Interquantile Expectation Regression

Sander Barendse

[1] *Erasmus University Rotterdam, The Netherlands*

Interquantile Expectation Regression

Sander Barendse*

Erasmus University Rotterdam

March 17, 2017

Abstract

We propose a semiparametric estimator to determine the effects of explanatory variables on the conditional interquantile expectation (IQE) of the random variable of interest, without specifying the conditional distribution of the underlying random variables. IQE is the expected value of the random variable of interest given that its realization lies in an interval between two quantiles, or in an interval that covers the range of the distribution to the left or right of a quantile. Our so-called interquantile expectation regression (IQER) estimator is based on the GMM framework. We derive consistency and the asymptotic distribution of the estimator, and provide a consistent estimator of the asymptotic covariance matrix. Our results apply to stationary and ergodic time series. In a simulation study we show that our asymptotic theory provides an accurate approximation in small samples. We provide an empirical illustration in finance, in which we use the IQER estimator to estimate one-step-ahead daily expected shortfall conditional on previously observed daily, weekly, and monthly aggregated realized measures.

Keywords: quantile, interquantile expectation, regression, generalized method of moments, risk management, expected shortfall

JEL classification: C13, C14, C32, C58, G32

1   Introduction

We propose a semiparametric estimator to determine the effects of explanatory variables on the conditional interquantile expectation (IQE) of the random variable of interest. IQE is the expected value of the random variable of interest given that its realization lies in an interval between two quantiles, or in an interval that covers the range of the distribution to the left or right of a quantile. We refer to the latter two special cases as lower-quantile expectation (LQE) and upper-quantile expectation (UQE). Our so-called interquantile expectation regression (IQER) estimator estimates the interquantile expectation jointly with the quantiles and does not require the specification of the conditional distribution of the variable of interest. We show that our estimator is consistent and asymptotically normal for stationary and ergodic time series. To the best of our knowledge we are the first to study a semiparametric estimator of the conditional interquantile expectation and its asymptotic properties.

Our estimator is useful in several settings. In various academic disciplines LQE and UQE are used to measure risk, because they capture the expected value of rare events beyond a certain (tail) quantile. In finance the LQE of an asset return is referred to as expected shortfall (ES). Our estimator is the first to facilitate the estimation of ES conditional on several explanatory variables without specifying a full multivariate model. As a consequence it is easy to estimate systemic risk measures such as CoES of Adrian and Brunnermeier (2016), which relates the ES of the financial system to the return of an individual financial asset and several state variables, whereas current approaches require multivariate modelling and neglect important state variables.

IQE can be seen as the expectation of a truncated distribution. Our estimator is therefore related to estimators of trimmed or truncated conditional means, such as the trimmed least squares estimator of Ruppert and Carroll (1980). This estimator aims to robustly, but inefficiently, estimate the parameters of a linear conditional mean model by removing outliers of the dependent variable in a first stage. Our estimator can be used for the same purpose, but improves on such estimators because it provides standard errors that do not depend on a nuisance parameter. Moreover, it is not restricted to robustly estimating conditional means. We can estimate the parameters of IQEs on any interquantile interval, such that we can study and compare the relation of the explanatory variables and the variable of interest on different intervals of the support of the conditional distribution. Our estimator can therefore be used to

quantify nonlinearities in the relation between the variable of interest and explanatory variables.

IQER is closely related to linear regression and quantile regression (Koenker and Bassett, 1978), since these methods estimate the parameters of linear models for the conditional mean and the conditional quantile, whereas we estimate the parameters of a linear model for the conditional IQE. We pursue a linear IQE specification because it is suitable in many scenarios. This follows from a linear quantile specification implying a linear IQE specification, and the widespread use of such linear quantile specifications, see for instance Koenker and Hallock (2001) for an overview. We leave the development of a nonlinear IQER estimator for future research.

To develop our estimator we use a recently proposed strictly consistent scoring function for the quantile and LQE pair introduced in Fissler et al. (2016b) to obtain the moment conditions required in the generalized method of moments (GMM) framework of Hansen (1996). The strict consistency of this scoring function, or loss function as it is often referred to in econometrics, implies that its expectation is uniquely minimized at the true quantile and LQE, such that we can use the first order conditions as moment conditions. The moment conditions obtained from the scoring function are nonsmooth, such that we use results on nonsmooth GMM estimation in Newey and McFadden (1994) to obtain the consistency and asymptotic distribution of the IQER estimator, and a consistent estimator of the asymptotic covariance matrix.

By construction the estimation of IQE requires the estimation of the quantiles that are used in the definition of the IQEs, either jointly or in a preceding stage. The IQER estimator therefore also estimates the quantile parameters. Moreover, we allow for the joint estimation of IQEs and quantiles at many quantile levels, such that we can test hypotheses that relate to IQEs at different quantile levels. We show that estimation error of the quantile parameter estimates does not affect the estimation error of the IQE parameter estimates. Moreover, unlike the standard errors of the quantile parameters, the standard errors of the IQE parameters obtained with the IQER estimator do not contain a nuisance parameter in terms of the conditional density function. As such, we do not require numerical approximation or bootstrap techniques to overcome the nuisance parameter problem when we only consider IQE parameter estimates.

We propose a multi-stage estimation procedure in which we (i) obtain quantile parameter estimates in a first stage through GMM quantile estimation, and (ii) obtain the IQE estimates in the subsequent stages that consist of auxiliary regressions. This procedure is computationally fast and we show that it results in an asymptotically efficient estimator. Finally, we show that in the case of iid data we can substitute GMM quantile estimation with quantile regression, for

which optimization routines are available in most statistical computing packages.

When we only consider LQEs we can relate the IQER estimator to several estimators of ES. Our estimator differs from the others in that these are usually parametric or nonparametric and univariate in nature, whereas our method allows for multiple relevant explanatory variables and does not require the specification of the conditional distribution. A large number of such estimators is summarized in Nadarajah et al. (2014). The two methods that are closest in spirit to our estimator are firstly Taylor (2008a), who proposes an estimator based on expectiles, a generalized form of quantiles introduced by Newey and Powell (1987). Secondly, Taylor (2008b) provides an estimator of LQE that is equivalent to the IQER estimator in the unconditional case, i.e. when we do not consider explanatory variables.

In simulation experiments we show that our asymptotic theory provides a good approximation in small samples. We use a data generating process in which an explanatory variable linearly affects the conditional mean and volatility of the dependent variable. This data generating process has an exact solutions for the true IQE parameters, and allows for autocorrelation and conditional heteroskedasticity.

In an empirical illustration we provide an application of the IQER estimator in finance. We estimate one-step-ahead ES of the daily return of a S&P 500 exchange traded fund using the IQER estimator with previously observed daily, weekly, and monthly aggregated realized volatility as explanatory variables. We compare our method to estimates of one-step-ahead ES that first estimate one-step-ahead realized volatility with the same variables, as in the HAR model of Corsi (2009), and subsequently impose a distributional form on the standardized errors to map realized volatility to ES. The estimates of the IQER estimator are on average more conservative than other models, and react more strongly to realized volatility observed in the recent past. Moreover, we can reject a significant relation between monthly aggregated realized volatility and one-step-ahead ES.

The remainder is structured as follows. Section 2 introduces the IQE models and develops the estimation method and asymptotic theory of the IQER estimator. Section 3 uses Monte Carlo experiments to study the small-sample properties of our estimators. Section 4 provides two empirical illustrations of our method. Section 5 concludes.

2 Theory

2.1 Definitions

Consider a stochastic process $\{(Y_t, \mathbf{Z}_t')' : \Omega \longrightarrow \mathbb{R}^{l+1}, 1 \leq l < \infty, t = 1, \ldots, T\}$ defined on a complete probability space $(\Omega, \mathcal{F}, P)$, where $\mathcal{F} \equiv \{\mathcal{F}_t, t = 1, \ldots, T\}$ and $\mathcal{F}_t$ is defined as the σ-algebra $\mathcal{F}_t = \sigma\{(Y_{s-1}, \mathbf{Z}_s')', s \leq t\}$. $Y_t : \Omega \longrightarrow \mathbb{R}$ denotes the random variable of interest, whereas $\mathbf{Z}_t : \Omega \longrightarrow \mathbb{R}^l$ defines a vector of random variables of length l. We select k, $1 \leq k < \infty$, elements from $(Y_{t-1}, \mathbf{Z}_t', Y_{t-2}, \mathbf{Z}_{t-1}', \ldots)'$ to include in the explanatory variable vector $\mathbf{X}_t$, for all $t = 1, \ldots, T$.

Define $F_t(\cdot)$, $F_t^{-1}(\cdot)$, $f_t(\cdot)$, and $E_t[\cdot]$ as the distribution function, the inverse distribution function, the probability density function, and the expectation of Y_t conditional on $\mathcal{F}_t$. We assume that $f_t(\cdot)$ is continuous and strictly positive, such that $F_t(\cdot)$ has unique quantiles. This assumption is common in the quantile estimation literature.

Consider two levels α_i and α_j, such that $0 < \alpha_i < \alpha_j < 1$. Then the conditional quantile, $F_t^{-1}(\alpha_i)$, will be referred to as $Q_t(\alpha_i)$, the conditional lower-quantile expectation, $E_t[Y_t | Y_t < Q_t(\alpha_i)]$, as $L_t(\alpha_i)$, the interquantile expectation $E_t[Y_t | Q_t(\alpha_i) < Y_t < Q_t(\alpha_j)]$, as $I_t(\alpha_i, \alpha_j)$, and the upper-quantile expectation, $E_t[Y_t | Y_t > Q_t(\alpha_i)]$, as $U_t(\alpha_i)$.

