Journal of Empirical Finance 16 (2009) 264–279

Contents lists available at ScienceDirect

Journal of Empirical Finance

journal homepage: www.elsevier.com/locate/jempfin

Quantile regression analysis of hedge fund strategies

Loukia Meligkotsidou [a], Ioannis D. Vrontos [b,*], Spyridon D. Vrontos [c]

[a] *Department of Mathematics, University of Athens, Athens, Greece*
[b] *Department of Statistics, Athens University of Economics and Business, Patission 76, GR-10434 Athens, Greece*
[c] *Department of Statistics and Insurance Science, University of Piraeus, Piraeus, Greece*

A R T I C L E I N F O

Article history:
Received 11 July 2007
Received in revised form 11 September 2008
Accepted 5 October 2008
Available online 15 October 2008

Jel classification:
G11
G12
C11

Keywords:
Conditional quantiles
Model selection techniques
Model uncertainty
Hedge funds
Bayesian model averaging
Risk factors
Style portfolio construction

A B S T R A C T

Extending previous work on hedge fund pricing, this paper introduces the idea of modelling the conditional quantiles of hedge fund returns using a set of risk factors. Quantile regression analysis provides a way of understanding how the relationship between hedge fund returns and risk factors changes across the distribution of conditional returns. We propose a Bayesian approach to model comparison which provides posterior probabilities for different risk factor models that can be used for model averaging. The most relevant risk factors are identified for different quantiles and compared with those obtained for the conditional expectation model. We find differences in factor effects across quantiles of returns, which suggest that the standard conditional mean regression method may not be adequate for uncovering the risk-return characteristics of hedge funds. We explore potential economic impacts of our approach by analysing hedge fund single strategy return series and by constructing style portfolios.

1. Introduction

Hedge funds have received a vast amount of attention over the last decades. They are alternative investment vehicles which have highly dynamic nature and great flexibility by using leverage, short-selling and derivatives. These characteristics allow for investment strategies that differ significantly from traditional investments, such as mutual funds, which usually employ a non-leveraged, static buy-and-hold strategy. There exist various hedge fund strategies which exhibit different statistical properties and different risk return characteristics (see Fung and Hsieh, 1999). The increasing interest in the hedge fund market has led to a need for sophisticated statistical/econometric models which are able to capture the particular characteristics of hedge fund return series.

Linear regression models have been widely used in the hedge fund literature to describe the causal relationship of hedge fund returns with a set of covariates usually referred to as risk factors. A plethora of articles have been presented to investigate the ability of various linear and/or nonlinear risk factors to explain hedge fund returns, and to evaluate hedge fund performance using different asset pricing models; see, for example, Capocci and Hubner (2004) and the references therein. Although estimating standard multifactor regression models is straightforward, identification of the most important risk factors to be included in the model is difficult. To a large extent, this is due to the dynamic nature of hedge fund investments and the lack of transparency of the hedge fund managers' activities. Furthermore, existing equilibrium pricing theories are not explicit about which factors should enter the regression model.

* Corresponding author. Tel.: +30 210 8203927.
E-mail address: vrontos@aueb.gr (I.D. Vrontos).

0927-5398/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jempfin.2008.10.002

Several variable selection strategies have been used in the context of linear regression. Specifically, the method of stepwise regression (STEP hereafter) has been widely used in hedge fund pricing (see, for example, Liang, 1999; Fung and Hsieh, 2000, 2001; Agarwal and Naik, 2004) to identify valuable pricing factors. The use of information based criteria, such as Akaike's (1973) information criterion (AIC hereafter) and Schwarz's (1978) Bayesian information criterion (BIC hereafter), is also a standard approach to variable selection (Bossaerts and Hillion, 1999). Recent empirical studies have considered the problem of variable/model uncertainty using Bayesian techniques (see, for example, Avramov, 2002; Cremers, 2002; Vrontos et al., 2008). In particular, Vrontos et al. (2008) have compared STEP, AIC, BIC and Bayesian techniques using hedge fund return series and have found that the latter techniques outperform the other selection methods in terms of predictive ability and performance of selected portfolios.

Existing regression specifications for hedge funds model the conditional *expectation* of returns. That is standard regression models describe only the average relationship of hedge fund returns with the set of risk factors. However, this approach might not be adequate due to the particular characteristics of hedge fund returns. Recent research (see, for example, Amin and Kat, 2003) has revealed that, due to their highly dynamic complex nature, hedge fund returns may exhibit a high degree of non-normality, fat tails, excess kurtosis and skewness. In the presence of these characteristics the conditional mean approach may not capture the effect of risk factors to the entire distribution of returns and may provide estimates which are not robust.

The aim of this study is to explore the impact of a number of observable factors on the entire conditional distribution of hedge fund returns by modelling a set of *quantiles* of returns. This approach helps in uncovering potential differences in factor effects across quantiles of returns. In the analysis of hedge funds it is of great interest to examine how various risk factors affect hedge fund returns not only on average but also at high and low values. Looking at just the conditional mean of the hedge fund return series can 'hide' interesting risk-return characteristics. It can lead us to conclude that a risk factor should not be included in the model, while it is actually needed for explaining the lower or/and the upper conditional quantiles of returns. In fact, this is the case in the empirical application in the present paper. Especially in cases where the error distribution deviates from normality, i.e. when the distribution is characterised by skewness, has outliers or fat tails, or in general if there is some uncertainty about the shape of the distribution generating the sample, the standard conditional mean regression method may not be adequate, and the quantile regression approach provides more robust and more efficient estimates/results.

A large amount of theoretical and practical work has been done in the area of quantile regression. Since the seminal paper of Koenker and Bassett (1978), who first proposed a class of linear regression models for conditional quantiles, several papers have been written which suggest new estimation techniques and consider applications of and extensions to the original models (for details, see the review papers of Buchinsky, 1998; Yu et al., 2003). Quantile regression models can be estimated using either a non-parametric approach (Koenker and Bassett, 1978), where parameter estimates can be obtained as the solution to a simple optimisation problem, or a parametric approach which arises if some specific error distribution is used. The most common choice in the latter approach is the asymmetric Laplace error distribution. The asymmetric Laplace density yields a likelihood function that can be easily used for inference within a classical or a Bayesian setting (Yu and Moyeed, 2001). Moreover, it is flexible and therefore can be used for model selection based on the likelihood (see Section 3.2). Recently, a general family of distributions, the tick-exponential family which includes the asymmetric Laplace as its special case, has been proposed for quantile regression models and a quasi-maximum likelihood approach to inference has been developed (see Komunjer, 2005). Several methods for estimating standard errors for the model parameters have been also proposed in the literature (for a comparative study, see Buchinsky, 1995a).

The quantile regression approach has been widely used in many areas of applied economics and econometrics such as the investigation of wage structure (Buchinsky, 1994, 1995b), earnings mobility (Trede, 1998; Eide and Showalter, 1999) and educational quality issues (Eide and Showalter, 1998; Knight et al., 2000; Levin, 2001). There is also growing interest in employing quantile regression methods in the financial literature. Applications in this field include work on Value at Risk (Taylor, 1999; Chernozhukov and Umantsev, 2001; Engle and Manganelli, 2004), option pricing (Morillo, 2000) and the analysis of the cross section of stock market returns (Barnes and Hughes, 2002). Also, Bassett and Chen (2001) proposed the quantile regression method as an appropriate way for the characterization of mutual fund investment styles.

In our analysis it is also of particular interest to identify the risk factors associated with hedge fund investing. To this end, we consider various model selection approaches to investigate which subset of factors is the most relevant to describe each quantile of returns. To our knowledge there is no variable selection strategy available for quantile regression models. We consider a parametric approach to inference for quantile regression models, based on the asymmetric Laplace density, which enables us to use several likelihood based approaches to model selection. Specifically, we can use AIC and BIC to select the best model in the classical sense. Furthermore, we consider the Bayesian approach to model selection which is based on calculating the posterior model probabilities of all possible models. Then, inferences based on a single 'best' model can be drawn conditional on the most probable model (MP hereafter). Moreover, model uncertainty can be taken into account by using Bayesian model averaging (BMA hereafter). That is inferences can be based on all possible models (or a subset of most probable models) weighted by their respective posterior model probabilities.

The proposed quantile regression approach can lead to an alternative measure of performance for hedge fund strategies, similar in nature to Jensen's alpha. That is we can use the estimated alphas of the best (according to some model selection method) median regression multi-factor models, or even use the alphas from several regression quantiles, in order to assess the performance of different investment strategies. In the present paper we investigate whether the above approach has economic significance by constructing style allocation portfolios based on the ranking of the estimated alphas. This investigation should be of interest to investors who are considering investing in indices of hedge funds. We construct equally weighted portfolios based on the best performing strategies and evaluate their relative performance in an out-of-sample fashion.

Several important results emerge from our analysis. First, we have found strong evidence for model uncertainty in all regression models considered for hedge fund strategies, which suggests that model averaging may be a more appropriate inferential approach than model selection. Furthermore, we have found differences in the most relevant subsets of factors both between mean and median regression and among different regression quantiles. The former result indicates that the distributions of hedge fund returns are characterised by skewness rather than being symmetric. In terms of the different regression quantiles, a larger number of factors are generally needed to explain the lower and upper quantiles than the central ones. Finally, our results have shown that using conditional quantile regression improves our ability to construct style portfolios. This is reflected in the higher values of the Sharpe ratio and Sortino ratio of the portfolios constructed using the median/quantile regression approach relative to the mean regression based portfolios.

The contributions of this work are several. First, we propose modelling the entire distribution of hedge fund returns using the quantile regression approach. Second, we propose a Bayesian approach to model selection for conditional quantile regression models. To our knowledge there is no model selection strategy available for this class of models. In the context of hedge funds, little is known about the factors that affect the conditional distribution of hedge fund returns. This fact creates considerable model uncertainty which should be accounted for. Our approach accounts for model uncertainty by calculating posterior model probabilities for all possible models for a set of quantile regressions and producing inferences based on model averaging. Third, we carry out an extensive empirical analysis to explore several aspects of hedge fund returns by comparing different model selection strategies and also by detecting the differences between the quantile regression and the conditional mean approach. The results of our analysis provide useful insights to finance researchers and practitioners. Finally, we show that the quantile regression approach and the Bayesian methods provide a more powerful framework for constructing style portfolios.

The remainder of the paper is organised as follows. Section 2 describes briefly the conditional mean regression model and introduces the quantile regression approach. Section 3 considers inference conditional on a specific model and presents Bayesian inference under model uncertainty. Section 4 describes the data and presents the empirical application, while Section 5 summarizes and concludes.

2. The quantile regression model

In this section we present the quantile regression models we use to analyse hedge fund returns. In our analysis it is of great interest to identify the risk factors which are most relevant to explain the distribution of hedge funds. In particular, we examine how the relevant subset of factors or/and the risk exposures vary across different quantiles of returns. Before deriving the conditional quantile regression model we briefly describe the standard conditional mean regression model which is used as a benchmark in our analysis.

