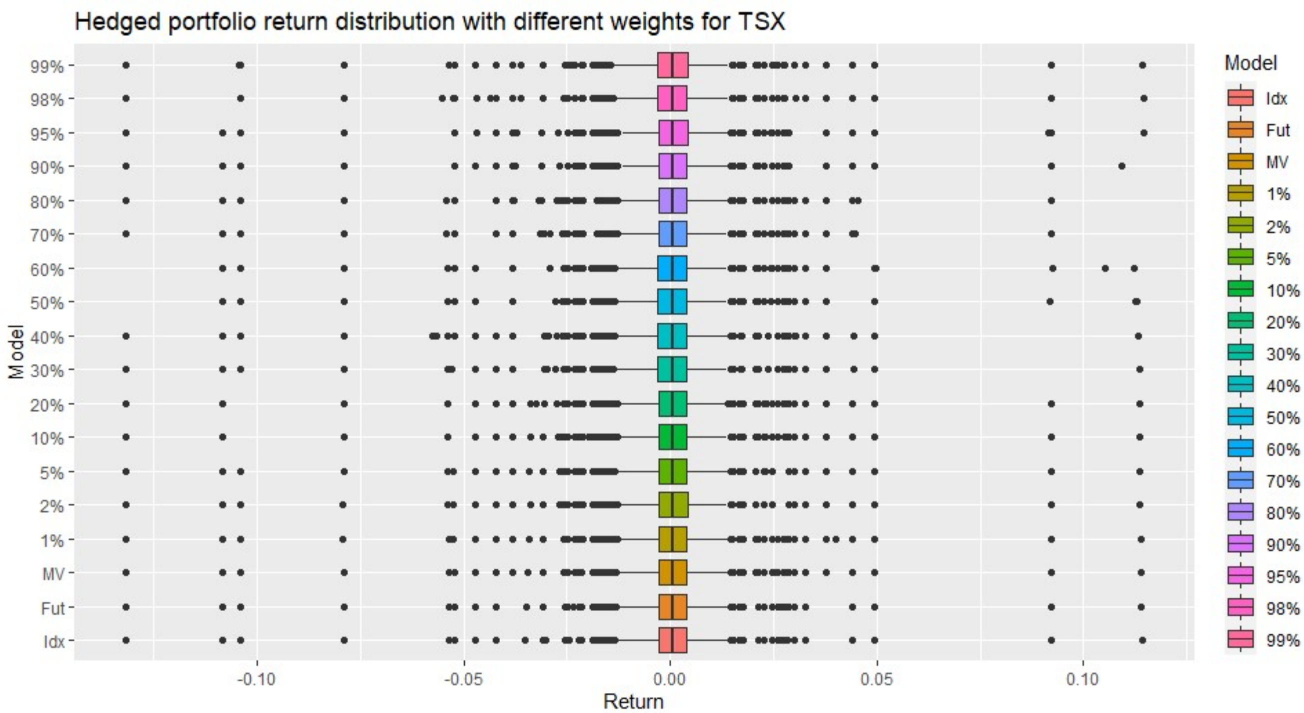
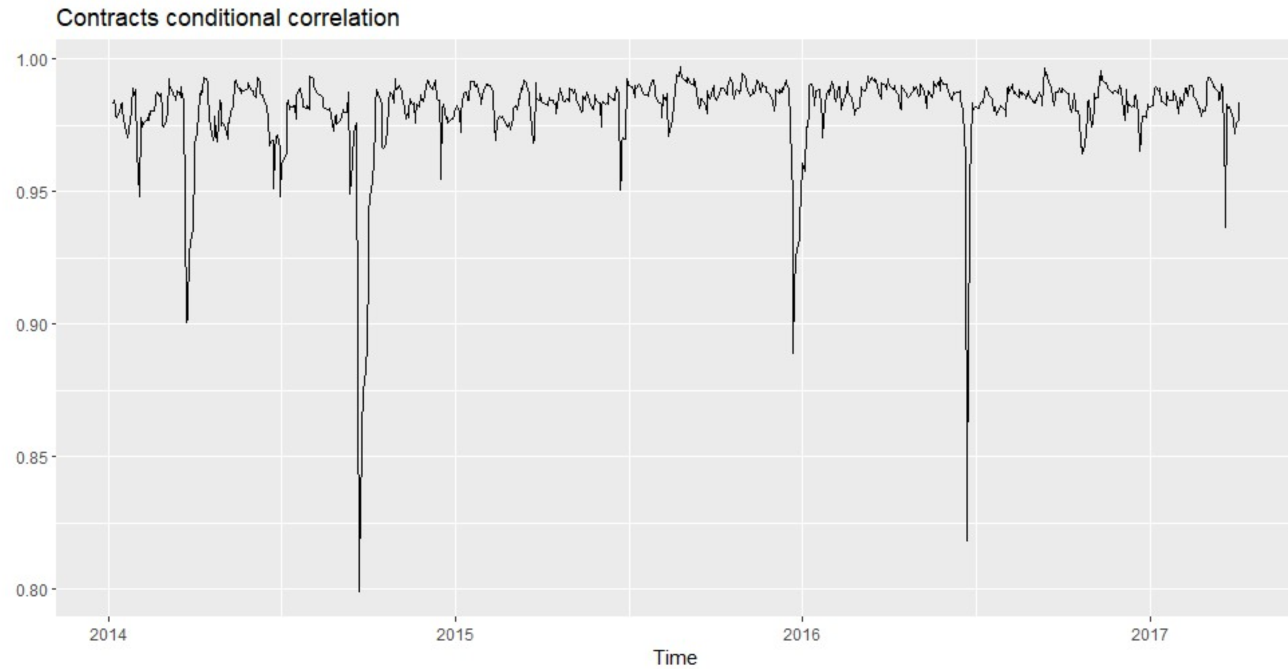


Figure 1.40: Box-plot of the return distribution of S&P Toronto Stock Exchange Index, its front month future, OLS Mean-Variance hedged portfolio, and the conditional quantile regression hedged portfolios.

	Left Tail					Right Tail					Student-t Scale	
	k	γ_H	γ_M	γ_{Adj}	ρ_H	k	γ_h	γ_M	γ_{Adj}	ρ_H	σ	σ_{Err}
MV	79	0.66232	0.65710	0.66232	-0.94459	81	0.58548	0.57619	0.58548	-0.88075	0.00400	0.00016
1%	79	0.66773	0.64420	0.66774	-0.74786	81	0.58983	0.56191	0.58984	-0.60704	0.00519	
2%	79	0.66771	0.64431	0.66772	-0.74968	81	0.59223	0.56041	0.59225	-0.55504	0.00494	0.00033
5%	79	0.66767	0.64452	0.66768	-0.75318	81	0.59215	0.56129	0.59217	-0.57113	0.00523	
10%	79	0.66757	0.64503	0.66758	-0.76124	81	0.59691	0.55905	0.59689	-0.48356	0.00422	0.00017
20%	79	0.66756	0.64510	0.66757	-0.76233	81	0.59931	0.55731	0.59927	-0.43037	0.00410	0.00016
30%	79	0.66741	0.64571	0.66742	-0.77190	81	0.60526	0.55399	0.60515	-0.31896	0.00415	0.00017
40%	79	0.66736	0.64590	0.66737	-0.77471	81	0.60540	0.55429	0.60530	-0.32319	0.00413	0.00017
50%	79	0.66732	0.64605	0.66733	-0.77709	81	0.60482	0.55511	0.60475	-0.34240	0.00417	0.00017
60%	79	0.66722	0.64644	0.66723	-0.78295	81	0.60346	0.55703	0.60343	-0.38763	0.00420	0.00017
70%	79	0.66713	0.64677	0.66714	-0.78793	81	0.60236	0.55858	0.60235	-0.42407	0.00408	0.00016
80%	79	0.66724	0.64636	0.66725	-0.78178	81	0.60372	0.55665	0.60369	-0.37881	0.00412	0.00017
90%	79	0.66714	0.64674	0.66715	-0.78748	81	0.60246	0.55844	0.60244	-0.42086	0.00412	0.00017
95%	79	0.66675	0.64805	0.66676	-0.80702	81	0.59860	0.56385	0.59859	-0.54745	0.00407	0.00016
98%	79	0.66788	0.64824	0.66788	-0.79879	81	0.60551	0.56144	0.60542	-0.44330	0.00405	0.00016
99%	79	0.66782	0.64850	0.66783	-0.80219	81	0.60628	0.56164	0.60620	-0.43971	0.00413	0.00017

Table 1.41: Shows different Tail Index estimators value and the Student-t scale parameter over the testing set for the model of the TSX against its front month future



Parameters

	Estimate	Std. Error	t value	Pr(>|t|)
μ^{Idx}	0.000307	0.000344	0.893388	0.371649
θ^{Idx}	-0.03003	0.036932	-0.813111	0.416154
ω^{Idx}	1.6e-05	3e-06	6.165012	0
α^{Idx}	0.106723	0.014956	7.135942	0
β^{Idx}	0.75669	0.033508	22.582146	0
μ^{Fut}	0.000301	0.000366	0.823303	0.410336
θ^{Fut}	-0.047948	0.038348	-1.250329	0.21118
ω^{Fut}	2e-05	1e-05	1.906911	0.056532
α^{Fut}	0.10052	0.047244	2.127671	0.033364
β^{Fut}	0.735077	0.117895	6.235005	0
α^{Cor}	0.153398	0.066524	2.30589	0.021117
β^{Cor}	0.741154	0.207492	3.571958	0.000354

Figure 1.41: GARCH-DCC Results for US2000 Index and Future correlation.

	$\hat{\beta}$	$\hat{\sigma}_\beta$	5%	95%	t statistic	Total Loss
1%	-0.42461	0.00002	-0.42465	-0.42457	-24530.24118	4.69412
2%	-0.39586	0.00002	-0.39590	-0.39581	-20826.55046	4.60152
5%	-0.37297	0.00004	-0.37305	-0.37288	-10335.42295	4.53263
10%	-0.34739	0.00002	-0.34743	-0.34734	-18773.59076	4.45895
20%	-0.29940	0.00002	-0.29944	-0.29936	-16147.71958	4.32102
30%	-0.24651	0.00001	-0.24653	-0.24648	-23438.09919	4.16520
40%	-0.08304	0.00001	-0.08306	-0.08302	-8573.84447	3.63681
50%	-0.16136	0.00001	-0.16138	-0.16133	-14915.17351	3.91816
60%	-0.20158	0.00001	-0.20160	-0.20156	-20241.05096	4.07309
70%	-0.20372	0.00001	-0.20374	-0.20369	-21331.35760	4.09969
80%	-0.19860	0.00001	-0.19863	-0.19858	-17058.32405	4.10157
90%	-0.17649	0.00002	-0.17653	-0.17646	-10659.57443	4.04489
95%	-0.16782	0.00002	-0.16788	-0.16776	-6793.43085	4.02451
98%	-0.14927	0.00001	-0.14929	-0.14924	-14124.55608	3.96620
99%	-0.12679	0.00001	-0.12681	-0.12677	-14556.05230	3.89057

	$\hat{\beta}$	$\hat{\sigma}_\beta$	5%	95%	t statistic	Total Loss
1%	-0.42461	0.00003	-0.42468	-0.42453	-13582.56729	5.71851
2%	-0.39586	0.00004	-0.39594	-0.39577	-11159.76035	5.60277
5%	-0.37297	0.00006	-0.37311	-0.37282	-5962.35346	5.51260
10%	-0.34739	0.00004	-0.34749	-0.34728	-7724.41381	5.41317
20%	-0.29940	0.00002	-0.29944	-0.29936	-17257.59527	5.22674
30%	-0.24651	0.00001	-0.24654	-0.24648	-19120.00594	5.02069
40%	-0.08304	0.00001	-0.08306	-0.08302	-11020.90339	4.37513
50%	-0.16136	0.00001	-0.16138	-0.16134	-16757.50762	4.69442
60%	-0.20158	0.00001	-0.20160	-0.20156	-25466.61430	4.86227
70%	-0.20372	0.00001	-0.20374	-0.20369	-19208.52224	4.87780
80%	-0.19860	0.00001	-0.19864	-0.19857	-14120.51158	4.86424
90%	-0.17649	0.00002	-0.17654	-0.17645	-8826.68970	4.78221
95%	-0.16782	0.00005	-0.16794	-0.16771	-3435.08823	4.75069
98%	-0.14927	0.00002	-0.14932	-0.14921	-6284.94773	4.67797
99%	-0.12679	0.00004	-0.12688	-0.12670	-3332.14352	4.58811

Table 1.42: Shows $\hat{\beta}$ inference results over the testing and training set for the model of the Russell 2000 Index against its front month future, $n = 819$.

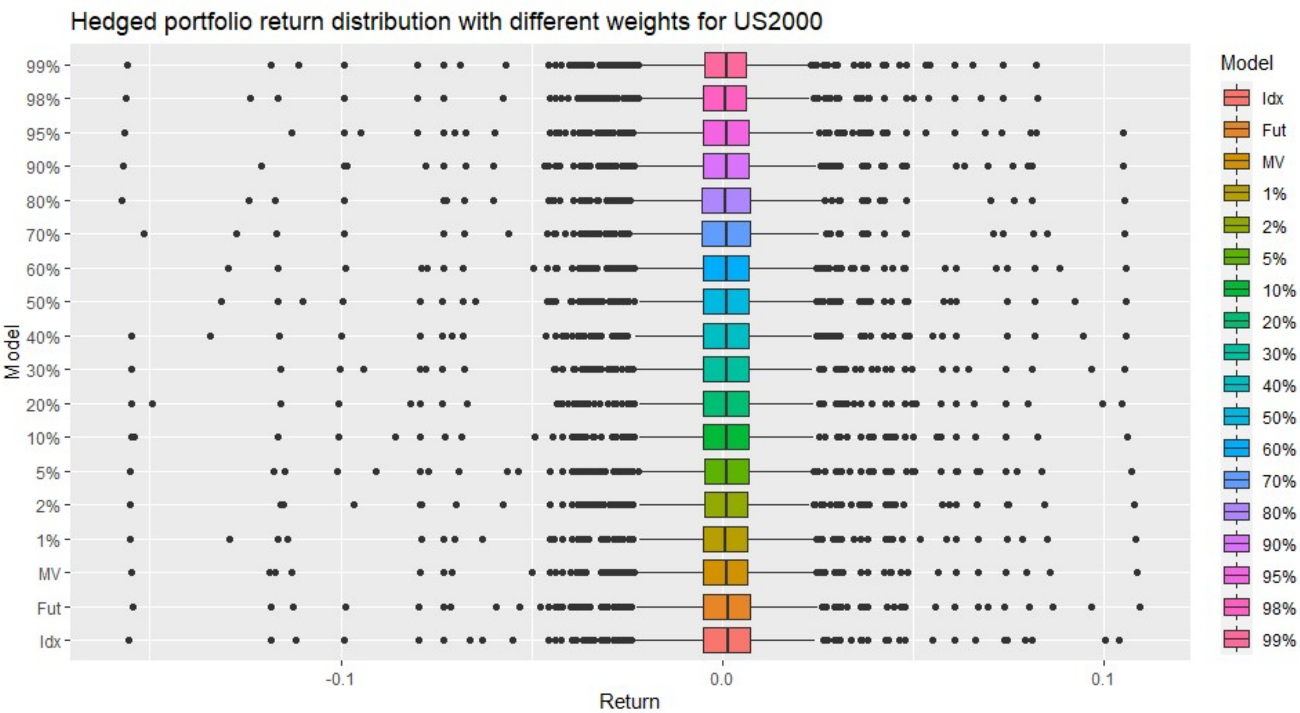
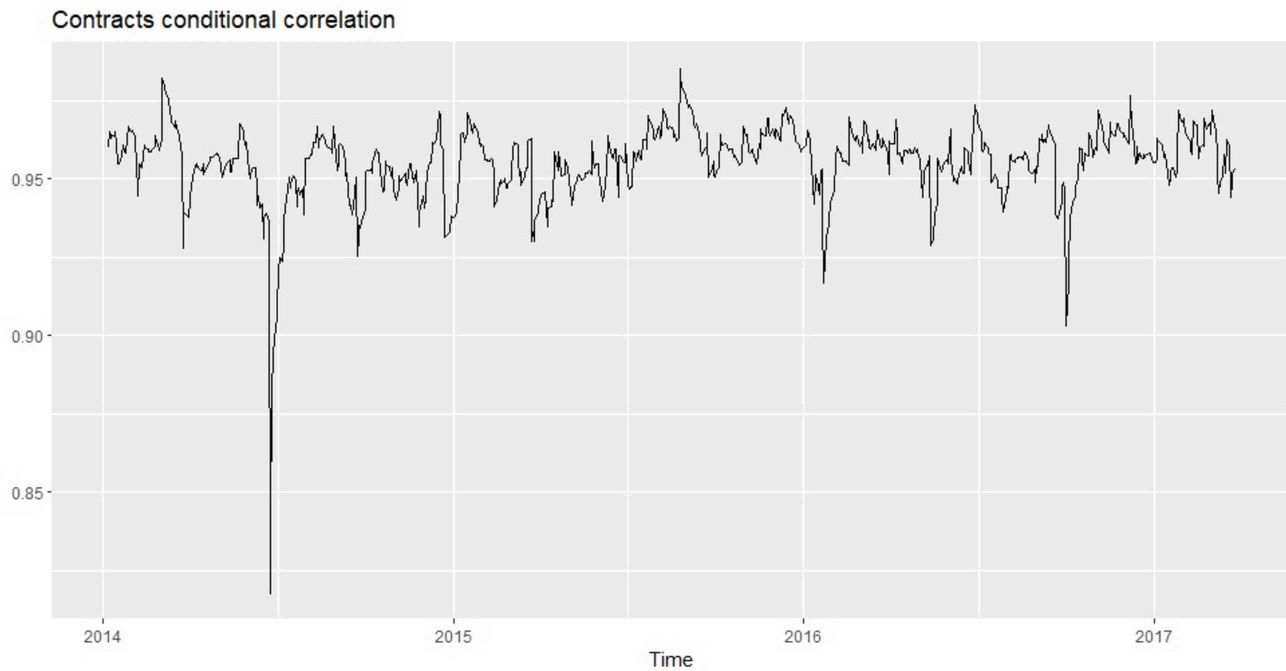


Figure 1.42: Box-plot of the return distribution of Russell 2000 Index, its front month future, OLS Mean-Variance hedged portfolio, and the conditional quantile regression hedged portfolios.