From the definitions above we see that $L_t(\alpha_i)$ and $U_t(\alpha_i)$ denotes the expectation of Y_t conditional on its realization lying below and above the conditional quantile of Y_t given the realization. These quantities are sometimes also referred to as first order lower and upper partial moments. The quantity $I_t(\alpha_i, \alpha_j)$ denotes the expectation of Y_t conditional on its realization lying between two conditional quantiles given the realization. We can therefore use these quantities to study the average effect of $\mathcal{F}_t$-measurable explanatory variables on Y_t on a specific subset of realizations defined by open intervals in which realizations of Y_t must fall.

We propose the following linear models to study the effect of the explanatory variables collected in $\mathbf{X}_t$ on the conditional quantiles and interquantile expectations of Y_t.

$$Q_t(\alpha_i) = \mathbf{X}_t'\boldsymbol{\beta}_{Q,i}, \tag{1}$$

$$L_t(\alpha_i) = \mathbf{X}_t'\boldsymbol{\beta}_{L,i}, \tag{2}$$

$$I_t(\alpha_i, \alpha_j) = \mathbf{X}_t'\boldsymbol{\beta}_{I,i,j}, \tag{3}$$

$$U_t(\alpha_i) = \mathbf{X}_t'\boldsymbol{\beta}_{U,i}, \tag{4}$$

where $\boldsymbol{\beta}_{Q,i}$ and $\boldsymbol{\beta}_{L,i}$, $\boldsymbol{\beta}_{I,i,j}$, and $\boldsymbol{\beta}_{U,i}$ denote the $(k \times 1)$ parameter vectors at quantile level α_i, and between α_i and α_j.

The Q_t model in (1) was introduced in the literature by Koenker and Bassett (1978). The linear models (2)-(4) have, to the best of our knowledge, not been studied before.

Whether these linear models are appropriate is ultimately determined by the research question at hand. It should be noted, however, that empirical studies that employ quantile regression methods generally use linear quantile regression specifications. This suggests that linear models of (2)-(4) should be appropriate as well. Indeed a linear model of Q_t naturally suggests a linear model for $L_t(\alpha_i)$, $I_t(\alpha_i, \alpha_j)$, and $U_t(\alpha_i)$. This can be seen by, for instance, rewriting $L_t(\alpha_i)$ as the integral $L_t(\alpha_i) = \frac{1}{\alpha_i} \int_0^{\alpha_i} Q_t(u) du = \mathbf{X}_t'(\frac{1}{\alpha_i} \int_0^{\alpha_i} \boldsymbol{\beta}_{Q,u} du) = \mathbf{X}_t'\boldsymbol{\beta}_{L,i}$.

Another benefit of linear models concerns the the transition from conditional IQEs to marginal IQEs, which is straightforward for linear models. Without loss of generality consider the marginal LQE given by $E[L_t(\alpha_i)] = E[Y_t | Y_t < Q_t(\alpha_i)]$. We find the linear relation $E[L_t(\alpha_i)] = E[\mathbf{X}_t | Y_t < Q_t(\alpha_i)]'\boldsymbol{\beta}_{L,i}$. We can therefore study hypothetical changes to $L_t(\alpha_i)$ by changing $\boldsymbol{\beta}_{L,i}$ and estimating $E[\mathbf{X}_t | Y_t < Q_t(\alpha_i)]$ from the data. One caveat here is that we assume constant conditional quantiles. If this assumption is too restrictive we must resort to other methods, because the transition of conditional to marginal quantiles requires more advanced methods, see for instance Chernozhukov et al. (2013). However, when we consider underlying changes that affect tails but not the body of the distribution, or *vice versa*, the assumption is justified.

2.2 Moment conditions

The basis of our estimation theory is the existence of a strictly consistent scoring function for the pair $(Q_t(\alpha_i), L_t(\alpha_i))$ as shown in Fissler et al. (2016b). The expectation of strictly consistent scoring function, or loss function, is uniquely optimized at the true values of the statistics that it takes as arguments. Examples of well-known scoring functions are the mean squared prediction error function, which is strictly consistent for the mean, and the tick-loss function, which is strictly consistent for the quantile. We refer to Gneiting and Raftery (2007) and Gneiting (2011) for a detailed treatment of strictly consistent scoring functions.

For our linear model we consider the following specification of the strictly consistent scoring

6

function:

$$
\begin{aligned}
S(Y_t, \mathbf{X}'_t \boldsymbol{\beta}_{Q,i}, \mathbf{X}'_t \boldsymbol{\beta}_{L,i}) \equiv\ & (\alpha_i - \mathbb{1}[Y_t - \mathbf{X}'_t \boldsymbol{\beta}_{Q,i} < 0])(Y_t - \mathbf{X}'_t \boldsymbol{\beta}_{Q,i}) \\
& - \frac{1}{\alpha_i} \exp(\mathbf{X}'_t \boldsymbol{\beta}_{L,i}) \mathbb{1}[Y_t - \mathbf{X}'_t \boldsymbol{\beta}_{Q,i} < 0](Y_t - \mathbf{X}'_t \boldsymbol{\beta}_{Q,i}) \\
& + \exp(\mathbf{X}'_t \boldsymbol{\beta}_{L,i})(\mathbf{X}'_t \boldsymbol{\beta}_{L,i} - \mathbf{X}'_t \boldsymbol{\beta}_{Q,i}) - \exp(\mathbf{X}'_t \boldsymbol{\beta}_{L,i}).
\end{aligned}
\tag{5}
$$

This specification is not unique. The entire set of strictly consistent scoring functions for $(Q_t(\alpha_i), L_t(\alpha_i))$ is given in Corollary 5.5 in Fissler et al. (2016b). We use (5) because it is closely related to the tick-loss function and it is recommended in Fissler et al. (2016a), which accompanies Fissler et al. (2016b) as an empirical note.

The existence of this scoring function facilitates the estimation of $(Q_t(\alpha_i), L_t(\alpha_i))$ by optimization of its expectation, because the strict consistency of (5) implies that the conditional expectation of the scoring function is optimal at the true values, i.e.:

$$
E_t[S(Y_t, \mathbf{X}'_t \boldsymbol{\beta}_{0,Q,i}, \mathbf{X}'_t \boldsymbol{\beta}_{0,L,i})] \leq E_t[S(Y_t, \mathbf{X}'_t \boldsymbol{\beta}^*_{Q,i}, \mathbf{X}'_t \boldsymbol{\beta}^*_{L,i})],
\tag{6}
$$

for all $(\boldsymbol{\beta}^*_{Q,i}{}', \boldsymbol{\beta}^*_{L,i}{}')'$ in the parameter space, and where $\mathbf{X}_t \boldsymbol{\beta}_{0,Q,i} = Q_t(\alpha_i)$ and $\mathbf{X}'_t \boldsymbol{\beta}_{0,L,i} = L_t(\alpha_i)$ equal the true values.

Fissler et al. (2016b) show that no strictly consistent scoring function exists for $L_t(\alpha_i)$ alone. We must therefore either estimate $L_t(\alpha_i)$ and $Q_t(\alpha_i)$ jointly or estimate $Q_t(\alpha_i)$ in advance. Here we consider the joint estimation of the LQEs, IQEs, and UQEs of interest, and the respective quantiles that we must necessarily estimate to find these quantities.

We therefore derive moment conditions for $Q_t(\alpha_i)$, $L_t(\alpha_i)$, $I_t(\alpha_i, \alpha_j)$, and $U_t(\alpha_i)$ from conditional moment restrictions such that we can obtain an estimator using the GMM framework. We obtain conditional moment restrictions for $Q_t(\alpha_i)$, $L_t(\alpha_i)$, and $U_t(\alpha_i)$ from the first order conditions of $E_t[S(Y_t, \mathbf{X}'_t \boldsymbol{\beta}_{Q,i}, \mathbf{X}'_t \boldsymbol{\beta}_{L,i})]$. The conditional moment restriction for $I_t(\alpha_i, \alpha_j)$ is subsequently derived from a generalization of the conditional moment restrictions of $L_t(\alpha_i)$ and $U_t(\alpha_i)$.

Let $\boldsymbol{\beta}_{0,Q,i}$, $\boldsymbol{\beta}_{0,L,i}$, $\boldsymbol{\beta}_{0,I,i,j}$, and $\boldsymbol{\beta}_{0,U,i}$ denote the true parameter vectors of the linear models at quantile level α_i. The following lemma shows that $\boldsymbol{\beta}_{0,Q,i}$, $\boldsymbol{\beta}_{0,L,i}$, and $\boldsymbol{\beta}_{0,U,i}$ satisfy the following first order conditions.

Lemma 1. *(First order conditions) Assume that for every t, $1 \leq t \leq T$, Y_t has a continuously*

differentiable conditional distribution function F_t. Then, for any level α_i, $0 < \alpha_i < 1$, the optimal parameter vector $(\boldsymbol{\beta}'_{0,Q,i}, \boldsymbol{\beta}'_{0,L,i})'$ satisfies the following first order conditions:

$$E_t \begin{bmatrix} g_{Q,i}(\boldsymbol{\beta}_{0,Q,i}, Y_t, \mathbf{X}_t) \\ g_{L,i}(\boldsymbol{\beta}_{0,Q,i}, \boldsymbol{\beta}_{0,L,i}, Y_t, \mathbf{X}_t) \end{bmatrix} = \mathbf{0} \ \ a.s.\text{-}P, \tag{7}$$

where $g_{Q,i}(\boldsymbol{\beta}_{Q,i}, Y_t, \mathbf{X}_t) \equiv \mathbb{1}(Y_t - \mathbf{X}'_t\boldsymbol{\beta}_{Q,i} < 0) - \alpha_i$, and $g_{L,i}(\boldsymbol{\beta}_{Q,i}, \boldsymbol{\beta}_{L,i}, Y_t, \mathbf{X}_t) \equiv \mathbf{X}'_t\boldsymbol{\beta}_{L,i} - \mathbf{X}'_t\boldsymbol{\beta}_{Q,i} - \frac{1}{\alpha_i}\mathbb{1}(Y_t - \mathbf{X}'_t\boldsymbol{\beta}_{Q,i} < 0)(Y_t - \mathbf{X}'_t\boldsymbol{\beta}_{Q,i})$.