Suppose that we observe T consecutive realisations of hedge fund returns, $\mathbf{r}=(r_1,\ldots,r_T)'$, and that we use k risk factors, $\mathbf{f}_i=(f_{i1},\ldots,f_{iT})'$, $i=1,\ldots,k$, to explain this series of returns. The risk factors are assumed to be observed. The standard linear regression model has the form

$$r_t = \alpha + \sum_{i=1}^{k} \beta_i f_{it} + \varepsilon_t, \tag{1}$$

where r_t is the hedge fund excess return at time t, f_{it}, $i=1,\ldots,k$, is the return of factor i at time t, α is the intercept and β_i is the loading of risk factor i. The usual assumption for the errors ε_t is that they are independent and identically distributed with mean equal to 0. Model (1) suggests that the conditional mean of the random variable R_t given a set of k factors $f_{1t},\ldots,f_{kt}$ is equal to $E(R_t|f_{1t},\ldots,f_{kt}) = \alpha + \sum_{i=1}^{k} \beta_i f_{it}$. That is the model provides a decomposition of the hedge fund's expected returns into a part that can be replicated by the risk factors and the residual (alpha) which is attributed to the hedge fund manager's skill.

The above conditional mean regression method models only the expectation and not the whole conditional distribution of returns. That is, this model assumes that the risk factors' effects are constant across different return levels, although it is reasonable to believe that the factors' effects differ across quantiles of returns, especially between the median and the tails of their distribution. For this reason, quantile regression methods have been introduced in the literature (see, for example, the seminal paper of Koenker and Bassett, 1978) to describe the relationship between asset returns and risk factors via modelling a set of conditional quantiles. The quantile regression approach provides a natural generalisation of the standard conditional mean method and produces more robust inferences in some cases of non-Gaussian, especially skew, error distributions, or in the presence of outliers.

Therefore, in this paper we adopt quantile regression models for the analysis of hedge funds. Consider the following pth quantile linear regression model

$$r_t = \alpha^{(p)} + \sum_{i=1}^{k} \beta_i^{(p)} f_{it} + \varepsilon_t, \tag{2}$$

where r_t is the excess return of a hedge fund investment at time t, f_{it} is the return of factor i at time t, $\alpha^{(p)}$ is the intercept, $\beta_i^{(p)}$ is the loading of risk factor i and the errors ε_t are assumed to be independent from an error distribution $g_p(\varepsilon)$ with pth quantile equal to 0, i.e. $\int_{-\infty}^{0} g_p(\varepsilon)d\varepsilon = p$. The error distribution can be left unspecified.

Model (2) suggests that the pth conditional quantile of the random variable R_t given $f_{1t},\ldots,f_{kt}$ is $Q_{Rt}(p|f_{1t},\ldots,f_{kt}) = \alpha^{(p)} + \sum_{i=1}^{k} \beta_i^{(p)} f_{it}$, where the intercept and the regression coefficients depend on p. Therefore, model (2) describes the distributional dependence of the returns on the risk factors as p varies within the range (0,1). The factor loading $\beta_i^{(p)}$ shows how the ith risk factor affects the returns at the level of the pth quantile. The $\beta_i^{(p)}$'s are likely to be different for different p's, therefore this extra information can be used to analyse not only the median return but also its entire distribution. When $\alpha^{(p)}$, $\beta_i^{(p)}$, $i=1,\ldots, k$, vary with p, the model specifies a form of heteroscedasticity in which the conditional return distribution depends on the associated risk factors. Clearly, quantile regression reveals a larger amount of information about returns than conditional mean regression, especially if the error distribution is not symmetric (since asymmetric distributions result in differences between the mean and the median).

Suppose now that there are K available risk factors. A topic of growing interest in the hedge fund literature is that of selecting the subset of most relevant risk factors to be included in a regression model. This variable selection problem can be considered as a model selection problem, where each model corresponds to a different subset of the factors f_{kt}, $k=1,\ldots, K$. It is convenient to indicate each of these 2^K possible choices of subsets by the vector $\gamma=(\gamma_1,\ldots, \gamma_K)'$, where $\gamma_i=1$ if the ith factor is included and $\gamma_i=0$ if not. Thus, the vector γ plays the role of a model identifier.

3. Inference on quantile regression models

We consider inference about conditional quantile regression models given a set of K risk factors. First we conduct inference conditional on a specific model, i.e. subset of risk factors, and then we consider the problem of model comparison and inference in the presence of model uncertainty using model averaging.

3.1. Inference for a given model

In this section we consider the problem of estimating a vector of unknown regression parameters from a sample of independent observations $\{r_t, f_{1t},\ldots, f_{kt}\}_{t=1}^T$, that is given a subset of k risk factors. The conditional mean regression model can be estimated using the ordinary least squares method without making any particular assumptions for the error distribution. Least squares estimation is based on the fact that the expectation of a random variable R with distribution function F arises as the point estimate of R corresponding to the quadratic loss function $\rho(u)=u^2$, that is as the value of $\bar{r}$ which minimises the expected loss

$$E\rho(R-\bar{r}) = \int\rho(r-\bar{r})dF(r).$$

Therefore, the parameters $\boldsymbol{\theta}=(\alpha, \beta_1,\ldots, \beta_k)'$ in the conditional mean regression model (1) can be estimated by minimising the sample estimate of the quadratic expected loss $T^{-1}\sum_{t=1}^T \rho\left(r_t-a-\sum_{i=1}^k \beta_i f_{it}\right)$, with respect to $\boldsymbol{\theta}$, or equivalently by minimising the sum of squares

$$\sum_{t=1}^T \left(r_t-a-\sum_{i=1}^k \beta_i f_{it}\right)^2.$$

A parametric approach to inference on the regression parameters can be followed if the functional form of the error distribution is specified. Most commonly the errors are assumed to follow the normal distribution with zero mean and variance σ^2. Then, inference on the model parameters, including quantifying the uncertainty associated with them, can be based on the arising likelihood function using either classical or Bayesian methods. Note that the maximum likelihood estimates (MLEs) of the regression parameters in this case are identical to the ordinary least squares estimates.

Similarly to the expectation of the random variable R, its pth quantile arises as the solution to a decision theoretic problem; that of obtaining the point estimate of R corresponding to the asymmetric linear loss function, usually referred to as the check function,

$$\rho_p(u) = u(p-I(u<0)) = \frac{1}{2}[|u| + (2p-1)u]. \tag{3}$$

That is minimisation of the expected loss

$$E\rho_p\left(R-\bar{r}^{(p)}\right) = \int\rho_p\left(r-\bar{r}^{(p)}\right)dF(r),$$

with respect to $\bar{r}^{(p)}$ leads to the pth quantile. In the symmetric case of the absolute loss function ($p=1/2$) we obtain the median. If a sample, $r_1,\ldots, r_T$, is drawn from F, an estimate of the pth quantile of R is obtained by minimising the sample estimate of the expected loss, i.e. the function $T^{-1}\sum_{t=1}^T \rho_p\left(r_t-\bar{r}^{(p)}\right)$. The above idea can be used to estimate the parameters of the linear quantile regression model (2). In this model a conditional quantile function is specified as $Q_R(p|f_1,\ldots,f_k) = \alpha^{(p)} + \sum_{i=1}^k \beta_i^{(p)} f_i$ and the parameters $\beta^{(p)}=(\alpha^{(p)}, \beta_1^{(p)},\ldots, \beta_k^{(p)})$ need to be estimated. This can be done by minimising the sum

$$\sum_{t=1}^T \rho_p\left(r_t-\alpha^{(p)}-\sum_{i=1}^k \beta_i^{(p)} f_{it}\right), \tag{4}$$

where the check function $\rho_p(u)$ has been given in Eq. (3).

A parametric approach to inference on the quantile regression parameters arises if the error distribution $g_p(\varepsilon)$ is specified. The error distribution that has been widely used for parametric inference in the quantile regression literature is the asymmetric Laplace distribution (for details, see Yu and Moyeed, 2001; Yu and Zhang, 2005) with propability density function

$$g_p(\varepsilon) = \frac{p(1-p)}{\sigma^{(p)}}\exp\left[-\frac{|\varepsilon| + (2p-1)\varepsilon}{2\sigma^{(p)}}\right], 0 < p < 1, \sigma > 0, \tag{5}$$

For $p = \frac{1}{2}$, corresponding to the median regression, Eq. (5) becomes the symmetric Laplace density. A likelihood function can be formed by combining T independent asymmetric Laplace densities of the form of Eq. (5), i.e.

$$L\left(\boldsymbol{r}|\beta^{(p)}, \sigma^{(p)}, \boldsymbol{f}_i, \dots, \boldsymbol{f}_k\right) = \left(\frac{p(1-p)}{\sigma^{(p)}}\right)^T \exp\left\{-\frac{1}{\sigma^{(p)}}\sum_{t=1}^{T}\rho_p\left(\boldsymbol{r}_t - \alpha^{(p)} - \sum_{i=1}^{k}\beta_i^{(p)}\boldsymbol{f}_{it}\right)\right\}, 0 < p < 1, \sigma^{(p)} > 0. \tag{6}$$

Then Eq. (6) can be used for likelihood based inference for the parameters $\beta^{(p)}$, $\sigma^{(p)}$, for example for maximum likelihood estimation. The maximisation of this likelihood function with respect to $\beta^{(p)}$ is equivalent to minimising Eq. (4), therefore, the MLEs of $\beta^{(p)}$ are identical for any value of $\sigma^{(p)}$. Note that the likelihood in Eq. (6) does not necessarily correspond to the distribution of r. However, it has the advantage that the respective MLEs are identical to the estimates obtained by minimising Eq. (4) and, therefore, inherit their assymptotic properties (for details on these properties see Koenker, 2005). In this paper we follow the parametric approach to inference and adopt the asymmetric Laplace error distribution given in Eq. (5).

In this work we are not only interested in inference on the quantile regression coefficients but also in model comparison. The parametric approach to inference enables us to compare a number of different quantile regression models, corresponding to different subsets of factors, using criteria based on the likelihood function (see Section 3.2). The asymmetric Laplace density (5) apart from the mean has an extra parameter, σ, which makes it quite flexible and therefore appropriate for model selection based on the likelihood. To our knowledge, apart from the asymmetric Laplace density, the only other distributional assumption that has been proposed in the literature for conditional quantile models is the tick-exponential distribution family of Komunjer (2005). This family includes the asymmetric Laplace density as a special case. However, Komunjer (2005) gives only two example members of this family; the asymmetric Laplace with fixed $\sigma^{(p)}$ and one more density for which one of its parameters also has to be fixed. That is, the quasi-maximum likelihood estimation approach of that paper is an alternative approach for estimating the quantile regression coefficients, but since the other parameters of the distributions are not estimated it cannot be used for model comparison based on the likelihood.