	Left Tail					Right Tail					Student-t Scale	
	k	γ_H	γ_M	γ_{Adj}	ρ_H	k	γ_h	γ_M	γ_{Adj}	ρ_H	σ	σ_{Err}
MV	78	0.56655	0.42876	0.56146	-0.09338	82	0.51595	0.41206	0.51454	-0.26666	0.00747	0.00035
1%	78	0.59413	0.46735	0.58681	-0.00056	82	0.55168	0.41287	0.54245	-0.08104	0.00720	0.00035
2%	78	0.58284	0.46934	0.57824	-0.07651	82	0.55828	0.39663	0.54657	-0.06448	0.00719	0.00035
5%	78	0.57746	0.46885	0.57372	-0.10519	82	0.56246	0.38617	0.55820	-0.15994	0.00684	0.00032
10%	78	0.59094	0.44746	0.58752	-0.13886	82	0.55956	0.38637	0.55541	-0.16502	0.00730	0.00035
20%	78	0.61362	0.40491	0.61337	-0.56129	82	0.55382	0.39197	0.54830	-0.13396	0.00731	0.00035
30%	78	0.60588	0.39441	0.60578	-0.67914	82	0.54697	0.40679	0.45628	-0.00646	0.00736	0.00035
40%	78	0.56896	0.41741	0.56847	-0.45616	82	0.55779	0.40378	0.51090	-0.01442	0.00732	0.00035
50%	78	0.58258	0.40773	0.58234	-0.56033	82	0.54328	0.42140	0.53888	-0.15113	0.00738	0.00035
60%	78	0.59324	0.39990	0.59310	-0.63885	82	0.54727	0.41486	0.53729	-0.07652	0.00737	0.00035
70%	78	0.59382	0.39955	0.59369	-0.64189	82	0.54848	0.41323	0.53565	-0.05924	0.00737	0.00035
80%	78	0.59243	0.40039	0.59229	-0.63441	82	0.54573	0.41694	0.53812	-0.09894	0.00737	0.00035
90%	78	0.58652	0.40450	0.58633	-0.59463	82	0.54540	0.41812	0.53909	-0.11406	0.00738	0.00035
95%	78	0.58425	0.40631	0.58404	-0.57570	82	0.54376	0.42051	0.53890	-0.14083	0.00738	0.00035
98%	78	0.57948	0.41053	0.57918	-0.52865	82	0.54391	0.42118	0.53959	-0.15022	0.00738	0.00035
99%	78	0.57395	0.41582	0.57358	-0.46572	82	0.55325	0.40930	0.52986	-0.03111	0.00731	0.00035

Table 1.43: Shows different Tail Index estimators value and the Student-t scale parameter over the testing set for the model of the US2000 against its front month future



Parameters

	Estimate	Std. Error	t value	Pr(>|t|)
μ^{Idx}	-0.000135	0.000383	-0.352378	0.724555
θ^{Idx}	0.072341	0.038025	1.902457	0.057111
ω^{Idx}	3e-06	4e-06	0.893224	0.371737
α^{Idx}	0.052675	0.010671	4.936172	1e-06
β^{Idx}	0.920228	0.013954	65.945907	0
μ^{Fut}	-0.000133	0.000382	-0.347782	0.728004
θ^{Fut}	0.054598	0.038144	1.431352	0.152329
ω^{Fut}	3e-06	4e-06	0.657131	0.511096
α^{Fut}	0.041956	0.021221	1.977164	0.048023
β^{Fut}	0.936844	0.009991	93.766626	0
α^{Cor}	0.05145	0.023477	2.191532	0.028413
β^{Cor}	0.818447	0.070814	11.557632	0

Figure 1.43: GARCH-DCC Results for WIG20 Index and Future correlation.

	$\hat{\beta}$	$\hat{\sigma}_\beta$	5%	95%	t statistic	Total Loss
1%	-0.59670	0.00002	-0.59675	-0.59666	-31524.08475	5.22501
2%	-0.58465	0.00002	-0.58469	-0.58461	-35379.94226	5.18386
5%	-0.53962	0.00004	-0.53971	-0.53953	-13427.82126	5.03169
10%	-0.50993	0.00003	-0.50999	-0.50987	-20049.60423	4.92686
20%	-0.47771	0.00002	-0.47775	-0.47766	-23130.73764	4.80696
30%	-0.45200	0.00001	-0.45202	-0.45197	-37465.16573	4.70914
40%	-0.39320	0.00001	-0.39322	-0.39317	-34183.87496	4.50636
50%	-0.33075	0.00001	-0.33077	-0.33073	-34937.40056	4.29379
60%	-0.36213	0.00001	-0.36215	-0.36210	-34384.69074	4.37947
70%	-0.35568	0.00001	-0.35571	-0.35565	-28676.88235	4.34523
80%	-0.33263	0.00001	-0.33266	-0.33260	-24288.80186	4.25910
90%	-0.30037	0.00002	-0.30043	-0.30031	-12227.81615	4.14490
95%	-0.27480	0.00002	-0.27485	-0.27474	-11238.87037	4.05887
98%	-0.23973	0.00001	-0.23975	-0.23971	-27056.58544	3.94657
99%	-0.22848	0.00001	-0.22850	-0.22846	-23396.17651	3.91059

	$\hat{\beta}$	$\hat{\sigma}_\beta$	5%	95%	t statistic	Total Loss
1%	-0.59670	0.00004	-0.59679	-0.59662	-16759.54838	6.27765
2%	-0.58465	0.00003	-0.58471	-0.58459	-23021.49287	6.22690
5%	-0.53962	0.00005	-0.53973	-0.53951	-11418.05668	6.03999
10%	-0.50993	0.00003	-0.50999	-0.50987	-18949.91275	5.90805
20%	-0.47771	0.00002	-0.47775	-0.47766	-26823.02491	5.75227
30%	-0.45200	0.00002	-0.45203	-0.45196	-29828.83054	5.62315
40%	-0.39320	0.00001	-0.39323	-0.39317	-31486.09625	5.36719
50%	-0.33075	0.00001	-0.33077	-0.33072	-28075.53499	5.09983
60%	-0.36213	0.00002	-0.36217	-0.36209	-22288.35468	5.19325
70%	-0.35568	0.00001	-0.35571	-0.35565	-25458.96297	5.14174
80%	-0.33263	0.00002	-0.33267	-0.33260	-21097.45248	5.02775
90%	-0.30037	0.00002	-0.30042	-0.30032	-12859.68095	4.88007
95%	-0.27480	0.00004	-0.27488	-0.27471	-7639.50641	4.77137
98%	-0.23973	0.00002	-0.23977	-0.23969	-13921.70947	4.63276
99%	-0.22848	0.00002	-0.22852	-0.22844	-12683.75875	4.58828

Table 1.44: Shows $\hat{\beta}$ inference results over the testing and training set for the model of the Warsaw Stock Index against its front month future, $n = 810$.

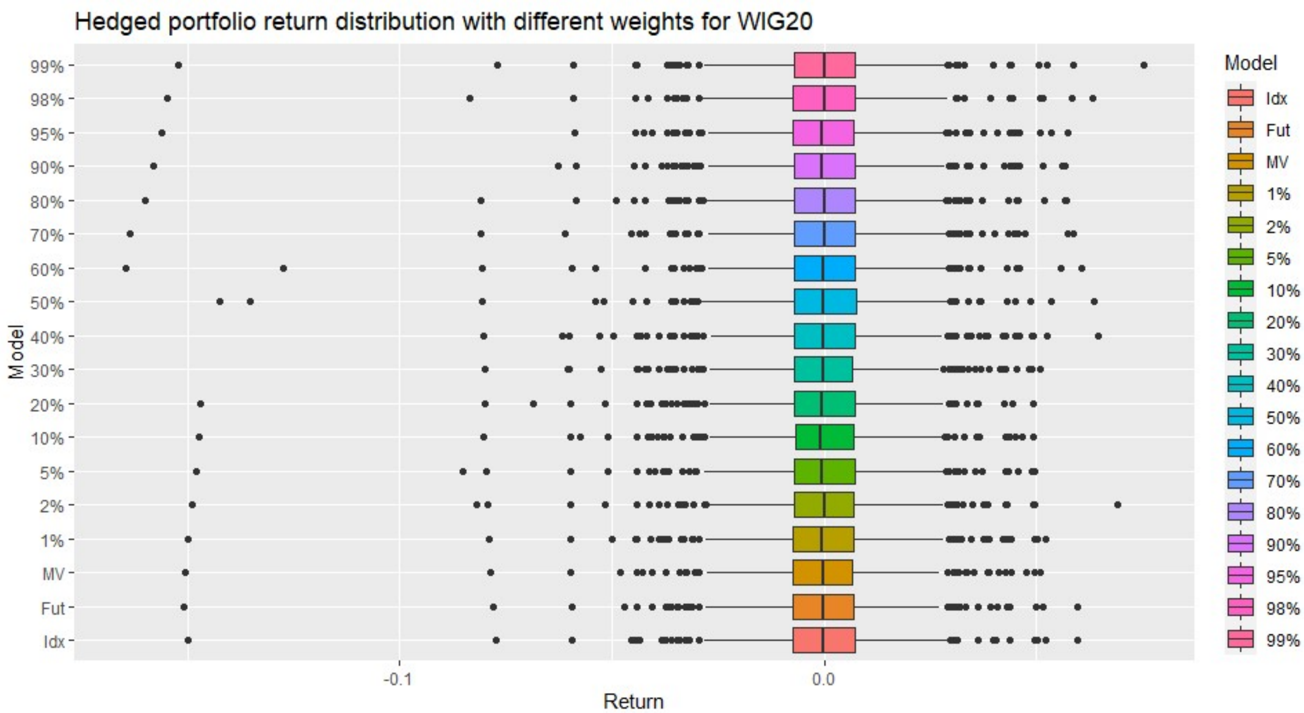


Figure 1.44: Box-plot of the return distribution of Warsaw Stock Index, its front month future, OLS Mean-Variance hedged portfolio, and the conditional quantile regression hedged portfolios.

	Left Tail					Right Tail					Student-t Scale	
	k	γ_H	γ_M	γ_{Adj}	ρ_H	k	γ_h	γ_M	γ_{Adj}	ρ_H	σ	σ_{Err}
MV	79	0.50330	0.29440	0.50330	-2.66110	81	0.39952	0.16902	0.39847	-0.29836	0.00913	0.00039
1%	80	0.37330	0.40062	0.37338	-0.39662	81	0.46287	-0.05710	0.46264	-0.60379	0.01012	0.00043
2%	80	0.37762	0.39693	0.37763	-0.56196	81	0.46385	-0.06216	0.46365	-0.62244	0.01019	0.00044
5%	80	0.37838	0.40674	0.37872	-0.24266	81	0.41683	0.07676	0.41565	-0.34038	0.00984	0.00042
10%	80	0.38778	0.40069	0.38778	-0.60657	81	0.41783	0.06576	0.41722	-0.42718	0.00978	0.00041
20%	80	0.39689	0.39702	0.39689	-0.99480	81	0.40970	0.08549	0.40906	-0.41227	0.00971	0.00041
30%	80	0.39140	0.41292	0.39179	-0.09471	81	0.40754	0.09032	0.40700	-0.43254	0.00976	0.00042
40%	80	0.40428	0.40883	0.40427	-0.51290	81	0.41668	0.05905	0.41656	-0.63780	0.00955	0.00040
50%	80	0.39508	0.42944	0.39508	-7.74266	81	0.40079	0.10870	0.40055	-0.52793	0.00946	0.00040
60%	80	0.40098	0.41759	0.40098	-1.42212	81	0.40376	0.09789	0.40354	-0.53963	0.00949	0.00040
70%	80	0.40152	0.41759	0.40152	-1.51687	81	0.40356	0.09892	0.40336	-0.54261	0.00962	0.00040
80%	80	0.39555	0.42857	0.39555	-6.95477	81	0.40149	0.10646	0.40127	-0.53597	0.00947	0.00040
90%	80	0.40986	0.41302	0.40986	-2.43990	81	0.39275	0.13546	0.39228	-0.42454	0.00943	0.00039
95%	80	0.41374	0.41032	0.41374	-1.42945	81	0.38184	0.16936	0.38050	-0.25052	0.00941	0.00039
98%	80	0.41562	0.40718	0.41562	-9.62404	81	0.38111	0.17687	0.37953	-0.21320	0.00938	0.00039
99%	80	0.41214	0.41212	0.41214	-0.96721	81	0.38187	0.17693	0.38025	-0.21297	0.00937	0.00039

Table 1.45: Shows different Tail Index estimators value and the Student-t scale parameter over the testing set for the model of the WIG20 against its front month future

Chapter 2

P-Spline FPCR Portfolio Selection

In this chapter I investigate the implementation of an index tracking portfolio strategy in an high-dimensional setting, by mean of Functional Principal Component Regression (FPCR) of the smoothed stock price time series, attempting to overcome some flaws affecting some state of the art techniques for endogenous index tracking portfolio selection.

2.1 Introduction

An asset allocation is passive if it aims to reproduce the risk-return profile of some specified benchmark, while in the past these strategies were advisable nor desirable, there is evidence that they have gained increasing popularity both to practitioners and customers. One of the most famous passive strategy families in the scientific literature is indexing, trying to replicate the results of a specific index by investing in its constituents.
The easiest way to track an index is to hold all its assets in the same relative quantities, this is the so-called *full replication* approach, which has several drawbacks due to the complexity of the index composition, that sometimes includes thousands of stocks, and this implies the need to frequent revision the portfolio weights thus incurring in high transaction costs. Another relevant approach is to synthesize the index through equity derivatives (such as ETF) and future contracts, this is preferable since one usually buys only one contract and is able to fully replicate the index behaviour for short period of time, nevertheless these contracts are negotiated on a range of maturities (CME, for example, offers quarterly contracts for five consecutive quarters) and rolling contracts to dynamically track the underlying index is expensive and risky.
Another approach, relies on the selection of a smaller subset of index's components to replicate its behaviour, this is usually a less effective strategy compared either to the *full replication* or *synthetic replication* approach, but allows

to overcome the formers alternatives flaws, through the application of more or less sophisticated statistical methods. In this chapter I solve the stock selection problem to achieve an index tracking portfolio with the application of Principal Component Regression over the P-spline smoothed price series of all the stocks included in the S&P500 index during the observation window, and provide a comparison with the solution provided by a recent implementation of the cointegration method.
Next section gives a review of more or less recent methodological proposal for the solution of the stated problem, with special consideration to the branches of literature applying cointegration analysis and unsupervised learning techniques. Section 3 describes my proposal and provides descriptions of the solution to technical issues related to its implementation. Section 4 is about data and results, while Section 5 contains conclusion, considerations and further development.

2.2 Literature Review

2.2.1 Traditional econometrics approaches

Traditionally, quantitative methods for portfolio selection rely on the economic theory of investor's optimal portfolio choice, pionereed by Markowitz[106], Merton[107], Fama[52] and Samuelson [124], that is based on a mathematical programming approach where, after some assumptions on the market structure, one models the optimizing behaviour of the agent according to some specified family of utility functions. From the merely statistical point of view these approaches led to two branches of econometric literature on the subject: *plug-in estimation* and *decison theory*.
Under the plug-in estimation approach, the analyst draws inference about decision maker's optimal portfolio weights to make descriptive statements, in the decision theory approach the analyst takes the role of the investor and draws inferences about the return distribution to choose portoflio weights that are optimal with respect to these inferences.

Plug-in Estimation

Much of the portfolio choice literature falls under the plug-in estimation (or calibration) label, meaning that the analyst has a numerical or analytical solution to the investor's problem and plugs in the estimated parameters of the data generating process, if the analyst treats the parameter as estimates, the portfolio weights are estimated, otherwise (if the parameter are assumed to be true) the portfolio weights are calibrated.
Another distinction that can be stated in this branch is related to the time horizon of the problem, if one considers the single-period choice problem, the investor's solution maps the preference parameters ϕ, the state vector z_t and the parameter of the data generating process θ into the optimal portfolio weights

w_t

$$w_t^* \equiv w(\phi, z_t, \theta),$$

where ϕ is specified ex-ante, z_t is observed and θ is estimated from the data $Y_T \equiv \{y_t\}_{t=0}^T$, plugging $\hat{\theta}$ in the w_t^* expression one obtains $\hat{w}_t^* \equiv w(\phi, z_t, \hat{\theta})$. Assuming $\hat{\theta}$ consistency with asymptotic distribution $\sqrt{T}(\hat{\theta} - \theta) \sim N[0, V_\theta]$, the asymptotic distribution of the estimator $\hat{w}_t^*$ can be computed using delta method

$$\sqrt{T}(\hat{w}_t^* . w_t^*) \sim N[0, w_3(\cdot) V_\theta w_3(\cdot)'].$$

For example in the mean-variance case, assuming i.i.d excess returns with constant risk premia μ and covariance matrix Σ the optimal portfolio weights are $w^* = (1/\gamma)\Sigma^{-1}\mu$, where γ is the prespecified risk aversion coefficient for CRRA utility. Thus, given excess return data $\{r_{t+1}\}_{t=1}^T$, the moments can be estimated from the sample analog

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} r_{t+1} \text{ and } \hat{\Sigma} = \frac{1}{T-N-2}\sum_{t=1}^{T}(r_{t+1} - \hat{\mu})(r_{t+1} - \hat{\mu})'$$

and the optimal plug-in portfolio weights estimates are $\hat{w}_t^* = (1/\gamma)\hat{\Sigma}^{-1}\hat{\mu}$ and are unbiased due to normality and standard independence assumptions. Without normality, or without standard covariance matrix normalization, the estimator is generally biased but still consistent.
There is a long literature branch documenting the shortcomings of plug-in estimates, especially in the context of large-scale mean-variance problems [81, 82, 109, 13, 14, 32]. The general conclusion is that plug-in estimates are extremely imprecise and that the asymptotic approximations are unreliable. Moreover the precision of plug-in estimation deteriorates with the number of assets held in the portfolio. This motivated a huge branch of literature to suggest different, or complementary, methods for improving plug-in estimation for practical applications.
Shrinkage estimation "shrinks" the sample means toward a common value, a convex combination of the sample means, that dominates those of the random variables in terms of joint mean-squared error,

$$\mu_S = \delta_0 \mu_0 + (1 - \delta)\bar{\mu},$$

thereby reducing the extreme estimation errors that occur in the cross section of individual means and resulting in a lower overall variance of the estimators. This technique, that has been applied to portfolio choice problems by [83, 86] among the others, leads to estimates that dominate, in terms of expected utility, those provided by the plug-in methods constructed with the usual sample means. Shrinking estimation has been also applied to covariance matrices [98, 99, 100],

$$\hat{\Sigma}_s = \delta \hat{S} + (1 - \delta)\hat{\Sigma},$$

showing reduced sampling error, that guarantees a positive definite estimate also when the sample covariance matrix is itself singular (N > T).

$$r_{i,t} = \alpha_i + \beta_i' f_t + \varepsilon_{i,t}$$
$$\Sigma = B\Sigma_f B' + \Sigma_\varepsilon$$

General K-factor model.