Similarly, the optimal parameter vector $(\boldsymbol{\beta}'_{0,Q,i}, \boldsymbol{\beta}'_{0,U,i})'$ satisfies the following first order conditions:

$$E_t \begin{bmatrix} g_{Q,i}(\boldsymbol{\beta}_{0,Q,i}, Y_t, \mathbf{X}_t) \\ g_{U,i}(\boldsymbol{\beta}_{0,Q,i}, \boldsymbol{\beta}_{0,U,i}, Y_t, \mathbf{X}_t) \end{bmatrix} = \mathbf{0} \ \ a.s.\text{-}P, \tag{8}$$

where $g_{U,i}(\boldsymbol{\beta}_{Q,i}, \boldsymbol{\beta}_{U,i}, Y_t, \mathbf{X}_t) \equiv \mathbf{X}'_t\boldsymbol{\beta}_{U,i} - \mathbf{X}'_t\boldsymbol{\beta}_{Q,i} - \frac{1}{1-\alpha_i}\mathbb{1}(Y_t - \mathbf{X}'_t\boldsymbol{\beta}_{Q,i} > 0)(Y_t - \mathbf{X}'_t\boldsymbol{\beta}_{Q,i})$.

In $g_{Q,i}(\cdot)$ we recognize the derivative of the tick-loss function, which is the strictly consistent scoring function for $Q_t(\alpha_i)$. Notice that it only depends on $\boldsymbol{\beta}_{Q,i}$. A similar consistent moment condition for the quantile is given in Giacomini and Komunjer (2005), who use a GMM quantile estimator to obtain an encompassing test of two quantile forecasts.

Moreover, we find that $g_{L,i}(\cdot)$ and $g_{U,i}(\cdot)$ depend linearly on $\boldsymbol{\beta}_{L,i}$ and $\boldsymbol{\beta}_{U,i}$, respectively. Alternative conditional moment restrictions for $\boldsymbol{\beta}_{L,i}$ and $\boldsymbol{\beta}_{U,i}$ can be directly derived from the definitions of $L_t(\alpha_i)$ and $U_t(\alpha_i)$ in (2) and (4), e.g. $E_t[\mathbf{X}'_t\boldsymbol{\beta}_{L,i} - \frac{1}{\alpha_i}\mathbb{1}(Y_t - \mathbf{X}'_t\boldsymbol{\beta}_{Q,i} < 0)Y_t] = 0$ for $L_t(\alpha_i)$. However, this specification leads to standard errors that contain $f_t(\cdot)$ as nuisance parameter, which is a considerable disadvantage.

The conditional moment condition restriction for $I_t(\alpha_i, \alpha_j)$ cannot be obtained from a strictly consistent scoring function, because this scoring function has not been developed yet. We can, however, induce a conditional moment restriction for $I_t(\alpha_i, \alpha_j)$ from the functions $g_{L,i}(\cdot)$ and $g_{U,i}(\cdot)$. Specifically, consider the function

$$g_{I,i,j}(\boldsymbol{\beta}_{Q,i}, \boldsymbol{\beta}_{Q,j}, \boldsymbol{\beta}_{I,i,j}, Y_t, \mathbf{X}_t) \equiv \mathbf{X}'_t\boldsymbol{\beta}_{U,i} - \frac{1}{\alpha_j - \alpha_i}\Big[\alpha_j\mathbf{X}'_t\boldsymbol{\beta}_{Q,j} + (1 - \alpha_i)\mathbf{X}'_t\boldsymbol{\beta}_{Q,i}$$
$$+ \mathbb{1}(\mathbf{X}'_t\boldsymbol{\beta}_{Q,i} < Y_t < \mathbf{X}'_t\boldsymbol{\beta}_{Q,j})Y_t - \mathbb{1}(Y_t - \mathbf{X}'_t\boldsymbol{\beta}_{Q,i} > 0)\mathbf{X}'_t\boldsymbol{\beta}_{Q,i} - \mathbb{1}(Y_t - \mathbf{X}'_t\boldsymbol{\beta}_{Q,j} < 0)\mathbf{X}'_t\boldsymbol{\beta}_{Q,j}\Big]. \tag{9}$$

Using the definition of $I_t(\alpha_i, \alpha_j)$ and under the assumptions of Lemma 1 we immediately

find the following conditional moment restriction for $I_t(\alpha_i, \alpha_j)$:

$$E_t[g_{I,i,j}(\boldsymbol{\beta}_{0,Q,i}, \boldsymbol{\beta}_{0,Q,j}, \boldsymbol{\beta}_{0,I,i,j}, Y_t, \mathbf{X}_t)] = 0. \tag{10}$$

The moment conditions now follow by multiplying $g_{Q,i}(\cdot)$, $g_{L,i}(\cdot)$, $g_{I,i,j}(\cdot)$, and $g_{U,i}(\cdot)$ by a $\mathcal{F}_t$-measurable vector.

2.3 Estimation

We consider the joint estimation of many $Q_t(\alpha_i)$, $L_t(\alpha_i)$, $I_t(\alpha_i, \alpha_j)$, and $U_t(\alpha_i)$ parameter vectors. Let $0 < \alpha_1 < \alpha_2 < \ldots < \alpha_{d_Q} < 1$ denote the d_Q quantile levels necessary to find all quantities of interest. Since we do not necessarily consider all possible $L_t(\alpha_i)$, $I_t(\alpha_i, \alpha_j)$, and $U_t(\alpha_i)$ under the d_Q quantile levels, we must carefully specify which quantities we consider at which levels. To do so we introduce some notation concerning the particular choice of $L_t(\alpha_i)$, $I_t(\alpha_i, \alpha_j)$, and $U_t(\alpha_i)$. However, this extra notation can be skipped when we only consider estimation at a single quantile level. In Section 2.6 we treat the special case of LQE estimation at a single quantile level.

Let d_L denote the number of $L_t(\alpha_i)$ we consider, d_I the number of $I_t(\alpha_i, \alpha_j)$, and d_U the number of $U_t(\alpha_i)$, such that we consider $d = d_Q + d_U + d_I + d_U$ parameter vectors in total. Now let $\{\underline{n_L} \equiv n_{L,1}, n_{L,2}, \ldots, \overline{n_L} \equiv n_{L,d_L}\}$ denote a finite sequence of length d_L of unique integers on $[1, d_Q]$, such that we consider lower-quantile expectations $L_t(\alpha_{\underline{n_L}}), \ldots, L_t(\alpha_{\overline{n_L}})$. Similarly, let $\{\underline{n_U} \equiv n_{U,1}, n_{U,2}, \ldots, \overline{n_U} \equiv n_{U,d_U}\}$ denote a finite sequence of length d_U of unique integers on $[1, d_Q]$, such that we consider upper-quantile expectations $U_t(\alpha_{\underline{n_U}}), \ldots, U_t(\alpha_{\overline{n_U}})$. Finally, let $\{\{\underline{n_{I,1}}, \underline{n_{I,2}}\} \equiv \{n_{I,1,1}, n_{I,1,2}\}, \{n_{I,2,1}, n_{I,2,2}\} \ldots, \{\overline{n_{I,1}}, \overline{n_{I,2}}\} \equiv \{n_{I,d_I,1}, n_{I,d_I,2}\}\}$ denote a finite sequence of unique pairs on $[1, d_Q]^2$, with $n_{I,i,1} < n_{I,i,2}$, such that we consider interquantile expectations $I_t(\alpha_{\underline{n_{I,1}}}, \alpha_{\underline{n_{I,2}}}), \ldots, I_t(\alpha_{\overline{n_{I,1}}}, \alpha_{\overline{n_{I,2}}})$.

We can now stack all model parameters in the $(kd \times 1)$ parameter vector

$$\boldsymbol{\beta} = (\boldsymbol{\beta}'_{Q,1}, \ldots, \boldsymbol{\beta}'_{Q,d_Q}, \boldsymbol{\beta}'_{L,\underline{n_L}}, \ldots, \boldsymbol{\beta}'_{L,\overline{n_L}}, \boldsymbol{\beta}'_{I,\underline{n_{I,1}},\underline{n_{I,2}}}, \ldots, \boldsymbol{\beta}'_{I,\overline{n_{I,1}},\overline{n_{I,2}}}, \boldsymbol{\beta}'_{U,\underline{n_U}}, \ldots, \boldsymbol{\beta}'_{U,\overline{n_U}})'. \tag{11}$$

In what follows we denote the true parameter vector as $\boldsymbol{\beta}_0$ and the estimator as $\hat{\boldsymbol{\beta}}_T$. Subvectors of $\boldsymbol{\beta}$, $\boldsymbol{\beta}_0$ and $\hat{\boldsymbol{\beta}}_T$ are given equivalent notation. We will refer to the full estimator $\hat{\boldsymbol{\beta}}_T$ as the IQER estimator, whereas we will refer to the subvector $\hat{\boldsymbol{\beta}}_{Q,T} = (\boldsymbol{\beta}'_{Q,1}, \ldots, \boldsymbol{\beta}'_{Q,d_Q})'$ as the GMM quantile estimator.

Consider the $(kd \times 1)$ vector

$$\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t) = (g_{Q,1}, \dots, g_{Q,d}, g_{L,\underline{n_L}}, \dots, g_{L,\overline{n_L}}, g_{I,\underline{n_{I,1}},\underline{n_{I,2}}}, \dots, g_{I,\overline{n_{I,1}},\overline{n_{I,2}}}, g_{U,\underline{n_U}}, \dots, g_{U\overline{n_U}},)' \otimes \mathbf{X}_t, \tag{12}$$

with $\otimes$ denoting the Kronecker product operator and where we use abbreviated notation $g_{Q,i} = g_{Q,i}(\boldsymbol{\beta}_{Q,i}, Y_t, \mathbf{X}_t)$, $g_{L,i} = g_{L,i}(\boldsymbol{\beta}_{Q,i}, \boldsymbol{\beta}_{L,i}, Y_t, \mathbf{X}_t)$, $g_{I,i,j} = g_{I,i,j}(\boldsymbol{\beta}_{Q,i}, \boldsymbol{\beta}_{I,i,j}, Y_t, \mathbf{X}_t)$, and $g_{U,i} = g_{U,i}(\boldsymbol{\beta}_{Q,i}, \boldsymbol{\beta}_{U,i}, Y_t, \mathbf{X}_t)$.

