Asset Allocation by Penalized Least Squares

Simone Manganelli*

25 October 2006 (First Draft: 8 August 2005)

Abstract

This paper shows how the problem of mean-downside risk portfolio allocation can be cast in terms of penalized least squares (PLS). The penalty is given by a power function of the returns below a certain threshold. We derive the asymptotic properties of the PLS estimator, allowing for possible nonlinearities and misspecification of the model. We illustrate the usefulness of this new class of estimators with two empirical applications. First, we estimate an autoregressive model, in the spirit of the GARCH literature. Second, we suggest a simple strategy to derive the optimal portfolio weights associated to a mean-downside risk model.

*I would like to thank Bruno Gerard and the seminar participants at the ECB. All errors remain my own. DG-Research, European Central Bank, Kaiserstrasse 29, D-60311 Frankfurt am Main, Germany. E-mail: simone.manganelli@ecb.int

Keywords: Portfolio otpimization, mean-risk utility model, stochastic dominance, asymmetric least squares, expectile.

JEL classification: C14, C22, G11.

1 Introduction

The classic paradigm for portfolio allocation is [14] Markowitz's (1952) mean-variance model. By using variance as a measure of risk, Markowitz formalized the intuition that investors optimize the trade off between returns and risks. The use of variance accounts for the great success and endurance of this model, as it made the portfolio allocation problem analytically tractable. The fact that the variance of a portfolio involves all covariance terms added economic intuition and allowed to draw rich empirical implications.

The mean-variance model has been criticized on several grounds (see, e.g., [5] Bawa 1975 and [10] Fishburn 1977 for an early review of this criticism). The model rests on the assumptions of quadratic preferences and/or elliptical symmetric return distributions. Quadratic utility seems highly implausible, as it implies increasing absolute risk aversion and negative marginal utility beyond a certain threshold. Elliptical symmetric distributions are not realistic as well in that they rule out asymmetry and skewness typical of financial market returns. In addition, the use of variance as a measure of risk is not very intuitive, as it weighs equally positive and negative returns. As [15] Markowitz (1959) himself recognizes, investors will typically associate risk to failure of attaining a target return.

[10] Fishburn (1977) suggests an alternative mean-risk paradigm for portfolio allocation, where risk is defined as a probability-weighted function of deviations below a specific target return. Fishburn's model is usually referred to as an (α, t) model, where α represents the degree of risk aversion of following below the target return t. This model not only builds on a more appealing definition of risk, but is also compatible with the standard expected utility model and - unlike mean-variance - with stochastic dominance relationships. Fishburn's class of risk measures includes the Safety First criterion of [19] Roy (1952) and the semivariance of [15] Markowitz (1959). It is also the building block of the Limited Expected Losses Risk Management model of [3] Basak and Shapiro (2001), who show how a Value-at-Risk based risk management might have perverse effects on the stability of the financial system in the most adverse states of the world. Value-at-Risk and the Tail Conditional Expectation recently proposed by [2] Artzner et al. (1999) are closely related as well.

Although more appealing from a theoretical point of view, the (α, t) model hasn't been as successful as the mean-variance model, essentially because when dealing with lower partial moments, portfolio optimization becomes analytically intractable and computationally problematic. The reason is that there is no one to one correspondence between lower partial moments of individual securities and the aggregated portfolio. This implies that in general there is no analytical solution to the (α, t) model of portfolio selection, making the optimization problem cumbersome and computationally intensive.

This paper suggests an econometric strategy to estimate the general (α, t) model, deriving the optimal portfolio allocation associated to it. We start by observing that the expected utility of the (α, t) model with $\alpha = 1$ is given by the expectile of the portfolio return distribution, where the expectile is the Asymmetric Least Squares estimator of [17] Newey and Powell (1987). Rewriting the Asymmetric Least Squares objective function in terms of penalized least squares, we show that the corresponding estimator can be generalized to give the expected utility of the (α, t) model for any α. Since the proposed estimator solves a least squares problem, we call it **projectile**.

Recently, several independent contributions have shown that variants of the (α, t) model with $\alpha = 1$ can be efficiently solved by formulating it in terms of a linear programming problem (see [18] Rockafellar and Uryasev 2000, [20] Ruszczynski and Vanderbei 2003, [1] Acerbi 2004, [4] Bassett, Koenker and Kordas 2004, and [6] Bertsimas, Lauprete and Samarov 2004). However, it is not clear how these approaches can accommodate time varying moments. Furthermore, the fact that $\alpha = 1$ implies risk neutrality for returns below target, limits the economic plausibility of this case. As shown by [5] Bawa (1975) and [10] Fishburn (1977), it is only for $\alpha \geq 2$ that the (α, t) efficient set is a subset of third order stochastically non-dominated portfolios (i.e., the set of portfolios chosen by individuals with increasing, concave utility functions which also display decreasing absolute risk aversion).

The paper is structured as follows. In the next section, we show how the Asymmetric Least Squares estimator can be generalized in terms of Penalized Least Squares, and how these relate to Fishburn's (α, t) model. In section 3, we derive the large sample properties of nonlinear projectiles under possible misspecification. Sections 4 and 5 contain two empirical applications. Section 4 proposes a model for conditional projectiles and shows how they relate to standard GARCH models. We propose an autoregressive specification and estimate it on a sample of daily returns. Section

5 suggests a computationally simple strategy to maximize investor's expected utility as a function of portfolio weights. We first notice that (α, t) model optimization is equivalent to maximize the projectile with respect to the portfolio weights. Next, applying the implicit function theorem to the first order conditions of the Penalized Least Squares problem, we derive the analytical first and second derivatives of the projectile, which are subsequently provided as input into a standard optimization algorithm. The convexity of (α, t) model's expected utility as a function of the weights guarantees that any local maximum is also global, greatly simplifying the corresponding optimization problem. Using monthly data from the thirty stocks of the Dow Jones Industrial Average index, we illustrate how the empirical distribution of the mean-risk optimal portfolio is characterized by a significantly shorter left tail (and longer right tail) than that of the mean-variance optimal portfolio. Section 6 concludes.

2 Penalized Least Squares

For a given sample $\{y_t\}_{t=1}^{T}$ of realizations from a random variable Y with c.d.f. $F(y)$, consider the estimator $\hat{b}(k,p,q)$ which minimizes the function

$$Q_T(b) \equiv T^{-1} \sum_{t=1}^{T} \phi_{kpq}(y_t - b) \tag{1}$$

over b for fixed values of k, p and q, where $\phi_{kpq}(\cdot)$ is a convex loss function of the form

$$\phi_{kpq}(\lambda) = |\lambda|^p + kI(\lambda < 0)|\lambda|^q \qquad\qquad k \geq 0 \;\; p, q \geq 1 \tag{2}$$

and $I(A)$ denotes the indicator function for the event A. The loss functions behind regression quantiles ([12] Koenker and Bassett 1978) and asymmetric least squares ([17] Newey and Powell 1987) estimators are special cases of this loss function. When $p = q = 1$, $\phi_{k11}(\lambda)$ is the regression quantile loss function. When $p = q = 2$, $\phi_{k22}(\lambda)$ is the asymmetric least squares loss function.[1]

Writing the loss function as in (2) highlights how both regression quantiles and asymmetric least squares can be seen as a type of penalized L_p-norm, whose penalty is represented by the term $kI(\lambda < 0)|\lambda|^q$. The penalty is associated only to negative values of the argument (positive values for $\tau > 0.5$ - see footnote 1) and is proportional to a power function of the argument. *Ceteris paribus* the higher the parameter k, the higher the penalty and the more extreme the quantile or expectile associated to

[1] The loss function for regression quantiles and asymmetric least squares is typically written as $r_\tau(\lambda) \equiv |\tau - I(\lambda < 0)| \cdot |\lambda|^s$, $s = 1, 2$. Just set $k = (1 - 2\tau)/\tau$ for $\tau \in (0, 0.5)$ and notice that the loss functions are identical (up to a factor of proportionality). The case $\tau \in (0.5, 1)$ is covered by $\tilde{\phi}_{kpq}(\lambda) = |\lambda|^p + kI(\lambda > 0)|\lambda|^q$, where $k \geq 0$ and $p, q \geq 1$.

it.

This motivates our definition of **penalized least squares (PLS)**. The PLS loss function is obtained by setting $p = 2$ in expression (2):

$$\rho_{kq}(\lambda) = \lambda^2 + kI(\lambda < 0)|\lambda|^q \qquad k \geq 0 \qquad q \geq 1 \tag{3}$$

and the corresponding class of estimator $\hat{m}(k, q)$ is defined to minimize over m

$$R_T(m) \equiv T^{-1} \sum_{t=1}^{T} \rho_{kq}(y_t - m) \tag{4}$$

To determine the class of estimators generated by PLS, consider the parameter $m^0(k, q)$ which minimizes the function $E[\rho_{kq}(Y - m)]$ over m. From the first order conditions of this minimization problem, $m^0(k, q)$ is the solution to the equation

$$m^0(k, q) = E[Y] - 0.5kq \int_{-\infty}^{m^0(k,q)} |m^0(k, q) - y|^{q-1} dF(y) \tag{5}$$

Since $q \geq 1$ and $k \geq 0$, the second order derivative is always positive, which guarantees that the solution to (5) is unique. Since $m^0(k, q)$ solves a least squares problem, we will call it **projectile**, as originally suggested by Gary Chamberlain (see footnote 3 in [17] Newey and Powell 1987).