Another approach to reduce the statistical error in plug-in estimates is to impose a factor structure for the covariation among assets. Sharpe(1963)[131] has been the first to propose the use of a single-factor marker model covariance matrix in a mean-variance problem, reducing the dimensionality of the portfolio problem to $3N + 1$ terms, with the drawback that one single factor may not be able to capture all the covariation among assets, leading to potentially biased estimates of the return covariance matrix.

This problem can be faced with an increasing number of factors, that translate to an increased number in the degrees of freedom. To avoid this in the literature there is an established preference towards common factors model. Typically one can approach this problem in three ways. First, one can choose factors based on economic theory, as those proposed by Sharpe(1963) (aggregate wealth portfolio) or aggregate investment opportunity set as in Merton(1973)[108](ICAPM). Second, the choice can be based on empirical evidence, thus including macroeconomic factors[29], industry factors, firm characteristic-based factors[53] and their combinations. Third, the factors can be obtained from returns using statistical procedures as factor analysis or principal component analysis [36].

Decision Theory

According to the second traditional econometric approach the analyst takes the role of the investor and chooses portfolio weights optimal with regards to the subjective belief about the true return distribution. Due to statistical uncertainity about parameters or the parametrization of the data generating process, the subjective return distribution may be different from the results of plug-in approach estimates leading to different optimal portfolio weights. I consider the expected utility maximization problem stated as:

$$\max_{w_t} \int u(w_t' r_{t+1} + R^f) p(r_{t+1}|\theta) \mathrm{r}_{t+1},$$

In the previous exposed approach it was implicitly assumed that the problem was well posed, meaning that all the information required to solve it were available to the decision maker. If one supposes instead that the investor doesn't know the distributions' true value parameters the problem can't be solved. In these situations one can proceed in three different ways, first one may naively use estimates of the parameters as in the plug-in approach, except that now is the decision of the investor to be modeled. Alternatively one may consider worst case outcome under some prespecified set of possible parameter values, as in a

robust control framework. Finally one can eliminate the optimization problem dependence from the unknown parameters, replacing the true distribution with a subjective one, leading to irrelevant sub-optimality due to the unknowability of the truth[143, 90, 19].
In this context the most popular way to specify a prior is to rely on theoretical implications of an economic model [15, 35, 116].

2.2.2 Cointegration Analysis

Cointegration analysis is an econometric technique developed to analyze a particular class of vector unit root processes known as *cointegrated* processes. Such specification had been already implicitly defined in "error-correction" model (such those advocated by Davidson, Hendry, Srba and Yeo(1978)[38]), but the formal key concept hadn't been developed in the field until the groundbreaking work by Granger(1983)[65] and Engle and Granger(1987)[50]. The simplest example of cointegrated vector process is the bivariate system:

$$\begin{aligned} y_{1,t} &= \gamma y_{2,t} + u_{1,t}, \\ y_{2,t} &= y_{2,t-1} + u_{2,t}, \end{aligned}$$

which matrix polynomial moving average operator has a root at unity, hence is non-invertible, and this makes the finite-order VAR in differences a poor approximation due to the information about y_1 contained in the *level* of y_2, the introduction of the lagged levels along with the lagged differences brings a stationary representation of the process and leads to the definition of cointegrated process as a vector of time series, which individually are nonstationary with a unit root, but with a linear combination (*the lagged levels*) $\mathbf{a'y_t}$ that is stationary for some (n $\times$ 1) vector $\mathbf{a}$, this can be interpreted as a common stochastic trend shared by two (or more) time series.
The first application of cointegration analysis to asset allocation relies on the observations on common trend by Stock and Watson[137], which justified the application of cointegration analysis for optimal portfolio selection by Alexander [2], that achieved the identification of optimal trading pairs, gained enhanced weight stability and a better mining of the information contained in the stock price series, allowing him to build levered and self-financing index tracking and long-short market neutral trading strategies. The same author used the same technique to construct cointegration-based portfolio to search for potential "*alpha*" sources concluding that cointegration analysis can improve traditional models [4, 5, 43].
In this framework one assumes that stock price series are $I(1)$, therefore being $P_{1,t}, P_{2,t}, ..., P_{k,t}$ a sequence of $I(1)$ time series if there are nonzero real numbers $\beta_1, \beta_2, ..., \beta_k$ such that

$$\beta_1 P_{1,t} + \beta_2 P_{2,t} + ... + \beta_k P_{k,t}$$

becomes an $I(0)$ series, then one can say that the former series are cointegrated, that they share a stationary long-run stable relationship with the property of mean reversion. One can thus assume that the index tracking model can be stated in the following way:

$$log(I_t) = \beta_0 + \sum_{i=1}^{K} \beta_i log(P_{i,t}) + \varepsilon_t,$$

where I_t is the index value at time t. Then, normalizing the cointegration coefficients β_i to sum up to one, the analyst determines the proportional weights for each stock. If β catches the effect of the cointegration relationship, then the residuals are supposed to be stationary. So, one defines the loss functions as

$$L(\varepsilon_t) = \frac{\hat{\rho} - 1}{\hat{\sigma}_\rho},$$

where ρ stands for the autocorrelation coefficients (and $\hat{\sigma}_\rho$ for its relative standard error) in the dynamic error correction relationship, formally

$$\varepsilon_t = \alpha + \rho\varepsilon_{t-1} + \sum_{i=1}^{d} \gamma\Delta\varepsilon_{t-1} + u_t,$$

and d is the considered lag-order.

2.2.3 Unsupervised learning technique

One of the most common scientific application of unsupervised statistical learning techniques is portfolio selection, this is because this class of methods aim to exploit data patterns to identify homogeneous groups (thought as latent categorical variable) in large datasets, a task that is very close to the portfolio selection process and that inspired scholar from different fields and with different background.
A traditional application is the factor covariance matrix decomposition for the identification of additional "hidden" factors (or "uncertainty structure") to enhance the results of factor models [63, 120, 87, 97], this is usually done by Principal Component Analysis (PCA) or Independent Component Analysis, Alexander and Dumitru(2004)[3] applies the same technique to select a portfolio tracking the first principal component of a group of stocks, thus capturing only the common trend in stock returns.
Fabozzi and Focardi (2004)[56] discusses the problem of implementing optimal investment strategy when full replication is not deemed suitable, discovering correlation and cointegration structure of the index components through cluster analysis, Pattarin et al. (2004)[117] combines PCA and evolutionary clustering algorithm to discover mutual funds style by analysing the time series of their return. Fang and Wang (2005)[54] applies the fuzzy logic to a bi-objective programming model for the selection on index tracking portfolio problem while Gaivoronsky et al.(2005)[60] produces an algorithm which determines whether

or not to rebalance a given portfolio based on transaction costs with an application to an index tracking portfolio for the Oslo stock exchange. Dose and Cincotti(2005)[42] successfully combines stochastic-optimization technique with time series cluster analysis in a two step procedure to achieve the construction of an enhanced index tracking portfolio and Basalto et al.(2007)[11] groups stock price time series according to their Hausdorff distance[1] to discover common trend, Monfort et al. (2008)[110] develops an optimizing sampling algorithm to construct portfolio that tracks an index *"as accurately as possible"* and Jeurissen and van den Berg(2008)[80] investigates an approach for tracking the Dutch AEX index using hybrid genetic algorithm which chromosome represents a specific subset of the stocks from the index, the fitness function to the minimized achievable tracking error for that subset and defines the tracking portfolio as the highest fitness achievable. Caiado and Crato (2010)[23] proposes volatility and spectral based methods for the cluster analysis of stock returns, looking to the hierarchical structure tree, something similar is also done by other authors [114, 77, 115].

Bruni et al.(2012)[20] proposes the application of large-size optimization model for Enhanced Index Tracking that selects the optimal portfolio according to a new stochastic dominance criterion and solves this problem with an efficient constraint generation technique. Guastaroba and Speranza (2012)[66] introduces mixed-integer linear programming formulations for the index tracking portfolio selection problem and solves this through the Kernel Search heuristic framework, a similar procedure is proposed in Chen and Kwon (2012)[28] that develops a robust portfolio selection model for tracking a market index using subset of its assets, here the model is an integer program that maximize the similarity between selected assets and those of the target index. Edirisinghe (2013)[45] considers the index tracking portfolio selection problem for the S&P500 index and solves it with a tracking optimization model thought as an extension of the Mean-Variance model with constant adjustments to portfolio weights, dependent on the index variance and assets' return parameters. Wu et al. (2014) [139] proposes the nonnegative-lasso method for portfolio selection in high dimensional linear regression models, achieving smaller tracking error when compared to more traditional approaches.

[1]Hausdorff distance measure the distance between two subsets of a matric space by considering the distance between their closest elements, formally:

$$d_H(A,B) = max\left\{\sup_{a\in A} d(a,B), \sup_{b\in B} d(A,b)\right\},$$

where $d(a,B) = \inf_{b\in B} d(a,b)$ quantifies the distance from a point $a \in A$ to $b \in B$.

2.3 P-spline FPCR Portfolio selection

2.3.1 Principal Component Analysis and Principal Component Regression

First proposed by Pearson (1901)[118] PCA is an essential tool for multivariate data analysis and unsupervised dimension reduction. Its goal is to find the sequence of orthogonal components that explains in the most efficient way the overall variance of the observations. Its original version was (and still is) useful in the context of longitudinal studies to address singularity in the covariance matrix due to multicollinearity or high-dimensionality, Hotelling (1933) provided the description of the procedures to attend for principal components computation [70].
The main advantage of PCA is its ability to find a lower-dimensional representation of the original variables while preserving their amount of information. For centered data $\mathbf{X_0}$ on a $(N \times P)$ matrix PCA yields an orthogonal decomposition for a given number of principal components, given by

$$\underset{1\times N}{\beta} = \underset{1\times P}{\mathbf{\Phi}_1} \underset{N\times P}{\mathbf{X_0}'} \tag{2.1}$$

where $\mathbf{\Phi}_1$ is the first principal component and $\underset{1\times N}{\beta}$ is the set of principal component scores with mean zero. One finds $\mathbf{\Phi}_1$ by maximization of the variance of $\underset{1\times P}{\mathbf{\Phi}_1} \underset{N\times P}{\mathbf{X_0}'}$ and then obtains the other principal components by substitution of the reduced data matrix $\mathbf{X_k}$ to the original one. Another (and easier) algorithm to perform the PCA is through singular value decomposition (SVD), that for the centered data matrix $\mathbf{X_0}$ can be expressed as

$$\underset{N\times P}{\mathbf{X_0}} = \underset{N\times K}{\mathbf{U}} \underset{K\times K}{\mathbf{D}} \underset{K\times P}{\mathbf{V}}', \tag{2.2}$$

where $K \leq \min(N,P)$, $\mathbf{U}'\mathbf{U} = \mathbf{V}'\mathbf{V} = \mathbf{I}_K$ and $\mathbf{D}$ is a diagonal matrix with $d_1 > d_2 > ... > d_K$ on the diagonal, $\mathbf{UD}$ containing the principal components score.
Now, consider a multivariate linear regression model

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon,$$

where $\mathbf{y}$ is a vector of centered responses, $\mathbf{X}$ is an $(N \times P)$ matrix of predictors, β is a vector of uknown regression coefficients and ε is a vector of i.i.d. random errors, using the SVD of $\mathbf{X}$ in (2.2) the ordinary least squares coefficients can be written as

$$\begin{aligned}\hat{\beta}_{OLS} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \\ &= \left[(\mathbf{UDV}')'\mathbf{UDV}'\right]^{-1}(\mathbf{UDV}')'\mathbf{y}, \\ &= \mathbf{VD^{-1}U}'\mathbf{y} = \sum_{k=1}^{p} \frac{v_k u_k'}{d_k}\mathbf{y}. \end{aligned} \tag{2.2}$$

Principal Component Regression (PCR) starts by using the principal component of the predictor variables in place of predictors. Since the principal components are uncorrelated by construction it solves the problem arising in the presence of rank-deficient model matrix by deleting those components that have low variances. Mathematically the model is defined as

$$\mathbf{y} = \mathbf{\Phi}\beta_{PCR} + \varepsilon.$$

The principal component scores are calculated via OLS and given by

$$\begin{aligned}\hat{\beta}_{PCR} &= (\mathbf{\Phi}'\mathbf{\Phi})^{-1}\mathbf{\Phi}'\mathbf{y} \\ &= (\mathbf{L}^2)^{-1}\mathbf{\Phi}'\mathbf{y},\end{aligned}$$

where $\mathbf{L}^2$ represents the diagonal matrix whose k^{th} element is the k^{th} largest eigenvalue of $\mathbf{X}'\mathbf{X}$. This estimator has the advantage to "*shrink*" the expansion (2.2), thus

$$\hat{\beta}_{PCR} = \sum_{k=1}^{K} \frac{v_k u_k'}{d_k}\mathbf{y},\ K < \min(N, P)$$

Functional PCA and Functional PCR

In my application PCA is applied to time series objects, which are defined in a space defined by price and time, so can be considered as functional data.
Many authors realized that PCA runs many difficulties in the analysis of functional data due to the "curse of dimensionality", FPcA overcomes this difficulty and provides a more informative way of examining the covariance structure than PCA. FPCA finds the set of orthogonal principal component functions maximizing the variance along each component, namely

$$\beta_1 = \int_{x_1}^{x_p} \Phi_1(x)\mathbf{f}(x)\,dx,$$

as before, successive principal component functions are obtained iteratively by subtracting the first k principal component from $\mathbf{f^0}(x) = \mathbf{f}(x)$, that is

$$\mathbf{f^k}(x) = \mathbf{f^{k-1}}(x) - \beta_k\phi_k(x),$$

and then computing the next principal components scores

$$\beta_{k+1} = \int_{x_1}^{x_p} \Phi_{k+1}(x)\mathbf{f^k}(x)\,dx$$

which variance is maximized under the constraints

$$\int_{x_1}^{x_p} \phi_{k+1}^2(x)\,dx = \|\phi_{k+1}^2(x)\| = 1,$$

$$\int_{x_1}^{x_p} \phi_{k+1}(x)\phi_j(x)\,dx = 0 \text{ for } j = 1, ..., k.$$

The Functional principal component regression (FPCR) describes the relationship between the functional predictors and responses, where the response variable can be scalar or function, and can be expressed as follows

$$f_t(y) = \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\phi_k(x) + \varepsilon_t(x), \ t = 1, 2, \dots, n, \tag{2.3}$$

where $\mu(x) = \mathbb{E}[\mathbf{f}(x)]$ is the mean function and $\mathbf{f}(x)$ is a vector of n realizations of a stochastic process, $\phi_k(x)$ is the k^{th} orthonormal eigenfunction of $Var[\mathbf{f}(x)]$ and the $\beta_{\mathbf{k}}$ is the k^{th} functional principal component scores, given by the projection of $\mathbf{f}(x) - \mu(x)$ in the k^{th} eigenfunction direction, $\varepsilon_t(x)$ is the error function for the t^{th} observation (including the excluded functional principal component) and K is the number of retained functional principal components. In my framework the scope of the analysis is to find the eigenvectors of the covariance matrix that would describe the shape of the observed time series as in Cerioli et al. (2005)[26]. The problem is that these eigenvectors may bee too noisy, meaning that in high dimension the "space" between points stretches making the true covariance matrix look essentially uniform, thus very sensitive to noise. To overcome the computational difficulties of the integration in the the FPCA expression, one can rely on three approaches:

- Discretization: one performs FPCA similarly to PCA, except that after the decomposition one has to renormalize the eigenvectors and interpolate them with a suitable smoother.
- Basis function expansion: one can express each function (time series) as a linear combination of basis functions $f_t(x) \approx \sum_{k=1}^{K} \beta_{t,k}\phi_k(x)$, and approximating each function with a finite number of basis functions.
- Numerical Approximation: one uses quadrature rules to approximate FPCA.

In this application I rely on the second approach, applying Penalized Spline smoother to the stock price time series before performing PCA.