Since $\mathbf{X}_t$ is a $\mathcal{F}_t$-measurable vector, we find the moment conditions

$$\mathbf{g}_0(\boldsymbol{\beta}_0) = E[\mathbf{g}(\boldsymbol{\beta}_0; Y_t, \mathbf{X}_t)] = \mathbf{0}, \tag{13}$$

such that we obtain the GMM-estimator $\hat{\boldsymbol{\beta}}_T$ defined as

$$\hat{\boldsymbol{\beta}}_T = \arg\min_{\boldsymbol{\beta} \in \mathbf{B}} \ \bar{\mathbf{g}}_T(\boldsymbol{\beta})' \hat{\mathbf{W}}_T^{-1} \bar{\mathbf{g}}_T(\boldsymbol{\beta}), \tag{14}$$

with sample moment function $\bar{\mathbf{g}}_T(\boldsymbol{\beta}) = \frac{1}{T} \sum_{t=1}^T \mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)$, and a consistent estimator $\hat{\mathbf{W}}_T$ of some asymptotic weighting matrix $\mathbf{W}_0$.

We proceed with the asymptotic properties of $\hat{\boldsymbol{\beta}}_T$. To prove the consistency of $\hat{\boldsymbol{\beta}}_T$, we must show that $\boldsymbol{\beta}_0$ is the unique vector at which (13) holds. The following lemma shows that uniqueness holds under mild conditions.

Lemma 2. *(Uniqueness) Assume that for every t, $1 \le t \le T$, (a) the conditional distribution of Y_t has finite first moments and the conditional density f_t is continuous and strictly positive; (b) $E[(F_t(\mathbf{X}_t' \boldsymbol{\beta}_{Q,i}) - \alpha_i)\mathbf{X}_t] \ne \mathbf{0}$, for all $\boldsymbol{\beta} \in \mathbf{B} \setminus \boldsymbol{\beta}_0$ and $i = 1, \dots, d_Q$; and (c) $E[\mathbf{X}_t \mathbf{X}_t']$ is nonsingular. Then $\boldsymbol{\beta}_0$ is unique.*

Assumption (a) is standard in quantile regression estimation and ensures Y_t will have unique quantiles. Assumption (b) ensures that uniqueness under the conditional moment condition implies uniqueness under the unconditional moment condition (13). In the iid case we should therefore include an intercept when all explanatory variables have mean zero. Assumption (c) ensures that $\mathbf{X}_t$ does not contain linearly dependent explanatory variables.

We show the consistency of the estimator in the following theorem.

Theorem 1. *(Consistency) Let the assumptions (a)-(c) hold. Furthermore assume that (d)*

$\{Y_t, \mathbf{X}_t\}$ *is stationary and ergodic sequence; (e)* $\mathbf{B}$ *is compact; (f)* $E\left[\|\mathbf{X}_t\|^2 + |Y_t|\|\mathbf{X}_t\|\right] < \infty$; *and (g)* $\hat{\mathbf{W}}_T \xrightarrow{p} \mathbf{W}_0$, *with* $\mathbf{W}_0$ *a positive semi-definite matrix. Then, as* $T \to \infty$, $\hat{\boldsymbol{\beta}}_T \xrightarrow{p} \boldsymbol{\beta}_0$.

We let the $\|\cdot\|$ operator denote to the L^2-norm. Assumption (d) precludes heterogeneity in the data, such as breaks and unconditional heteroskedasticity. Assumption (f) contains a boundedness condition on the unconditional mean of the moment conditions. We can ignore assumption (f) when $d_L = d_U = 0$, because it coincides with assumption (c) in this case. This follows from the realizations of Y_t that lie outside the conditional quantiles not entering in levels in $\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)$. Assumptions (e) and (g) are relatively standard assumptions for GMM estimation on the parameter space and the weighting matrix.

We provide the asymptotic distribution of the estimator in the following theorem.

Theorem 2. *(Asymptotic normality) Let assumptions (a)-(e) and (g) hold. Furthermore assume that (h) the conditional density of* Y_t, f_t, *is bounded; (i)* $\boldsymbol{\beta}_0$ *is an interior point of* $\mathbf{B}$; *(j)* $E\left[\|\mathbf{X}_t\|^4 + |Y_t|\|\mathbf{X}_t\|^3 + |Y_t|^2\|\mathbf{X}_t\|^2 + |Y_t|\|\mathbf{X}_t\|^2\right] < \infty$; *(k) for any* $\delta_T \to 0$, $\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0 \leq \delta_T\|} \sqrt{T} \times \|\bar{\mathbf{g}}_T(\boldsymbol{\beta}) - \bar{\mathbf{g}}_T(\boldsymbol{\beta}_0) - \mathbf{g}_0(\boldsymbol{\beta})\|/(1 + \sqrt{T}\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|) \xrightarrow{p} 0$; *(l)* $\boldsymbol{\Sigma}_0 \equiv E[\mathbf{g}_0(\boldsymbol{\beta}_0)\mathbf{g}_0(\boldsymbol{\beta}_0)']$ *is a positive definite matrix; and (m)* $E[f(\mathbf{X}_t'\boldsymbol{\beta}_{0,Q,i})\mathbf{X}_t\mathbf{X}_t']$ *is nonsingular for all* $i = 1, \dots, d$. *Then, as* $T \to \infty$, $\hat{\boldsymbol{\beta}}_T$ *is asymptotically normal,*

$$\boldsymbol{\Omega}_0^{-1/2}\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{I}), \tag{15}$$

with $\boldsymbol{\Omega}_0 = (\mathbf{G}_0'\mathbf{W}_0\mathbf{G}_0)^{-1}\mathbf{G}_0'\mathbf{W}_0\boldsymbol{\Sigma}_0\mathbf{W}_0\mathbf{G}_0(\mathbf{G}_0'\mathbf{W}_0\mathbf{G}_0)^{-1}$, *and* $\mathbf{G}_0$ *is a diagonal matrix with* d $(k \times k)$ *diagonal blocks*

$$\mathbf{G}_{0,ii} \equiv E \begin{cases} f_t(\mathbf{X}_t'\boldsymbol{\beta}_{0,Q,i})\mathbf{X}_t\mathbf{X}_t', & \text{if } i \leq d_Q, \\ \mathbf{X}_t\mathbf{X}_t', & \text{if } i > d_Q. \end{cases} \tag{16}$$

Assumption (j) imposes more stringent moment conditions as required by the central limit theory. When $d_L = d_U = 0$ we can replace assumption (j) with the substantially weaker assumption $E\|\mathbf{X}_t\|^4 < \infty$, such that we do not impose bounds on the moments of Y_t. Again, this follows from the realizations of Y_t that lie outside the conditional quantiles not entering in levels in $\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)$. In this scenario Y_t can therefore be Cauchy-distributed.

Additionally, when we study financial time series data, or fat-tailed data in general, we must be confident that assumption (j) holds. Imposing more structure on the data helps. For instance, Bai et al. (2003) provide conditions on GARCH processes with non-normal standard

errors such that the fourth moments of the process and its conditional volatility are bounded. These conditions amount to boundedness of the fourth moment of the standardized errors and some parameter restrictions. We can therefore use the results in Bai et al. (2003) to verify assumption (j) when Y_t is a GARCH process and $\mathbf{X}_t$ contains the conditional volatility.

We impose assumption (h) to ensure that the matrix $\mathbf{G}_0$ is bounded. Assumption (k) ensures stochastic equicontinuity of the nonsmooth sample moment condition. More primitive conditions for this property are difficult to obtain for the amount of dependence we allow in the data, i.e. asymptotical independence, because the sample moment conditions are unbounded, and partly non-Lipschitz continuous. Andrews (1994) provides more primitive conditions that are easily applied to our sample moment condition vector under the stronger condition of m-dependence, which covers the iid case. Assumption (i) and (l) are fairly standard for GMM estimation. Finally, we impose assumption (m) to obtain invertibility of the outmost factors of the asymptotic covariance matrix. Assumption (m) follows trivially from assumption (c) when the data is iid.

The first d_Q diagonal blocks of the matrix $\mathbf{G}_0$, which belong to $\boldsymbol{\beta}_{0,Q,i}$, contain an infinitely dimensional nuisance parameter in the form of conditional density f_t. Giacomini and Komunjer (2005) propose an approximation matrix that does not require knowledge of f_t. We apply their method to arrive at the block-diagonal approximation matrix $\hat{\mathbf{G}}_{T,\tau}$, with $(k \times k)$ diagonal blocks $(\hat{\mathbf{G}}_{T,\tau})_{ii}$, given by

$$(\hat{\mathbf{G}}_{T,\tau})_{ii} \equiv \begin{cases} \frac{1}{T}\sum_{t=1}^{T}[\frac{1}{\tau}\exp(\frac{1}{\tau}(Y_t - \mathbf{X}_t'\hat{\boldsymbol{\beta}}_{T,Q,i}))]\mathbb{1}\left(Y_t - \mathbf{X}_t'\hat{\boldsymbol{\beta}}_{T,Q,i}\right)\mathbf{X}_t\mathbf{X}_t', & \text{if } i \leq d_Q, \\ \frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t\mathbf{X}_t', & \text{if } i > d_Q. \end{cases} \tag{17}$$

The following theorem establishes convergence in probability of the components of the sample covariance matrix. Let $\hat{\boldsymbol{\Sigma}}_T \equiv \frac{1}{T}\sum_{t=1}^{T}\mathbf{g}(\hat{\boldsymbol{\beta}}_T; Y_t, \mathbf{X}_t)\mathbf{g}(\hat{\boldsymbol{\beta}}_T; Y_t, \mathbf{X}_t)'$.