The maximum likelihood estimates of the parameters $\beta^{(p)}$ and $\sigma^{(p)}$ of the quantile regression model are, respectively, given by

$$\widehat{\beta}^{(p)} = \arg\min_{\theta}\left[\sum_{t=1}^{T}\rho_p\left(r_t - \alpha^{(p)} - \sum_{i=1}^{k}\beta_i^{(p)}f_{it}\right)\right], \quad \text{and} \quad \widehat{\sigma}^{(p)} = \frac{1}{T}\sum_{t=1}^{T}\rho_p\left(r_t - \alpha^{(p)} - \sum_{i=1}^{k}\beta_i^{(p)}f_{it}\right).$$

In our model comparison exercise, it is useful to reparameterise $\sigma^{(p)}$ as $\phi^{(p)} = \log\sigma^{(p)}$. Since the MLEs are invariant in parameter transformations we can easily obtain the MLE of $\phi^{(p)}$ as $\widehat{\phi}^{(p)} = \log\widehat{\sigma}^{(p)}$. Let $\boldsymbol{\theta}^{(p)} = (\beta^{(p)}, \phi^{(p)})$ denote the parameter vector of interest. To quantify the uncertainty associated with the MLEs we need an estimate of their asymptotic standard errors. In order to estimate the covariance matrix of the parameter estimates we employ the Design Matrix Bootstrap method (DMB) initially introduced by Efron (1979, 1982) and applied to the quantile regression setting by Buchinsky (1994, 1995a). In particular, we use the vector of estimates obtained from the original sample as the pivotal value required to estimate the covariance matrix, which results to the DMBE estimator. Monte Carlo simulations have shown that the DMBE method provides a valid and consistent estimator of the asymptotic covariance matrix of the quantile regression estimates under general conditions (see Buchinsky, 1994, 1995a). Furthermore, in a comparative study among various estimators of the covariance matrix Buchinsky (1995a) has shown that the DMBE method is the most accurate even for relatively small sample sizes. Although being the more accurate method, the DMBE is computationally demanding. Therefore, if a very large number of models need to be estimated, alternative numerical methods of estimating the asymptotic covariance matrix of the MLEs can be employed.

3.2. Model comparison

Here we consider the problem of model comparison which deals with the uncertainty about the subset of factors that should be included in each quantile regression model. Under the parametric approach to inference, criteria based on the likelihood function can be used as tools for model comparison. For example, after obtaining the MLEs of a model's parameters, Akaike's information criterion and Schwarz's Bayesian information criterion can be computed as simple functions of the likelihood evaluated at the MLE's. Specifically, for a model γ with m_γ parameters and maximum log-likelihood l_γ, AIC is defined as $\text{AIC}_\gamma = -2l_\gamma + 2m_\gamma$ and BIC as $\text{BIC}_\gamma = -2l_\gamma + m_\gamma \log T$. BIC is more conservative than AIC and favours more parsimonious models.

Apart from the above classical model selection criteria we also consider a Bayesian approach to model comparison. This is a probabilistic approach to inference which has several advantages and interesting features. Under the Bayesian approach, the unknown quantities in the model are treated as random variables. In our setting, the random quantity is the subset of risk factors which should be included in a quantile regression model. This quantity can be represented by the model identifier γ. The

Bayesian model comparison approach is based on calculating the posterior distribution of γ, which corresponds to obtaining the posterior model probabilities of all competing models and ranking the models according to these probabilities. Then, inferences can be based on a single model, considered to be the true one, or model uncertainty can be taken into account and inferences can be averaged over all competing models, or a subset of most probable models, weighted by their (normalised) posterior model probabilities. The latter approach is usually referred to as Bayesian model averaging. Note that, BIC can be regarded as an approximation to the log Bayes factor and, therefore, it is asymptotically equivalent to Bayesian model comparison under weak conditions.

To proceed with the Bayesian model comparison approach, which treats the model identifier as a parameter, γ has to be assigned a prior distribution. A natural prior distribution for γ, commonly used in variable selection problems, is based on specifying a prior probability λ_k for the inclusion of a factor in a model with k factors. Hence, the prior distribution we adopt for γ is of the form

$$\pi(\gamma) = \prod_{k=1}^{K} \lambda_k^{\gamma_k}(1-\lambda_k)^{1-\gamma_k},$$

where $\lambda_k \in [0,1]$ indicating that each risk factor f_k, $k=1,\ldots,K$ enters the model independently of the others with probability $\pi(\gamma_k=1)=\lambda_k$. In this paper we assume that we do not have any prior information for the inclusion of a specific factor in the model, thus we take $\lambda_1=\ldots=\lambda_K=0.5$. This implies that the investor is neutral about the factors that will enter in the model. Therefore, each model specification, i.e. each vector γ, is assigned equal prior probability $\pi(\gamma)=1/2^K$.

The posterior probability of the pth quantile regression model with risk factor identifier γ is given by

$$\Pr(\gamma|\boldsymbol{r},p) = \frac{\pi(\gamma)\Pr(\boldsymbol{r}|\gamma,p)}{\sum_{\delta\in\Gamma}\pi(\delta)\Pr(\boldsymbol{r}|\delta,p)}, \tag{7}$$

for $\gamma \in \boldsymbol{\Gamma}$, with $\boldsymbol{\Gamma}$ denoting the set of all risk factor identifiers, and where

$$\Pr(\boldsymbol{r}|\gamma,p) = \int L_\gamma\left(\boldsymbol{r}|\theta_\gamma^{(p)},\boldsymbol{f}_1,\ldots,\boldsymbol{f}_k\right)\pi\left(\theta_\gamma^{(p)}\right)d\theta_\gamma^{(p)} \tag{8}$$

is the marginal probability of the data r under model γ, obtained by integrating the asymmetric Laplace likelihood (6) for the pth quantile with respect to some prior distribution $\pi(\boldsymbol{\theta}_\gamma^{(p)})$ on the model parameters. In this paper we adopt the unit information prior proposed by Kass and Wasserman (1995) for the vector $\boldsymbol{\theta}_\gamma^{(p)}$. This is a multivariate Normal prior distribution, with covariance matrix equal to $T\widehat{\Sigma}_\gamma^{(p)}$, where T indicates the sample size and $\widehat{\Sigma}_\gamma^{(p)}$ is some estimate of the covariance matrix of the maximum likelihood estimates of the model parameters; here we use the DMBE with 150 boostrap samples.

Two difficulties arise in the calculation of the posterior model probabilities. First, the required high dimensional integrations in Eq. (8) cannot be computed analytically and, therefore, some numerical technique or approximation is needed. Second, the number of competing models, 2^K, increases dramatically with K and, hence, the summation in the denominator of Eq. (7) can be exhausting.

To deal with the first problem we use a variant of Laplace approximation to the marginal probability in Eq. (8), based on maximum likelihood

$$\Pr\left(\widehat{\boldsymbol{r}|\gamma},p\right) \simeq (2\pi)^{\frac{d_\gamma}{2}}\left|\widehat{\Sigma}_\gamma^{(p)}\right|^{\frac{1}{2}}L_\gamma\left(\boldsymbol{r}|\widehat{\theta}^{(p)},\boldsymbol{f}_1,\ldots,\boldsymbol{f}_k\right)\pi\left(\widehat{\theta}^{(p)}\right),$$

where d_γ is the dimension of the parameter vector $\boldsymbol{\theta}_\gamma^{(p)}$ of model γ, $\widehat{\boldsymbol{\theta}}^{(p)}$ is the vector of the MLEs, $\widehat{\Sigma}_\gamma^{(p)}$ is the DMBE of the covariance matrix of the MLEs, $L_\gamma(\mathbf{r}|\widehat{\boldsymbol{\theta}}^{(p)},\mathbf{f}_1,\ldots,\mathbf{f}_k$ and $\pi(\widehat{\boldsymbol{\theta}}^{(p)})$ are the likelihood and the prior, respectively, evaluated at the MLEs. This method of approximation for the marginal likelihood has been used by Vrontos et al. (2003b), Dellaportas and Vrontos (2007) in the analysis of multivariate time varying volatility models, and by Vrontos et al. (2008) for the analysis of hedge fund returns using standard linear regression models.

To decide which is the 'best' model in the set of all competing models, on which inferences should be based, a decision theoretic approach can be followed. If we assume that the set of competing models includes the true one, the standard loss function to be adopted is the zero-one loss which assigns zero loss to selecting the true model and loss one to any other choice. The expected posterior loss in this case is minimised if the most probable model is selected (see, for example, Bernardo and Smith, 1994; Denison et al., 2002). Then, inferences are based on the most probable model. However, it is very likely not to have included the true model in the set of competing models. In particular, when constructing risk factor models for hedge funds we cannot expect to have considered all the relevant economic risk factors since they are virtually unknown. Due to the dynamic trading strategies followed by hedge fund managers and the lack of transparency in managerial activities, identification of the risk factors that affect hedge fund returns is extremely difficult. Therefore, the most appropriate approach is to consider none of the competing models to be true and try to produce inferences as close to the truth as possible. This gives rise to the use of Bayesian model averaging, an approach that produces inferences which are robust to model misspecification. BMA, unlike model selection, takes into account the uncertainty associated with the set of factors that should be included in the model, which is particularly important when different competing models score equally well (Kass and Raftery, 1995) or in cases where there are short histories of data as in the case of hedge funds (Vrontos et al., 2008).

The BMA posterior distribution of a quantity of interest, say Δ, given data r is given by

$$f(\Delta|\mathbf{r},p) = \sum_{\gamma\in\Gamma} f(\Delta|\mathbf{r},\gamma,p)\Pr(\gamma|\mathbf{r},p), \tag{9}$$

which is an average of the posterior distributions under each model weighted by the corresponding posterior model probabilities (see, for example, Raftery et al., 1996; Raftery et al., 1997; Vrontos et al., 2003a, for details). In Eq. (9) the predictive distribution of Δ given a particular pth quantile regression model with identifier γ can be found by integrating out the model parameters $\boldsymbol{\theta}_\gamma^{(p)}$, i.e.

$$f(\Delta|\mathbf{r},\gamma,p) = \int f\left(\Delta|\mathbf{r},\gamma,p,\boldsymbol{\theta}_\gamma^{(p)}\right)\pi\left(\boldsymbol{\theta}_\gamma^{(p)}\right)d\boldsymbol{\theta}_\gamma^{(p)}$$

or by using the maximum likelihood approximation:

$$f(\Delta|\mathbf{r},\gamma,p) \simeq f\left(\Delta|\mathbf{r},\gamma,p,\widehat{\boldsymbol{\theta}}^{(p)}\right).$$

Note that the optimal point estimate of a parameter of interest under the approach that none of the competing models is true is the BMA point estimate (Bernardo and Smith, 1994; Denison et al., 2002). For example, suppose that we wish to estimate $\alpha^{(p)}$. Under a decision theoretic approach we should minimise the difference between a model based estimate, $\widehat{\alpha}^{(p)}$, and the true value of $\widehat{\alpha}^{(p)}$. A common choice of the loss function considered in this setting is the squared error loss $(\widehat{\alpha}^{(p)}-\alpha^{(p)})^2$. The point estimate obtained by minimising the expected loss is given by

$$\widehat{\alpha}^{(p)} = \int \alpha^{(p)} f(\alpha^{(p)}|\mathbf{r},p)d\alpha^{(p)} = E(\alpha^{(p)}|\mathbf{r},p) = \sum_{\gamma\in\Gamma} E(\alpha^{(p)}|\mathbf{r},\gamma,p)\Pr(\gamma|\mathbf{r},p),$$

where $f(\alpha^{(p)}|\mathbf{r}, p)$ is the BMA posterior distribution of $\alpha^{(p)}$ and $E(.|.)$ denotes posterior expectation. In general, the BMA point estimator of $\alpha^{(p)}$ is given by a weighted average of the estimators $\widehat{\alpha}_\gamma^{(p)}$ over all competing models, i.e.

$$\widehat{\alpha}^{(p)} = \sum_{\gamma\in\Gamma} \widehat{\alpha}_\gamma^{(p)}\Pr(\gamma|\mathbf{r},p).$$

For example, $\widehat{\alpha}_\gamma^{(p)}$ can be the MLE of alpha under the model with identifier γ.