2.3.2 P-spline times series filtering

The main idea of smoothing (filtering) is to decompose the times series y_t in two components, one identifiable as a long phase variation (or trend) g_t, the other as residuals or unexplained short term variation ε_t, applying a suitable smoother (filter) to extract g_t and ε_t. In the approach applied in this chapter let $B(t)$ denote a rich spline basis with support over the observed time points t. A simple possible choice is to use the truncated polynomials in the form

$$B(t) = (1, t, \dots, t^q, (t - \tau_1)_+^q, \dots, (t - \tau_p)_+^q), \tag{2.4}$$

where q is the degree of the highest polynomial and $(t - \tau_i)_+ = t$ for $t > 0$ and $(t - \tau_i)_+ = 0$ otherwise and the knots are equidistantly chosen to cover the range of time points t. In this framework one smooths the time series y_t such as

$$y_t = g_t + \varepsilon_t = B(t)\theta + \varepsilon_t. \quad (2.5)$$

The Mixed Model interpretation of P-splines, that is very familiar to econometricians because it connects P-spline with others widely used filters like Hodrick-Prescott and Band-pass filters[88], is to furtherly decompose the basis function with low and high dimensional components, $B(t) = \{X(t), Z(t)\}$ reformulating (2.5) in the following way

$$y_t = B(t)\theta + \varepsilon_t = X(t)\beta + Z(t)u + \varepsilon_t,$$

with $\varepsilon \sim N(0, \sigma_\varepsilon^2 R_\epsilon$, where R_ϵ is a stationary correlation matrix. This means to impose a penalty on u leading to the penalized least square

$$l(\beta, u; h) = \{Y - B(t)\theta\}^T R_\varepsilon^{-1}\{Y - B(t)\theta\} + \frac{1}{2}\lambda u^t D u, \quad (2.6)$$

where D is a penalty matrix. The Lagrange penalty operator λ is the crucial parameter in this procedure; steering the amount of penalization, its selection provides an huge advantage of P-spline applications over other smoothing technique. Indeed, thinking about the penalty in (2.6) as a priori normal distribution and postulating normality for the residuals leads to a linear Mixed Model

$$Y|u \sim N(X\beta + Zu, \sigma_\varepsilon^2 R_\varepsilon), \; u \sim N(0, \sigma_u^2 D^-)$$

with X and Z as design matrices built from rows $X(t)$ and $Z(t)$ with $t = 1, 2, 3, \ldots$, D^- as generallized inverse of D and smoothing coefficient $\lambda = \sigma_\varepsilon^2/\sigma_u^2$. What does this mean? It means that if λ is well estimated, the estimate of g_t through $X(t)\hat{\beta}+Z(t)\hat{u}$ with $\hat{u}$ as the Best Linear Unbiased Predictor(BLUP)[96, 89]. This is an important advantage for P-splines smoothing, because it means to achieve good estimates results nevertheless the specification adopted, a property that doesn't hold for other smoothing techniques [113].

L- and V-curves for optimal λ selection

The λ selection is a crucial aspect of P-spline smoothing, in my application I make use of a recent development, the V-curve as described in Frasso and Eilers(2015)[57] because it handles very well serial correlation, thus having a preferential role in time series filtering. The best value of λ is determined from the data, Hansen (1992)[67] proposes the L-curve, a plot of $\log(\|y - X\theta\|^2)$ against $\log(\|\theta\|^2)$ for a grid of value of $\log(\lambda)$. If the spacing of the grid is fine, the plotted dots present a "curve" and Hansen advises to choose the λ corresponding to the corner and found good results. Frasso and Eilers (2015)[57] explores the L-curve for P-spline, plotting $\psi(\lambda) = \log(\|y - B\theta\|^2)$ against $\phi(\lambda) = \log(\|D\theta\|^2)$. They claim that no meaningful trend can be obtained with other selection techniques, such as leave-one-out cross-validation, generalized cross-validation or Akaike Information Criterion, and that the results of mixed model

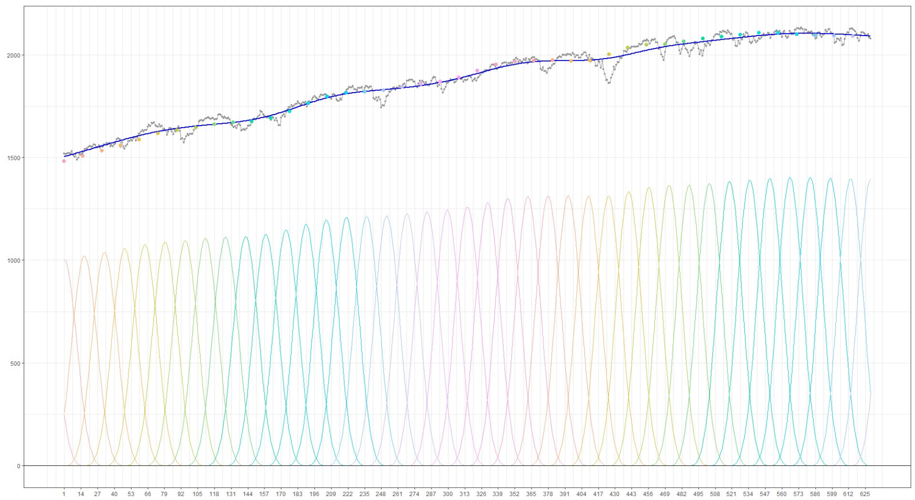


Figure 2.1: The core idea of P-spline: a sum of B-spline basis function with gradually changing heights. The grey dots show S&P500 value over the observation window, the large dots the B-spline coefficients (that have the same color as the splines) and the blue curve shows the P-spline fit.

based approaches are outperformed. Since the curvature of the L-curve can be computed using:

$$k(\lambda) = \frac{\psi'(\lambda)\phi''(\lambda) - \psi''(\lambda)\phi'(\lambda)}{[\psi'(\lambda)^2 + \phi'(\lambda)^2]^{3/2}}$$

in the end the V-curve is the function of the distance between points on the L-curve against the geometric mean of their lambdas.

2.4 Results

The efficacy of the proposal is tested on a dataset containing the price series of 471 actively traded stocks included in the S&P500 in the period comprised between 08/02/2013 and 10/08/2015 (8-th February 2013 and 10-th August 2015) including 629 trading days. The resulting portfolio performance were measured over a period comprised between 11/08/2015 and 07/02/2018 (11-th August 2015 and 7-th July 2018). The same experiment is conducted with an adaptation of a cointegration model presented in Sant'Anna et al.(2017)[125] and sketched in the appendix of this chapter. To summarize the content of section 3, the selection is implemented in a four step procedure:

1. P-spline smoothing over the log-transformed stock price time series with V-curve λ selection;

2. PCA over the smoothed dataset;

3. Selection of the principal components by evaluation of their effects on a linear regression model against the index to replicate;
4. Selection of the top absolute contributors to the principal component, according to their *loadings*[2] and inclusion in the portfolio according to an equally weighted scheme.

I performed PCA over the smoothed set of prices series, allowing the computation of 471 components, one for each series. Naturally, because the original data variance is sequentially decomposed over the different components, is not surprising that almost all the components have been discarded, since the first component alone explains the 67% of the variance in the dataset and first four components account for the 94%.

	PC1	PC2	PC3
Standard deviation	17.8472	8.6963	5.5133
Proportion of Variance	0.6777	0.1609	0.0647
Cumulative Proportion	0.6777	0.8386	0.9033

Table 2.1: Standard deviation of the fist three principal component, the original variance proportion showed by each component and its sum.

Variables	Model 1	Model 2	Model 3
	PC1	PC1+PC2	PC1+PC2+PC3
S&P500	1871.44 ***	1871.44 ***	1871.44 ***
	(1.971)	(1.453)	(1.450)
PC1	9.914 ***	9.914 ***	9.914 ***
	(0.111)	(0.082)	(0.081)
PC2		3.842 ***	3.842 ***
		(0.167)	(0.167)
PC3			-0.515
			(0.263)
N	629	629	629
R2	0.928	0.961	0.961
	*** $p < 0.001$;	** $p < 0.01$;	* $p < 0.05$.

Table 2.2: Results summury of the regression of the following three models on the *training* data.

The regressions of the S&P500 Index value against the selected principal com-

[2]Rearranging(2.1) one obtains

$$\mathbf{\Phi_1} = \beta_{1,1}x_1 + \beta_{1,2}x_2 + \cdots + \beta_{1,N}x_N,$$

then $\beta_{i,1}$ is the *loading* of the i-th variable on the first principal component.

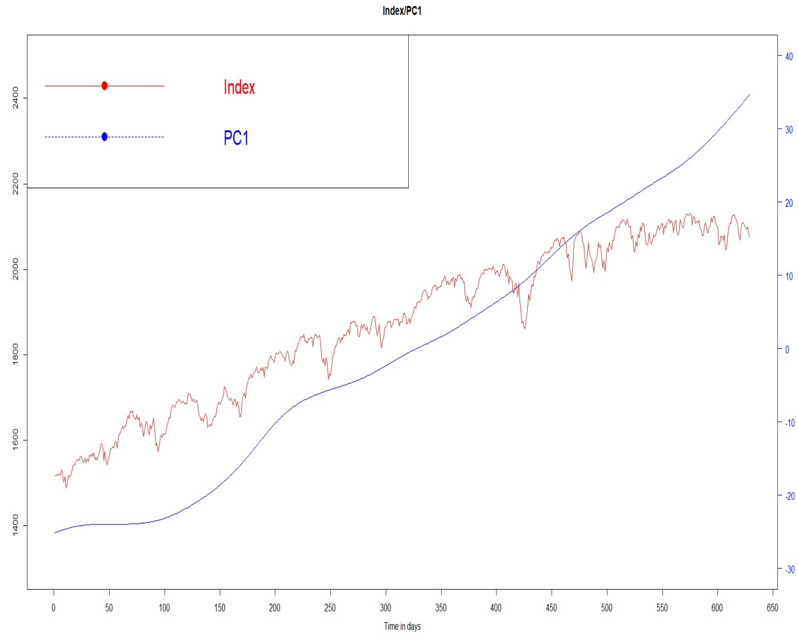


Figure 2.2: This plot overlaps the first principal component and the index value.

ponents suggest that the first principal component is able to capture almost the 97% of the variation in the dependent variable, this seems not reasonable since the dependent variable is raw while the components are the result of the linear combination of smoothed series, however the incredibly narrow confidence interval and a graphical inspection allows for a safe concordance statement between the two series. The same conclusion is supported by performing linear regression of the S&P500 Index over the three principal components on the testing period.

Variables	Model 1	Model 2	Model 3
	PC1	PC1+PC2	PC1+PC2+PC3
S&P500	2252.22 ***	2252.22 ***	2252.22 ***
	(2.493)	(2.185)	(2.113)
PC1	12.599 ***	12.599 ***	12.599 ***
	(0.140)	(0.123)	(0.119)
PC2		-3.471 ***	-3.471 ***
		(0.251)	(0.243)
PC3			2.559 ***
			(0.384)
N	629	629	629
R2	0.928	0.945	0.945
	*** $p < 0.001$;	** $p < 0.01$;	* $p < 0.05$.

Table 2.3: Results summury of the three principal components regression models on the *testing* data.

According to [76] the first principal component can be interpreted as a *long term trend component* in the dataset, so it fits well to the purpose of Index Replication, the same source states that the second component may have the interpretaion as a *shock component* and may be interestingin future research, to test is utility for replicating the second order moment of the Index distribution. The selection procedure continues by sorting the variables (in this case the stocks) in decreasing order of the absolute value of their *loadings* on the selected principal component, I decide to use the absolute value because the principal component is affected either by stocks with a positive loading than by stocks with a negative loading. I select the top four contributors to the first component.

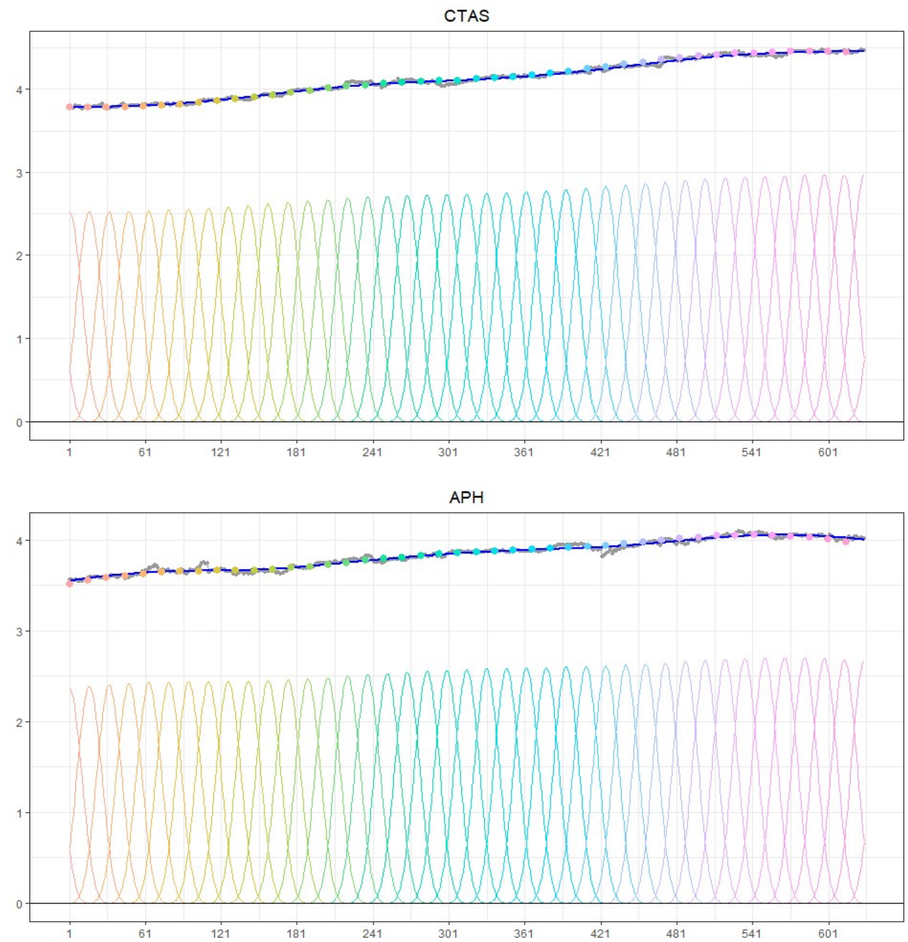


Figure 2.3: Plot showing the first two first principal component contributors, the connected grey dots show the row log-price values, the blue lines show the P-spline fitted value and the large dots the B-spline coefficients..

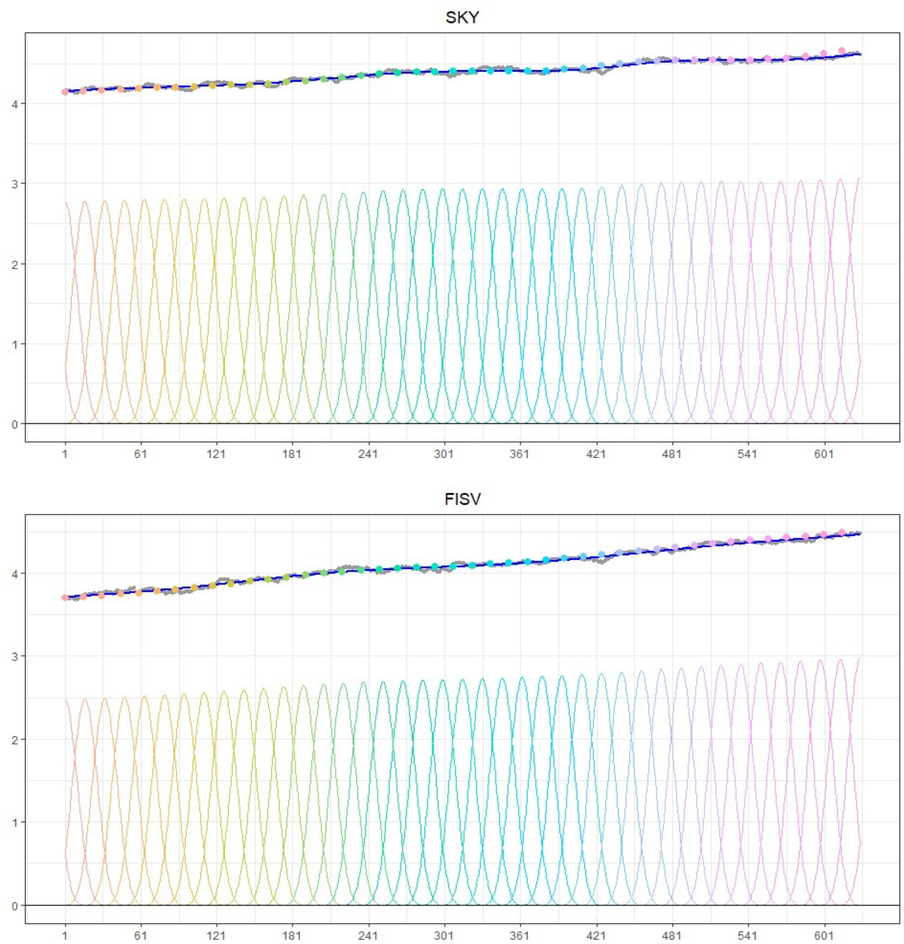


Figure 2.4: Plot showing third and fourth first principal component contributors, the connected grey dots show the row log-price values, the blue lines show the P-spline fitted value and the large dots the B-spline coefficients (they have the same colors as the corresponding splines). The horizontal locations of these dots correspond to the knots where the polynomial segments of the B-spline join.