Theorem 3. *(Asymptotic covariance) Under the assumptions of Theorem 2, $\hat{\boldsymbol{\Sigma}}_T \overset{p}{\to} \boldsymbol{\Sigma}_0$, and $\lim_{\tau \to 0}\hat{\mathbf{G}}_{T,\tau} \overset{p}{\to} \mathbf{G}_0$. Then, $\hat{\boldsymbol{\Omega}}_T \equiv \lim_{\tau \to 0}\left(\hat{\mathbf{G}}_{T,\tau}'\hat{\mathbf{W}}_T\hat{\mathbf{G}}_{T,\tau}\right)^{-1}\hat{\mathbf{G}}_{T,\tau}'\hat{\mathbf{W}}_T\hat{\boldsymbol{\Sigma}}_T\hat{\mathbf{W}}_T\hat{\mathbf{G}}_{T,\tau}\left(\hat{\mathbf{G}}_{T,\tau}'\hat{\mathbf{W}}_T\hat{\mathbf{G}}_{T,\tau}\right)^{-1} \overset{p}{\to} \boldsymbol{\Omega}_0$.*

We find the efficient covariance matrix $\boldsymbol{\Omega}_0 = \left(\mathbf{G}_0'\boldsymbol{\Sigma}_0^{-1}\mathbf{G}_0\right)^{-1}$ and its sample counterpart $\hat{\boldsymbol{\Omega}}_T = \left(\hat{\mathbf{G}}_{T,\tau}'\hat{\boldsymbol{\Sigma}}_T^{-1}\hat{\mathbf{G}}_{T,\tau}\right)^{-1}$ when we set $\mathbf{W}_0 = \boldsymbol{\Sigma}_0^{-1}$ and $\hat{\mathbf{W}}_T = \hat{\boldsymbol{\Sigma}}_T^{-1}$. Due to the invertibility of $\mathbf{G}_0$ by assumptions (c) and (m), however, we also find the efficient covariance matrix when we set $\mathbf{W}_0 = \mathbf{I}$, or any other positive definite weighting matrix. With this result we obtain an

efficient multi-stage estimator in the next subsection that is considerably less time-intensive to obtain. Moreover, for efficient estimators the particular block-diagonal structure of $\mathbf{G}_0$ implies that estimation error of $\hat{\boldsymbol{\beta}}_{T,Q,i}$ does not enter the estimation error of the $\hat{\boldsymbol{\beta}}_{T,L,i}$, and *vice versa*. There is also no estimation error spillover between parameter estimates at different levels α_i.

A test of linear null hypotheses of the form $H_0 : \mathbf{R}\boldsymbol{\beta}_0 = \mathbf{r}$ follows from the Wald test statistic

$$W_T = T(\mathbf{R}\hat{\boldsymbol{\beta}}_T - \mathbf{r})'(\mathbf{R}\hat{\boldsymbol{\Omega}}_T\mathbf{R}')^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}}_T - \mathbf{r}), \tag{18}$$

where $\mathbf{R}$ is a $(r \times n)$ matrix of full row rank, $\mathbf{r}$ is a vector of length r, and for which we reject the null hypothesis at a level q when $W_T > \chi^2_{2dk,1-q}$, where $\chi^2_{2dk,1-q}$ denotes the $(1-q)$ quantile of a χ^2-distribution with $2dk$ degrees of freedom.

2.4 Multi-stage estimation

When $\mathbf{W}_0 = \mathbf{I}$ we can restate optimization problem (14) as the multi-stage GMM problem

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_{T,Q,i} &= \underset{\boldsymbol{\beta}_{Q,i}\in\mathbf{B}_{Q,i}}{\arg\min}\ \bar{\mathbf{g}}'_{T,i}\bar{\mathbf{g}}_{T,i} \text{ for all } i = 1,\ldots,d_Q, \\
\hat{\boldsymbol{\beta}}_{T,L,i} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\tilde{Y}_{L,i} \text{ for all } i \in \{\underline{n_L},\ldots,\overline{n_L}\}, \\
\hat{\boldsymbol{\beta}}_{T,I,i,j} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\tilde{Y}_{I,i,j} \text{ for all } \{i,j\} \in \{\{\underline{n_{I,1}},\underline{n_{I,2}}\},\ldots,\{\overline{n_{I,1}},\overline{n_{I,2}}\}\}, \\
&\text{and} \\
\hat{\boldsymbol{\beta}}_{T,U,i} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\tilde{Y}_{U,i} \text{ for all } i \in \{\underline{n_U},\ldots,\overline{n_U}\},
\end{aligned}
\tag{19}
$$

where $\bar{\mathbf{g}}_{T,Q,i}$ refers to the ith element of $\bar{\mathbf{g}}_T$, $\mathbf{B}_{Q,i}$ denotes the part of domain $\mathbf{B}$ relating to $\boldsymbol{\beta}_{Q,i}$, $\mathbf{X} = [\mathbf{X}_1,\ldots,\mathbf{X}_T]'$, and $\tilde{Y}_{L,i}$, $\tilde{Y}_{I,i,j}$, and $\tilde{Y}_{U,i}$ denote $(T \times 1)$ auxiliary vectors with elements $\tilde{Y}_{L,i,t} = \mathbf{X}'_t\boldsymbol{\beta}_{Q,i} + \frac{1}{\alpha_i}\mathbb{1}(Y_t - \mathbf{X}'_t\boldsymbol{\beta}_{Q,i} < 0)(Y_t - \mathbf{X}'_t\boldsymbol{\beta}_{Q,i})$, $\tilde{Y}_{I,i,j,t} = \frac{1}{\alpha_j - \alpha_i}\Big[\alpha_j\mathbf{X}'_t\boldsymbol{\beta}_{Q,j} + (1-\alpha_i)\mathbf{X}'_t\boldsymbol{\beta}_{Q,i} +$ $\mathbb{1}(\mathbf{X}'_t\boldsymbol{\beta}_{Q,i} < Y_t < \mathbf{X}'_t\boldsymbol{\beta}_{Q,i})Y_t - \mathbb{1}(Y_t - \mathbf{X}'_t\boldsymbol{\beta}_{Q,i} > 0)\mathbf{X}'_t\boldsymbol{\beta}_{Q,i}) - \mathbb{1}(Y_t - \mathbf{X}'_t\boldsymbol{\beta}_{Q,j} < 0)\mathbf{X}'_t\boldsymbol{\beta}_{Q,j}\Big]$, and $\tilde{Y}_{U,i,t} = \mathbf{X}'_t\boldsymbol{\beta}_{Q,i} + \frac{1}{1-\alpha_i}\mathbb{1}(Y_t - \mathbf{X}'_t\boldsymbol{\beta}_{Q,i} > 0)(Y_t - \mathbf{X}'_t\boldsymbol{\beta}_{Q,i})$.

The equivalence of (14) and (19) follows from the cross-products of the quadratic objective function in (14) being zero when $\mathbf{W}_0 = \mathbf{I}$, and the last $k(d - d_Q)$ elements of $\bar{\mathbf{g}}_T(\boldsymbol{\beta})$ being linear in the lower-, inter-, and upperquantile expectation parameter vectors, such that, given any quantile parameter vector, a closed-form solution exists that sets these elements exactly to zero. Moreover, the parameter estimates obtained from (19) will be asymptotically efficient as shown in the previous subsection.

When $d_L = d_U = 0$ the multi-stage estimator is related to the trimmed OLS estimator of Ruppert and Carroll (1980). However, their estimator first removes all observations of Y_t and X_t lying outside the conditional quantiles and performs OLS on the remaining sample, whereas our multi-stage estimator can be viewed as an auxiliary regression on the full sample, with explanatory variables $\mathbf{X}_t$ and dependent variable $\frac{1}{\alpha_j - \alpha_i}\Big[\alpha_j \mathbf{X}_t'\boldsymbol{\beta}_{Q,j} + (1-\alpha_i)\mathbf{X}_t'\boldsymbol{\beta}_{Q,i} + \mathbb{1}(\mathbf{X}_t'\boldsymbol{\beta}_{Q,i} < Y_t < \mathbf{X}_t'\boldsymbol{\beta}_{Q,j})Y_t - \mathbb{1}(Y_t - \mathbf{X}_t'\boldsymbol{\beta}_{Q,i} > 0)\mathbf{X}_t'\boldsymbol{\beta}_{Q,i}) - \mathbb{1}(Y_t - \mathbf{X}_t'\boldsymbol{\beta}_{Q,j} < 0)\mathbf{X}_t'\boldsymbol{\beta}_{Q,j}\Big]$.

2.5 Substitution with quantile regression

In practice the linear conditional quantile parameters are often estimated by quantile regression instead of the GMM estimation procedure provided in this paper. Quantile regression optimization routines are included in many statistical computing packages and these routines are fast, because the quantile regression optimization problem can be restated as a linear programming problem.

Here we show that when $\mathbf{W}_0 = \mathbf{I}$ and we consider iid data we can replace the GMM quantile estimator with the quantile regression estimator and retain the preceding asymptotic distribution of the IQER estimator, because this estimator is consistent and satisfies an *asymptotic first order condition*.