2.1 Relationship with the Mean-Risk Utility Model

Terms of the type $\sigma(F) \equiv \int_{-\infty}^{t} |t-y|^{\alpha} dF(y)$ like those that appear on the right hand side of expression (5) have a long tradition in finance, as they are used to define downside risks of asset portfolios characterized by uncertain returns. They are special cases of [21] Stone's (1973) generalized risk measure and were employed by [10] Fishburn (1977) to develop a mean-risk model of portfolio choice with risk associated with below-target returns - referred to as an (α, t) model. This class of risk measures is motivated by the observation that portfolio managers usually associate risk with failure to attain a certain target return. t is the threshold with respect to which deviations are measured, while α measures the relative impact of small vs. large deviations.

[10] Fishburn (1977) (see his theorem 2) shows that when the (α, t) model is congruent with the standard expected utility model, the von Neumann-Morgenstern utility function can be written as

$$u(y) = \begin{cases} y & \text{if } y > t \\ y - h(t-y)^{\alpha} & \text{if } y \leq t \end{cases} \tag{6}$$

where h is a positive constant. In this context α can be interpreted as the parameter which describes the decision maker's attitude toward risk. Values of α greater than 1 imply risk aversion with regard to returns below target, $\alpha = 1$ implies risk neutrality, and $\alpha < 1$ is associated with risk seeking behavior.

Note that (5) is just the expectation of $u(y)$ as defined in (6), with $t = m^0(k, q)$, $\alpha = q - 1$ and $h = 0.5kq$. Therefore the projectile of a portfolio return distribution can be interpreted as the expected utility of an agent who wants to maximize expected returns and at the same time tries to avoid returns below a desired threshold.

An appealing feature of the (α, t) model is its consistency with stochastic dominance criteria. It is well known that mean-variance optimization may result in portfolio allocations that are dominated in the *second-order stochastic dominance* (SSD) sense. That is, there may exist portfolios which are preferred to the mean-variance optimal one by all risk averse agents. [5] Bawa (1975) argued for a rule even stricter than SSD for portfolio selection, the *third-order stochastic dominance* (TSD). The reason is that the TSD admissible set contains all those distributions (i.e., portfolio allocations) which are selected by agents with increasing, risk averse utility functions with positive third derivative. The restriction on the third derivative is motivated by the fact that positive third derivatives are implied by decreasing absolute risk aversion, a feature that seems consistent with observed economic behavior. [10] Fishburn (1977) (see his theorem

3) shows that the (α, t) efficient set is a subset of the *first-order stochastic dominance* efficient set for any $\alpha \geq 0$, a subset of the SSD efficient set for any $\alpha \geq 1$, and a subset of the TSD efficient set for any $\alpha \geq 2$, except for distributions with equal mean and equal risk. Therefore, asset allocation by (α, t) model will result in optimal portfolios that are not stochastically dominated.

3 Large Sample Properties of Penalized Least Squares Estimators

In this section we develop the asymptotic theory for nonlinear PLS estimators under possible misspecification. In the light of the discussion in the preceding section, we will limit ourselves to the case $q \geq 2$, to which corresponds a non risk seeking behavior of the optimizing agent.

Consider a sample of observations $\{y_t\}_{t=1}^T$ generated by the following model:

$$y_t = \mu_{yt}^{kq} + \varepsilon_t^{kq} \qquad \mu_{\varepsilon t}^{kq} = 0 \tag{7}$$

where $\mu_{\varepsilon t}^{kq}$ is the (k,q)-projectile of ε_t^{kq}, defined as the solution of $\mu_{\varepsilon t}^{kq} = E_t[\varepsilon_t^{kq}] - 0.5kq \int_{-\infty}^{\mu_{\varepsilon t}^{kq}} |\mu_{\varepsilon t}^{kq} - \varepsilon|^{q-1} dF_t(\varepsilon)$, $F_t(\varepsilon)$ is the c.d.f. of the error term ε_t^{kq}, conditional on all the past information Ω_t, and $E_t[\varepsilon_t^{kq}] \equiv E[\varepsilon_t^{kq}|\Omega_t]$ is the conditional expectation. The conditional projectile is given by:

$$\mu_{yt}^{kq} = \arg\min_{m} E_t \left[\rho_{kq}(y_t - m) \right] \tag{8}$$

where $\rho_{kq}(\lambda)$ is defined in (3).

We will develop the large sample properties of projectiles under possible model misspecification. It is well-known that misspecification may bias confidence intervals and invalidate hypothesis tests based on conventional variance-covariance matrices (see, for instance, [22] White 1994). It is therefore desirable to develop a theory for inference that is robust to it.

Denote with $f_t(\beta) \equiv f(W_t; \beta_{kq})$ the proposed projectile specification, where $W_t \in \Omega_t$, $\beta \in \mathbb{R}^p$, and we suppressed the subscripts k and q from β_{kq} for notational convenience. We give the following definition of correct model specification (see [11] Kim and White 2003 for an analogous definition in the regression quantile context):

Definition 1 (Correct Specification of the Projectile Model) - *A conditional projectile model* $\{f_t(\beta) : \mathbb{R}^{h_t} \times B \to \mathbb{R}, \beta \in \mathbb{R}^p, h_t, p \in \mathbb{N}, t = 1, 2, ...\}$ *is correctly specified for* μ_{yt}^{kq}*, if and only if there exists a vector* $\beta^0 \in \mathbb{R}^p$ *such that* $f_t(\beta^0) = \mu_{yt}^{kq}$ *almost surely, for a given choice of explanatory variables* $\{W_t\}_{t=1}^{T}$.

We impose the following projectile version of the orthogonality condition, which allows for possible misspecification of the conditional projectile:

Assumption M (Misspecification) - *There exists* β^* *such that* $E[\psi_{kq}(u_t)f_t(\beta)] = 0$, *for all* $\beta \in B$ *(a compact subset of* $\mathbb{R}^p$*), where* $u_t \equiv y_t - f_t(\beta^*)$ *and* $\psi_{kq}(\lambda) \equiv -2\lambda + kqI(\lambda < 0)|\lambda|^{q-1}$.

The standard projectile orthogonality condition would be $E_t[\psi_{kq}(u_t)] = 0$. The following theorem shows that assuming that $E_t[\psi_{kq}(u_t)] = 0$, which is stronger than assumption M, is equivalent to correct model specification. All the proofs are in Appendix B.

Theorem 2 (Correct Model Specification) - $E_t[\psi_{kq}(u_t)] = 0$ *if and only if the conditional projectile model is correctly specified.*

Therefore it is possible under assumption M that the conditional projectile model may be misspecified in the sense of definition 1.

The following theorems establish consistency and asymptotic normality of the PLS estimator, under possible misspecification.

Theorem 3 (Consistency) - *Under assumptions M and C0-C6 in Appendix A and for any* $k \geq 0$, $\hat{\beta}_T \xrightarrow{p} \beta^*$, *where* $\hat{\beta}_T = \arg\min_{\beta} T^{-1}\sum_{t=1}^{T}\rho_{kq}(y_t - f_t(\beta))$.

To prove consistency we verify that the conditions of theorem 3.5 in

[22] White (1994) are satisfied. Assumption C1 in Appendix A requires that the proposed parameterization for the projectile is continuous in the parameter space. Assumptions C2 and C3 are dominance conditions, while assumption C5 is a standard identification condition.

Define:

$$w_t(k,q) \equiv 2 + kq(q-1)I(u_t<0)|u_t|^{q-2} \tag{9}$$

$$Z_T \equiv E[T^{-1}\sum_{t=1}^{T} w_t(k,q)\nabla_\beta f_t(\beta^*)\nabla'_\beta f_t(\beta^*)] \tag{10}$$

$$V_T \equiv E[T^{-1}\sum_{t=1}^{T} \psi_{kq}(u_t)^2\nabla_\beta f_t(\beta^*)\nabla'_\beta f_t(\beta^*)] \tag{11}$$

where $\nabla_\beta f_t(\beta^*) \equiv \frac{\partial}{\partial\beta} f_t(\beta)|_{\beta=\beta^*}$.

Theorem 4 (Asymptotic Normality) - *If $q \geq 2$, under the assumptions of theorem 3 and assumptions AN1-AN5 in Appendix A, $\sqrt{T}V_T^{-1/2}Z_T(\hat{\beta}_T - \beta^*) \xrightarrow{d} N(0,I)$.*

Although the PLS objective function in (3) is continuous and differentiable, the presence of the indicator function implies that the first derivative is not differentiable. Therefore the asymptotic distribution cannot be obtained via the typical Taylor expansion applied to the first order conditions. We apply, instead, the techniques for nonsmooth objective functions as described by [16] Newey and McFadden (1994), based on stochastic equicontinuity.