When the number of candidate risk factors K is large, the number of possible models 2^K is enormous and analytic evaluation of Eqs. (7) and (9) might be computationally infeasible. In those cases we may use a stochastic search algorithm which jumps between models of the same or different dimensionality, i.e. number of explanatory factors, to generate a sample from the posterior distribution of γ. That is we may use a Markov chain Monte Carlo algorithm to simulate a Markov chain consisting of models/subsets $\gamma^{(1)}, \gamma^{(2)}, \gamma^{(3)},\ldots$, which, under weak conditions (see, for details Smith and Roberts, 1993; Tierney, 1994), converge to the limiting distribution $\Pr(\gamma|\mathbf{r}, p)$. The stochastic search algorithm for simulating from the posterior distribution of γ has been described in detail by Vrontos et al. (2008).

4. Application to hedge funds

4.1. The data

We illustrate the proposed quantile regression approach using hedge fund indices data from Hedge Fund Research (HFR). The HFR indices are equally weighted average returns of hedge funds and are computed on a monthly basis. In our analysis, we use directional strategies that bet on the direction of the markets, as well as non-directional strategies whose bets are related to diversified arbitrage opportunities rather than to the movement of the markets. In particular, we consider thirteen HFR single strategy indices: Convertible Arbitrage (CA), Distressed Securities (DS), Event-Driven (ED), Equity Hedge (EH), Equity Market Neutral (EMN), Fixed Income Arbitrage (FIA), Macro (M), Merger Arbitrage (MA), Market Timing (MT), Relative Value Arbitrage (RVA), Emerging Markets Total (EMT), Equity Non-Hedge (ENH) and Short Selling (SS). Our study of these hedge funds uses net-of-fee monthly excess returns (in excess of the three month US Treasury Bill) from April 1990 to December 2005.[1] This period includes a number of crises and market events which affected hedge funds returns, and caused large variability in the return series.

We model the hedge fund returns by using different information variables — pricing factors in the lines of Agarwal and Naik (2004). We consider the factors used by Vrontos et al. (2008). These include: returns on the Russel 3000 equity index (RUS), the Morgan Stanley Capital International (MSCI) world excluding the USA index (MXUS), the MSCI emerging markets index (MEM), the Salomon Brothers world government and corporate bond index (SBGC), the Salomon Brothers world government bond index (SBWG), the Lehman high yield index (LHY), the Goldman Sachs commodity index (GSCI), the Federal Reserve Bank competitiveness weighted dollar-index (FRBI); Fama and French's (1993) ‘size’ (SMB) and ‘book-to-market’ (HML) as well as Carhart's (1997) ‘momentum’ factors (MOM); the difference between the yield on the BAA-rated corporate bonds and the 10-year Treasury bonds (DEFSPR); and the change in equity implied volatility index VIX. This dataset also covers the period April 1990 to December 2005. Certainly, we cannot claim that this set of factors is complete, or that it includes the true subset of relevant factors

[1] Details can be found in Hedge Fund Research, www.hedgefundresearch.com.

for each analysed index and, for this reason, we advocate the use of model averaging instead of using a single 'best' model for inference (as discussed in Section 3.2).

4.2. Empirical analysis

In the following analysis we develop quantile regression risk factor models with the aim to model the entire distribution of single strategy hedge fund returns and we use a bayesian model averaging approach in order to identify the relevant risk factors that affect each conditional quantile of returns. We expect that the quantile regression approach will be more valuable for understanding the underlying structure of the different single strategy hedge fund indices and more useful at explaining the distributional dependence of hedge fund returns on risk factors compared to the standard conditional mean regression method. Next we attempt to investigate whether the proposed quantile regression approach has economic significance by performing an evaluation exercise. The objectives of this exercise are the construction of optimal style allocation portfolios under two different modeling approaches (quantile regression and mean regression) and under the different variable selection techniques of Section 3.2 (various approaches to model selection, as well as model averaging), and the comparison of the constructed portfolios in terms of their performance.

In the following analysis we split our sample into an estimation period, from April 1990 to December 2003, which is used to estimate our models, and an evaluation period, from January 2004 to December 2005, which is used for portfolio construction and performance evaluation.

4.2.1. Quantile regression analysis of hedge fund strategies

A comprehensive discussion on the characteristics of different hedge fund strategies follows. Table 1 presents summary statistics for the series of hedge fund returns calculated over the estimation period. We observe that the hedge fund strategies are very heterogeneous: there are some strategies with relatively high average returns and high volatilities, such as Equity Hedge, Equity Non-Hedge and Macro, while Equity Market Neural, Convertible Arbitrage and Merger Arbitrage have relatively low average returns and standard deviations. Almost all the hedge fund strategies (except for Short Selling) have positive returns. The standard deviations of hedge fund returns indicates major differences between strategies; in particular, the variability of returns of Short Selling, Emerging Markets and Equity Non-Hedge is higher compared to other hedge funds such as Equity Market Neutral and Convertible Arbitrage.

Differences are also apparent in higher order moments. In particular, Merger Arbitrage, Fixed Income Arbitrage, Relative Value Arbitrage, Event Driven and Convertible Arbitrage have high negative skewness and high kurtosis, while Equity Hedge, Market Timing, Equity Market Neutral, Macro and Short Selling have low positive skewness and relatively small kurtosis. Eight out of the thirteen hedge fund strategies have negative skewness. This suggests that extreme negative price falls are more likely than extreme price increases for the respective hedge fund strategies. Also, seven hedge fund strategies have large kurtosis indicating fat tails. The above reported statistics show that hedge fund returns may exhibit a high degree of non-normality which can be attributed to the special nature of the investment strategies adopted. The deviation form normality is confirmed by the results on the Jarque-Bera normality tests, presented in the last column of Table 1; for ten hedge fund return series the normality hypothesis is rejected at 5% level of significance.

For analysing each of the thirteen HFR single strategy indices we use the fourteen available risk factors discussed above. For each of the series of returns we have tested the hypothesis of no autocorrelation being present in the series. To those strategies for which the null was rejected we have added an extra pricing factor, namely the lagged net-of-fee excess return r_{t-1}, to the set of risk factors used. In our Bayesian model comparison setting, identifying the most relevant subset of risk factors involves comparing all possible models which include different subsets of the available factors. This results in the comparison of $2^{14}=16,384$ models in the case of no autocorrelation and $2^{15}=32,768$ models in the presence of autocorrelation. For all possible models we need to calculate their respective marginal likelihoods. These, weighted by the prior model probabilities and normalised, result to the corresponding posterior model probabilities. The primary focus in our analysis is to show that the relationship between hedge fund returns and the available risk factors changes across the distribution of conditional returns and therefore the quantile regression approach provides a better way to understand this relationship compared to the standard conditional mean regression method.

We have applied the approach described in Section 3.2 to compute the marginal likelihoods and hence the posterior probabilities of all competing models for different regression quantiles (10th, 25th, 50th, 75th, 90th) and for the conditional mean regression. We have found that there is a considerable amount of model uncertainty in the regression models considered. The posterior probabilities of the most probable models in most of the cases are quite small (below 0.25), while the average total posterior probability of the ten most probable models is 0.38. In order to take account of this uncertainty some Bayesian model averaging approach should be used. Hence, for each regression model, instead of just taking the most probable model specification as the subset of most relevant risk factors, we have used BMA for inference on the model identifier γ. To this end we have calculated the posterior probability of including each factor in each regression model, taking into account all possible model specifications. As pointed out in Section 3.2, BMA provides optimal inferences in the presence of model uncertainty if there is no reason to believe that the true model specification is included in the set of specifications considered.

The posterior probability of inclusion of each factor is computed as the sum of the posterior probabilities of all models including that particular factor. This quantity forms a statistic that can be used to investigate the robustness of risk factors in regression models (see Avramov, 2002). In Table 2 we report the most relevant risk factors for the analysed HFR single strategy hedge fund indices, for each regression model (i.e. the 10th, 25th, 50th, 75th and 90th regression quantiles and the conditional mean regression) according to BMA. Specifically, we report those factors that have posterior probability of inclusion greater than 0.5.

Table 1
Summary statistics for the HFR single strategy indices

Strategy	Mean	StD	Kurtosis	Skewness	25th Perc.	Median	75th Perc.	Jarque-Bera
CA	0.58	0.96	6.26	−1.33	0.16	0.76	1.19	117.5*
DS	0.84	1.83	8.13	−0.70	−0.14	0.81	1.76	187.7*
ED	0.85	1.87	8.05	−1.36	−0.01	1.04	1.90	218.8*
EH	1.08	2.59	4.33	0.12	−0.64	1.05	2.51	11.6*
EMN	0.41	0.90	3.37	0.08	−0.18	0.32	0.96	0.9
FIA	0.33	1.30	11.92	−1.69	−0.15	0.45	0.93	606.6*
M	1.01	2.50	3.32	0.24	−0.52	0.68	2.45	2.1
MA	0.54	1.10	14.32	−2.57	0.14	0.67	1.17	1033.2*
MT	0.70	1.97	2.45	0.10	−0.94	0.65	2.04	2.6
RVA	0.65	1.07	13.93	−0.99	0.13	0.65	1.17	821.8*
EMT	0.93	4.48	6.42	−0.82	−1.88	1.35	3.67	94.8*
ENH	1.04	4.23	3.51	−0.54	−1.60	1.72	3.57	9.4*
SS	−0.09	6.41	4.16	0.12	−4.03	−0.42	4.02	8.9*

This table presents summary statistics of monthly returns for thirteen HFR indices from April 1990 through to December 2003. The summary statistics include the mean, standard deviation (StD), kurtosis, skewness, median, 25th and 75th percentiles, and the Jarque-Bera test statistic for the normality assumption of the return series. *indicates rejection of the hypothesis of normality at 5% level of significance (the corresponding critical value is 5.992). CA is Convertible Arbitrage, DS is Distressed Securities, ED is Event Driven, EH is Equity Hedge, EMN is Equity Market Neutral, FIA is Fixed Income Arbitrage, M is Macro, MA is Merger Arbitrage, MT is Market Timing, RVA is Relative Value Arbitrage, EMT is Emerging Markets Total, ENH is Equity Non-Hedge, and SS is Short Selling.