This asymptotic first order condition $\bar{g}_T(\hat{\boldsymbol{\beta}}_T)'\hat{\mathbf{W}}_T\bar{g}_T(\hat{\boldsymbol{\beta}}_T) \leq \inf_{\boldsymbol{\beta}\in\mathbf{B}}\bar{\mathbf{g}}_T(\boldsymbol{\beta})'\hat{\mathbf{W}}_T\bar{\mathbf{g}}_T(\boldsymbol{\beta}) + o_P(T^{-1})$ ensures that the GMM objective function is optimal asymptotically at the estimator $\hat{\boldsymbol{\beta}}_T$, see Newey and McFadden (1994, p. 2187). This condition holds trivially for the GMM estimator proposed in this paper. Lemma A.1 and A.2 in Ruppert and Carroll (1980) show that the quantile regression estimator satisifies $\sqrt{T}\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{Q,i}(\hat{\boldsymbol{\beta}}_{T,KB,i};Y_t,\mathbf{X}_t)\otimes\mathbf{X}_t \xrightarrow{p} \mathbf{0}$, for all $i = 1,\ldots,d$, where $\hat{\boldsymbol{\beta}}_{T,KB,i}$ denotes the quantile regression estimator at quantile level α_i. In combination with the IQE moment conditions being exactly equal to zero this implies that the asymptotic first order condition is satisfied as well for the estimator that replaces the GMM quantile estimator with the quantile regression estimator in multi-stage optimization problem (19). Finally, Koenker and Bassett (1978) show the consistency of the quantile regression estimator for iid data.

For more general data settings the GMM quantile estimator provided in this paper can be used. The quantile regression estimator can then provide a valuable starting point in the optimization procedure.

2.6 LQE estimation at a single quantile level

When we are only interested in the LQE at a single quantile level, α_1, the notation simplifies considerably. Let $\boldsymbol{\beta}_0 = (\boldsymbol{\beta}'_{0,Q,1}, \boldsymbol{\beta}'_{0,L,1})'$ denote the true parameter vector relating to $Q_t(\alpha_1)$ and $L_t(\alpha_1)$. We can estimate $\hat{\boldsymbol{\beta}}_T = (\hat{\boldsymbol{\beta}}'_{T,Q,1}, \hat{\boldsymbol{\beta}}'_{T,L,1})'$ as in problem (14) or multi-stage problem (19), with $\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t) = \big(g_{Q,1}(\boldsymbol{\beta}_{Q,1}, Y_t, \mathbf{X}_t), g_{L,1}(\boldsymbol{\beta}_{Q,1}, \boldsymbol{\beta}_{L,1}, Y_t, \mathbf{X}_t)\big)' \otimes \mathbf{X}_t$. Our estimator $\hat{\boldsymbol{\beta}}_T$ converges in distribution as in Theorem 2, and we obtain an estimate of the asymptotic covariance matrix as in Theorem 3, with

$$
\hat{\mathbf{G}}_{T,\tau} = \begin{bmatrix} \frac{1}{T}\sum_{t=1}^T [\frac{1}{\tau}\exp(\frac{1}{\tau}(Y_t - \mathbf{X}'_t\hat{\beta}_{T,Q,1}))]\mathbb{1}\left(Y_t - \mathbf{X}'_t\hat{\beta}_{T,Q,1}\right)\mathbf{X}_t\mathbf{X}'_t, & \mathbf{0} \\ \mathbf{0} & \frac{1}{T}\sum_{t=1}^T \mathbf{X}_t\mathbf{X}'_t \end{bmatrix}. \tag{20}
$$

3 Simulation experiment

We conduct a Monte Carlo experiment to study the small sample properties of the IQER estimator. We will consider LQEs at tail quantile levels $\alpha = 0.01, 0.025, 0.05, 0.10$, and IQEs between quantile levels 0.10 and 0.20, 0.45 and 0.55, 0.10 and 0.90, and 0.01 and 0.99. We do not consider UQEs, since these have equivalent properties to LQEs for symmetric distributions. Using these quantile levels we can evaluate whether the asymptotic distribution provides a good approximation to the small sample distribution of the estimator, and how this approximation worsens as we consider more extreme tail quantile levels or smaller intervals between quantile levels.

We consider a data generating process that allows for conditional heteroskedasticity as well as dependence in the error term of the dependent variable and explanatory variables. It is defined as follows:

$$
\begin{aligned}
Y_t &= \gamma_0 + \gamma_1 X_t + \frac{1}{\sqrt{1+\gamma_2^2}}(1 + \gamma_2 X_t)\varepsilon_t, \\
X_t &= \phi_X X_{t-1} + \nu_t, \\
\varepsilon_t &= \phi_\varepsilon \varepsilon_{t-1} + \zeta_t, \\
\nu_t &\sim \text{iid } \mathcal{N}\left(0, (1 - \phi_X^2)\right), \\
\zeta_t &\sim \text{iid } \mathcal{N}\left(0, (1 - \phi_\varepsilon^2)\right), \\
\text{cov}(\epsilon_t, X_t) &= 0.
\end{aligned} \tag{21}
$$

We choose this data specification, because the regressor X_t can have a nonlinear effect on Y_t

by affecting its volatility. As a direct consequence of this the quantile and IQE parameters will differ across quantile levels. Moreover, for the normal distribution and several other distributions the true values of these parameters are known in closed form.

The data generating process allows for persistence in X_t and ε_t by setting non-zero AR(1) parameters ϕ_X and ϕ_ε. Especially in macroeconomic and financial applications this setting is important because explanatory variables are usually highly persistent in these scenarios, i.e. ϕ_X is close to one. Values of ϕ_ε that are different from zero indicate that the model specification is not able to correct for all autocorrelation in Y_t. We assume that the researcher is able to propose a model that explains most of the autocorrelation in Y_t, such that we consider values of ϕ_ε that are close to zero. Through the autocorrelation in X_t and ε_t the dependent variable Y_t will be autocorrelated as well. Y_t is stationary when $|\phi_X| < 1$ and $|\phi_\varepsilon| < 1$. The iid case follows when $\phi_X = \phi_\varepsilon = 0$. Moreover, the variances of ν_t and ζ_t, and the division by the factor $\sqrt{(1 + \gamma_2^2)}$ ensure that the unconditional variances of X_t, ε_t, and the errors $Y_t - \gamma_0 - \gamma_1 X_t$ equal one, such that the signal-to-noise ratio is equivalent across parameter settings in which γ_0 and γ_1 are fixed.

For any two quantile level α_i, α_j, $0 < \alpha_i < \alpha_j < 1$, the DGP in (21) implies the following values for the quantile, LQE, and IQE parameters:

$$\beta_{Q,i,1} = \gamma_0 + \Phi^{-1}(\alpha_i), \tag{22}$$

$$\beta_{Q,i,2} = \gamma_1 + \frac{\gamma_2}{\sqrt{1 + \gamma_2^2}} \Phi^{-1}(\alpha_i), \tag{23}$$

$$\beta_{L,i,1} = \gamma_0 - \frac{\phi(\Phi^{-1}(\alpha_i))}{\alpha_i}, \tag{24}$$

$$\beta_{L,i,2} = \gamma_1 - \frac{\gamma_2}{\sqrt{1 + \gamma_2^2}} \frac{\phi(\Phi^{-1}(\alpha_i))}{\alpha_i}, \tag{25}$$

$$\beta_{I,i,j,1} = \gamma_0 - \frac{\phi(\Phi^{-1}(\alpha_i) - \phi(\Phi^{-1}(\alpha_j))}{\alpha_j - \alpha_i}, \tag{26}$$

$$\beta_{I,i,j,2} = \gamma_1 - \frac{\gamma_2}{\sqrt{1 + \gamma_2^2}} \frac{\phi(\Phi^{-1}(\alpha_i) - \phi(\Phi^{-1}(\alpha_j))}{\alpha_j - \alpha_i}, \tag{27}$$

where Φ^{-1} and ϕ denote the inverse cdf and the density of the standard normal distribution, respectively.

We consider three relations between Y_t and X_t. Firstly, we let Y_t be independent of X_t by setting $\gamma_0 = \gamma_1 = \gamma_2 = 0$. Secondly, we let X_t influence the volatility of Y_t by setting $\gamma_1 = 0.25$ and $\gamma_0 = \gamma_2 = 0$. Finally, we let X_t influence the mean and volatility of Y_t by setting $\gamma_1 = \gamma_2 = 0.25$

and $\gamma_0 = 0$. We pick these parameter values because they impose a moderate effect of X_t on Y_t by affecting the mean and volatility. However, experiments with other parameter values show we can pick any γ_1 parameter value for similar results. The parameter γ_2 must be chosen with more care because we find that values above 0.75 introduce bias into the estimation of the quantiles and subsequently the LQE estimates, even at larger sample sizes. Essentially the variance of $\gamma_2 X_t$ must be relatively small with respect to the error variance.

With regard to stationary GARCH processes (Bollerslev, 1987) we verify that sufficiently persistent conditional volatilities, i.e. where the GARCH coefficient has a value larger than 0.80, we do not encounter this bias problem in the estimation procedure. This covers the range of values we typically find for financial time series.

We study sample sizes $T = 50, 100, 250, 500, 1000, 2500$, such that we can draw conclusions for very small data sets, e.g. yearly macroeconomic data collected since the 1960s, to larger data sets, e.g. daily financial returns collected over a period of 10 years. Finally, we choose tuning parameters $\tau = 0.02, 0.05, 0.10, 0.20, 0.50$ for the approximate gradient matrix $\hat{\mathbf{G}}_{T,\tau}$. Our results are based on 10,000 simulations.

The tables below provide results for the scenario where X_t influences the mean and volatility of Y_t, and where $\phi_X = 0.85$ and $\phi_\varepsilon = 0$. The results for the iid scenario and the scenario X_t only influences the volatility of Y_t are similar, such that we do not show them here. In the final paragraphs of this Section we discuss several data settings we consider to study robustness of these results. These scenarios include the presence of autocorrelation in the error, and standardized t-distributed regressors and errors.