Note that when $q = 2$ and $f_t(\beta)$ is linear, the variance-covariance matrix is identical to the one derived in theorem 3 of [17] Newey and Powell (1987). If $k = 0$, we get the variance-covariance matrix of nonlinear least squares, as in [23] White and Domowitz (1984).

Consistent estimates of the variance-covariance matrices can be obtained by standard plug-in estimators:

Theorem 5 (Variance-Covariance Matrix Estimation) - *Under the assumptions of theorem 4,*

$$\hat{V}_T \xrightarrow{p} p\lim V_T$$

$$\hat{Z}_T \xrightarrow{p} p\lim Z_T$$

where $\hat{V}_T$ and $\hat{Z}_T$ are the empirical analogues of V_T and Z_T evaluated at the estimated parameter $\hat{\beta}_T$.

4 Modeling Time-Varying Conditional Projectiles

A defining feature of daily financial returns is that their second moments tend to be highly autocorrelated. Time-varying second moments have been successfully captured by GARCH models of [8] Engle (1982) and [7]

Bollerslev (1986). [9] Engle and Manganelli (2004) have shown that the typical autocorrelation found in the GARCH variance characterizes also the quantiles of financial return distributions. The intuition is that since the quantile is linked to the variance of a distribution, it ought to share similar empirical properties.

By the same token, since the projectile is associated with the lower partial moments of the distribution, it is reasonable to expect it to exhibit some degree of autocorrelation when estimated with daily asset returns. To gain an insight on how to model conditional projectiles, it is worthwhile to explore their relationship with the standard GARCH model.

Proposition 6 - *Consider the following GARCH(1,1) model:*

$$\begin{aligned} y_t &= \sigma_t \varepsilon_t \qquad\qquad \varepsilon_t \backsim i.i.d.(0,1) \\ \sigma_t^2 &= \gamma_0 + \gamma_1 y_{t-1}^2 + \gamma_2 \sigma_{t-1}^2 \end{aligned} \tag{12}$$

The corresponding (k, q)*-projectile is given by:*

$$f_t(\beta) = c_t[\gamma_0 + \gamma_1 y_{t-1}^2 + \gamma_2 (f_{t-1}/c_{t-1})^2]^{1/2}$$

where $c_t \equiv \int_{-\infty}^{\tilde{f}_t} |\tilde{f}_t - \varepsilon|^{q-1} dF(\varepsilon)$*,* $\tilde{f}_t$ *is the* $(\tilde{k}_t, q)$*-projectile of* ε *and* $\tilde{k}_t \equiv k\sigma_t^{q-2}$*. If* $q = 2$ *and/or if* $\gamma_1 = \gamma_2 = 0$ *(i.e. if there is no heteroscedasticity),* $c_t = c_{t-1} = c$*.*

The above proposition shows that, since c_t will be in general time varying, there is no one to one relationship between the parameters of a GARCH model and the parameters of the conditional projectile. However, there is no a priori reason to define the DGP starting from a GARCH model. One could define the DGP starting directly from the projectile model. Motivated by the projectile derived in the above proposition, we propose the following specification:

$$f_t(\beta) = \beta_0 + \beta_1 |y_{t-1}| + \beta_2 f_{t-1}(\beta) \tag{13}$$

Apart from neglecting the time varying c_t, this specification could be derived from a GARCH process where we model the standard deviation, rather than the variance. Analogously to the Conditional Autoregressive Value at Risk (CAViaR) model by [9] Engle and Manganelli (2004), the autoregressive term $\beta_2 f_{t-1}(\beta)$ ensures that the projectile changes smoothly over time. The idea is that the autocorrelation typical of second moments is reflected in autocorrelated lower partial moments. The role of $\beta_1 |y_{t-1}|$ is to link the time t projectile to past returns.

Other specifications allowing for asymmetries in the projectile reaction to news are possible (see, for instance, the different models proposed by [9] Engle and Manganelli 2004).

Similarly to the CAViaR, the projectile model is more general than the corresponding GARCH model, as it doesn't require to impose a full structure on the underlying distribution. Moreover, unlike CAViaR models, both the minimization problem and the estimation of the variance-covariance matrix are much simpler to perform, since the loss function is everywhere differentiable.

4.1 Empirical Application

We estimated model (13) on a time series of daily IBM log returns. The price series was downloaded from Datastream and ranges from November 14, 1997 to 14 July, 2005, for a total of 2000 observations. We initialized each projectile with the unconditional projectile of the first 200 observations. We experimented with different initial conditions for the β parameters using random numbers between 0 and 1. Tolerance levels for function and parameter values were set to 10^{-4}. We used the command *fminsearch* in MATLAB as optimization algorithm, which is based on the Nelder-Mead simplex. Convergence is fast and very robust to the choice of initial conditions.

In table 1 we report parameter estimates and related standard errors for different cases. We consider three degrees of risk aversion ($q = 2$, 3 and 4) and three values for the penalty k in the PLS loss function (3).

	$k=5$			$k=10$			$k=20$		
	$q=2$	$q=3$	$q=4$	$q=2$	$q=3$	$q=4$	$q=2$	$q=3$	$q=4$
β_0	-0.01	-0.06	-0.08	-0.04	-0.09	-0.11	-0.06	-0.11	-0.14
s.e.	*(0.00)*	*(0.12)*	*(0.18)*	*(0.01)*	*(0.18)*	*(0.18)*	*(0.01)*	*(0.27)*	*(0.17)*
β_1	-0.02	-0.17	-0.28	-0.06	-0.23	-0.28	-0.09	-0.28	-0.27
s.e.	*(0.00)*	*(0.20)*	*(0.25)*	*(0.01)*	*(0.27)*	*(0.21)*	*(0.01)*	*(0.34)*	*(0.17)*
β_2	0.98	0.89	0.92	0.93	0.89	0.93	0.92	0.89	0.94
s.e.	*(0.00)*	*(0.12)*	*(0.09)*	*(0.01)*	*(0.13)*	*(0.07)*	*(0.01)*	*(0.15)*	*(0.05)*

Table 1: Parameter estimates and standard errors for projectile with different degree of risk aversion q and different k.

The striking feature of these results is that the autoregressive coefficient β_2 associated to the lagged values of the projectile in (13) hovers around 0.90 for all estimated models. This indicates that conditional projectiles tend to be very persistent, reflecting the clustering of volatilities typically found in financial data. This finding is consistent with the results from the large GARCH literature and with the more recent results on conditional quantiles by [9] Engle and Manganelli (2004). For values of q greater than 2, the coefficients β_0 and β_1 become insignificant, but the autoregressive coefficient remains highly significant.

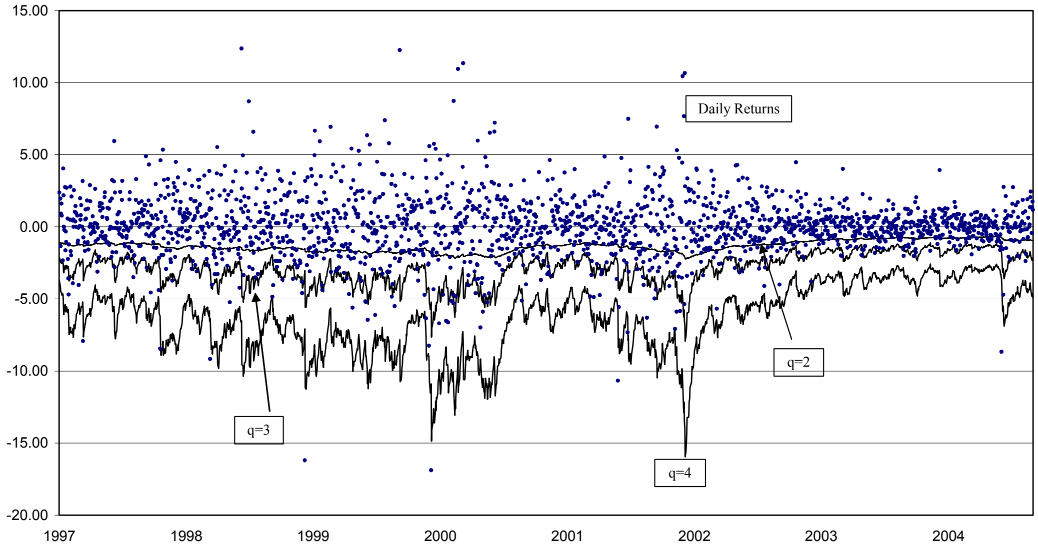

Figure 1: Figure 1: Projectiles for different risk aversion parameters q and $k = 5$. IBM daily returns.