The goodness of the selection procedure can be seen also with a regression exercise, as shown in table 4 and 5.
The selected stocks have been used as components of an index tracking portfolio with equally weighted scheme. To evaluate the performance of the resulting portfolio I computed, over the *testing* period the following indicators[3]:

- Annual average return;
- Cumulative return;
- Annual volatility;
- Average Tracking Error:

$$\overline{TE}_t = \frac{\sum_{t=1}^{T} \sum_{i}^{N} (w_i r_{i,t} - R_t)}{T};$$

- Tracking Error Variance:

$$\sigma^2_{TE} = \frac{\sum_{t=1}^{T} \left[TE_t - \overline{TE}_t\right]^2}{T};$$

- Sharpe Ratio[4]:

$$SR = \frac{\sum_{t=1}^{T} (R_{p,t} - R_t)}{T\sqrt{\sigma_p^2 - \sigma_R^2}};$$

[3]The expressions are general formulas for portfolios of N assets, trying to replicate the returns of a benchmark R_t, which performance has been measured over a period of length T.

[4]A modified version of the Sharpe Ratio to account for the volatility of the tracking benchmark

Variables	Model 1
(Intercept)	4.811 ***
	(0.049)
CTAS	-0.135 ***
	(0.027)
APH	0.347 ***
	(0.019)
SYK	0.275 ***
	(0.025)
FISV	0.180 ***
	(0.024)
N	629
R^2	0.972
*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.	

Table 2.4: Results of the regression on training data.

Variables	Model 1
(Intercept)	4.972 ***
	(0.033)
CTAS	0.198 ***
	(0.016)
APH	0.193 ***
	(0.021)
SYK	-0.026
	(0.021)
FISV	0.239 ***
	(0.022)
N	630
R2	0.972
*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.	

Table 2.5: Results of the regression on testing data.

Table 6 shows the performance of the resulting portfolio and compares them with those obtained with the cointegration approach [5] and with the tracked benchmark itself.

	$\bar{R}_p^y$	Cum. Ret.	$\bar{\sigma}_p^y$	$\overline{TE}$	σ_{TE}	SR
PS-FPCR	18.76 %	47.00 %	14 %	0.033 %	0.005	0.07
Coint	7.54 %	18.83 %	25.26 %	0.00	2.86	0
S&P500	10.43 %	26.00 %	10%	–	–	–

Table 2.6

2.5 Conclusions

In this chapter I have presented the application of a P-Spline Functional Principal Component Regression to the portfolio selection for index tracking. The results show that, while the P-Spline filtering allows a meaningful principal component extraction and the selected principal component is reliably able to track the performance of the selected benchmark, the selected portfolio is not sharply achieving its purpose because while the tracking performance shows a good tracking, financial performance shows huge divergences between the portfolio and the benchmark. In this case this doesn't seem to be a problem because the selected portfolio almost double the financial performance of the index (even if with a slight increase in volatility), but this maybe the result of the general state of the market, implying that if the benchmark had performed negatively during the observed period, then the selected portfolio may had doubled its loss. What clearly can be understood by my analysis is that the cointegration approach as presented in [4, 125] is not a valid approach for benchmark tracking. First of all because even if the residuals are stationary, this doesn't remove the effect of multicollinearity on the regression coefficients and their normalization for weights selection is not based on information gained from the data, due to the ambiguous ripartition of the observed variation between collinear covariates. Second, the stock picking selection is too random and results in a set of portfolio which presents too much variation, making the selection of the right stocks dangerous and unreliable. As Table 7 shows, the worst 25% of the portfolios selected with the cointegration method yielded average yearly return lower than -28.50% and cumulated return over the whole period lower than -71.25%.

[5] those are the average results of the approach, because it randomly generates several subsets.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
$\bar{R}_p^y$	-308.70%	-28.50%	5.78%	7.54%	43.16%	264.70%
Cum. Ret.	-771.00%	-71.25%	14.00%	18.83%	108.00%	661.00%

Table 2.7: The average yearly return and the cumulative return over the testing period of all the portfolios generated with the cointegration method.

2.6 Appendix I: Cointegration Based Portfolio Selection Algorithm

Here I present the cointegration algorithm that I mutated from [4, 125].

1. Estimation, for each stock included in the sample, of the model

$$\log(I_t) = \beta_0 + \log(P_{i,t})\beta_i + \varepsilon_t$$

   where I_t is the index value at time t and $P_{i,t}$ is the price of stock i at time t.

2. Augmented Dicky Fuller test performance for all the residuals of the former models, with the null of stationarity. Exclusion from the sample of all the stocks that don't cointegrate with the index.

3. Extraction of 1000[6] subsample, each including 10 elements from the principal sample.

4. Estimation for each subsample of the models

$$\log(I_t) = \beta_0 + \sum_{i=1}^{10} \log(P_{i,t})\beta_i + \varepsilon_t$$

$$\epsilon_t = \alpha + \rho\varepsilon_{t-1} + \sum_{i=1}^{d} \gamma\Delta\varepsilon_{t-1} + u_t \tag{2.7}$$

   Where $\Delta = 1$.

5. Augmented Dicky Fuller test on the residuals u_t and exclusion of the failing models.

6. Normalization of the absolute value[7] of the regression coefficient and portfolio construction according to these weights.

[6] The original procedure requires 100000 generations but due to hardware constraints I have to reduce this number.

[7] The original procedure allows coefficient to be negative, but in this context these generated portfolios with negative values.

Explicit

Noise has a central role either in statistics than in economics, the main purpose of this scritp has been to practically show the effectivness of P-spline smoothing in the solution of practical relevance econometrics issues in portfolio selection problems.
Chapter 1 provided a detailed description of the hedge ratio estimation under mean-variance and pessimistic frameworks and the results of the application of P-spline quantile regression in this task, from the statistical point of view results showed the difference in the behaviour of spot and future return distributions at different quantile, but the results were poor when applied to predict a useful hedge ratio for future periods.
Chapter 2 provided a detailed description of statistical portfolio estimation procedure using different methodologies in the attempt to produce an index tracking portfolio in an high-dimensional context. The results of the P-spline Functional Principal Component Regression exercise were solid from the statistical point of view, meaning that I was able to identify a small subset of index components able to track the index behaviour in the training period, however from the economics point of view, the result of the portfolio selection are not sharp, because I achieved an *enhanced index tracking* performance.
For both the research lines the question is yet to be answered and I provided what I believe to be their further developments at the end of each chapter.

Bibliography

[1] Carlo Acerbi and Dirk Tasche. On the coherence of expected shortfall. *Journal of Banking & Finance*, 26(7):1487–1503, 2002.

[2] Carol Alexander. Optimal hedging using cointegration. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 357(1758):2039–2058, 1999.

[3] Carol Alexander and Anca Dimitriu. Sources of outperformance in equity markets. *The Journal of Portfolio Management*, 30(4):170–185, 2004.

[4] Carol Alexander and Anca Dimitriu. Indexing and statistical arbitrage. *The Journal of Portfolio Management*, 31(2):50–63, 2005.

[5] Carol Alexander and Anca Dimitriu. Indexing, cointegration and equity market regimes. *International Journal of Finance & Economics*, 10(3):213–231, 2005.

[6] Yudhie Andriyana, Irène Gijbels, and Anneleen Verhasselt. P-splines quantile regression estimation in varying coefficient models. *Test*, 23(1):153–194, 2014.

[7] Yudhie Andriyana, Irène Gijbels, and Anneleen Verhasselt. Quantile regression in varying-coefficient models: non-crossing quantile curves and heteroscedasticity. *Statistical Papers*, 59(4):1589–1621, 2018.

[8] Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measures of risk. *Mathematical finance*, 9(3):203–228, 1999.

[9] Richard T Baillie and Robert J Myers. Bivariate garch estimation of the optimal commodity futures hedge. *Journal of Applied Econometrics*, 6(2):109–124, 1991.

[10] Massimiliano Barbi and Silvia Romagnoli. A copula-based quantile risk measure approach to estimate the optimal hedge ratio. *Journal of Futures Markets*, 34(7):658–675, 2014.

[11] Nicolas Basalto, Roberto Bellotti, Francesco De Carlo, Paolo Facchi, Ester Pantaleo, and Saverio Pascazio. Hausdorff clustering of financial time

series. *Physica A: Statistical Mechanics and its Applications*, 379(2):635–644, 2007.

[12] Gilbert W Bassett Jr, Roger Koenker, and Gregory Kordas. Pessimistic portfolio allocation and choquet expected utility. *Journal of financial econometrics*, 2(4):477–492, 2004.

[13] Michael J Best and Robert R Grauer. On the sensitivity of mean-variance-efficient portfolios to changes in asset means: some analytical and computational results. *The review of financial studies*, 4(2):315–342, 1991.

[14] Michael J Best and Robert R Grauer. Sensitivity analysis for mean-variance portfolio problems. *Management Science*, 37(8):980–989, 1991.

[15] Fischer Black and Robert Litterman. Global portfolio optimization. *Financial analysts journal*, 48(5):28–43, 1992.

[16] Kaatje Bollaerts, Paul HC Eilers, and Marc Aerts. Quantile regression with monotonicity restrictions using p-splines and the l1-norm. *Statistical Modelling*, 6(3):189–207, 2006.

[17] Tim Bollerslev. Modelling the coherence in short-run nominal exchange rates: a multivariate generalized arch model. *The review of economics and statistics*, pages 498–505, 1990.

[18] Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[19] Stephen J Brown. The portfolio choice problem: Comparison of certainty equivalence and optimal bayes portfolios. *Communications in Statistics-Simulation and Computation*, 7(4):321–334, 1978.

[20] Renato Bruni, Francesco Cesarone, Andrea Scozzari, Fabio Tardella, et al. A new stochastic dominance approach to enhanced index tracking problems. *Economics Bulletin*, 32(4):3460–3470, 2012.

[21] Zongwu Cai. Regression quantiles for time series. *Econometric theory*, pages 169–192, 2002.

[22] Zongwu Cai and Xian Wang. Nonparametric estimation of conditional var and expected shortfall. *Journal of Econometrics*, 147(1):120–130, 2008.

[23] Jorge Caiado and Nuno Crato. Identifying common dynamic features in stock returns. 10(7):797–807. Publisher: Routledge _eprint: https://doi.org/10.1080/14697680903567152.

[24] Zhiguang Cao, Richard DF Harris, and Jian Shen. Hedging and value at risk: A semi-parametric approach. *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, 30(8):780–794, 2010.

[25] Stephen G Cecchetti, Robert E Cumby, and Stephen Figlewski. Estimation of the optimal futures hedge. *The Review of Economics and Statistics*, pages 623–630, 1988.

[26] Andrea Cerioli, Fabrizio Laurini, and Aldo Corbellini. Functional cluster analysis of financial time series. In *New Developments in Classification and Data Analysis*, pages 333–341. Springer, 2005.

[27] Arthur Charpentier and Emmanuel Flachaire. Pareto models for top incomes. 2019.

[28] Chen Chen and Roy H Kwon. Robust portfolio selection for index tracking. *Computers & Operations Research*, 39(4):829–837, 2012.

[29] Nai-Fu Chen, Richard Roll, and Stephen A Ross. Economic forces and the stock market. *Journal of business*, pages 383–403, 1986.

[30] Sheng-Syan Chen, Cheng-Few Lee, and Keshab Shrestha. On a mean—generalized semivariance approach to determining the hedge ratio. *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, 21(6):581–598, 2001.

[31] C Sherman Cheung, Clarence CY Kwan, and Patrick CY Yip. The hedging effectiveness of options and futures: A mean-gini approach. *Journal of Futures Markets*, 10(1):61–73, 1990.

[32] Vijay K Chopra and William T Ziemba. The effect of errors in means, variances, and covariances on optimal portfolio choice. In *Handbook of the fundamentals of financial decision making: Part I*, pages 365–373. World Scientific, 2013.

[33] Gustave Choquet. Theory of capacities. In *Annales de l'institut Fourier*, volume 5, pages 131–295, 1954.

[34] WL Chou, KK Fan Denis, and Cheng F Lee. Hedging with the nikkei index futures: The convential model versus the error correction model. *The Quarterly Review of Economics and Finance*, 36(4):495–505, 1996.

[35] Gregory Connor. Sensible return forecasting for portfolio management. *Financial Analysts Journal*, 53(5):44–51, 1997.

[36] Gregory Connor and Robert A Korajczyk. Risk and return in an equilibrium apt: Application of a new test methodology. *Journal of financial economics*, 21(2):255–289, 1988.

[37] Jon Danielsson, Lerby Murat Ergun, Laurens de Haan, and Casper G de Vries. Tail index estimation: Quantile driven threshold selection. *Available at SSRN 2717478*, 2016.

[38] James EH Davidson, David F Hendry, Frank Srba, and Stephen Yeo. Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the united kingdom. *The Economic Journal*, pages 661–692, 1978.

[39] Cristina Davino, Marilena Furno, and Domenico Vistocco. *Quantile regression: theory and applications*, volume 988. John Wiley & Sons, 2013.

[40] Arnold LM Dekkers, John HJ Einmahl, and Laurens De Haan. A moment estimator for the index of an extreme-value distribution. *The Annals of Statistics*, pages 1833–1855, 1989.

[41] David A Dickey and Wayne A Fuller. Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica: journal of the Econometric Society*, pages 1057–1072, 1981.

[42] Christian Dose and Silvano Cincotti. Clustering of financial time series with application to index and enhanced index tracking portfolio. *Physica A: Statistical Mechanics and its Applications*, 355(1):145–151, 2005.

[43] Christian L Dunis and Richard Ho. Cointegration portfolios of european equities for index tracking and market neutral strategies. *Journal of Asset Management*, 6(1):33–52, 2005.

[44] Louis H Ederington. The hedging performance of the new futures markets. *The journal of finance*, 34(1):157–170, 1979.

[45] NCP Edirisinghe. Index-tracking optimal portfolio selection. *Quantitative Finance Letters*, 1(1):16–20, 2013.

[46] Ward Edwards. The theory of decision making. *Psychological bulletin*, 51(4):380, 1954.

[47] Paul HC Eilers and Renée X De Menezes. Quantile smoothing of array cgh data. *Bioinformatics*, 21(7):1146–1153, 2005.

[48] Robert Engle. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3):339–350, 2002.

[49] Robert Engle. *Anticipating correlations*. Princeton University Press, 2009.

[50] Robert F Engle and Clive WJ Granger. Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, pages 251–276, 1987.

[51] Robert F Engle and Simone Manganelli. Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of business & economic statistics*, 22(4):367–381, 2004.

[52] Eugene F. Fama. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2):383–417, 1970.

[53] Eugene F. Fama and Kenneth R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56, 1993.

[54] Yong Fang and Shou-Yang Wang. A fuzzy index tracking portfolio selection model. In *International Conference on Computational Science*, pages 554–561. Springer, 2005.

[55] Laurent Favre and José-Antonio Galeano. Mean-modified value-at-risk optimization with hedge funds. *The journal of alternative investments*, 5(2):21–25, 2002.

[56] Sergio M Focardi and Frank J Fabozzi 3. A methodology for index tracking based on time-series clustering. *Quantitative Finance*, 4(4):417–425, 2004.

[57] Gianluca Frasso and Paul HC Eilers. L-and v-curves for optimal smoothing. *Statistical Modelling*, 15(1):91–111, 2015.

[58] Milton Friedman and Leonard J Savage. The utility analysis of choices involving risk. *Journal of political Economy*, 56(4):279–304, 1948.

[59] Anqi Fu, Balasubramanian Narasimhan, and Stephen Boyd. Cvxr: An r package for disciplined convex optimization. *arXiv preprint arXiv:1711.07582*, 2017.

[60] Alexei A Gaivoronski, Sergiy Krylov, and Nico Van der Wijst. Optimal portfolio selection and dynamic benchmark tracking. *European Journal of operational research*, 163(1):115–131, 2005.

[61] Marco Geraci and Matteo Bottai. Quantile regression for longitudinal data using the asymmetric laplace distribution. *Biostatistics*, 8(1):140–154, 2007.

[62] Marco Geraci and Matteo Bottai. Linear quantile mixed models. *Statistics and computing*, 24(3):461–479, 2014.

[63] Donald Goldfarb and Garud Iyengar. Robust portfolio selection problems. *Mathematics of operations research*, 28(1):1–38, 2003.

[64] M Ivette Gomesa and M João Martins. "asymptotically unbiased" estimators of the tail index based on external estimation of the second order parameter. *Extremes*, 5(1):5–31, 2002.

[65] Clive WJ Granger. *Co-integrated variables and error-correcting models.* PhD thesis, Discussion Paper 83-13. Department of Economics, University of California at . . . , 1983.

[66] G. Guastaroba and M. G. Speranza. Kernel search: An application to the index tracking problem. 217(1):54–68.

[67] Per Christian Hansen. Analysis of discrete ill-posed problems by means of the l-curve. *SIAM review*, 34(4):561–580, 1992.

[68] Richard DF Harris and Jian Shen. Hedging and value at risk. *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, 26(4):369–390, 2006.

[69] Bruce M Hill. A simple general approach to inference about the tail of a distribution. *The annals of statistics*, pages 1163–1174, 1975.

[70] Harold Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417, 1933.

[71] Charles T Howard and Louis J D'Antonio. A risk-return measure of hedging effectiveness. *Journal of Financial and Quantitative Analysis*, pages 101–112, 1984.

[72] Charles T Howard and Louis J D'Antonio. Multiperiod hedging using futures: A risk minimization approach in the presence of autocorrelation. *The Journal of Futures Markets (1986-1998)*, 11(6):697, 1991.

[73] Charles T Howard and Louis J D'Antonio. The cost of hedging and the optimal hedge ratio. *The Journal of Futures Markets (1986-1998)*, 14(2):237, 1994.

[74] Jinbo Huang, Ashley Ding, and Yong Li. Nonparametric kernel method to hedge downside risk. *International Review of Finance*, 19(4):929–944, 2019.

[75] Jui-Cheng Hung, Chien-Liang Chiu, and Ming-Chih Lee. Hedging with zero-value at risk hedge ratio. *Applied Financial Economics*, 16(3):259–269, 2006.

[76] Salvatore Ingrassia and G Damiana Costanzo. Functional principal component analysis of financial time series. In *New developments in classification and data analysis*, pages 351–358. Springer, 2005.