Table 1 provides descriptive statistics of the LQE and IQE parameter estimates that we have estimated for the simulations at sample sizes $T = 50, 100, 1000$. We first discuss the LQE parameter estimates at tail quantiles levels $\alpha = 0.01, 0.025, 0.05, 0.10$. We observe that the estimates have positive bias in small samples, which is larger at more extreme tail levels. The LQE parameter estimates are biased toward the respective quantile parameter estimates, and when the sample size does not contain any observations that exceed the quantile, e.g. $T = 50$ for $\alpha = 0.01$, or few observations, e.g. $T = 50$ for $\alpha = 0.025$, the LQE parameter estimates are respectively exactly equal or close to the quantile parameter estimates. This is a direct consequence of the definition of $g_{L,i}(\cdot)$. This bias vanishes as the sample size increases and we obtain more information on realizations below the quantile. We see that the bias is mostly negligible at $T = 1000$, even for the extreme tail quantile levels. Dispersion of the parameter

17

estimates, as described by the standard deviation and 10%-90% range, is larger for more extreme tail quantile levels and reduces as the sample size increases. The dispersion of LQE parameter estimates is comparable to the dispersion of the respective quantile parameter estimates, as can be observed from Table 3 in the Appendix.

The descriptive statistics of the IQE parameter estimates show that the bias is negligible for all considered sample sizes and across all quantile level combinations that we consider. Moreover, the dispersion of the IQE parameter estimates is substantially smaller than those of the LQE parameter estimates. We see that the dispersion of the LQE and IQEs on equally-sized quantile intervals decreases as we move towards the center of the distribution. This can be observed from the LQE at $\alpha = 0.10$ and the IQEs lying between quantile intervals of size 0.10. Moreover, the dispersion of IQE parameter estimates decreases as the quantile interval on which the IQE lies increases. The distinctions weaken as the sample size increases.

[Table 1 about here.]

Table 2 provides the coverage rates of the standard errors of the LQE and IQE parameter estimate that we have estimated at sample sizes $T = 50, 100, 250, 500, 1000, 2500$. The coverage rates are calculated as the percentage of simulations for which the true parameter value lies within the 95% confidence interval based on the estimated parameters and asymptotic covariance matrix.

We observe that the coverage rates of the LQE parameter estimates converge to their nominal values when the sample size increases. For the LQEs at tail quantile levels $\alpha = 0.01, 0.025$ the coverage rates are around 90% at sample size $T = 2500$. This suggests that we need quite large sample sizes to obtain accurate standard errors of LQEs at extreme tail quantile levels. Moreover, the coverage rates are inaccurate when the sample has no observations below a certain quantile level, i.e. when $T = 50$ for $\alpha = 0.01$.

Table 4 in the Appendix presents the coverage rates of the GMM quantile estimator. Coverage rates based on the true asymptotic covariance matrix are provided in brackets, whereas the approximated asymptotic covariance matrix based on Theorem 3 are provided outside brackets. The true asymptotic covariance matrix assumes knowledge of the specification of the conditional density function. The approximated asymptotic covariance matrix results are those calculated with tuning parameter value $\tau = 0.05$, which is the best performing value for $T = 2500$. We observe that the *true* coverage rates of the quantile parameter estimates converge more quickly to

18

their nominal values than the LQE parameter estimates at equivalent quantile leves. However, we also find evidence of the nuisance parameter problem, since the *approximated* coverage rates do not agree precisely with the nominal coverage rates at $T = 2500$.

Coverage rates of the IQE parameter estimates converge quickly to their nominal values. Sample sizes of $T = 250$ are generally sufficient. We observe that coverage rates converge more quickly for IQEs on quantile intervals closer to the center of the distribution.

[Table 2 about here.]

We consider several alternative data settings to study the robustness of these results. First, we allow for some persistence in the errors ε_t by setting $\phi_\varepsilon = 0.10$, to capture residual auto-correlation in Y_t that is not be explained by X_t. As such, we assume that the model defined by the researcher is able to capture a substantial amount of the autocorrelation in Y_t, but not all. Second, we consider regressors and errors that are independently standardized t-distributed with four and ten degrees of freedom. As such, the regressors and errors have fat tails, which increases the probability of extreme events. We study these cases because fat tails are commonly found in many time series. The four degrees of freedom case concerns the fattest tails that are allowed by our theory. This minimum is imposed by assumption (j) of Theorem 2, because a t-distributed random variable has as many bounded moments as it has degrees of freedom. The error ε_t is allowed to have fatter tails, when we only consider IQE parameters.

We briefly summarize the results for the LQE and IQE estimators. Setting $\phi_\varepsilon = 0.10$ instead of $\phi_\varepsilon = 0$ does not meaningfully change our results in terms of bias or dispersion of parameter estimates in both experiments. Coverage rates are not affected. When we compare t-distributed and normally distributed regressors and errors we find that the standard normal and standardized t with ten degrees of freedom cases are similar. Bias and dispersion are slightly worse for the LQE parameter estimates, and coverage rates converge to nominal values at a similar pace for both experiments. When we compare the normal and standardized t with four degrees of freedom case we find larger bias in LQE parameters estimates, larger dispersion for all parameter estimates, and coverage rates that are about five percentage points lower for the LQE parameter estimates.

4   Empirical illustration

In finance LQE is known as expected shortfall. There exists a well-documented relation between volatility and expected shortfall, see for instance Nadarajah et al. (2014) for an overview. In time-varying settings estimates of one-step-ahead expected shortfall that consider volatility are generally obtained as a function of a previously estimated one-step-ahead volatility estimate. The IQER estimator obviates this two-stage approach, since we can estimate one-step-ahead expected shortfall directly from current and prior period volatility, such that we circumvent the intermediate step of estimating one-step-ahead volatility.

We study the differences between our direct estimator and the estimators implied from using the previously obtain one-step-ahead volatility estimate and imposing the standard normal distribution, the standardized t distribution, or the empirical distribution function (EDF) to describe the distribution of the standardized errors of the model. We use realized volatility as an estimate of volatility, which is a volatility estimator based on intra-day data that has become prevalent in finance. Andersen et al. (2006) provide an overview of realized volatility and other volatility estimators, and their relation to expected shortfall forecasting.

We study the daily returns of the Spyder exchange traded fund, which tracks the S&P 500 index, and its five-minute realized volatility estimate. We obtain the data from the Oxford-Man Institute of Quantitative Finance Realized Library for the period February 2000 to 5 December 2016 for a total of 4,137 daily observations.[1] We consider open-to-close returns, because realized volatilities are estimated for this intraday-period.

We study quantile level vector $\alpha = (0.025, 0.05, 0.10)'$ to cover frequently used quantile levels in financial risk management, and to assess the impact of estimation error when moving further into the left tail of the distribution.

We consider the simple conditionally heteroskedastic model

$$Y_{t+1} = \sigma_{t+1}\varepsilon_{t+1}, \tag{28}$$

where Y_{t+1} denotes the Spyder daily log return, σ_{t+1} denotes the conditional volatility of Y_{t+1}, and ε_{t+1} denotes the error term at time $t+1$ with mean zero and unit variance. Moreover, under certain conditions it follows that realized volatility is a good estimators of σ_t, i.e. $\sigma_t \approx RV_t$,

---

[1]The data is available at `http://realized.oxford-man.ox.ac.uk/data/download`. We remove trading days with unusual trading hours.

where RV_t denotes the daily realized volatility at time t, calculated from five-minute returns.

Corsi (2009) proposes the HAR model to estimate the one-step-ahead realized volatility based on currently observed realized volatilities. The HAR model is a linear model using this period's realized volatility, the average of the realized volatilities of the last five periods, and the average of the realized volatilities of the last 25 periods to cover daily, weekly, and monthly effects.

The HAR model estimate of one-step-ahead volatility is given by

$$\widehat{RV}_{t+1} = \hat{\phi}_1 + \hat{\phi}_2 RV_t + \hat{\phi}_3 RV_t^{(w)} + \hat{\phi}_4 RV_t^{(m)}, \tag{29}$$

where $RV_t^{(w)} = 1/5 \sum_{i=1}^{5} RV_{t-i+1}$ is the weekly aggregated realized volatility, and $RV_t^{(m)} = 1/25 \sum_{i=1}^{25} RV_{t-i+1}$ is the monthly aggregated realized volatility.

This linear structure allows us to compare a one-step-ahead expected shortfall obtained with the IQER estimator, given by

$$\hat{L}_{t+1} = \hat{\beta}_{L,1} + \hat{\beta}_{L,2} RV_t + \hat{\beta}_{L,3} RV_t^{(w)} + \hat{\beta}_{L,4} RV_t^{(m)}, \tag{30}$$

with estimates based on the two-stage approach, where we first estimate the HAR model and subsequently estimate one-step-ahead expected shortfall as

$$\hat{L}_{t+1}(\alpha_i) = \hat{h}_i \widehat{RV}_{t+1} = \hat{h}_i \left(\hat{\phi}_1 + \hat{\phi}_2 RV_t + \hat{\phi}_3 RV_t^{(w)} + \hat{\phi}_4 RV_t^{(m)} \right), \tag{31}$$

where $\hat{h}_i$ denotes an LQE estimate at quantile level α_i of the standardized errors $\varepsilon_{t+1} = Y_{t+1}/RV_{t+1}$. For instance, when ε_{t+1} is standard normally distributed $\hat{h}_i = -\frac{\phi(\Phi^{-1}(\alpha_i))}{\alpha_i}$, whereas if $\hat{\varepsilon}_{t+1}$ is standardized t-distributed $\hat{h}_i = -\sqrt{\frac{\hat{\nu}-2}{\hat{\nu}}} \frac{g_{\hat{\nu}}(t_{\hat{\nu}}^{-1}(\alpha_i))}{\alpha_i} \frac{\hat{\nu}+(t_{\hat{\nu}}^{-1}(\alpha_i))^2}{\hat{\nu}-1}$, as shown in McNeil et al. (2015, p. 71), with $\hat{\nu}$ denoting the degrees of freedom estimate, and t_ν and g_ν denoting the cumulative distribution function and density function of a t-distribution. We also derive $\hat{h}_i$ nonparametrically from the EDF function. We can therefore compare the IQER estimator coefficient $\hat{\beta}_{L,i,j}$ to $\hat{h}_i \hat{\phi}_j$, for all $i = 1, \ldots, 3$ and $j = 1, \ldots, 4$.