In figure 1, we provide an illustration of the dynamic behaviour of the conditional projectile for different degrees of risk aversion q and $k = 5$. The plots resemble the typical GARCH variances and CAViaR quantiles. The projectile tends to be very persistent, reacting to large realizations of the previous day returns. Increasing the coefficient of risk aversion q results as expected in a lower projectile.

5 Asset Allocation

As explained in the introduction, the main drawback of [10] Fishburn's (1977) (α, t) model is that when dealing with lower partial moments of a distribution, the associated portfolio optimization problem becomes computationally intractable. In this section, we show how the asset allocation problem associated to the projectile framework developed in this paper is actually computationally trivial.

Consider a portfolio with $n+1$ assets. Denote with a the n-vector of weights associated to the first n assets entering a given portfolio, and denote with $y_t(a)$ the portfolio return at time t, where the dependence on the individual assets weights has been made explicit. Since all the weights must sum to one, note that $\sum_{i=1}^{n} a_i = 1 - a_{n+1}$, where a_{n+1} is the weight associated to the $(n+1)^{th}$ asset of the portfolio.

As noted in section 2, the projectile gives the expected mean-risk utility associated to a given portfolio a. Agents will choose the portfolio allocation a^* that maximizes their expected utility. The portfolio allocation problem can therefore be reformulated as follows:

$$\max_{a} \mu_{yt}^{kq}(a) \tag{14}$$

where μ_{yt}^{kq} has been defined in (7). A nice feature of projectiles is that they are convex functions of portfolio weights as shown in the following theorem.

Theorem 7 (Convexity) - *$\mu_{yt}^{kq}(a)$ with $k \geq 0$ and $q \geq 2$ is convex with respect to a.*

Convexity greatly simplifies the numerical optimization problem, as it guarantees that any local maximum is also global. Note that when in the loss function (3) we set $q = 2$, the first order conditions are piecewise linear (when modeling unconditional projectiles). Variants of this special case can be framed as a linear programming problem and have been studied by [18] Rockafellar and Uryasev (2000), [20] Ruszczynski and Vanderbei (2003), [1] Acerbi (2004), [4] Bassett, Koenker and Kordas (2004), and [6] Bertsimas, Lauprete and Samarov (2004).

To solve the optimization problem in (14), rewrite the projectile specification as $f_t(a, \hat{\beta}_T(a))$, to highlight its dependence on the portfolio weights. Next, note that the methodology of [13] Manganelli (2004) allows one to compute the analytical first and second derivatives of $f_t(a, \hat{\beta}_T(a))$ w.r.t. a, which can then be fed to a standard optimization algorithm.

The key insight is to recognize that the projectile is a function of the portfolio weights not only through the portfolio returns that enter its specification, but also through the estimated parameters, which are function of the portfolio returns and therefore also of the portfolio weights:

$$\nabla_a f_t(a, \hat{\beta}_T(a)) = \nabla_a f_t(a, \hat{\beta}_T) + \nabla_a \hat{\beta}_T'(a) \nabla_\beta f_t(a, \hat{\beta}_T) \tag{15}$$

$$\begin{aligned} \nabla_{aa'} f_t(a, \hat{\beta}_T(a)) &= \nabla_{aa'} f_t(a, \hat{\beta}_T) + \nabla_{a\beta'} f_t(a, \hat{\beta}_T) \nabla_{a'} \hat{\beta}_T(a) + \\ &+ \nabla_{a'} vec\left(\nabla_a \hat{\beta}_T'(a)\right)\left(\nabla_\beta f_t(a, \hat{\beta}_T) \otimes I_n\right) + \\ &+ \nabla_a \hat{\beta}_T'(a) \nabla_{\beta a'} f_t(a, \hat{\beta}_T) \end{aligned} \tag{16}$$

where $\nabla_a f_t(a, \hat{\beta}_T) \equiv \frac{\partial}{\partial a} f_t(a, \beta)|_{\beta=\hat{\beta}_T}$, $\nabla_{a\beta'} f_t(a, \hat{\beta}_T) \equiv \frac{\partial^2}{\partial a \partial \beta'} f_t(a, \beta)|_{\beta=\hat{\beta}_T}$, I_n is an (n,n) identity matrix and vec and $\otimes$ denote the vec and Kronecker operator, respectively. To evaluate equations (15) and (16), it is necessary to derive $\nabla_a \hat{\beta}_T'(a)$ and $\nabla_{a'} vec\left(\nabla_a \hat{\beta}_T'(a)\right)$, the other terms being easily obtained. These derivatives can be computed by applying the implicit function theorem to the first order conditions of the PLS maximization problem. The first order conditions of the PLS maximization problem are:

$$\varphi_{kq}(a, \hat{\beta}_T(a)) \equiv T^{-1} \sum_{t=1}^{T} \psi_{kq}(y_t - f_t(\hat{\beta}_T)) \nabla_\beta f_t(\hat{\beta}_T) = 0 \tag{17}$$

Application of the implicit function theorem as in theorem 1 of [13] Manganelli (2004) gives:[2]

$$\nabla_a \hat{\beta}'_T(a) = -[\nabla_a \varphi'_{kq}(a, \hat{\beta}_T)][\nabla_{\beta'} \varphi_{kq}(a, \hat{\beta}_T)]^{-1} \tag{18}$$

$$\begin{aligned} \nabla_{a'} vec\left(\nabla_a \hat{\beta}'_T(a)\right) &= -\left([\nabla_{\beta'} \varphi_{kq}(a, \hat{\beta}_T)]^{-1} \otimes \nabla_a \hat{\beta}'_T(a)\right) \nabla_{a'} vec[\nabla_{\beta'} \varphi_{kq}(a, \hat{\beta}_T)] \\ &\quad - \left([\nabla_{\beta'} \varphi_{kq}(a, \hat{\beta}_T)]^{-1} \otimes I_n\right) \nabla_{a'} vec[\nabla_{a'} \varphi_{kq}(a, \hat{\beta}_T)] \end{aligned} \tag{19}$$

5.1 Empirical Application

We apply our methodology to monthly log returns of the 30 stocks of the Dow Jones Industrial Average (DJIA) index, as of July 15, 2005. The sample runs from January 1, 1987 to July 1, 2005, for a total of 223 observations.

To compare our results with standard mean-variance optimizations, we model the projectile as a constant, i.e. $f_t(\beta) = \beta$. This simplifies the calculations of first and second derivatives considerably. Conceptually, however, the same framework would work with time-varying conditional projectiles as well.

When $q > 2$, applying the formulae for the first and second derivatives,

[2]Note that $\varphi_{kq}(a, \hat{\beta}_T(a))$ is not differentiable whenever $y_t = f_t(\hat{\beta}_T)$. However, the points over which this condition is satisfied form a set of measure zero.

we get:

$$\nabla_a \varphi'_{kq}(a, \hat{\beta}_T) = -T^{-1} \sum_{t=1}^{T} \nabla_a y_t(a)[2 + kq(q-1) I(y_t < \hat{\beta}_T) (\hat{\beta}_T - y_t)^{q-2}] \tag{20}$$

$$\nabla_{\beta'} \varphi_{kq}(a, \hat{\beta}_T) = T^{-1} \sum_{t=1}^{T} [2 + kq(q-1) I(y_t < \hat{\beta}_T)(\hat{\beta}_T - y_t)^{q-2}] \tag{21}$$

$$\nabla_{a'} vec[\nabla_{a'} \varphi_{kq}(a, \hat{\beta}_T)] = T^{-1} \sum_{t=1}^{T} [2 + kq(q-1)(q-2) I(y_t < \hat{\beta}_T) (\hat{\beta}_T - y_t)^{q-3}] \nabla_a y_t(a) \nabla'_a y_t(a) \tag{22}$$

$$\nabla_{a'} vec[\nabla_{\beta'} \varphi_{kq}(a, \hat{\beta}_T)] = -T^{-1} \sum_{t=1}^{T} [2 + kq(q-1)(q-2) I(y_t < \hat{\beta}_T) (\hat{\beta}_T - y_t)^{q-3}] \nabla_a y_t(a) \tag{23}$$

We computed the optimal portfolio allocation for two coefficients of risk aversion, $q = 2.5$ and $q = 4.5$. The penalty weight k was set equal to 10. Convergence is very fast and robust to different initial conditions. This is not surprising, given the convexity result of theorem 7. In figure 2, we report the cumulative distribution functions of the two optimal mean-risk portfolios, together with the cumulative distribution functions of the DJIA portfolio (using the weights as of July 15, 2005) and of the standard optimal mean-variance portfolio. Summary statistics for the distributions of the different portfolios are reported in table 2.