Firstly, from Table 2 we may observe that there are differences between the median regression and the conditional mean regression. A larger number of factor's are generally used to explain the conditional median than those needed in the conditional mean regression (see, for example, EH, ED, EMT, M, and RVA), while some of the factors are usually different between the two models (see EH, EMN, MA, MT and EMT). The estimates of the model parameters (i.e. the alphas and the beta coefficients) are also quite different.[2] Similar conclusions are drawn if, instead of using BMA, we consider the best models obtained under different model selection methods, such as bayesian model comparison, AIC or BIC (results not shown for reasons of space). The above findings may be attributed to the fact that the median regression provides more robust and more efficient estimates/results, because the hedge fund return distribution deviates from normality, i.e. is characterised by skewness, or has fat tails.

Most importantly, from Table 2 it can be seen that there exist substantial differences among the most relevant subsets of factors obtained for different regression quantiles. In particular, a larger number of factors is needed in the tails of the distributions of returns, especially in the 10th and/or the 90th quantile; for example, RVA and ENH require a much larger number of factors for the 10th conditional quantile compared to the number of factors used in the other conditional quantiles and the conditional mean regression, MA and DS require a large number of factors for the 90th conditional quantile, while for EMN, EMT and SS a large number of factors are included in both extreme regression quantiles. That is there exist many cases where some factors are only needed to explain certain quantiles in the tails but not in the center of the distribution of conditional returns, and hence these factors are not included in the respective conditional mean regression model.

The above findings seem reasonable since different hedge fund strategies employ different trading tools each involving certain financial assets and, therefore, each strategy is expected to be on average explained by a number of strategy specific risk factors. However, in extreme economic conditions (for example, in financial distress, political instability or in periods of economic prosperity) the financial asset classes affect each other. Hence, in such cases hedge fund strategies, which are flexible with respect to the types of securities they hold and the types of positions they take, are likely to be affected by a larger variety of economic risk factors. This is reflected on the dependence of higher and/or lower hedge fund returns on a wide class of factors.

For example, consider the Merger Arbitrage strategy which includes funds that invest in the securities of companies involved in a merger or aquisition usually by buying the stocks of the target (the company being acquired) and going short on the stocks of the acquirer, but occasionally in reverse if the manager believes that the deal may fail. We have found that, on average, MA returns are explained by RUS, RUS(−1) and MEM, i.e. equity oriented factors. This is consistent with the findings of previous studies; see for example Mitchell and Pulvino (2001) and Meligkotsidou and Vrontos (2008). However, our quantile regression analysis has also revealed that the factor GSCI affects both the extreme quantiles of MA returns, while the factors MOM, LHY, DEFSPR, and FRBI are also used to explain the 90th quantile. That is commodity and interest rate oriented factors have been also used to explain the tails of the conditional distribution of MA returns. Indeed, changes in energy costs and in row material prices, as well as abrupt changes (rises or falls) in interest rates are likely to affect acquisition activity (such as leveraged buyouts) and hence the opportunities of investing in mergers or acquisitions; see Nicholas (2000), Hedges (2005) and Stefanini (2006). It is interesting to note that the GSCI factor, which is included only in the two extreme regression quantiles, has beta coefficient of opposite sign at opposite ends of the distribution of conditional returns; its BMA estimate is equal to 0.0257 in the 10th quantile and equal to −0.0183 in the 90th. Hence, an increase in GSCI (keeping the values of the other factors fixed) will lead to a decrease in both the profits and the losses of MA managers. This can be attributed to the fact that an increase in energy costs and row material prices results in economic distress and therefore in a reduction in the number of mergers that the average MA manager is involved in. On the other hand, GSCI factor's beta becomes zero around the center of the distribution of conditional returns and therefore the conditional mean regression does

[2] The parameter estimates of the different models arising from the different model selection techniques, for the analysed hedge fund strategies, are not reported for reasons of space.

Table 2
The most relevant risk factors for the quantile and mean regression models for the HFR single strategy indices

HFR	Quantiles	Models								HFR	Quantiles	Models								
CA	10th	1	3	4	5	10	11	13	15	DS	10th	4	8	10	13	14	15			
	25th	1	5	8	10	11	13	15			25th	4	5	6	8	10	14	15		
	50th	10	11	13	15						50th	4	5	8	10	12	14	15		
	75th	10	11	15							75th	4	5	8	10	12	14	15		
	90th	4	6	10	11	15					90th	2	3	4	5	8	9	12	14	15
	CM	4	8	10	11	15					CM	4	5	8	10	12	14	15		
ED	10th	1	3	4	5	8	9	15		EH	10th	1	2	3	4	5	6	10	13	
	25th	1	4	5	8	10	15				25th	1	2	4	5	6	7	13	14	
	50th	1	4	5	8	10	13	15			50th	1	2	5	7	13	14			
	75th	1	4	5	6	7	10	15			75th	1	5	6	8	13				
	90th	1	4	5	6	8	12				90th	1	5	6	8	13				
	CM	1	4	5	8	10	15				CM	1	5	6	13	14				
EMN	10th	1	4	5	6	7	8	10		FIA	10th	1	3	4	11	12	15			
	25th	1	4	5	7	8					25th	1	3	4	11	12	15			
	50th	5	7	8							50th	1	4	9	12	15				
	75th	1	7	8							75th	1	4	7	9	12	15			
	90th	1	2	7	11	13					90th	1	4	7	9	10	12			
	CM	1	5	7							CM	1	4	11	12	15				
M	10th	3	4	5	7	8	9			MA	10th	1	2	4	13					
	25th	4	5	6	8						25th	1	2	4	5					
	50th	4	5	7	8	15					50th	1	2	5						
	75th	4	7	8	9	15					75th	1	2	5	12					
	90th	4	7	8	9	15					90th	1	2	5	7	10	11	12	13	
	CM	4	7	8	15						CM	1	2	4						
MT	10th	1	2	3	5	7	12	14		RVA	10th	3	5	8	9	10	13			
	25th	1	5	7	12	14					25th	1	5	8						
	50th	1	4	5	6	7	14				50th	1	2	5	8					
	75th	1	3	5	6	8	9	11			75th	5	7	8						
	90th	1	3	6	14						90th	4	5	6	7	8				
	CM	1	3	5	6	14					CM	2	4	5						
EMT	10th	1	3	4	7	8	9	11	15	ENH	10th	1	2	4	5	6	9	10	11	13
	25th	4	7	8	15						25th	1	4	5	6	9	10	14		
	50th	2	4	7	8	15					50th	1	4	5	6	14				
	75th	2	4	6	8	10	15				75th	4	5	6	14					
	90th	2	3	4	6	8	10	13	15		90th	1	4	5	6	12	13	14		
	CM	4	7	8	15						CM	1	4	5	6	14				
SS	10th	1	5	6	11	14														
	25th	1	5	6	11															
	50th	1	5	6																
	75th	1	5	6																
	90th	1	5	6	9	11														
	CM	1	5	6																

This table presents the most relevant risk factors for the mean and the quantile regression models (10th, 25th, 50th, 75th and 90th quantiles) obtained using BMA for inference on the model identifier γ (for the estimation period from April 1990 to December 2003). The factors with posterior probability greater than 0.5 are reported. Numbers are associated with the corresponding risk factors; 1 is the Russel 3000 equity index excess return (RUS), 2 is the Russel 3000 equity index excess return lagged once [RUS(−1)], 3 is the Morgan Stanley Capital International world excluding USA index excess return (MXUS), 4 is the Morgan Stanley Capital International emerging markets index excess return (MEM), 5 is the Fama and French's (1993) 'size' (SMB), 6 is the Fama and French's (1993) 'book-to-market' (HML), 7 is Carhart's (1997) 'momentum' factor (MOM), 8 is the Salomon Brothers world government and corporate bond index excess return (SBGC), 9 is the Salomon Brothers world government bond index excess return (SBWG), 10 is the Lehman high yield index excess return (LHY), 11 is the difference between the yield on the BAA-rated corporate bonds and the 10-year Treasury bonds (DEFSPR), 12 is the Federal Reserve Bank competitiveness weighted dollar-index excess return (FRBI), 13 is the Goldman Sachs commodity index excess returns (GSCI), 14 is the change in S&P 500 implied volatility index (VIX), 15 is the respective HFR strategy excess return lagged once [HFR(−1)].

not reflect the economic significance of the factor. Clearly, the quantile regression approach prevents us from drawing incorrect inferences with respect to the factor's effect on the distribution of returns.

The Equity Market Neutral strategy includes funds that seek to generate returns by exploiting pricing inefficiencies between related equity securities by combining long and short positions with the aim to neutralise exposure to market risk. However, it is not easy for such portfolios to be beta neutral with respect either to the market or to various risk factors because reactions to large market movements or market events are unpredictable as one side of the portfolios will behave differently than the other. Indeed, in our analysis we have found that EMN is not market neutral, while other variables have been also found to explain the conditional distribution of EMN returns. Specifically, although EMN is neutral with respect to the Russel equity index in the median, it has significant exposure to RUS in all other conditional quantiles. On the other hand, it can be seen that on average EMN appears not to be market neutral. Moreover, we have found that various equity related factors (MEM, SMB, HML, MOM) explain the whole conditional distribution of EMN returns, while some other economic variables (i.e. bond and commodity related factors; SBGC, LHY, DEFSPR, GSCI) have been found to affect the tails of the distribution. Note that the conditional mean model apart from RUS only includes the SMB and MOM factors. These findings suggest that our conditional quantile approach provides some additional insight to the study of neutrality breadth (the number of sources of 'market' risk) and depth (the 'completeness' of the neutrality to market risks). Our conclusions confirm and extend the results of previous analyses since we are able to identify the relevant risk factors and the impact of each one of them not only on average but also at different conditional quantiles. Patton (forthcoming) has analysed market neutral individual hedge fund returns and found that one quarter to one third of these funds were not market neutral. Furthermore, Foerster (2006) has found EMN strategy returns on average not to be market neutral, while being affected by other economic variables, both equity and bond related. Finally, Billio et al. (2006) have concluded that EMN is not market neutral during market up and down times, while they have also found exposures to bond and commodity related factors.