[77] Carmela Iorio, Massimo Aria, Antonio D'Ambrosio, and Roberta Siciliano. Informative trees by visual pruning. *Expert Systems with Applications*, 127:228–240, 2019.

[78] Carmela Iorio, Gianluca Frasso, Antonio D'Ambrosio, and Roberta Siciliano. Parsimonious time series clustering using p-splines. *Expert Systems with Applications*, 52:26–38, 2016.

[79] Carmela Iorio, Gianluca Frasso, Antonio D'Ambrosio, and Roberta Siciliano. A p-spline based clustering approach for portfolio selection. *Expert Systems with Applications*, 95:88–103, 2018.

[80] Roland Jeurissen and Jan van den Berg. Optimized index tracking using a hybrid genetic algorithm. In *2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence)*, pages 2327–2334. ISSN: 1941-0026.

[81] J David Jobson and Bob Korkie. Estimation for markowitz efficient portfolios. *Journal of the American Statistical Association*, 75(371):544–554, 1980.

[82] J David Jobson and Robert M Korkie. Putting markowitz theory to work. *The Journal of Portfolio Management*, 7(4):70–74, 1981.

[83] JD Jobson. Improved estimation for markowitz portfolios using james-stein type estimators. In *Proceedings of the American Statistical Association, Business and Economics Statistics Section*, volume 71, pages 279–284, 1979.

[84] Soren Johansen and Katarina Juselius. Maximum likelihood estimation and inference on cointegration—with appucations to the demand for money. *Oxford Bulletin of Economics and statistics*, 52(2):169–210, 1990.

[85] L.L. Johnson. The theory of hedging and speculation in commodity futures. *Review of Economic Studies*, 27:139–151, 1960.

[86] Philippe Jorion. Bayes-stein estimation for portfolio analysis. *Journal of Financial and Quantitative analysis*, pages 279–292, 1986.

[87] Dhanya Jothimani, Ravi Shankar, and Surendra S. Yadav. A PCA-DEA framework for stock selection in indian stock market. 12(3):386–403. Place: Bingley Publisher: Emerald Group Publishing Ltd WOS:000412339600004.

[88] Goeran Kauermann, Tatyana Krivobokova, and Willi Semmler. Filtering time series with penalized splines. *Studies in Nonlinear Dynamics & Econometrics*, 15(2), 2011.

[89] Göran Kauermann, Tatyana Krivobokova, and Ludwig Fahrmeir. Some asymptotic results on generalized penalized spline smoothing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):487–503, 2009.

[90] Roger W Klein and Vijay S Bawa. The effect of estimation risk on optimal portfolio choice. *Journal of financial economics*, 3(3):215–231, 1976.

[91] Roger Koenker. Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, 91(1):74–89, 2004.

[92] Roger Koenker and Kevin F Hallock. Quantile regression. *Journal of economic perspectives*, 15(4):143–156, 2001.

[93] Roger Koenker, Pin Ng, and Stephen Portnoy. Quantile smoothing splines. *Biometrika*, 81(4):673–680, 1994.

[94] Roger Koenker and Frank Schorfheide. Quantile spline models for global temperature change. *Climatic change*, 28(4):395–404, 1994.

[95] Robert W Kolb and John Okunev. An empirical evaluation of the extended mean-gini coefficient for futures hedging. *The Journal of Futures Markets (1986-1998)*, 43(1):177, 1992.

[96] Tatyana Krivobokova and Göran Kauermann. A note on penalized spline smoothing with correlated errors. *Journal of the American Statistical Association*, 102(480):1328–1337, 2007.

[97] Nathan Lassance and Frederic Vrins. Portfolio selection with parsimonious higher comoments estimation *. 126:106115. Place: Amsterdam Publisher: Elsevier WOS:000637969200006.

[98] Olivier Ledoit and Michael Wolf. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of empirical finance*, 10(5):603–621, 2003.

[99] Olivier Ledoit and Michael Wolf. Honey, i shrunk the sample covariance matrix. *The Journal of Portfolio Management*, 30(4):110–119, 2004.

[100] Olivier Ledoit, Michael Wolf, et al. Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics*, 40(2):1024–1060, 2012.

[101] Hsiang-Tai Lee and Jonathan Yoder. Optimal hedging with a regime-switching time-varying correlation garch model. *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, 27(5):495–516, 2007.

[102] Donald Lien and Xiangdong Luo. Estimating multiperiod hedge ratios in cointegrated markets. *Journal of Futures Markets*, 13(8):909–920, 1993.

[103] Donald Lien and Xiangdong Luo. Estimating the extended mean-gini coefficient for futures hedging. *Journal of Futures Markets*, 13(6):665–676, 1993.

[104] Donald Lien, Keshab Shrestha, and Jing Wu. Quantile estimation of optimal hedge ratio. *Journal of Futures Markets*, 36(2):194–214, 2016.

[105] Donald Lien, Ziling Wang, and Xiaojian Yu. Optimal quantile hedging under markov regime switching. *Empirical Economics*, pages 1–25, 2020.

[106] Harry Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

[107] Robert C Merton. Lifetime portfolio selection under uncertainty: The continuous-time case. *The review of Economics and Statistics*, pages 247–257, 1969.

[108] Robert C Merton. An intertemporal capital asset pricing model. *Econometrica: Journal of the Econometric Society*, pages 867–887, 1973.

[109] Richard O Michaud. The markowitz optimization enigma: Is 'optimized'optimal? *Financial analysts journal*, 45(1):31–42, 1989.

[110] Kees van Montfort, Elout Visser, and Laurens Fijn van Draat. Index tracking by means of optimized sampling. 34(2):143–152.

[111] Vito MR Muggeo, Federico Torretta, Paul HC Eilers, Mariangela Sciandra, and Massimo Attanasio. Multiple smoothing parameters selection in additive regression quantiles. *Statistical Modelling*, page 1471082X20929802, 2020.

[112] Robert J Myers and Stanley R Thompson. Generalized optimal hedge ratio estimation. *American Journal of Agricultural Economics*, 71(4):858–868, 1989.

[113] Jean Opsomer, Yuedong Wang, and Yuhong Yang. Nonparametric regression with correlated errors. *Statistical Science*, pages 134–153, 2001.

[114] Giuseppe Pandolfo, Carmela Iorio, Roberta Siciliano, and Antonio D'Ambrosio. Robust mean-variance portfolio through the weighted $\mathcal{L}_p$ depth function. *Annals of Operations Research*, 292(1):519–531, 2020.

[115] Giuseppe Pandolfo, Carmela Iorio, Michele Staiano, Massimo Aria, and Roberta Siciliano. Multivariate process control charts based on the lp depth. *Applied Stochastic Models in Business and Industry*, 37(2):229–250, 2021.

[116] L'uboš Pástor. Portfolio selection and asset pricing models. *The Journal of Finance*, 55(1):179–223, 2000.

[117] Francesco Pattarin, Sandra Paterlini, and Tommaso Minerva. Clustering financial time series: an application to mutual funds style analysis. 47(2):353–372.

[118] Karl Pearson. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901.

[119] Peter CB Phillips and Pierre Perron. Testing for a unit root in time series regression. *Biometrika*, 75(2):335–346, 1988.

[120] E. R. Pike and E. G. Klepfish. *The analysis of financial time series data by independent component analysis.* Springer-Verlag Tokyo. Pages: 174-180 Publication Title: Application of Econophysics, Proceedings WOS:000187421800024.

[121] John Quiggin. A theory of anticipated utility. *Journal of economic behavior & organization*, 3(4):323–343, 1982.

[122] R Tyrrell Rockafellar and Stanislav Uryasev. Conditional value-at-risk for general loss distributions. *Journal of banking & finance*, 26(7):1443–1471, 2002.

[123] David Ruppert. Selecting the number of knots for penalized splines. *Journal of computational and graphical statistics*, 11(4):735–757, 2002.

[124] Paul A Samuelson. Lifetime portfolio selection by dynamic stochastic programming. *Stochastic Optimization Models in Finance*, pages 517–524, 1975.

[125] Leonardo R Sant'Anna, Tiago P Filomena, and João F Caldeira. Index tracking and enhanced indexing using cointegration and correlation with endogenous portfolio selection. *The Quarterly Review of Economics and Finance*, 65:146–157, 2017.

[126] EJ Schlossmacher. An iterative technique for absolute deviations curve fitting. *Journal of the American Statistical Association*, 68(344):857–859, 1973.

[127] David Schmeidler. Subjective probability and expected utility without additivity. *Econometrica: Journal of the Econometric Society*, pages 571–587, 1989.

[128] Peter S Sephton. Hedging wheat and canola at the winnipeg commodity exchange. *Applied Financial Economics*, 3(1):67–72, 1993.

[129] Haim Shalit et al. Mean-gini hedging in futures markets. *Journal of Futures Markets*, 15(6):617–636, 1995.

[130] Haim Shalit and Shlomo Yitzhaki. Mean-gini, portfolio theory, and the pricing of risky assets. *The journal of Finance*, 39(5):1449–1468, 1984.

[131] William F Sharpe. A simplified model for portfolio analysis. *Management science*, 9(2):277–293, 1963.

[132] William F Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, 19(3):425–442, 1964.

[133] William F Sharpe. Mutual fund performance. *The Journal of business*, 39(1):119–138, 1966.

[134] William F Sharpe. The sharpe ratio. *Journal of portfolio management*, 21(1):49–58, 1994.

[135] Keshab Shrestha, Ravichandran Subramaniam, Yessy Peranginangin, and Sheena Sara Suresh Philip. Quantile hedge ratio for energy markets. *Energy Economics*, 71:253–272, 2018.

[136] Mohommed M Siddiqui. Distribution of quantiles in samples from a bivariate population. *J. Res. Nat. Bur. Standards B*, 64:145–150, 1960.

[137] James H Stock and Mark W Watson. Testing for common trends. *Journal of the American statistical Association*, 83(404):1097–1107, 1988.

[138] FEDERICO TORRETTA. P-spline quantile regression: a new algorithm for smoothing parameter selection. 2016.

[139] Lan Wu, Yuehan Yang, and Hanzhong Liu. Nonnegative-lasso and application in index tracking. 70:116–126.

[140] Menahem E Yaari. The dual theory of choice under risk. *Econometrica: Journal of the Econometric Society*, pages 95–115, 1987.

[141] Haixiang Yao, Zhongfei Li, and Yongzeng Lai. Mean–cvar portfolio selection: A nonparametric estimation framework. *Computers & Operations Research*, 40(4):1014–1022, 2013.

[142] Keming Yu and Jin Zhang. A three-parameter asymmetric laplace distribution and its extension. *Communications in Statistics—Theory and Methods*, 34(9-10):1867–1879, 2005.

[143] Arnold Zellner and V Karuppan Chetty. Prediction and decision problems in regression models from the bayesian point of view. *Journal of the American Statistical Association*, 60(310):608–616, 1965.

Appendix A

Splines

A.1 Introduction

In this section I am going to discuss the study of nonparametric regression by way of smoothing splines. The target is to estimate the function g_0, typically assumed to be smooth and defined in some kind of Sobolev space $W^{m,p}(a,b)$, and I want to accomplish this task using the penalized smoothing spline estimators.
A natural measure of smoothness associated with a function is $\int g^{(m)}(x)^2 dx$, a natural measure of goodness-of-fit to the data is the residual sum-of-squares $n^{-1}\sum_{i=1}^{n}(y_i - g(x_i))^2$, thus an overall measure of quality of the candidate estimator g is provided by the sum:

$$(1-q)n^{-1}\sum_{i=1}^{n}(y_i - g(x_i))^2 + q\int g^{(m)}(x)^2 dx, \tag{A.1}$$

for some $0 < q < 1$. An "optimal" estimator should then be the one obtained by minimization of this functional over the function space. Being $\lambda = q/(1-q)$ the former operation becomes equivalent to study the function g_λ minimizing

$$n^{-1}\sum_{i=1}^{n}(y_i - g(x_i))^2 + \lambda\int g^{(m)}(x)^2 dx,\ \lambda > 0. \tag{A.2}$$

The result is the smoothing spline estimator of the regression function.
The λ parameter governs the trade-off between smoothness and goodness-of-fit, usually referred to as the *smoothing parameter*, when its value is large a premium is being placed on smoothness and potential estimators with large $m-th$ derivatives are penalized, while small value of λ corresponds to more emphasis on goodness-of-fit.
In the context of polynomial regression, through the application of Taylor The-

orem, one can rewrite the model as

$$y_i = \sum_{j=1}^{m} \theta_j x_i^{j-1} + Rem(x_i) + \epsilon_i, \ i = 1, ..., n, \tag{A.3}$$

with constants $\theta_1, ..., \theta_m$ and

$$Rem(x_i) = [(m-1)!]^{-1} \int g^{(m)}(\xi)(x-\xi)_+^{m-1} d\xi. \tag{A.4}$$

Then by Cauchy-Schwarz inequality one has

$$\max_{1 \le i \le n} Rem(x_i)^2 \le \frac{J_m(g)}{(2m-1)[(m-1)!]^2}, \tag{A.5}$$

where the numerator is the smoothness measure criterion, that is providing a bound on how far the regression function departs from the model. Knowing this before one could minimize $n^{-1}RSS(g)+\lambda(J_m(g)-c)$ (c is this latter bound) with λ that becomes the Lagrange multiplier for the constraint, producing the same estimator as (A.1), this leads to the following theorem (by Schoenberg(1964)).

Assume that $n \ge m$ and let $g(\cdot, c)$ be the minimizer of the RSS in $W_2^m[0,1]$ subject to $J_m(g) \le c$. Let g_λ be the minimizer of (A.1) in $W_2^m[0,1]$, then there is a computable constant c_0 such that the sets $\{g(\cdot, c) : 0 \le c \le c_0\}$ and $\{g_\lambda(\cdot) : 0 \le \lambda \le \infty\}$ are identical in that any value of λ there is a unique c such that $g_\lambda(\cdot) = g(\cdot, c)$ and conversely. If $c \le c_0$ then $J(g(\cdot, c)) = c$.

This theorem has the consequence that the solution to the constrained problem is a smoothing spline estimator of the function corresponding to λ. Thus, the choice of a particular value for λ implies the assumption about $J_m(g) \le c_\lambda$, with $c_\lambda = J_m(g_\lambda)$ reflecting the beliefs about the magnitude of the reminder terms and therefore giving an extension of polynomial regression estimator avoiding departures from the idealized polynomial regression model, furthermore, since smoothing spline are minimax estimators, provides protection against "worst case" departures.
Smoothing splines estimators also have a Bayesian interpretation always on the strand of polynomial regression. Assume that we observe responses at distinct design points, conditional on the value of a parameter vector $\beta = (\beta_1, ..., \beta_m)^T$ that satisfies

$$y_i = \sum_{j=1}^{m} \beta_j x_{j,i} + \epsilon_i, \ i = 1, ..., n, \ j = 1, ..., m \tag{A.6}$$

for $\epsilon = (\epsilon_1, ..., \epsilon_n)^T$ a zero mean, normal random vector with covariance matrix $\sigma^2 I$ and $x_1, ..., x_m$ the Demmler-Reinsch basis for the natural splines of order 2m with knots $t_1, ..., t_n$, to complete the model specification β is taken to be m-variate normal with zero mean and covariance matrix

$$Var(\beta) = \frac{\sigma^2}{n\lambda} D_\nu^{-1} = \frac{\sigma^2}{n\lambda} diag(\underbrace{\nu, ..., \nu}_{\text{m}}, \gamma_1, ..., \gamma_{n-m})^{-1}, \tag{A.7}$$

with $\gamma_1, ..., \gamma_{n-m}$ the Demmler-Reinsch eigenvalues.[1] Thus we have that the joint density for the response vector $\mathbf{y}$ and the coefficient vector is proportional to

$$\begin{aligned}&\exp\left\{-\frac{1}{2\sigma^2}\left(\mathbf{y}-\mathbf{X}\beta\right)^T\left(\mathbf{y}-\mathbf{X}\beta\right)-\frac{n\lambda}{2\sigma^2}\beta^T\mathbf{D}_\nu\beta\right\}=\\&=\exp\left\{-\frac{1}{2\sigma^2}\mathbf{y}^T(\mathbf{I}-\mathbf{S}_{\lambda,\nu})\mathbf{y}-\frac{1}{2\sigma^2}(\mathbf{X}\beta-\mathbf{S}_{\lambda,\nu}\mathbf{y})^T\mathbf{S}_{\lambda,\nu}^{-1}(\mathbf{X}\beta-\mathbf{S}_{\lambda,\nu}\mathbf{y})\right\},\end{aligned} \tag{A.8}$$

where

$$\mathbf{S}_{\lambda,\nu}=\mathbf{X}(\mathbf{X^TX}+n\lambda\mathbf{D}_\nu)^{-1}\mathbf{X}^{-1}=n^{-1}\mathbf{X}(\mathbf{I}+\lambda\mathbf{D}_\nu)^{-1}\mathbf{X}^T. \tag{A.9}$$

Thus, we have $\mathbb{E}[g|\mathbf{y}]=\mathbf{S}_{\lambda,\nu}\mathbf{y}$ and $Var[g|\mathbf{y}]=\sigma^2\mathbf{S}_{\lambda,\nu}$ for $g=\mathbf{X}\beta$ the conditional mean vector for the response, and the unconditional $\mathbf{y}$ distribution is an n-variate normal with mean zero and variance $\sigma^2(\mathbf{I}-\mathbf{S}_{\lambda,\nu})$. Also, since S_λ is the smoothing spline hat matrix we have that

$$\begin{aligned}&\lim_{\nu\to 0}\mathbb{E}[g|y]=g_\lambda,\\&\lim_{\nu\to 0}Var[g|y]=\sigma^2\mathbf{S}_\lambda\lim_{\nu\to 0},\\&Var[(\mathbf{I}-\mathbf{S}_{\lambda,\nu})\mathbf{y}]=\sigma^2(\mathbf{I}-\mathbf{S}_\lambda).\end{aligned} \tag{A.10}$$

In this framework the first equation says that the smoothing splines fitted values are the posterior mean of $g=\mathbf{X}\beta$ while the other two equations give the covariance matrices for g and the residual vector, leaving only λ and σ^2 to estimate. Wahba(1990) developed maximum likelihood estimators for these parameters by splitting the design matrix in two uncorrelated parts and then relying on the orthogonality of the Demmler-Reinsch to find the distribution of the not polynomial part and therefore formulating the log-likelihood and by maximization obtains the following estimators:

$$\tilde{\sigma}^2=\frac{\mathbf{y}^T(\mathbf{I}-\mathbf{S}_\lambda)\mathbf{y}}{n-m}, \tag{A.11}$$

$$\tilde{\lambda}=argmin_\lambda\frac{\mathbf{y^T}(\mathbf{I}-\mathbf{S}_\lambda)\mathbf{y}}{|\mathbf{I}-\mathbf{S}_\lambda|_+^{1/(n-m)}}. \tag{A.12}$$

Thus, this show that the smoothing spline can be derived from a Bayesian regression model wherein the regression function is a random natural spline whose distribution is diffuse over polynomials of order m.
The origins of smoothing splines lies in the work on graduating data by Whittaker(1923) and remained mainly a numerical analysis method until Grace Wahba proved their usefulness in the solution of statistical estimation problems, making clear that they are a extremely flexible data analysis tool.