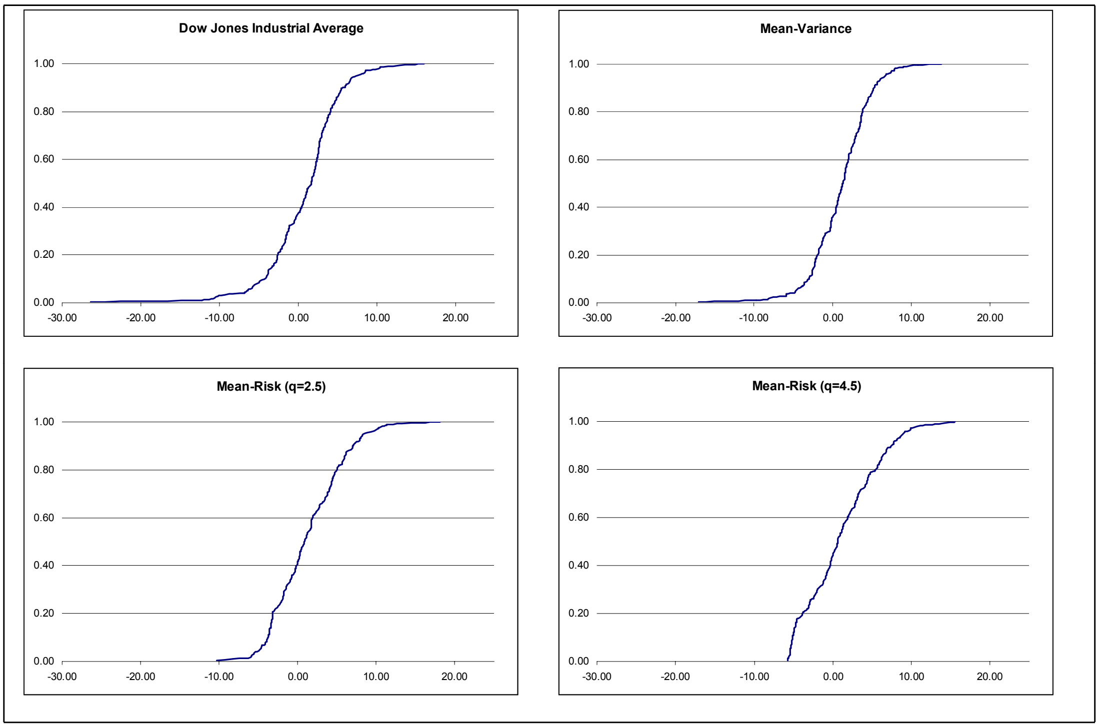

Figure 2: Empirical c.d.f. of DJIA and different mean-risk optimal portfolios.

It is obvious from the figure that the optimal mean-risk portfolio manages to reduce the occurrence of events in the left tail with respect to the mean-variance optimal portfolios. Consistently with economic intuition, higher risk aversion is associated to a shorter left tail. In the case of $q = 4.5$ the maximum loss is limited to less than 6%. This result seems to be particularly striking as the sample includes the crash of October 1987, which resulted in a monthly loss of more than 26% for the DJIA portfolio.

The second nice feature of the mean-risk portfolios is that limited

	Minimum	**Maximum**	**Mean**	**Variance**	**Skewness**	**Kurtosis**
DJIA	-26.43	15.99	0.92	22.35	-0.95	8.02
M-V	-17.11	13.87	1.09	13.71	-0.49	5.61
q=2.5	-10.37	18.12	1.30	19.70	0.50	3.40
q=4.5	-5.78	15.52	1.09	22.01	0.47	2.70

Table 2: Summary statistics for the different optimal portfolios and DJIA.

downside risk does not come at the expenses of upside opportunities. The maximum return of the mean-variance portfolio is lower than the maximum return of the portfolio with risk aversion equal to 4.5 and considerably lower than that for $q = 2.5$.

The average return for the mean-risk portfolios are both higher than the historical average of the DJIA (the average return of the mean-variance model is equal to that of the mean-risk with $q = 4.5$ by construction). Not surprisingly, the variance is lowest for the mean-variance portfolio. However, this comes at the expenses of negative skewness and much higher kurtosis with respect to mean-risk portfolios.

6 Conclusion

This paper developed an econometric framework to estimate the expected utility of an agent who wants to maximize the trade off between expected returns and downside risk of a portfolio. We showed how the estimation problem can be cast in terms of penalized least squares, where the penalty is associated to portfolio returns below a certain threshold. We derived the large sample properties of the estimator, allowing for possible nonlinearities and misspecification of the model. We illustrated the usefulness of this new class of estimators with two empirical applications. First, we modeled the daily behavior of the estimator using an autoregressive specification, in the spirit of the GARCH models. We showed how the process tends to be very persistent and characterized by high autoregressive coefficients, as typically found in the GARCH literature. Second, we proposed a simple strategy to derive the optimal asset allocation associated to a mean-downside risk expected utility. The results show that the empirical c.d.f. of the mean-downside risk optimal portfolio tends to have much shorter left tail and longer right tail than that associated to the standard mean-variance model.

7 Appendix A - Assumptions

Define $u_t \equiv y_t - f_t(\beta^*)$ and $\delta_t(\beta) \equiv f_t(\beta) - f_t(\beta^*)$.

Consistency Assumptions

C0. The observed data are a realization of a stochastic process $X \equiv \{X_t : \Omega \rightarrow \mathbb{R}^v, v \in \mathbb{N}, t = 1, 2, ...\}$ on a complete probability space (Ω, F, P), where $\Omega = \times_{t=1}^{\infty}\mathbb{R}^v$.

C1. $f_t(\beta) : \mathbb{R}^{h_t} \times B \rightarrow \mathbb{R}$ is such that is measurable-F for $\beta \in B$, a compact subset of $\mathbb{R}^p$, and is continuous in B for all ω^t, the realizations of a finite history of explanatory variables $W^t = (W_1', ..., W_t')'$.

C2. $E|u_t|^q < \infty$, $\forall t$.

C3. $E|f_t(\beta)|^q < \infty$ for each $\beta \in B$ and for all t.

C4. $\{\rho_{kq}(u_t - \delta_t(\beta))\}$ obeys the uniform law of large numbers.

C5. For any $\xi > 0$, there exists $\upsilon > 0$ and $T_0 \in \mathbb{N}$ such that, for all $T > T_0$, $\min_{||\beta-\beta^*||>\xi} T^{-1}\sum_{t=1}^{T} E([\delta_t(\beta)]^2) > \upsilon$.

Asymptotic Normality Assumptions

AN0. β^* is an interior point of B.

AN1. $f_t(\beta)$ is twice differentiable for each $\beta \in B$. Moreover, for all β and γ in a neighborhood of β^* such that $||\beta - \gamma|| < d$ for d sufficiently small and for all t:

(i) $||\nabla_\beta f_t(\beta)|| \leq F_1(W_t)$

(ii) $||\nabla_{\beta\beta} f_t(\beta)|| \leq F_2(W_t)$

where $F_1(W_t)$ and $F_2(W_t)$ are some possibly stochastich functions of W_t, such that $E[F_1(W_t)^q] < \infty$ and $E[F_1(W_t)^{q-1} F_2(W_t)] < \infty$.

AN2. There exists some (possibly) stochastic function of $X_t \equiv [Y_t, W_t']'$, $U_1(X_t)$, such that for all t $|u_t| < U_1(X_t)$, where $E[U_1(X_t)^{q-2} F_1(W_t)^2] < \infty$ and $E[U_1(X_t)^{q-1} F_2(W_t)] < \infty$.

AN3. $\{T^{-1} \sum_{t=1}^{T} \left[\psi_{kq}(u_t) \nabla_{\beta\beta} f_t(\beta) + w_t(k,q) \nabla_\beta f_t(\beta) \nabla'_\beta f_t(\beta)\right]\}$ satisfies the uniform law of large numbers, for β in a neighborhood of β^*, where $\psi_{kq}(\lambda)$ and $w_t(k,q)$ were defined in Assumption M and equation (9), respectively.

AN4. The sequence $\{\sqrt{T} \sum_{t=1}^{T} \psi_{kq}(u_t) \nabla_\beta f_t(\beta^*)\}$ obeys the central limit theorem.

8 Appendix B - Proofs of theorems in the text

Proof of theorem 2 (Correct Model Specification) - Define $\eta_t(\beta) \equiv f_t(\beta) - \mu_{yt}^{kq}$, so that $\varepsilon_t^{kq} = u_t + \eta_t(\beta^*)$. Then:

$$\begin{aligned}
E_t[\psi_{kq}(u_t)] &= E_t[-2u_t + kqI(u_t < 0)|u_t|^{q-1}] \\
&= E_t[-2(\varepsilon_t^{kq} - \eta_t(\beta^*)) + kqI(\varepsilon_t^{kq} < \eta_t(\beta^*))|\eta_t(\beta^*) - \varepsilon_t^{kq}|^{q-1}] \\
&= E_t[-2\varepsilon_t^{kq} + kqI(\varepsilon_t^{kq} < 0)|\varepsilon_t^{kq}|^{q-1} + \\
+2\eta_t(\beta^*) + \quad kq \quad & I(\varepsilon_t^{kq} < \eta_t(\beta^*))|\eta_t(\beta^*) - \varepsilon_t^{kq}|^{q-1} - kqI(\varepsilon_t^{kq} < 0)|\varepsilon_t^{kq}|^{q-1}]
\end{aligned}$$

Note that the term in the first row in the last equality is zero by (7). The term in the second row will be always greater than 0 when $\eta_t(\beta^*) > 0$ and less than 0 when $\eta_t(\beta^*) < 0$ (this follows by the properties of the integral). Therefore a necessary and sufficient condition for $E_t[\psi_{kq}(u_t)] = 0$ is $\eta_t(\beta^*) = 0$, which is equivalent to assuming that the conditional expectile model is correctly specified. ■

Proof of theorem 3 (Consistency) - We verify that the conditions of theorem 3.5 of [22] White (1994) are satisfied.