By analysing Short Selling hedge fund strategy we have found that the strategy returns are on average (and in the median) exposed to the market index (RUS) and to the Fama and French's size (SMB) and book-to-market (HML) factors. Note that these factors appear in all the regression quantiles, with negative betas on RUS and SMB and positive betas on HML. These results are sensible because as SS managers go short to overvalued securities expecting that their price will decrease, SS exposures are likely to be of opposite sign to those of strategies taking long positions. Similar results have been found by Agarwal and Naik (2004) who used a conditional mean regression model. However, our analysis has revealed that some other risk factors also affect the tails of the conditional distribution of SS returns. In particular, we have found that the factors DEFSPR and VIX affect the lower quantile, while the factors SBWG and DEFSPR affect the higher quantile. Note that the beta coefficient of DEFSPR is of opposite sign at opposite ends of the conditional distribution of returns, with BMA estimate of beta equal to −5.3073 at the 10th quantile and equal to 9.1330 at the 90th. That is the quantile regression analysis suggests that an increase (decrease) of 1% in DEFSPR results in a 5.3073 percentage point decline (increase) in the lower returns of SS index and in a 9.1330 percentage point increase (decline) in its upper returns. Billio et al. (2006) have drawn similar conclusions about how SS strategy is affected by credit spreads at different regimes of the conditional distribution of returns using a regime switching model. Our analysis has shown that credit spreads affect only the extreme conditional quantiles of returns, while DEFSPR's beta is zero in the median and mean regression.

Now, the analysis of Emerging Market strategy shows that on average there is (positive) exposure to the emerging market index (MEM), the momentum factor (MOM), the Salomon brothers world government and corporate bond index (SBGC) and to the lagged values of the hedge fund's returns due to autocorrelation. An additional factor, the lagged value of Russel index, appears in the median regression model, while a larger number of factors, equity and bond oriented, are needed to explain the extreme quantiles of the distribution of conditional returns. The factors that have been found to explain the series of returns are reasonable since Emerging Market funds involve equity as well as fixed income investments worldwide. Additional information from equity and bond factors is required to explain the conditional returns in extreme situations.

Similar results regarding additional factors used to explain the tail(s) of the distribution of conditional returns have been found for most of the analysed hedge fund strategies. For example, the factors MXUS, SBWG, LHY and GSCI only appear in the 10th regression quantile of RVA, and similarly the factors RUS(−1), SBWG, LHY and DEFSPR appear in the lower quantiles of ENH, i.e. the 10th and occasionally the 25th regression quantiles. On the other hand, the factors RUS(−1), MXUS and SBWG only appear in the 90th regression quantile of DS. Obviously, in all these cases, looking just at the conditional mean regression model one would conclude that some factors are not significant, while in fact there is some interesting economic relationship of the upper or lower values of conditional returns with the factors which are ignored. Analogous results can be drawn by analysing the value weighted hedge fund indices of Credit Suisse/Tremont database.[3]

To conclude, our analysis has shown that in the conditional mean regression models fewer factors are included than those required to explain the entire conditional distribution of returns via regression quantiles. It is evident that the quantile regression technique provides considerably stronger insight to the distributional dependence of hedge fund returns on factors compared to the standard conditional mean regression method. Indeed, our findings show that the quantile regression approach is very useful for the analysis of hedge funds as it is able to see through what the conditional mean model misses. Let us note at this point that the probabilities of including the various factors in the conditional mean regression model are roughly averages of the probabilities of inclusion over the different regression quantiles. This observation is consistent with the fact that the mean regression only reflects the relationship of conditional returns with factors on average, while the quantile regression approach captures the change of this relationship across the entire distribution of conditional returns. The latter approach to modelling hedge fund returns is able to capture the special characteristics of the data, and provides a more sophisticated decomposition of the returns into the part that can be replicated by related risk factors and the residual which is attributed to the managers' skill across different states represented by certain conditional quantiles. This may have an impact to the ranking of hedge fund strategies based on their alphas and the construction of style portfolios. In fact, it turns out in our empirical portfolio exercise that the portfolios constructed based on median/quantile regression alphas have superior performance.

4.2.2. Hedge fund style portfolio construction

Hedge funds are known to have superior performance to traditional investment vehicles. This superior performance is generally attributed to the skills of hedge fund managers. The evaluation of the performance of different hedge fund strategies is usually based on some measure of the managers' skill. The most commonly used measure is Jensen's alpha, introduced by Jensen (1968), that is the intercept in the mean regression of the fund's excess return on the excess return of some market index (which can be thought of as a benchmark). The intuition behind using alpha as a measure of performance is that, if the average hedge fund return is significantly higher than what expected, given the average benchmark return and the regression beta, then the performance is

[3] Details can be found in Credit Suisse Tremont Hedge Index, www.hedgeindex.com.

considered to be superior. Obvious extensions arise if we consider the alpha of multiple regression models, i.e. regressions of the fund's excess returns on several economic risk factors, built within the Arbitrage Pricing Theory context.

Although Jensen's alpha, or in general the alpha of a multi-factor conditional mean regression model, is a well established measure of performance, there is some recent scepticism with regard to its suitability for hedge fund evaluation (see, for example, Amin and Kat, 2003). Due to the highly dynamic complex trading strategies followed by hedge fund managers, the series of hedge fund returns may exhibit some special characteristics. Usually, the series of returns are characterized by a high degree of kurtosis and skewness and, therefore, deviate substantially from normality. Hence, the standard conditional mean regression method and the corresponding regression alpha as a measure of performance, which both assume normality, seem to be inadequate for modelling and evaluating the performance of hedge funds, respectively. To properly evaluate the performance of non-normal returns the entire distribution of conditional returns has to be considered.

Here we propose the use of an alternative measure of performance, similar in nature to Jensen's alpha, which is based on quantile regression. The quantile regression approach models the entire distribution of hedge fund returns without assuming normality. In particular, the conditional median regression provides more robust and more efficient estimates than those of the conditional mean regression in the case of distributions of returns that deviate from normality. To take advantage of these features, we consider using the alpha of a median regression multi-factor model as a measure of performance. If the median hedge fund return is significantly higher than what expected, given a set of factors and their respective risk exposures, then the performance is considered to be superior. Therefore, hedge fund strategies with a higher value of median alpha are preferable. We expect that this measure should perform better than Jensen's alpha in our application to hedge fund returns whose characteristics clearly show deviations from normality (see Table 1).

Markowitz (1952) argued that the rule an investor should follow is to consider the expected return as a desirable thing and the variance of return as an undesirable thing. In these lines, we propose a measure based on the conditional quantiles, which aims at maximising the median conditional return while keeping the variability of conditional returns low. To do so we consider the following quantity

$$a^* = a_{0.5} - \frac{\lambda}{2}(a_{0.75} - a_{0.25}),$$

where $a_{0.5}$ is the alpha of a median multi-factor regression model estimated by maximum likelihood or by BMA, $a_{0.75}$ and $a_{0.25}$ are the respective estimates of the alphas of a 75th and a 25th quantile multi-factor regression model, and λ is some prespecified constant which gives weight to the impact of the variability of conditional returns or risk, as measured by the range $a_{0.75} - a_{0.25}$, to the value of the measure a^*. According to this measure hedge fund strategies with a higher value of a^* are preferable. The constant λ can be thought of as a risk aversion parameter; high values of λ characterise risk averse investors. A similar risk aversion parameter is used in classical mean-variance portfolio construction. In this literature several values of λ have been used; for example, Carlappi, Uppal and Wang (2007) have used $\lambda = 1$, while Han (2006) considered different values of λ varying from 2 to 10. In this study we take $\lambda = 2$.

The analysis in this section investigates the implications of the proposed quantile regression approach in terms of hedge fund style portfolio construction. This should be of interest to investors who are considering investing in indices of hedge funds (see Amenc and Martellini, 2002; McFall Lamm, 2003; Agarwal and Naik, 2004; Morton et al., 2006; Giamouridis and Vrontos, 2007). Our main objective is to evaluate the relative performance of the constructed style portfolios based on the best, according to several model selection strategies, multi-factor conditional mean and conditional quantile models in an out-of-sample fashion. For the conditional mean regression method we consider Bayesian model selection techniques, namely MP and BMA, which have been

Table 3
Ranking of HFR single indices using different model selection strategies

Median approach				Quantile approach		Mean approach	
BMA	MP	AIC	BIC	BMA	MP	BMA	MP
ENH	ENH	ENH	EH	RVA	RVA	EH	EH
EH	RVA	EH	ENH	MA	MA	M	ENH
RVA	EH	RVA	RVA	ED	ED	ENH	MT
MA	MA	MA	MA	CA	CA	RVA	ED
EMT	SS	EMT	EMT	DS	DS	MT	RVA
SS	EMT	M	ED	ENH	ENH	EMT	M
M	ED	MT	M	EMN	EMN	ED	EMT
ED	M	ED	SS	FIA	FIA	MA	MA
MT	DS	SS	DS	EH	EH	DS	DS
DS	FIA	FIA	FIA	MT	MT	SS	EMN
FIA	MT	DS	MT	EMT	EMT	EMN	SS
EMN	EMN	CA	EMN	M	M	CA	FIA
CA	CA	EMN	CA	SS	SS	FIA	CA

This table presents the ranking of HFR single strategies during the estimation period April 1990 to December 2003. After estimating the best (according to different model selection techniques) conditional mean and quantile regression alphas (median alphas and α^*) for the thirteen HFR strategies considered, the strategies have been ranked on the basis of their estimated alphas.

Table 4
Correlation Measures of alphas for different model selection methods

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: correlations									
BMA median	(1)	1.000							
BMA mean	(2)	0.766	1.000						
MP median	(3)	0.978	0.638	1.000					
MP mean	(4)	0.761	0.985	0.632	1.000				
AIC median	(5)	0.932	0.812	0.891	0.790	1.000			
BIC median	(6)	0.979	0.745	0.969	0.752	0.909	1.000		
BMA quantile	(7)	0.099	0.026	0.067	0.101	0.040	0.176	1.000	
MP quantile	(8)	0.098	0.015	0.072	0.085	0.040	0.176	0.998	1.000
Panel B: probabilities of agreement									
BMA median	(1)	1.000							
BMA mean	(2)	0.782	1.000						
MP median	(3)	0.949	0.731	1.000					
MP mean	(4)	0.769	0.885	0.718	1.000				
AIC median	(5)	0.910	0.821	0.859	0.782	1.000			
BIC median	(6)	0.936	0.795	0.936	0.782	0.897	1.000		
BMA quantile	(7)	0.500	0.462	0.526	0.500	0.513	0.539	1.000	
MP quantile	(8)	0.513	0.474	0.539	0.513	0.526	0.551	0.987	1.000

This table presents the correlation measures of alphas for different model selection methods using the median and mean regression approach during the estimation period April 1990 to December 2003. Panel A, shows the correlation coefficients of alphas; Panel B, shows the probability of agreement of the rankings obtained for different pairs of hedge fund strategies.

found to outperform standard model selection strategies (AIC and BIC) by Vrontos et al. (2008). For the conditional quantile regression approach we further consider AIC and BIC in order to investigate their relative performance in the conditional quantiles setting.