[1] The appeal of Demmler-Reinsch basis is a technicality for allowing the design matrix to have properties that will be clearer in the formal definition of the estimator.

A.2 Formal definition of smoothing spline estimator

The formal and explicit expression of the smoothing spline estimator requires the definition of the knots. Those depends from the order of the smoothing spline, which defines the minimum support of this kind of function. A natural spline of order $2m$ is a $2m$-th order piecewise polynomial with $2m-2$ continuous derivatives consisting of different polynomial segments over each of the intervals defined by a sequence of (for the sake of clarity, distinct) knots $[t_i, t_{i+1}], i = 1, ..., n-1$ and is a polynomial of order m outside of $[t_1, t_n]$.
Thus, being g_λ a natural spline, the problem of minimizing (A.2) over all functions in $W_2^m[0,1]$ reduces to the finite dimensional problem of minimization over the n dimensional set of natural splines, allowing the proof of the following theorem that gives a closed form for the estimator.
Let $x_1, ..., x_n$ be a basis for the set of natural splines of order $2m$ with knots at $t_1, ..., t_n$ and define $\mathbf{X}$=$\{x_j(t_i)\}_{i,j=1,n}$. If $n \geq m$ then the unique minimizer of (A.2) is $g_\lambda = \sum_{j=1}^{n} b_{\lambda,j} x_j$, where $\mathbf{b}_\lambda = (b_{\lambda,1}, ..., b_{\lambda,n})^T$ is the unique solution with respect to $\mathbf{c} = (c_1, ..., c_n)^T$ of the system:

$$(\mathbf{X^T X} + n\lambda\Omega)\mathbf{c} = \mathbf{X^T y}, \tag{A.13}$$

with

$$\Omega = \left\{ \int_0^1 x_i^{(m)}(t) x_j^{(m)}(t) dt \right\}_{i,j=1,n}. \tag{A.14}$$

The vector of fitted values corresponding to the smoothing spline estimator has seen to be

$$g_\lambda = (g_\lambda(t_1), ..., g_\lambda(t_n))^T = \mathbf{S}_\lambda \mathbf{y}, \tag{A.15}$$

with

$$\mathbf{S}_\lambda = \mathbf{X}(\mathbf{X^T X}+n\lambda\Omega)^{-1}\mathbf{X^T} \tag{A.16}$$

The reader with a deeper knowledge about regularization methods in regression analysis can recognize a similarity with ridge regression, this is due to the common Bayesian heritage of both the methods.
As I have anticipated in the last section an insightful representation of the estimator requires a judicious choice of the basis elements, in order to simultaneously diagonalize $\mathbf{X}^T\mathbf{X}$ and Ω in S_λ expression.
For $m = 1$ and a uniform design, the Demmler-Reinsch basis functions admit closed form. Suppose that we have data at points $t_i = (2i - i)/2n, i = 1, ..., n$ and we estimate $g \in W_2^1[0,1]$ by minimization of (A.2). In this setting we have a linear smoothing spline estimator and $g_\lambda = \sum_{i=1}^{n} b_{\lambda,i} x_i$ where the x_i are basis and $\mathbf{b}_\lambda$ is the solution to (A.13). The x_i functions are all natural splines that interpolate the constant and the functions $\sqrt{2}cos(j\pi t), j = 1, ..., n-1at$ the

design points, given explicitly by $x_1(t) \equiv 1$ and

$$x_{j+1}(t) = \begin{cases} \sqrt{2}\cos(j\pi t_1), & 0 \leq t < t_1, \\ \\ \sqrt{2}cos(j\pi t_i) + & \\ \sqrt{2}\frac{t-t_i}{t_{i+1}-t_i}[\cos(j\pi t_{i+1}) - \cos(j\pi t_i)], & t_i \leq t < t_{i+1}, \\ & i = 1, ..., n-1, \\ \sqrt{2}\cos(j\pi t_n), & t_n \leq t \leq 1. \end{cases} \tag{A.17}$$

The x_j in this case are all natural linear splines since they are all continuous, constant outside the design points and linear over each subinterval of the latter. Also this choice implies that $\mathbf{X}^T\mathbf{X} = \mathbf{X}\mathbf{X}^T = n\mathbf{I}$. In this context, let the Demmler-Reinsch eigenvalues be

$$\gamma_j = (2n\sin(j\pi/2n))^2, \text{ j=1,...,n-1}, \tag{A.18}$$

in this way one can define

$$\int_0^1 x'_{i+1}(t)x'_{j+1}dt = 2n\sum_{r=1}^{n-1}[\cos(i\pi t_{r+1}) - \cos(i\pi t_r)] \times [\cos(j\pi t_{r+1}) - \cos(j\pi t_r)]$$
$$= \delta_{i,j}\gamma_j, \text{ i,j=1,...,j-1}, \tag{A.19}$$

thus the x_j are the Demmler-Reinsch basis functions under which the (A.13) becomes

$$[n\mathbf{I} + n\lambda diag(0, \gamma_1, ..., \gamma_{n-1})]\mathbf{c} = n\mathbf{b}, \tag{A.20}$$

where b is the vector of the sample cosine Fourier coefficients, that is $b_1 = \bar{y}$, the average response, and

$$b_j = \frac{\sqrt{2}}{n}\sum_{j=2}^{n} y_i \cos((j-1)\pi t_i), \tag{A.21}$$

such that for any specific value of λ the linear smoothing spline is given by

$$g_\lambda = b_1 + \sum_{j=2}^{n}\frac{b_j}{1+\lambda\gamma_{j-1}}, j = 2, ..., n. \tag{A.22}$$

So, at the design points one has

$$g_\lambda(t_i) = b_1 + \sum_{j=2}^{n}\frac{b_j}{1+\lambda\gamma_{j-1}}\sqrt{2}\cos((j-1)\pi t_i), i = 1, ..., n. \tag{A.23}$$

Now it is clear that the linear smoothing spline is essentially a weighted series estimator that smooths the data in a similar manner to that of kernel estimator, relying on the information in the sample Fourier coefficients and weighting them by a damping factor (i.e. $(1+\lambda\gamma_j)^{-1}$), so the smoothing parameter controls the mix of high and low frequency information that is used in the estimation of $g(\cdot)$, as this goes to infinity damping becomes severe and the estimator reduces to the sample average, while when it goes to zero the interpolation touches any data point so no smoothing is performed and no damping appears, leading to a behaviour of the estimator that is very similar to the kernel estimator, and to an asymptotic equivalence between the two types of estimators. To continue this discussion, is useful to rearrange the latter equation in the following way

$$\begin{aligned} g_\lambda &= n^{-1}\sum_{i=1}^{n} y_i K_n(t, t_i; \lambda), \\ K_n(t,s;\lambda) &= 1+\sqrt{2}\sum_{j=1}^{n-1}\frac{cos(j\pi s)x_{j+1}(t)}{1+\lambda\gamma_j} \end{aligned} \tag{A.24}$$

with x_j and γ_j the now (almost) usual Demmler-Reinsch basis functions and eigenvalues. Thus one can expect for large n the smoothing spline can be approximately given by $1+2\sum_{j=1}^{\infty}\cos(j\pi s)\cos(j\pi t)/(1+\lambda(j\pi^2))$ and through some simplifications achieve[2]

$$\begin{aligned} \sum_{k=0}^{\infty}\frac{\cos(kx)}{a^2+k^2} &= \frac{\pi}{2a}\frac{e^{-a(\pi-|x|)}+e^{a(\pi-|x|)}}{e^{a\pi}-e^{-a\pi}}+\frac{1}{2a^2},\ |x|\le 2\pi, \\ \sum_{k=1}^{\infty}\frac{\cos(kx)}{a^2+k^2} &= \frac{\pi}{2a}\frac{e^{-a(\pi-|x|)}+e^{a(\pi-|x|)}}{e^{a\pi}-e^{-a\pi}}-\frac{1}{2a^2},\ |x|\le 2\pi. \end{aligned} \tag{A.25}$$

Thus with $a=(\sqrt{\lambda}\pi)^{-1}$ one see that $K_n(t,s;\lambda)$ is approximately equal to

$$K^{+}(t,s;\lambda)=\frac{e^{-|t-s|\sqrt{\lambda}}+e^{-2\sqrt{\lambda}}e^{-|t-s|\sqrt{\lambda}}+e^{-(t+s)\sqrt{\lambda}}+e^{(t+s-2)\sqrt{\lambda}}}{2\sqrt{\lambda}(1-e^{-2/\sqrt{\lambda}})} \tag{A.26}$$

and leading to the following theorem. Assume that $n\to\infty, \lambda\to 0$ in such a way that $n\lambda\to\infty$.
Then,

$$K_n(t,s;\lambda)=K^{+}(t,s;\lambda)+\mathcal{O}\left(\frac{1}{n\lambda}\right) \tag{A.27}$$

uniformly for $t,s\in[0,1]$[3]. Thus asymptotically K_n is the sum of a weight function $e^{-|t-s|/\sqrt{\lambda}}$ and terms $e^{-(t+s)/\sqrt{\lambda}}/(2\sqrt{\lambda})$ and $e^{(t-1+s-1)/\sqrt{\lambda}}/(2\sqrt{\lambda})$, thus for large n and fixed t the functions behave like a kernel estimator with Laplace kernel $K(u)=e^{-|u|}$ and bandwidth $\sqrt{\lambda}$.
For general m or nonuniform designs there isn't, at the best of my knowledge, any simple form of the Demmler-Reinsch basis functions and eigenvalues. However their properties are known and one can make some comparison with the

[2]The following approximation results are from Gradshteyn and Ryzhik (2007)
[3]The proof of the theorem is in Eubank(2000) p.248.

former framework.
Demmler-Reinsch showed that a natural spline basis $x_1, ..., x_n$ may be chosen so that:
1) $x_1, ..., x_m$ span the space of polynomials of order m,
2) the function x_j has at least $j-1$ sign changes over (0,1),
3) $\mathbf{X}^T\mathbf{X} = n\mathbf{I} = \mathbf{X}\mathbf{X}^T$,
4) $\Omega = diag(\underbrace{0, ..., 0}_{\text{m}}, \gamma_1, ..., \gamma_{n-m})$,
5) $\gamma_j = C(j\pi)^{2m}(1 + o(1))$ fir C a constant that depends only on m and the design.
So the general representation for the m-th order smoothing spline is

$$g_\lambda = \sum_{j=1}^{m} b_j x_j + \sum_{j=m+1}^{n} \frac{b_j}{1 + \lambda_{\gamma_{j-m}}} x_j, \tag{A.28}$$

with $b_j = n^{-1} \sum_{j=1}^{m} b_j x_j + \sum_{j=m+1}^{n} \frac{b_j}{1+\lambda_{\gamma_{j-m}}} x_j$, the Demmler-Reinsch Fourier coefficients, not the cosine Fourier coefficients, but they still can be compared to the cosine functions due to the sign change property, thus still providing a partitioning of the frequency content of the data with larger values of the coefficient index signifying higher frequencies. With this interpretation in mind and with property 5) one has essentially the same conclusion as for the linear smoothing spline case, therefore a smoothing spline is a type of damped series estimator with λ controlling the relative amount of low and high frequency information that is used in estimating g.
Another version of the Theorem 3.2.1 can be derived for a more general smoothing crieterion as

$$n^{-1} \sum_{i=1}^{n} w_i (y_i - g(t_i))^2 + \lambda \int_0^1 g^{(m)}(t)^2 dt, \tag{A.29}$$

with positive weights $w_i > 0$, taking $w_i = [Var(y_i)]^{-1}, i = 1, ..., n$, this criterion becomes useful for heteroskedastic observations.

A.3 Large Sample Properties

Theorem 3.2.2 allows the parallel between kernel estimators and smoothing spline, but it is also the starting point to analyze the point-wise variance and bias of the linear smoothing spline using techniques similar to those employed for kernel estimators. Define $K(t, s; \lambda)$ in (A.27) as

$$K(t, s; \lambda) = \frac{1}{2\sqrt{\lambda}} \left\{ e^{-|t-s|/\sqrt{\lambda}} + e^{-(t+s)/\sqrt{\lambda}} + e^{(t-1+s-1)/\sqrt{\lambda}} \right\}, \tag{A.30}$$

if t is a lower boundary pont such as $t = \sqrt{\lambda} q$ for some $q > 0$ then,

$$K(t, s; \lambda) = \frac{1}{2\sqrt{\lambda}} \left\{ e^{-|t-s|/\sqrt{\lambda}} + e^{-2q} + e^{(t-s)/\sqrt{\lambda}} \right\} + \mathcal{O}\left((n\lambda)^{-1}\right), \tag{A.31}$$

gives a first order boundary correction which makes the integral $K(t,\cdot;\lambda)$ asymptotically the same as for t point, formally

$$\int_0^1 (2\sqrt{\lambda})^{-1}\left\{e^{-|t-s|/\sqrt{\lambda}} + e^{-2q}e^{(t-s)/\sqrt{\lambda}}\right\} ds = 1 + \mathcal{O}(e^{-1/\sqrt{\lambda}}). \tag{A.32}$$

Then one can show that

$$\mathbb{E}\, g_\lambda(t) = n^{-1}\sum_{i=1}^{n} g(t_i)K_n(t,t_i;\lambda) = \int_0^1 g(s)K(t,s;\lambda)ds + \mathcal{O}((n\lambda)^{-1}) \tag{A.33}$$

and

$$\begin{aligned} Var(g_\lambda(t)) &= \frac{\sigma^2}{n^2}\sum_{i=1}^{n} K_n^2(t,t_i;\lambda) \\ &= \frac{\sigma^2}{4n\sqrt{\lambda}}\{1 + e^{2(t-1)/\sqrt{\lambda}}(1-2(t-1)(\lambda)^{-1/2} + \\ &\quad + e^{-2t/\sqrt{\lambda}}(1+2t/\sqrt{\lambda}) + o(1)\}. \end{aligned} \tag{A.34}$$

In order to derive the point-wise approximation to the bias of the linear smoothing spline, one can use a Taylor expansion in (A.33) to see that if g'' satisfies a Lipschitz condition[4] of order 2η then

$$\begin{aligned} \mathbb{E}\, g_\lambda(t) =& g(t) - \sqrt{\lambda}g'(t)\{e^{(t-1)/\sqrt{\lambda}} - e^{-t/\sqrt{\lambda}}\} \\ &+ \lambda g''(t)\left(1 + \frac{t-1}{\sqrt{\lambda}}e^{(t-1)/\sqrt{\lambda}} - \frac{t}{\sqrt{\lambda}}e^{-t/\sqrt{\lambda}}\right) \\ &+ \mathcal{O}\left(\frac{1}{n\lambda} + \lambda^{1+\eta}\right), \end{aligned} \tag{A.35}$$

or, when $t \in [0,1]$

$$\mathbb{E}\, g_\lambda(t) = g(t) + \lambda g''(t) + \mathcal{O}\left(\frac{1}{n\lambda} + \lambda^{1+\eta}\right), \tag{A.36}$$

therefore given the variance approximation provided in (A.34) one can state that g_λ is a point-wise second order estimator since $\mathbb{E}(g_\lambda(t) - g(t))^2$ is of order $\eta^{-4/5}$ if λ is of order $n^{-4/5}$. This is not true at boundary points, since while the variance will be of order $(n\sqrt{\lambda})^{-1}$, the bias becomes

$$\mathbb{E}\, g_\lambda(t) - g(t) = \sqrt{\lambda}g'(0)e^{-q} + \mathcal{O}\left(\lambda + \frac{1}{n\lambda}\right), \tag{A.37}$$

[4] A regularity condition stating that, given a function $g : X \times Y \to \mathbb{R}$ there is a constant $L > 0$ such that for a point of its domain

$$\|g(x,y_i) - g(x,y_j)\| \le L\|y_i - y_j\|_{2\eta}, \text{ for any } x \in X \text{ and any } y_i, y_j \in Y.$$

thus if $n\lambda^{3/2} \to \infty$, g_λ is only first order in boundary regions unless $g'(0) = g'(1) = 0$.
Globally the performance of the estimator can be assessed as the approximation $\mathbb{E}(g_\lambda(t_i) - g(t_i))^2, i = 1, ..., n,$, by (A.34) and (A.35), averaged over the design to approximate the following criterion

$$R_n(\lambda) = n^{-1} \sum_{i=1}^{n} \mathbb{E}(g(t_i) - g_\lambda(t_i))^2. \tag{A.38}$$