Assumption 2.1 in [22] White (1994) is assumption C0. Assumption 2.3 is also satisfied, given that $f_t(\beta)$ is continuous by assumption C1 and

$\rho_{kq}(y_t - f_t(\beta))$ is also continuous in β.

Assumption 3.1(a) in [22] White (1994) requires to show that $E[\rho_{kq}(y_t - f_t(\beta))]$ exists and is finite.

$$\begin{aligned} E[\rho_{kq}(y_t - f_t(\beta))] &= E[\rho_{kq}(u_t - \delta_t(\beta))] \\ &= E[(u_t - \delta_t(\beta))^2 + kI(u_t < \delta_t(\beta))|\delta_t(\beta) - u_t|^q] \\ &\leq E[|u_t - \delta_t(\beta)|^2 + k|u_t - \delta_t(\beta)|^q] \\ &\leq 2[E|u_t|^2 + E|\delta_t(\beta)|^2] + k2^{q-1}[E|u_t|^q + E|\delta_t(\beta)|^q] \end{aligned}$$

because of the inequality $E|X+Y|^r \leq c_r[E|X|^r + E|Y|^r]$, where $c_r = 2^{r-1}$, $r \geq 1$ (see, e.g., Zellner, p.111, Handbook of Econometrics, Vol.1). The result follows from assumptions C2 and C3.

Assumption 3.1(b) in [22] White (1994) (the continuity of $E[\rho_{kq}(y_t - f_t(\beta))]$ in B) follows from the continuity of $\rho_{kq}(\lambda)$ in λ and the continuity of $f_t(\beta)$ in β (assumption C1). Assumption 3.1(c) in [22] White (1994) is simply C4.

It remains to verify Assumption 3.2 of theorem 3.5 in [22] White (1994), that is that $E[T^{-1}\sum_{t=1}^{T}\rho_{kq}(u_t - \delta_t(\beta))]$ has identifiably unique minimizers β_T^*, that is we need to show that $E[T^{-1}\sum_{t=1}^{T}\rho_{kq}(u_t - \delta_t(\beta))] - E[T^{-1}\sum_{t=1}^{T}\rho_{kq}(u_t)] > 0$ if $\min\limits_{||\beta-\beta^*||>\xi} T^{-1}\sum_{t=1}^{T}E[\delta_t(\beta)^2] > \upsilon$. Consider each element at time t:

$$\begin{aligned} E[\rho_{kq}(u_t - \delta_t(\beta)) - \rho_{kq}(u_t)] &= E[(u_t - \delta_t(\beta))^2 + kI(u_t < \delta_t(\beta))|\delta_t(\beta) - u_t|^q - \\ -u_t^2 - kI(u_t < 0)|u_t|^q] \\ &= E[u_t^2 + \delta_t(\beta)^2 - 2\delta_t(\beta)u_t + kI(u_t < \delta_t(\beta))|\delta_t(\beta) - u_t|^q - \\ -u_t^2 - kI(u_t < 0)|u_t|^q + kq & \; I(u_t < 0)|u_t|^{q-1}\delta_t(\beta) - kqI(u_t < 0)|u_t|^{q-1}\delta_t(\beta)] \end{aligned}$$

Now note that by assumption M, $E[\{-2u_t + kqI(u_t < 0)|u_t|^{q-1}\}\delta_t(\beta)] =$ $E[\psi_{kq}(u_t)(f_t(\beta) - f_t(\beta^*))] = 0$. Therefore:

$$\begin{aligned} E[\rho_{kq}(u_t - \delta_t(\beta)) - \rho_{kq}(u_t)] &= E[\delta_t(\beta)^2 + k\{I(u_t < \delta_t(\beta))|\delta_t(\beta) - u_t|^q - \\ -I(u_t < 0)|u_t|^q - qI & \; (u_t < 0)|u_t|^{q-1}\delta_t(\beta)\}] \\ &\equiv E[\delta_t(\beta)^2 + kA_t] \end{aligned}$$

If we show that $A_t \geq 0 \; \forall t$, the result follows. We need to consider two cases.

i) $\delta_t(\beta) > 0$

$$\begin{aligned} A_t &= [I(u_t < 0) + I(0 \leq u_t \leq \delta_t(\beta))]|\delta_t(\beta) - u_t|^q - I(u_t < 0)[|u_t|^q + q|u_t|^{q-1}\delta_t(\beta)] \\ &\geq I(u_t < 0)[|\delta_t(\beta) - u_t|^q - |u_t|^q - q|u_t|^{q-1}\delta_t(\beta)] \\ &\equiv I(u_t < 0)B_t \end{aligned}$$

ii) $\delta_t(\beta) < 0$

$$A_t = I(u_t < \delta_t(\beta))[|\delta_t(\beta) - u_t|^q - |u_t|^q - q|u_t|^{q-1}\delta_t(\beta)] - \\ - I(\delta_t(\beta) \leq u_t \leq 0)|u_t|^{q-1}[-u_t + q\delta_t(\beta)]$$

The last term is positive if $-u_t + q\delta_t(\beta) < 0$, which is true whenever $q \geq 1$ and $\delta_t(\beta) \leq u_t \leq 0$. Therefore:

$$\begin{aligned} A_t &\geq I(u_t < \delta_t(\beta))[|\delta_t(\beta) - u_t|^q - |u_t|^q - q|u_t|^{q-1}\delta_t(\beta)] \\ &\equiv I(u_t < \delta_t(\beta))B_t \end{aligned}$$

Showing that $A_t \geq 0$ is therefore equivalent to showing that $B_t \geq 0$ whenever $u_t < 0$:

$$\begin{aligned} B_t &= |u_t|^q[| - \delta_t(\beta)/u_t + 1|^q - 1 + q\delta_t(\beta)/u_t] \\ &\equiv |u_t|^q C_t \end{aligned}$$

C_t is a function of the type $f(a;q) = (a+1)^q - (1+aq)$, where $q \geq 1$ and $-1 < a < \infty$ (this follows from the fact that $u_t < 0$ and $u_t < \delta_t(\beta)$ when $\delta_t(\beta) < 0$). Since $f(a;q)$ is globally concave and achieves a minimum at $f(0;q) = 0$, we have shown that $C_t \geq 0 \Longrightarrow B_t \geq 0 \Longrightarrow A_t \geq 0$. Therefore $E[\rho_{kq}(u_t - \delta_t(\beta)) - \rho_{kq}(u_t)] \geq E[\delta_t(\beta)^2]$ $\forall t$ and:

$$E[T^{-1}\sum_{t=1}^{T}\rho_{kq}(u_t - \delta_t(\beta))] - E[T^{-1}\sum_{t=1}^{T}\rho_{kq}(u_t)] \geq E[T^{-1}\sum_{t=1}^{T}\delta_t(\beta)^2] > 0$$

if $\min_{||\beta-\beta^*||>\xi} T^{-1}\sum_{t=1}^{T} E[\delta_t(\beta)^2] > \upsilon$, by assumption C5. ■

Proof of theorem 4 (Asymptotic Normality) - Since the PLS objective function is not twice differentiable, we need to resort to asymptotic normality results for nonsmooth objective function. We show that the conditions of theorems 7.2 and 7.3 of [16] Newey and McFadden (1994) hold. Define the following:

$$\begin{aligned} g(x_t;\beta) &\equiv \psi_{kq}(y_t - f_t(\beta))\nabla_\beta f_t(\beta) \text{ where } x_t \equiv (y_t, \omega_t')' \\ g_0(\beta) &\equiv E[g(x_t;\beta)] \\ \hat{g}_T(\beta) &\equiv T^{-1}\sum_{t=1}^{T} g(x_t;\beta) \end{aligned}$$

To apply theorem 7.2 of [16] Newey and McFadden (1994), we need to check the following conditions:

1. $g_0(\beta^*) = 0$
2. $r(x_t;\beta) = ||g(x_t;\beta) - g(x_t;\beta^*) - \Delta(x_t)(\beta-\beta^*)||/||\beta-\beta^*|| \to 0$ as $\beta \to \beta^*$, where $\Delta(x_t)$ is some function of x_t.
3. $E\left[\sup_{||\beta-\beta^*||<\varepsilon} r(x_t;\beta)\right] < \infty$ for $\varepsilon > 0$
4. $T^{-1}\sum_{t=1}^{T}\Delta(x_t) \xrightarrow{p} Z \equiv E\left[\Delta(x_t)\right]$
5. β^* is an interior point of B.