We use the period from April 1990 to December 2003 to estimate all competing models with different subsets of factors and we apply the model selection methods to obtain the best models under each method. Next we obtain the estimated alphas (based on the estimation period) for the thirteen hedge fund strategies considered, for the best conditional mean and quantile regression models, and we rank the hedge fund strategies on the basis of their estimated alphas. The different rankings of the strategies are shown in Table 3. It can be seen that there is some amount of disagreement on the ranking among different model selection techniques; the amount of disagreement between the conditional mean method and the conditional median approach is quite large, while the disagreement of the quantile approach (based on a^*) with the other approaches is the largest. We further calculate the correlations of the alphas and the probabilities of agreement[4] for pairs of different methods. In Table 4 we present results on these quantities for all pairs of methods considered. In general, the correlations of alphas as well as the probabilities of agreement between pairs of methods within the median and mean regression approaches are relatively high. It can be seen, however, that the correlations of the quantile approach (which uses the measure a^*) with the other approaches are very low, while the corresponding probabilities of agreement are moderate. This is due to the quite large value of λ used in our study, which gives considerable weight to the variability of hedge fund returns measured by the range $a_{0.75}-a_{0.25}$. Lower values of λ give less penalty to this variability and, therefore, the ranking obtained by using a^* is more similar to that of the median regression approach and the respective correlations become much higher.

For all methods we formulate equally weighted portfolios based on the top 40% alpha hedge fund strategies. In our style allocation portfolio we use the top five alpha sorted hedge fund strategies since, in the literature, investing in at least five different strategies is considered to be a reasonable rule of thumb, which provides a minimal diversification in the constructed portfolio (see, for example, Hamza et al., 2006). To evaluate the performance of our portfolios in a two-year out-of-sample period we proceed as follows. The estimation period is redefined iteratively — it is augmented by one monthly observation at each step in order to utilize all the available information. At each iteration we re-estimate the best models' parameters, rank the estimated alphas and allocate wealth to the top five hedge fund strategies. The procedure ends when all 24 out-of-sample monthly observations have been processed.

We evaluate the different methods using unconditional (out-of-sample) measures. In particular, we consider the realised returns, the portfolio risk and the risk adjusted realised returns. We calculate the mean excess return within the out-of-sample period and the cumulative excess return at the end of the period. As a measure of risk we compute the standard deviation of excess returns. Now, in risk adjusted performance measurement the realised returns are related to a suitable risk measure. In the hedge fund performance literature the most commonly used risk adjusted measure is the Sharpe ratio (see, for example, Ackermann et al.,

[4] The probability of agreement is calculated as in Amenc and Martellini (2003). It essentially measures the probability that two methods, k, l will agree on the rank order of a randomly chosen pair of hedge funds. Mathematically, the 'probability of agreement', $A_{k,l}$ is computed through: $A_{k,l}=\frac{\sum_{i=1}^{N}\sum_{j=i+1}^{N} g_{k,l}(i,j)}{N(N-1)/2}$, where $g_{k,l}(i,j)$ is 1 if k, l agree on the rank order, i.e. $\alpha_k(i)>\alpha_k(j)$ and $\alpha_l(i)>\alpha_l(j)$ or $\alpha_k(i)<\alpha_k(j)$ and $\alpha_l(i)<\alpha_l(j)$, and 0 otherwise.

Table 5
Realised returns, portfolio risk and risk-adjusted performance measures

	Mean Ret.	Port. StD.	Downs. Risk	Cum. Ret.	ShR	SoR
BMA median	0.60%	1.43%	0.79%	14.42%	0.421	0.757
BMA mean	0.54%	1.51%	0.88%	12.93%	0.356	0.615
MP median	0.56%	1.31%	0.67%	13.35%	0.425	0.829
MP mean	0.59%	1.56%	0.87%	14.05%	0.376	0.673
AIC median	0.60%	1.48%	0.83%	14.37%	0.406	0.719
BIC median	0.63%	1.50%	0.83%	15.20%	0.422	0.761
BMA quantile	0.40%	0.82%	0.40%	9.57%	0.485	0.988
MP quantile	0.46%	1.00%	0.55%	10.93%	0.454	0.827

This table presents performance measures for style allocation portfolios comprising of the top 40% 'alpha' ranked hedge fund strategies. Portfolios are constructed for a period of two years using the models obtained for the estimation period, re-estimating alpha each month, and allocating wealth in the five top-performing strategies. The monthly mean excess return (Mean Ret.), the portfolio standard deviation (Port. StD.), the portfolio downside risk (Downs. Risk), the cummulative excess 2-year return (Cum. Ret.), the Sharpe ratio (ShR) and the Sortino ratio (SoR) are reported. The ranking of alphas is obtained using the mean and quantile regression approaches and different model selection methods, namely Bayesian model averaging (BMA), Bayesian most probable model (MP), Akaike's (1973) information criterion (AIC), and Schwartz (1978) Bayesian information criterion (BIC).

1999; Edwards and Liew, 1999; Liang, 1999; Schneeweis et al., 2002). The Sharpe ratio (Sharpe, 1966, 1994) is calculated as the ratio of the average portfolio excess return, $E(r_p)$, and the portfolio's standard deviation of excess returns, σ, i.e.

$$\mathrm{ShR} = \frac{E(r_p)}{\sigma}.$$

Thus, it can be interpreted as a measure of the reward per unit of risk. However, the use of Sharpe ratio as an evaluation measure for hedge funds has been criticised (see Amin and Kat, 2003) due to its dependence on the normality assumption which is not valid in the case of hedge fund returns. Another feature of Sharpe ratio is the use of the standard deviation as a measure of risk. Instead of using the volatility as a risk measure it would be preferable to consider some measure of the, so called, downside risk, which only measures the negative deviations from some reference value, since positive deviations from this value are considered to be desirable. The downside risk, δ, is given by

$$\delta = \sqrt{\frac{1}{T}\sum_{t=1}^{T} \min(0, r_{pt} - \mathrm{RV})^2},$$

where RV is the reference value, which is taken to be zero in our study. The reference value can be thought of as a minimum acceptable return. An alternative risk adjusted measure of performance, which measures risk by the downside deviation, is the Sortino ratio (Sortino and van der Meer, 1991; Sortino and Price, 1994) which is calculated as the ratio of the average excess return and the downside risk, i.e.

$$SoR = \frac{E(r_p) - RV}{\delta}.$$

In Table 5 we report the unconditional (out-of-sample) measures considered. Specifically, we have calculated and present the average portfolios' excess returns, the portfolios' standard deviations and downside risks, the cumulative excess returns and the risk adjusted performance measures, namely the Sharpe ratio and the Sortino ratio, for all the methods of portfolio construction. First we examine the realised excess returns of the constructed portfolios. Looking at the average portfolio returns and the cumulative portfolio returns we observe that the median regression approach based on BIC and BMA rank first and second, respectively. The portfolio constructed under the BIC Median approach outperforms the BMA Mean and MP Mean portfolios by 2.27% and 1.15% in excess returns, respectively. Similar conclusions are drawn if we look at the measures of portfolio risk. The lowest values of the portfolio standard deviation and downside risk are obtained for the quantile approach, which is based on the measure a^*. On the other hand, the largest values of the measures of risk are those obtained for the mean regression approach based on MP and BMA.

Combining the information on the average portfolio excess return and some measure of risk we compute risk adjusted measures of performance. From Table 5 it can be seen that the lowest values of both the Sharpe ratio and the Sortino ratio are those obtained for the mean regression approach; the Sharpe ratio is 0.356 and 0.376 for BMA Mean and MP Mean, respectively, while the Sortino ratio values are 0.615 and 0.673, respectively. It is evident that all the median and quantile regression approaches outperform the mean regression methods in terms of the risk adjusted performance measures, with the quantile approaches based on a^* being the best method. These findings may be attributed to the fact that, firstly, the quantile regression approach models the entire conditional distribution of hedge fund returns and is able to capture the special characteristics of the return series. Secondly, the measure a^* is constructed with the aim to exploit this information in order to construct style allocation portfolios. Hence, if the proposed approach based on regression quantiles is used to rank the hedge fund strategies, the constructed portfolios appear to be superior in terms of risk return performance measures.

5. Discussion

In this paper we have proposed a conditional quantile regression approach for the analysis of hedge fund single strategy return series. The aim of our analysis was to explore the impact of a number of risk factors on the entire conditional distribution of hedge fund returns. Unlike the standard conditional mean regression method, which only examines how the risk factors affect the returns on average, our quantile regression approach is able to uncover how this dependence varies across quantiles of returns. Thus, the approach provides useful insights into the distributional dependence of hedge fund returns on risk factors.

In our analysis it was also of great interest to identify the risk factors associated with hedge fund investing, in the context of quantile regression. To this end, we have proposed a parametric approach to inference on regression quantiles which enabled us to develop several model selection methods. In particular, we have proposed a Bayesian model comparison method which provides posterior probabilities for all competing models (i.e. subsets of risk factors) and the use of Bayesian model averaging in order to produce robust inferences for the quantities of interest.

Our quantile regression approach is particularly useful in cases where the distribution of returns is characterised by large skewness, kurtosis, fat tails, or in general deviates from normality. In those cases, the conditional mean regression method may not be adequate, while the quantile regression approach provides more robust and more efficient estimates/results. Deviations from normality are very common in hedge fund returns data. Due to the highly dynamic complex trading strategies of hedge funds, the return series usually display special characteristics, such as skewness, kurtosis and non-normality. Therefore, the quantile regression approach provides a useful tool for the analysis of hedge funds.

We have applied our proposed approach to thirteen hedge fund single strategy indices and compared it with the standard conditional mean regression method. We have found substantial differences with respect to the most relevant subsets of factors between the two methods and, in particular, we have found that a larger number of factors are required to model the upper and lower quantiles of hedge fund returns than those needed to explain the conditional mean or median. To investigate potential economic impacts of our approach we have constructed style allocation portfolios based on the best performing (according to the ranking of alphas for the best, under different model selection methods, mean and quantile regression models) hedge fund strategies. We have shown that the conditional quantile regression models provide improved style portfolios compared to the mean regression based portfolios.

We believe that there exist several potential applications of our conditional quantile regression approach in the area of hedge fund investing. For example, an interesting application of the proposed method could be the management of risk of hedge fund strategies, by calculating the Value at Risk (VaR) of a strategy using the 5th or 10th regression quantile. Rather than just providing a quantitative risk measure, the method is also able to qualify the nature of risk a hedge fund strategy is exposed to by identifying the relevant risk factors and associating the VaR with them through the estimated risk exposures. Clearly, many interesting questions remain open and various topics for future research arise in the context of conditional quantile regression for the analysis of hedge funds.