Now, since $n\sqrt{\lambda} \to \infty$

$$\begin{aligned} n^{-1} \sum_{i=1}^{n} Var(g_\lambda(t_i)) &= \frac{\sigma^2}{4n\sqrt{\lambda}} \left\{ 1 + \int_0^1 \left[e^{2(t-1)/\sqrt{\lambda}} \left(1 - \frac{2(t-1)}{\sqrt{\lambda}} \right) + \right. \right. \\ &\qquad \left. \left. + e^{-2t/\sqrt{\lambda}} \left(1 + \frac{2t}{\sqrt{\lambda}} \right) \right] dt + o(1) \right\} \\ &= \frac{\sigma^2}{4n\sqrt{\lambda}} (1 + o(1)), \end{aligned} \tag{A.39}$$

thus

$$n^{-1} \sum_{i=1}^{n} (\mathbb{E}\, g_\lambda(t_i) - g(t_i))^2 = \frac{\lambda^{3/2}}{2} [g'(0)^2 + g'(1)^2] + o(\lambda^{3/2}),$$

while in the case where $g'(0) = g'(1) = 0$

$$n^{-1} \sum_{i=1}^{n} (\mathbb{E}\, g_\lambda(t_i) - g(t_i))^2 = \lambda \int_0^1 g''(t)^2 dt + o(\lambda^2).$$

So combining these expressions one can obtain global risk of smoothing spline under different assumptions about the boundary properties of the regression function. For example, if g' is bounded and at least one of g'(0) or g'(1) is not zero one achieves an estimator that, with an optimized smoothing parameter, decay at the rate of $n^{-3/4}$, the same of a second order kernel estimator but without using any boundary correction.
Similar asymptotic results have been established for more general cases aside from the uniform design and $m = 1$ case. Nychka (1995) showed that under conditions similar to the latter case, a general smoothing spline behaves as a second order kernel estimator with bandwidth $\sqrt{\lambda/w(t)}$ for any w which empirical distribution is "enough" close to a continuous one with a strictly positive density function. This bandwidth has the property to be easily expandable or contractable to adjust to rich and sparse regions of the design, extending the kernel approximations developed by Silverman(1984b), Messer(1991) and Nychka(1995).
A smoothing spline with penalty function J_m can generally attain the $\mathcal{O}(n^{-2m/(2m+1)})$ optimal decay rate for its risk when the regression function is in $W_2^m[0, 1]$, with the advantage of a faster convergence due to a less computational intensive

boundary adjustment.[5]
What is left to explain regarding smoothing spline is the comparison of their estimation risk related to other nonparametric estimators. Carter,Eagleson and Silverman(1992) compare the the risk behavior of smoothing splines with the one of the minimax spline estimator of Speckman(1985), which is the best possible in terms of average risk over all regression functions in $W_2^m[0,1]$ for which $J_m(g) \leq \rho$, they show that when $m = 2$ and under optimal levels of smoothing for both estimators, the cubic smoothing spline is only $8,3\%$ less efficient than the fully efficient minimax estimator, therefore the cubic smoothing spline is very nearly optimal as a second order estimator.

A.4 Penalized B-Spline

In the last section I have showed off some asymptotic results about the smoothing spline estimator, during such discussion I have highlighted several times that these results are valid mostly for uniform design and always with an already fixed smoothing parameter λ, meaning that in order to achieve gratifying results the analyst has to select the knots which the spline should pass through, and has to rely on some strategy for the selection of the smoothing parameter by usually optimizing an information criterion. While this latter topic will be discussed in the next section, since the scientific debate has already shrinked the range of alternative produced in the last decades to a narrow set of established selection techniques, the knots location problem has not been solved yet.
The problem is that for some fixed K knots there are $\sum_{i=0}^{K} \binom{K}{i} = 2^K$ possible models and, because the locations has additionally a marked effect on the fit, the usual selection procedures become unfeasible. This instance caused the blooming of several approaches for the selection both of the amount and position of knots, each of them has revealed complicated and computationally intensive. Instead of developing some variety of spline smoother a growing branch of literature relied on a combination of B(asis)-spline and difference penalties (on the estimated coefficients), which emerged with the name of P(enalized)-splines. Despite the first attempts are dated back to the papers of Parker and Rice(1985) and O'Sullivan(1986), this estimation technique became popular after that Eilers and Marx(1996) illuminated the numerical practicability and flexibility of this approach.
A B-spline consists of polynomial pieces connected in a special way, at the joining points not only the ordinates of the pieces match, but their first derivatives are equal. Since these basis overlap each other in the joining points, the degree of the B-splines explains how much they overlap. De Boor(1978) gives a simple

[5] More information on the topic is available in Rice and Rosenblatt(1983) and Cox(1983).

recursive formula for defining B-splines based on a set of knots

$$\begin{aligned} B_j^0(x) &= I_{[t_j,t_{j+1}]}(x), \\ B_j^m(x) &= \frac{x-t_j}{t_{j+m}-t_j} B_j^{m-1}(x) + \frac{t_{j+1}-x}{t_{j+m+1}-t_{j+1}} B_{j+1}^{p-1}, \end{aligned} \tag{A.40}$$

where $B_j^m(x)$ denotes the j-th B-spline of degree m and $t_j, j = 1, ..., K$ are the knots. Eiler and Marx(2004) showed that B-splines can be computed by differencing of the correspondent truncated polynomials, with the following general formula

$$B_j^m(x) = (-1)^{m+1} \Delta^{m+1} Z_j^m(x)/(h^m m!), \tag{A.41}$$

where $h = t_{j-1} - t_j$, $Z_j^m(x) = (x - t_j)_+^m$ and Δ^m is the difference operator applied to the spline coefficients at the m-th order, thus a complete B-spline matrix of degree m for n observations based on K knots has dimension $n \times (K + 1 + m)$. Now consider the regression of n data points (y_i, x_i) on a set of m B-splines $B_j(\cdot)$. The least square objective functions to minimize is

$$Q = \sum_{i=1}^{n} \left\{ y_i - \sum_{j=1}^{m} a_j B_j(x_i) \right\}^2 . \tag{A.42}$$

Let the number of knots be relatively large, such that the fitted curve will be more variable than how much the data would justify. O'Sullivan(1986,1988) introduced a penalty on the second derivative of the fitted curve and so formed the objective function

$$Q_{O'S} = \sum_{i=1}^{n} \left\{ y_i - \sum_{j=1}^{m} a_j B_j(x_i) \right\}^2 + \lambda \int_{x_{\min}}^{x_{\max}} \left\{ \sum_{j=1}^{m} a_j B_j''(x) \right\}^2 dx. \tag{A.43}$$

Eilers and Marx(1996) proposed to base the penalty on (higher-order) finite differences of the coefficients of adjacent B-splines

$$Q_{E\&M} = \sum_{i=1}^{n} \left\{ y_i - \sum_{j=1}^{m} a_j B_j(x_i) \right\}^2 + \lambda \sum_{j=k+1}^{m} (\Delta^k a_j)^2, \tag{A.44}$$

thus reducing the dimensionality of the problem to m, the order of the spline, therefore obtaining robustness to the placement of the knots.
The system of equation that one has to solve in the minimization of (A.44) can be written as:

$$B^T y = (B^T B + \lambda D_k^T D_k) a, \tag{A.45}$$

where D_k represent the matrix of the difference operator Δ^k, and the elements of B are $b_{i,j} = B_j(x_i)$. When $0 < \lambda < \infty$ (the burden cases have been already discussed) the penalty only influences the main diagonal and k sub-diagonals (on both sides of the main diagonal) of the system, giving him a banded structure.

In a generalized linear model (GLM) we introduce a linear predictor $\eta_i = \sum_{j=1}^{n} b_{i,j} a_j$ and a link function $\eta_i = g(\mu_i)$ where μ_i is the expectation of y_i, the penalty is subtracted from the likelihood function

$$L = l(y; a) - \frac{\lambda}{2} \sum_{j=k+1}^{m} (\Delta^k a_j)^2, \tag{A.46}$$

and the subsequent optimization leads to the following system of equation

$$B^T (y - \mu) = \lambda D_k^T D_k a \tag{A.47}$$

that can be solved with the system

$$B^T \tilde{W} (y - \tilde{\mu}) + B^T \tilde{W} B \tilde{a} = (B^T \tilde{(W)} B + \lambda D_k^T D_k) a, \tag{A.48}$$

where $\tilde{a}$ and $\tilde{\mu}$ are the approximations to the solution and $\tilde{W}$ is a diagonal matrix of weights

$$w_{i,i} = \frac{1}{v_i} \left(\frac{\partial \mu_i}{\partial \eta_i} \right). \tag{A.49}$$

P-spline have a number of useful properties, in first place P-splines have no boundary effects, meaning that there is no problem in the spreading the fitted curve outside of the (physical) domain of the data. Also P-splines can fit polynomial data exactly, then if y_i are a polynomial in x of degree k, the B-splines of the same degree (or even higher) will exactly fit the data, and this is true also for P-splines, if the order of the penalty is $k + 1$ or higher, whatever the value of λ.
P-spline conserve the moments of the data, such that for $GLM's$ with canonical links it holds that

$$\sum_{i=1}^{n} x^k y_i = \sum_{i=1}^{n} x^k \hat{y}_i, \tag{A.50}$$

for all values of λ, with $\hat{y}_i = \sum_{j=1}^{m} b_{i,j} \hat{a}_j$, leading to a substantial advantage related to many kernel smoothers that inflate the variance increasingly with stronger smoothing.

In conclusion, as the smoothing is controlled by the penalty parameter, for the P-spline the number of knots is not a crucial one. However, simple simulation studies showed in Ruppert (2002) showed that there must be enough knots to fit features in the data, thus there is a minimum necessary number of knots to reach. Also there are specific situation where an higher number of knots may increase the MSE by a moderate amount. Thus, he suggest the application of a GCV-like procedure to verify the right number of knots.

A.5 Smoothing parameter selection

Given the minimum acceptable number of knots, the P-spline estimator achieves the same results as the natural smoothing spline estimator, given a fixed smoothing parameter, and outperforms the latter in non-uniform situations. This last

section will be devoted to a brief discussion of smoothing parameter selection procedures emerged and established in the recent literature about this kind of estimators.
Following the scheme of Kauermann(2005) one has to think to the spline estimation through the Bayesian interpretation, in connection with Linear Mixed Models and thus one has to think that the basis coefficients are considered as random effects and the penalization as a priori distribution imposed on the basis coefficients, leading to the equivalence of the spline smoothing to the maximum posterior Bayes estimation. In this scenario the smoothing parameter will plays the role of the a priori variance of the basis coefficients and this interpretations allows its estimation through Maximum Likelihood estimators or Residual Maximum Likelihood estimators.
In this context P-spline estimation is pursued by replacing $g(\cdot)$ by the parametric form

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \mathbf{b} + \epsilon_i, \tag{A.51}$$

where x_i is a low-dimensional parametric basis, the linear basis $\mathbf{x}_i = (1, x_i)^T$, and $\mathbf{z}_i$ is a high-dimensional basis inliearly indipendent of x_i, Kauermann suggest to choose the latter in a "lush" and "generous" way to achieve that the difference $\delta(x_i) = g(x_i) - \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \mathbf{b}$ is negligible, thus the introduction of the penalty parameter lead to the following penalized likelihood

$$l(\beta; b; \lambda) = -\frac{1}{2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b})^T(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b}) - \frac{1}{2}\mathbf{b}^T\mathbf{D}_K\mathbf{b}/\lambda, \tag{A.52}$$

where $\mathbf{D}_K$ is a $K \times K$ dimensional penalty matrix, differentiating the former equation leads to the following estimating equations:

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{Y} - \mathbf{Z}\hat{\mathbf{b}}), \\ \hat{\mathbf{b}} &= (\mathbf{Z}^T\mathbf{Z} + \mathbf{D}_K/\lambda)^{-1}\mathbf{Z}^T(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}). \end{aligned} \tag{A.53}$$

In this context a reasonable choice for λ is obtained by minimizing the Mean-Squared Error (MSE) leading to the following optimal value

$$\lambda_{MSE} = \frac{\mathbf{b}^T\mathbf{D}_K\mathbf{F}_{Z,X}\mathbf{D}_K\mathbf{b} + 3\sigma_\epsilon^2 tr(\mathbf{F}_{Z,X}\mathbf{D}_K\mathbf{F}_{Z,X}\mathbf{D}_K)/n}{\sigma_\epsilon^2 tr(\mathbf{F}_{Z,X}\mathbf{D}_K)} + \mathcal{O}(n^{-2}), \tag{A.54}$$

where $\mathbf{F}_{Z,X}$ is the Fisher information matrix.
If one makes the following distributional assumption, then (A.52) appears as the likelihood of the Linear Mixed Model:

$$\mathbf{b} \sim N(0, \sigma_b^2\mathbf{D}_K^{-1}), \quad \mathbf{Y}|\mathbf{b} \sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}, \sigma_\epsilon^2\mathbf{I}_n). \tag{A.55}$$

Then if one considers $\mathbf{b}$ as random effect, on can marginalize the former model and obtain

$$\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma_\epsilon^2\mathbf{V}_\lambda), \tag{A.56}$$

where $\mathbf{V}_\lambda = \mathbf{I}_n + \lambda \mathbf{Z}\mathbf{D}_K^{-1}\mathbf{Z}^T$ and $\lambda = \sigma_b^2/\sigma_\epsilon^2$, that can be estimated through REML

$$l_{REML}(\beta;\lambda) = -\frac{(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^T\mathbf{V}_\lambda^{-1}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})}{\sigma_\epsilon^2} - \log|\mathbf{V}_\lambda| - \log|\mathbf{X}^T\mathbf{V}_\lambda^{-1}\mathbf{X}|, \tag{A.57}$$

thus to the following optimal λ equation

$$\hat{\lambda}_{REML} = \frac{\hat{\mathbf{b}}^T\mathbf{D}_K\hat{\mathbf{b}}/\sigma_\epsilon^2 + tr(\mathbf{F}_{\mathbf{Z},\mathbf{X}}\mathbf{D}_K/n)}{K} + \mathcal{O}(n^{-2}). \tag{A.58}$$

Another strand of literature about the smoothing parameter selection is based upon the Generalized Cross Validation technique, thus a data-driven approach, as defined by Craven Whaba(1979)

$$GCV(\lambda) = n^{-1}\sum_{i=1}^{n}\left\{\frac{y_i-\hat{y}_i}{1-tr(S_\lambda)/n}\right\} = \frac{RSS/n}{(1-tr(S_\lambda)/n)^2} \tag{A.59}$$

where S_λ is the smoothing matrix.
Notice that in expectation the GCV approximate the average MSE, indeed

$$\mathbb{E}[GCV(\lambda)] \approx \frac{1}{n}\left\{\sigma_\epsilon^2 tr(S_\lambda^2)\left[1-2\frac{tr(S_\lambda)}{n}\right] + \|m(x)(I-S_\lambda)\|^2\left[1+2\frac{tr(S_\lambda)}{n}\right]\right\} + \sigma_\epsilon^2 = MASE(\lambda) + \sigma_\epsilon^2 + o(n^{-1}).$$

A similar approach is the famous Mallow's C_p, in order to motivate this approach let make a step backward in the GCV definition, indeed

$$\mathbb{E}[RSS/n] = MASE(\lambda) + \sigma_\epsilon^2 - 2\sigma_\epsilon^2 tr(S_\lambda)/n$$

In this expression if one substitutes the σ_ϵ^2 with its estimates achieves the C_p statistic:

$$C(\lambda) = RSS(\lambda)/n + 2tr\hat{\sigma_\epsilon^2}tr(S_\lambda)/n,$$
$$\hat{\lambda}_{C_p} = \frac{\hat{\mathbf{b}}^T\mathbf{D}_K\mathbf{F}_{Z,X}\mathbf{D}_K\hat{b}}{\hat{\sigma_\epsilon}^2 tr(\mathbf{F}_{Z,X}\mathbf{D}_K)}\{1+\mathcal{O}_p(n^{-1})\}, \tag{A.60}$$

i.e. something like a plug-in estimate of (A.54).

A relatively new and active branch of literature focuses on the problem on the numerical side, so treating it as an ill-posed problem to be solved through a regularization method, thus trying to obtain solutions that are robust to small perturbation of the problem. The approach taken is to use generalized singular value decomposition (GSVD) to overcome the problems associated with the condition's number by replacing the problem with a "nearby" well-conditioned problem whose solution approximates the required solution and is more satisfactory than the one obtained with ordinary least squares. The idea, firstly illustrated in the book by Lawson and Hanson, and then developed in Hansen(1992)

is to display the plot of the norm of the regularized solution , $\|\mathbf{D}\hat{\mathbf{a}}(\lambda)\|$, versus the norm of the corresponding residual vector, $\|\mathbf{y} - \mathbf{B}\hat{\mathbf{a}}(\lambda)\|$, obtaining the L-curve, that is the relation of this two measure with respect to a λ that has value on $[0, \infty)$. When some regularity conditions are satisfied the L-curve exhibits a "corner" behavior as a function of λ, wherefore is the optimal one. In fact the "corner" λ yields a good balance between a small residual norm $\|\mathbf{y} - \mathbf{B}\hat{\mathbf{a}}(\lambda)\|$ and a small solution semi-norm $\|\mathbf{D}\hat{\mathbf{a}}(\lambda)\|$, and also tend to balance the regularization and perturbation errors.
The V-curve criterion simplifies this selection by requiring the minimization of the Euclidean distance between adjacent points lying on the L-curve, thus obtaining the following expression

$$\lambda_{V-curve} = \underset{\lambda}{argmin} \sqrt{\{\Delta \log \|\mathbf{y} - \mathbf{B}\hat{\mathbf{a}}(\lambda)\|\}^2 + \{\Delta \log \|\mathbf{D}\hat{\mathbf{a}}(\lambda)\|\}^2}, \qquad \text{(A.61)}$$

whereas the Δ is the first order difference operator.