6. $\sqrt{T}\hat{g}_T(\beta^*) \xrightarrow{d} N(0, V)$

If these conditions are satisfied, then theorems 7.2 and 7.3 of [16] Newey and McFadden (1994) imply that $\sqrt{T}(\hat{\beta}_T - \beta^*) \xrightarrow{d} N(0, Z^{-1}VZ^{-1})$.

Let $\Delta(x_t) \equiv \left[\psi_{kq}(u_t)\nabla_{\beta\beta}f_t(\beta^*) + w_t(k,q)\nabla_\beta f_t(\beta^*)\nabla'_\beta f_t(\beta^*)\right]$. Conditions 4, 5 and 6 are automatically satisfied by assumptions AN3, AN0 and AN4, respectively.

For condition 1, we first check that $E|\psi_{kq}(u_t)\nabla_\beta f_t(\beta)| < \infty$ and then apply the Lebesgue Dominated Convergence Theorem (LDCT).

$$
\begin{aligned}
E|\psi_{kq}(u_t)\nabla_\beta f_t(\beta)| &= E|[-2u_t + kqI(u_t < 0)|u_t|^{q-1}]\nabla_\beta f_t(\beta)| \\
&\leq E[(2|u_t| + kq|u_t|^{q-1})||\nabla_\beta f_t(\beta)||] \\
&\leq E[\max[2, kq]|u_t|^{q-1}||\nabla_\beta f_t(\beta)||] \\
&\leq E[\max[2, kq]U_1(X_t)F_1(W_t)||] \\
&< \infty
\end{aligned}
$$

Having found a dominating function for $\psi_{kq}(u_t)\nabla_\beta f_t(\beta)$ with finite expected value, we can apply the LDCT:

$$
\begin{aligned}
E|\psi_{kq}(u_t)\nabla_\beta f_t(\beta^*)| &= \frac{\partial}{\partial\beta}E[\psi_{kq}(u_t)f_t(\beta)]|_{\beta=\beta^*} \\
&= 0
\end{aligned}
$$

by assumption M.

To establish condition 2, consider the components of $r(x_t;\beta)$:

$$g(x_t;\beta) - g(x_t;\beta^*) \equiv \psi_{kq}(u_t - \delta_t(\beta))\nabla_\beta f_t(\beta) - \psi_{kq}(u_t)\nabla_\beta f_t(\beta^*)$$

$$= -2(u_t - \delta_t(\beta))\nabla_\beta f_t(\beta) + kqI(u_t < \delta_t(\beta))|\delta_t(\beta) - u_t|^{q-1}$$

$$+2u_t\nabla_\beta f_t(\beta^*) - kqI(u_t < 0)|u_t|^{q-1}\nabla_\beta f_t(\beta^*)$$

$$= -2u_t(\nabla_\beta f_t(\beta) - \nabla_\beta f_t(\beta^*)) + 2\delta_t(\beta)\nabla_\beta f_t(\beta) +$$

$$+ kq\{I(u_t < \delta_t(\beta))|\delta_t(\beta) - u_t|^{q-1}(\nabla_\beta f_t(\beta) - \nabla_\beta f_t(\beta^*)) +$$

$$+ [I(u_t < \delta_t(\beta))|\delta_t(\beta) - u_t|^{q-1} - I(u_t < 0)|u_t|^{q-1}]\nabla_\beta f_t(\beta^*)\}$$

$$= [-2u_t + kqI(u_t < \delta_t(\beta))|\delta_t(\beta) - u_t|^{q-1}]\nabla_{\beta\beta} f_t(\tilde{\beta})(\beta - \beta^*) +$$

$$+ 2\nabla_\beta f_t(\beta)\nabla'_\beta f_t(\tilde{\beta})(\beta - \beta^*) +$$

$$+ (q-1)kqI(u_t < \delta_t(\beta))|\delta_t(\beta) - u_t|^{q-2}\nabla_\beta f_t(\beta^*)\nabla'_\beta f_t(\tilde{\beta})(\beta - \beta^*)$$

where $\tilde{\beta}$ comes from the mean value theorem.

Substituting everything into $r(x_t, \beta)$, we get:

$$r(x_t, \beta) \leq ||[-2u_t + kqI(u_t < \delta_t(\beta))|\delta_t(\beta) - u_t|^{q-1}]\nabla_{\beta\beta} f_t(\tilde{\beta}) - \psi_{kq}(u_t)\nabla_{\beta\beta} f_t(\beta^*)|| +$$

$$+ ||2\nabla_\beta f_t(\beta)\nabla'_\beta f_t(\tilde{\beta}) - 2\nabla_\beta f_t(\beta^*)\nabla'_\beta f_t(\beta^*)|| +$$

$$+ ||(q-1)kqI(u_t < \delta_t(\beta))|\delta_t(\beta) - u_t|^{q-2}\nabla_\beta f_t(\beta^*)\nabla'_\beta f_t(\tilde{\beta}) -$$

$$- (q-1)kqI(u_t < 0)|u_t|^{q-2}\nabla_\beta f_t(\beta^*)\nabla'_\beta f_t(\beta^*)||$$

$$\equiv r_1(\tilde{\beta}) + r_2(\tilde{\beta}) + r_3(\tilde{\beta})$$

These functions are continuous in β and obviously converge to zero almost surely as $\beta \to \beta^*$.

We next prove condition 3. From the previous step we have:

$$\sup_{||\beta-\beta^*||<\varepsilon} r(x_t;\beta) \leq \sup_{||\beta-\beta^*||<\varepsilon} \left[r_1(\tilde{\beta}) + r_2(\tilde{\beta}) + r_3(\tilde{\beta}) \right]$$

We check element by element. For $r_1(\tilde{\beta})$, note first that if $\delta_t(\beta) < 0$, $I(u_t < \delta_t(\beta))|\delta_t(\beta) - u_t|^{q-1} \leq |u_t|^{q-1}$ and if $\delta_t(\beta) > 0$, $I(u_t < \delta_t(\beta))|\delta_t(\beta) - u_t|^{q-1} \leq I(-\delta_t(\beta) < u_t < \delta_t(\beta))|2\delta_t(\beta)|^{q-1} + I(u_t < -\delta_t(\beta))|u_t|^{q-1} \leq |2\delta_t(\beta)|^{q-1} + |u_t|^{q-1}$. Therefore, $I(u_t < \delta_t(\beta))|\delta_t(\beta) - u_t|^{q-1} \leq |2\delta_t(\beta)|^{q-1} + |u_t|^{q-1} = |2\nabla'_\beta f_t(\tilde{\beta})(\beta - \beta^*)|^{q-1} + |u_t|^{q-1}$.

$$\begin{aligned} r_1(\tilde{\beta}) &\leq 4|u_t| \cdot ||\nabla_{\beta\beta} f_t(\beta)|| + kq\varepsilon^{q-1}||2\nabla_\beta f_t(\beta)||^{q-1} \cdot ||\nabla_{\beta\beta} f_t(\beta)|| + 2kq|u_t|^{q-1} \cdot ||\nabla_{\beta\beta} f_t(\beta)|| \\ &\leq 4U_1(X_t)F_2(W_t) + kq(2\varepsilon)^{q-1}F_1(W_t)^{q-1}F_2(W_t) + 2kqU_1(X_t)^{q-1}F_2(W_t) \end{aligned}$$

whose expectation is finite by assumptions AN1 and AN2.

$$\sup_{||\beta-\beta^*||<\varepsilon} r_2(\tilde{\beta}) \leq 4F_1(W_t)^2$$

For $r_3(\tilde{\beta})$, note that reasoning as before $I(u_t < \delta_t(\beta))|\delta_t(\beta) - u_t|^{q-2} \leq |2\nabla'_\beta f_t(\tilde{\beta})(\beta - \beta^*)|^{q-2} + |u_t|^{q-2}$.

$$\begin{aligned} r_3(\tilde{\beta}) &\leq (q-1)kq(2\varepsilon)^{q-2}||\nabla_\beta f_t(\beta)||^q + 2(q-1)kq|u_t|^{q-2}||\nabla_\beta f_t(\beta)||^2 \\ &\leq (q-1)kq(2\varepsilon)^{q-2}F_1(W_t)^q + 2(q-1)kqU_1(X_t)^{q-2}F_1(W_t)^2 \end{aligned}$$

whose expectation is finite by assumptions AN1 and AN2.