Acknowledgements

We would like to thank the editor and two anonymous referees for numerous constructive comments that improved the quality of the paper. We also thank Mark Carhart for giving us access to the data of the 'momentum' factor; the Fama and French (1993) 'size' and 'book-to-market' factors' data were obtained from Kenneth French's homepage.

References

Ackermann, C., McEnally, R., Ravenscraft, D., 1999. The performance of hedge funds: risk, return, and incentives. Journal of Finance 54, 833–874.
Agarwal, V., Naik, N.Y., 2004. Risks and portfolio decisions involving hedge funds. Review of Financial Studies 17, 63–98.
Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petrox, B.N., Caski, F. (Eds.), Second International Symposium on Information Theory. Akademiai Kiado, Budapest, pp. 267–281.
Amenc, N., Martellini, L., 2002. Portfolio optimization and hedge fund style allocation decisions. Journal of Alternative Investments 5, 7–20.
Amenc, N., Martellini, L., 2003. The alpha and omega of hedge fund performance measurement. Discussion paper, EDHEC Business School.
Amin, G.S., Kat, H.M., 2003. Hedge fund performance 1990–2000: do the "money machines" really add value? Journal of Financial and Quantitative Analysis 38, 251–274.
Avramov, D., 2002. Stock return predictability and model uncertainty. Journal of Financial Economics 64, 423–458.
Barnes, M.L., Hughes, A.W., 2002. A quantile regression analysis of the cross section of stock market returns. Working Paper 02-2. Federal Reserve Bank of Boston.
Bassett, W.G., Chen, H.L., 2001. Portfolio style: return-based attribution using quantile regression. Empirical Economics 26, 293–305.
Bernardo, J.M., Smith, A.F.M., 1994. Bayesian theory. Wiley Series in Probability and Mathematical Statistics.
Billio, M., Getmansky, M., Pelizzon, L., 2006. Time-varying risk exposure of hedge funds. Working Paper.
Bossaerts, P., Hillion, P., 1999. Implementing statistical criteria to select return forecasting models: what do we learn? Review of Financial Studies 12, 405–428.
Buchinsky, M., 1994. Changes in U.S. wage structure 1963–1987: an application of quantile regression. Econometrica 62, 405–458.
Buchinsky, M., 1995a. Estimating the asymptotic covariance matrix for quantile regression models: a Monte Carlo study. Journal of Econometrics 68, 303–338.
Buchinsky, M., 1995b. Quantile regression box-cox transformation model, and the U.S. wage structure, 1963–1987. Journal of Econometrics 65, 109–154.
Buchinsky, M., 1998. Recent advances in quantile regression models: a practical guideline for empirical research. Journal of Human Resources 33, 88–126.
Capocci, D., Hubner, G., 2004. Analysis of hedge fund performance. Journal of Empirical Finance 11, 55–89.
Carhart, M.M., 1997. On the persistence in mutual fund performance. Journal of Finance 52, 57–82.
Carlappi, L., Uppal, R., Wang, T., 2007. Portfolio selection with parameter and model uncertainty: a multi-prior approach. Review of Financial Studies 20, 41–81.
Chernozhukov, V., Umantsev, L., 2001. Conditional value-at-risk: aspects of modeling and estimation. Empirical Economics 26, 271–292.
Cremers, M.K.J., 2002. Stock return predictability: a Bayesian model selection perspective. Review of Financial Studies 15, 1223–1249.
Dellaportas, P., Vrontos, I.D., 2007. Modelling volatility asymmetries: a Bayesian analysis of a class of tree structured multivariate GARCH models. Econometrics Journal 10, 503–520.

Denison, D.G.T., Holmes, C.C., Mallick, B.K., Smith, A.F.M., 2002. Bayesian methods for non-linear classification and regression. Wiley Series in Probability and Statistics.
Edwards, F.R., Liew, J., 1999. Hedge funds versus managed futures as Asset classes. Journal of Derivatives 6, 45–64.
Efron, B., 1979. Bootstrap methods: another look at the jackknife. Annals of Statistics 7, 1–26.
Efron, B., 1982. The jackknife, the bootstrap and other resampling plans. Society for Industrial and Applied Mathematics, Philadelphia, PA.
Eide, E., Showalter, M.H., 1998. The effect of school quality on student performance: a quantile regression approach. Economic Letters 58, 345–350.
Eide, E., Showalter, M.H., 1999. Factors affecting the transmission of earnings across generations: a quantile regression approach. Journal of Human Resources 34, 253–267.
Engle, R.F., Manganelli, S., 2004. CAViaR: conditional autoregressive value at risk by regression quantiles. Journal of Business and Economic Statistics 22, 367–381.
Fama, E., French, K., 1993. Common risk factors in the returns of stocks and bonds. Journal of Financial Economics 33, 3–56.
Foerster, S., 2006. What drives equity market neutral hedge fund returns? Working Paper.
Fung, W., Hsieh, D.A., 1999. A primer on hedge funds. Journal of Empirical Finance 6, 309–331.
Fung, W., Hsieh, D.A., 2000. Performance characteristics of hedge funds and commodity funds: natural vs spurious biases. Journal of Financial and Quantitative Analysis 35, 291–307.
Fung, W., Hsieh, D.A., 2001. The risk in hedge fund strategies: theory and evidence from trend followers. Review of Financial Studies 14, 313–341.
Giamouridis, D., Vrontos, I.D., 2007. Hedge fund portfolio construction: a comparison of static and dynamic approaches. Journal of Banking and Finance 31, 199–217.
Hamza, O., Kooli, M., Roberge, M., 2006. Further evidence on hedge fund return predictability. Journal of Wealth Management 9, 68–79.
Han, Y., 2006. Asset allocation with a high dimensional latent factor stochastic volatility model. Review of Financial Studies, 19 1, 237–271.
Hedges IV, J.R., 2005. Hedges on Hedge Funds: How to Successfully Analyze and Select an Investment. John Wiley & Sons.
Jensen, M., 1968. The performance of mutual funds in the period 1945–1964. Journal of Finance 23, 389–416.
Kass, R.E., Raftery, A.E., 1995. Bayes factors. Journal of the American Statistical Association 90, 773–795.
Kass, R.E., Wasserman, L., 1995. A reference Bayesian test for nested hypothesis and its relationship to the Schwarz criterion. Journal of the American Statistical Association 90, 928–934.
Knight, K., Bassett, G. and Tam, M.S. (2000). Comparing Quantile Estimators for the Linear Model. Preprint.
Koenker, R., 2005. Quantile regressions. Cambridge University Press.
Koenker, R., Bassett, G., 1978. Regression quantiles. Econometrica 46, 33–50.
Komunjer, I., 2005. Quasi-maximum likelihood estimation for conditional quantiles. Journal of Econometrics 128, 137–164.
Levin, J., 2001. For whom the reductions count: a quantile regression analysis of class size on scholastic achievement. Empirical Economics 26, 221–246.
Liang, B., 1999. On the performance of hedge funds. Financial Analysts Journal 55, 72–85.
Markowitz, H., 1952. Portfolio selection. Journal of Finance 7, 77–91.
McFall Lamm, R., 2003. Asymmetric returns and optimal hedge fund portfolios. Journal of Alternative Investments 6 (2), 9–21.
Meligkotsidou, L., Vrontos, I.D., 2008. Detecting structural breaks and identifying risk factors in hedge fund returns: a bayesian approach. Journal of Banking and Finance 32, 2471–2481.
Mitchell, M., Pulvino, T., 2001. Characteristics of risk and return in risk arbitrage. Journal of Finance 56, 2135–2175.
Morillo, D.S. (2000). Monte Carlo American Option Pricing with Nonparametric Regression. In Essays in Nonparametric Econometrics, Dissertation, University of Illinois.
Morton, D.P., Popova, E., Popova, I., 2006. Efficient fund of hedge funds construction under downside risk measures. Journal of Banking and Finance 30, 503–518.
Nicholas, J.G., 2000. Market Neutral Investing: Long/Short Hedge Fund Strategies. Bloomberg Press.
Patton, A.J. (forthcoming). Are "Market Neutral" hedge funds really Market Neutral?. Review of Financial Studies.
Raftery, A.E., Madigan, D., Volinsky, C.T., 1996. Accounting for model uncertainty in survival analysis improves predictive performance (with discussion). In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics, 5. Oxford University Press, pp. 323–349.
Raftery, A.E., Madigan, D., Hoeting, J.A., 1997. Bayesian model averaging for linear regression models. Journal of the American Statistical Association 92, 179–191.
Schneeweis, T., Kazemi, H., Martin, G., 2002. Understanding hedge fund performance: research issues revisited-part I. Journal of Alternative Investments 5, 6–22.
Schwarz, G., 1978. Estimating the dimension of a model. Annals of Statistics 6, 461–464.
Sharpe, W.F., 1966. Mutual fund performance. Journal of Business 1, 119–138.
Sharpe, W.F., 1994. The Sharpe ratio. Journal of Portfolio Management 21, 49–58.
Smith, A.F.M., Roberts, G.O., 1993. Bayesian computation via the Gibbs Sampler and related Markov Chain Monte Carlo Methods. Journal of the Royal Statistical Society B 55, 3–23.
Sortino, F.A., van der Meer, R., 1991. Downside risk. Journal of Portfolio Management 17, 27–31.
Sortino, F.A., Price, L.N., 1994. Performance measurement in a downside risk framework. Journal of Investing 3, 50–58.
Stefanini, F., 2006. Investment Strategies of Hedge Funds. John Wiley & Sons.
Taylor, J., 1999. A quantile regression approach to estimating the distribution of multiperiod returns. Journal of Derivatives 7, 64–78.
Tierney, L., 1994. Markov chains for exploring posterior distributions (with discussion). Annals of Statistics 22, 1701–1762.
Trede, M., 1998. Making mobility visible: a graphical device. Economics Letters 59, 77–82.
Yu, K., Moyeed, R.A., 2001. Bayesian quantile regression. Statistics and Probability Letters 54, 437–447.
Yu, K., Zhang, J., 2005. A three-parameter asymmetric laplace distribution and its extension. Communications in Statistics — Theory and Methods 34, 1867–1879.
Yu, K., Lu, Z., Stander, J., 2003. Quantile regression: applications and current research areas. The Statistician 52, 331–350.
Vrontos, I.D., Dellaportas, P., Politis, D.N., 2003a. Inference for some multivariate ARCH and GARCH models. Journal of Forecasting 22, 427–446.
Vrontos, I.D., Dellaportas, P., Politis, D.N., 2003b. A full-factor multivariate GARCH model. Econometrics Journal 6, 312–334.
Vrontos, S.D., Vrontos, I.D., Giamouridis, D., 2008. Hedge fund pricing and model uncertainty. Journal of Banking and Finance 32, 741–753.