To conclude the proof, note that

$$\begin{aligned}
& T^{-1}\sum_{t=1}^{T}\psi_{kq}(u_t)\nabla_{\beta\beta}f_t(\beta^*) \xrightarrow{p} T^{-1}\sum_{t=1}^{T}E[\psi_{kq}(u_t)\nabla_{\beta\beta}f_t(\beta^*)] \\
= \; & T^{-1}\sum_{t=1}^{T}\frac{\partial^2}{\partial\beta\partial\beta'}E[\psi_{kq}(u_t)f_t(\beta)]_{\beta=\beta^*} \\
= \; & 0
\end{aligned}$$

by the LDCT and assumption M. ■

Proof of theorem 5 (Variance-Covariance Matrix Estimation) - Define $\tilde{V}_T \equiv T^{-1}\sum_{t=1}^{T}\psi_{kq}(u_t)^2\nabla_\beta f_t(\beta^*)\nabla'_\beta f_t(\beta^*)$. It suffices to show that

$\hat{V}_T - \tilde{V}_T \xrightarrow{p} 0$, as$\tilde{V}_T - V_T \xrightarrow{p} 0$ is guaranteed by assumption AN4.

$$\begin{aligned}
\hat{V}_T - \tilde{V}_T \;\equiv\; & T^{-1}\sum_{t=1}^{T}\{\psi_{kq}(\hat{u}_t)^2\nabla_\beta f_t(\hat{\beta})\nabla'_\beta f_t(\hat{\beta}) - \psi_{kq}(u_t)^2\nabla_\beta f_t(\beta^*)\nabla'_\beta f_t(\beta^*)\} \\
= \; & T^{-1}\sum_{t=1}^{T}\{\psi_{kq}(\hat{u}_t)^2\nabla_\beta f_t(\hat{\beta})[\nabla'_\beta f_t(\hat{\beta}) - \nabla'_\beta f_t(\beta^*)] + \\
& +\psi_{kq}(\hat{u}_t)^2[\nabla_\beta f_t(\hat{\beta}) - \nabla_\beta f_t(\beta^*)]\nabla'_\beta f_t(\beta^*) + \\
& +[\psi_{kq}(\hat{u}_t)^2 - \psi_{kq}(u_t)^2]\nabla_\beta f_t(\beta^*)\nabla'_\beta f_t(\beta^*)\}
\end{aligned}$$

Consistency of $\hat{\beta}$ guarantees that the expressions in the last three lines converge to zero in probability. The proof for $\hat{Z}_T$ is similar.

Proof of proposition 6 - The (k,q)-projectile of y_t is

$$\begin{aligned} f_t &= -0.5kq\int_{-\infty}^{f_t}|f_t-y|^{q-1}dF(y) \\ &= -0.5\sigma_t^{q-1}kq\int_{-\infty}^{f_t/\sigma_t}|f_t/\sigma_t-\varepsilon|^{q-1}dF(\varepsilon) \end{aligned}$$

Rearranging:

$$f_t/\sigma_t = -0.5\sigma_t^{q-2}kq\int_{-\infty}^{f_t/\sigma_t}|f_t/\sigma_t-\varepsilon|^{q-1}dF(\varepsilon)$$

Note that f_t/σ_t is the $(\tilde{k}_t,q)$-projectile of ε, where $\tilde{k}_t \equiv \sigma_t^{q-2}k$. The result follows from noting that $\tilde{k}_t$ is constant if $q=2$ and/or if σ_t is constant, and from the fact that ε is i.i.d. ■

Proof of theorem 7 (Convexity) - Let y, y^1 and y^2 denote the returns associated to portfolios a, a_1 and a_2, respectively, where $a=\lambda a_1+(1-\lambda)a_2$ and $0<\lambda<1$. Define also $\mu \equiv E[y-0.5kqI(\mu>y)|\mu-y|^{q-1}\}]$ and $\mu^i \equiv E[y^i-0.5kqI(\mu^i>y^i)|\mu^i-y^i|^{q-1}]$, $i=1,2$. Finally, define $x\equiv\mu-y$ and $x^i\equiv\mu^i-y^i$, $i=1,2$.

We need to show that $\mu \geq \lambda\mu^1+(1-\lambda)\mu^2$. Suppose, by contradiction, that $\mu \leq \lambda\mu^1+(1-\lambda)\mu^2$. Then, noticing that $-I(x>0)|x|^{q-1}$ is convex in x:

$$\begin{aligned}
\mu &\geq E[y - 0.5kqI(\lambda\mu^1 + (1-\lambda)\mu^2 > y)|\lambda\mu^1 + (1-\lambda)\mu^2 - y|^{q-1}] \\
&= E[y - 0.5kqI(\lambda x^1 + (1-\lambda)x^2 > 0)|\lambda x^1 + (1-\lambda)x^2|^{q-1}] \\
&\geq E[y - 0.5kq\{\lambda I(x^1 > 0)|x^1|^{q-1} + (1-\lambda)I(x^2 > 0)|x^2|^{q-1}\}] \\
&= \lambda E[y^1 - 0.5kq\lambda I(\mu^1 > y^1)|\mu^1 - y^1|^{q-1}] + (1-\lambda)E[y^2 - 0.5kq\lambda I(\mu^2 > y^2)|\mu^2 - y^2|^{q-1}] \\
&= \lambda\mu^1 + (1-\lambda)\mu^2
\end{aligned}$$

which contradicts the initial assumption. ■

References

[1] **Acerbi, C. (2004)**, Coherent Representations of Subjective Risk-Aversion, in *Risk Measures for the 21st Century*, edited by G. Szegö, John Wiley & Sons Ltd, Chichester, England.

[2] **Artzner, P. F. Delbaen, J.M. Eber and D. Heath (1999)**, Coherent Measures of Risk, *Mathematical Finance*, **9**: 203-228.

[3] **Basak, S. and A. Shapiro (2001)**, Value-at-Risk-Based Risk Management: Optimal Policies and Asset Prices, *The Review of Financial Studies*, **14(2)**: 371-405.

[4] **Bassett, G., R. Koenker and G. Kordas (2004)**, Pessimistic Portfolio Allocation and Choquet Expected Utility, *Journal of Financial Econometrics*, **2(4)**: 477-492.

[5] **Bawa, V.S. (1975)**, Optimal Rules for Ordering Uncertain Prospects, *Journal of Financial Economics*, **2**: 95-121.

[6] **Bertsimas, D., G. Lauprete and A. Samarov (2004)**, Shortfall as a Risk Measure: Properties, Optimization and Applications, *Journal of Economic Dynamics and Control*, **28**: 1353-1381.

[7] **Bollerslev, T. (1986)**, Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, **31**: 307-327.

[8] **Engle, R.F. (1982)**, Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica*, **50(4)**: 987-1006.

[9] **Engle, R.F. and S. Manganelli (2004)**, CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles, *Journal of Business and Economic Statistics*, **22(4)**: 367-381.

[10] **Fishburn, P.C. (1977)**, Mean-Risk Analysis with Risk Associated with Below-Target Returns, *The American Economic Review*, **67(2)**: 116-126.

[11] **Kim, T.H. and H. White (2003)**, Estimation, Inference, and Specification Testing for Possibly Misspecified Quantile Regressions, in *Advances in Econometrics: Maximum Likelihood Estimation of Misspecified Models - Twenty Years Later*, **17**: 107-132, edited by T.B. Fomby and R.C. Hill.

[12] **Koenker, R. and G. Bassett (1978)**, Regression Quantiles, *Econometrica*, **46**: 33-50.

[13] **Manganelli, S. (2004)**, Asset Allocation by Variance Sensitivity Analysis, *Journal of Financial Econometrics* **2(3)**: 370-389.

[14] **Markowitz, H.M. (1952)**, Portfolio Selection, *Journal of Finance*, **39**: 47-61.

[15] **Markowitz, H.M. (1959)**, *Porfolio Selection: Efficient Diversification of Investments*, Wiley, New York.

[16] **Newey, W. and D. McFadden (1994)**, Large Sample Estimation and Hypothesis Testing, in *Handbook of Econometrics*, Volume IV, edited by R.F. Engle and D.L. McFadden, Elsevier Science.

[17] **Newey, W. and J. Powell (1987)**, Asymmetric Least Squares Estimation and Testing, *Econometrica* **55(4)**: 819-847.

[18] **Rockafellar, R. and S. Uryasev (2000)**, Optimization of Conditional VaR, *Journal of Risk*, **2**: 21-41.

[19] **Roy, A.D. (1952)**, Safety First and the Holding of Assets, *Econometrica*, **20**: 431-449.

[20] **Ruszczynski, A. and R. Vanderbei (2003)**, Frontiers of Stochastically Nondominated Portfolios, *Econometrica* **71(4)**: 1287-1297.

[21] **Stone, B.K. (1973)**, A General Class of Three-Parameter Risk Measures, *Journal of Finance*, **28**: 675-685.

[22] **White, H. (1994)**, *Estimation, Inference and Specification Analysis*, Cambridge, UK: Cambridge University Press.

[23] **White, H. and I. Domowitz (1984)**, Nonlinear Regression with Dependent Observations, *Econometrica*, **52**: 143-161.