Figure 1 provides plots of the point estimates of the IQER estimator coefficients in blue, their 95% confidence intervals in red, and the point estimates obtained from the two-stage procedure with the standard normal distribution and EDF in green and black, respectively. When we impose a standardized t-distribution on the standardized residuals we find a very

21

large degrees of freedom estimate, such that expected shortfall estimates based on the standard normal and standardized t distribution coincide. We therefore only report expected shortfall estimates based on the standard normal distribution and the EDF.

Generally, we observe that the estimates based on the EDF and standard normal distribution are largely similar, with the EDF being slightly more negative. Our IQER estimates differs from the other two estimates in several ways.

First, the intercept is more negative for the IQER estimates, which suggests that they are more conservative. Indeed, in annualized percentages the expected shortfall estimates obtained with the IQER estimator are on average about 0.5 percentage points lower than the standard normal and EDF expected shortfall estimates. This difference vanishes when we remove the intercept from the models.

The coefficients of the daily realized volatility show a similar pattern, although the IQER estimates are lower. The IQER estimate is significantly different from zero and differences between the IQER estimate and the two-stages estimates are insignificant.

The weekly aggregated realized volatility coefficients are similar as well, although the IQER estimates are again lower. Again, the IQER estimates are significantly different from zero and differences between IQER estimates and the two-stage estimates are insignificant.

The monthly aggregated realized volatility coefficients are more distinct. The IQER estimate is close to zero, whereas the two-stage estimates suggest a negative coefficient. The IQER confidence interval suggests that the LQE coefficients are not significantly different from zero and differences between IQER estimates and the two-stage estimates are insignificant. We can draw the conclusion that the monthly coefficient is insignificant using the IQER estimator, as the standard errors follow directly from the estimation. Obtaining the standard errors of the two-stage estimators is less straightforward, because we must take into account the two-stage estimation procedure and potential spillover of the estimation error from the first stage into the second stage.

In short, using the IQER estimator we show that the monthly aggregated realized volatility estimate does not contain explanatory power for one-step-ahead daily expected shortfall for daily log returns of the Spyder fund, whereas IQER coefficients are significantly different from zero and lower for daily realized volatility and the weekly aggregated volatility. As such, one-step-ahead expected shortfall point estimates that are obtained with the IQER estimator react more forcefully to recently observed realized volatility estimates.

[Figure 1 about here.]

5   Concluding remarks

We propose a semiparametric estimator of the parameters in linear models of conditional LQE, IQE, and UQE. Our so-called IQER estimator is based on the GMM framework of Hansen (1996) where we use the strictly consistent scoring function for the quantile and LQE pair introduced by Fissler et al. (2016b) to obtain moment conditions. We provide conditions under which the estimator is consistent and derive its asymptotic distribution, and a consistent estimator of the asymptotic covariance matrix. We propose a convenient multi-stage optimization method that results in the efficient estimator and which is computationally fast. Our theory applies to stationary and ergodic time series.

In a simulation study we show that the asymptotic theory provides an accurate approximation in small samples for data settings that include autocorrelation in the regressors, conditional heteroskedasticity, and fat tails.

We provide an empirical illustration of our theory in finance, in which we estimate one-step-ahead daily tail-quantile LQE, or expected shortfall, of a S&P 500 exchange traded fund return by IQER estimation conditional on previously observed daily, weekly, and monthly aggregated realized volatility estimates of Corsi (2009) and compare this to commonly used two-stage models, in which we first estimate one-step-ahead realized volatility and then use an estimate of the distribution of the standardized errors separately to map the next period realized volatility to a next period expected shortfall. We find that the IQER point estimates differ from the other methods.

Several extensions to the theory in this paper come to mind. First, an LQE specification testing framework can be developed that in part uses our asymptotic theory. Second, we can extend our theory to include nonlinear specifications of IQE models.

23

References

Adrian, T. and Brunnermeier, M. K. (2016). CoVaR. *American Economic Review*, 106(7):1705–1741.

Andersen, T. G., Bollerslev, T., Christoffersen, P. F., and Diebold, F. X. (2006). Chapter 15 Volatility and Correlation Forecasting. volume 1 of *Handbook of Economic Forecasting*, pages 777–878. Elsevier.

Andrews, D. W. (1994). Chapter 37 Empirical Process Methods in Econometrics. volume 4 of *Handbook of Econometrics*, pages 2247–2294. Elsevier.

Bai, X., Russell, J. R., and Tiao, G. C. (2003). Kurtosis of GARCH and Stochastic Volatility Models with Non-Normal Innovations. *Journal of Econometrics*, 114(2):349–360.

Bollerslev, T. (1987). A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return. *The Review of Economics and Statistics*, 69(3):542–547.

Chernozhukov, V., Fernández-Val, I., and Melly, B. (2013). Inference on Counterfactual Distributions. *Econometrica*, 81(6):2205–2268.

Corsi, F. (2009). A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics*, 7(2):174–196.

Fissler, T., Fasciati-Ziegel, J., and Gneiting, T. (2016a). Expected Shortfall is Jointly Elicitable with Value at Risk - Implications for backtesting. *Risk Magazine*.

Fissler, T., Ziegel, J. F., et al. (2016b). Higher Order Elicitability and Osbands Principle. *The Annals of Statistics*, 44(4):1680–1707.

Giacomini, R. and Komunjer, I. (2005). Evaluation and Combination of Conditional Quantile Forecasts. *Journal of Business & Economic Statistics*, 23(4):416–431.

Gneiting, T. (2011). Making and Evaluating Point Forecasts. *Journal of the American Statistical Association*, 106(494):746–762.

Gneiting, T. and Raftery, A. E. (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, 102(477):359–378.

Hansen, B. E. (1996). Stochastic Equicontinuity for Unbounded Dependent Heterogeneous Arrays. *Econometric Theory*, 12(2):347–359.

Koenker, R. and Bassett, G. (1978). Regression Quantiles. *Econometrica*, 46(1):33–50.

Koenker, R. and Hallock, K. F. (2001). Quantile Regression. *Journal of Economic Perspectives*, 15(4):143–156.

McNeil, A. J., Frey, R., and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools.* Princeton University Press, Princeton, NJ.

Nadarajah, S., Zhang, B., and Chan, S. (2014). Estimation Methods for Expected Shortfall. *Quantitative Finance*, 14(2):271–291.

Newey, W. K. and McFadden, D. (1994). Chapter 36 Large Sample Estimation and Hypothesis Testing. volume 4 of *Handbook of Econometrics*, pages 2111–2245. Elsevier.

Newey, W. K. and Powell, J. L. (1987). Asymmetric Least Squares Estimation and Testing. *Econometrica*, 55(4):819–847.

Ruppert, D. and Carroll, R. J. (1980). Trimmed Least Squares Estimation in the Linear Model. *Journal of the American Statistical Association*, 75(372):828–838.

Taylor, J. W. (2008a). Estimating Value at Risk and Expected Shortfall Using Expectiles. *Journal of Financial Econometrics*, 6(2):231–252.

Taylor, J. W. (2008b). Using Exponentially Weighted Quantile Regression to Estimate Value at Risk and Expected Shortfall. *Journal of Financial Econometrics*, 6(3):382–406.

White, H. (1996). *Estimation, Inference and Specification Analysis.* Cambridge University Press, Cambridge, UK.

White, H. (2001). *Asymptotic Theory for Econometricians.* Academic Press, Cambridge, MA.

A Additional results

Lemma A.1. *Let assumptions (a), (c), (d), and (j) hold. Then, as $T \to \infty$, the sample covariance matrix $\hat{\boldsymbol{\Sigma}}_T(\boldsymbol{\beta}) \equiv \frac{1}{T}\sum_{t=1}^{T} \mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)' \xrightarrow{p} E[\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)'].$*

Proof: We apply a LLN for stationary and ergodic sequences to $\hat{\boldsymbol{\Sigma}}_T(\boldsymbol{\beta})$ to establish that for every $\boldsymbol{\beta} \in \mathbf{B}$ $\hat{\boldsymbol{\Sigma}}_T(\boldsymbol{\beta}) \xrightarrow{p} E[\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)']$. It suffices to show that the assumptions of Theorem 3.34 in White (2001) hold.

The function $\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)'$ is measurable. Moreover, by assumption (d) $\{Y_t, \mathbf{X}_t\}$ is a stationary and ergodic sequence. Therefore, $\{\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)'\}$ is stationary and ergodic by Theorem 3.35 of White (2001).

We can now apply Theorem 3.34 in White (2001) to all elements of $\{\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)'\}$ to obtain the desired result if $E\|\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)'\|_\infty < \infty$. The operator $\|\cdot\|_\infty$ denotes the (matrix) maximum norm, i.e. for a given matrix $\mathbf{A}$ with ijth element $A_{(i,j)}$, $\|\mathbf{A}\|_\infty = \max_{i,j}|A_{(i,j)}|$.

By norm equivalence,

$$\|\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)'\|_\infty \le c_2 \cdot \|\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)\|^2 \tag{32}$$

$$\le c_3 \cdot \left((1 + \|\mathbf{X}_t\| + |Y_t|) \cdot \|\mathbf{X}_t\| \right)^2 \text{ a.s.-}P, \tag{33}$$

for finite constants c_2, and c_3. By assumption (j) we then have the required result

$$E\|\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)'\|_\infty \le c_2 \cdot E\left[\max\left\{ 1\,,\, \left[(1 + \|\mathbf{X}_t\| + |Y_t|) \cdot \|\mathbf{X}_t\| \right]^2 \right\} \right] \tag{34}$$

$$\le c_3 \cdot \left(1 + E\left\{ (1 + \|\mathbf{X}_t\| + |Y_t|) \cdot \|\mathbf{X}_t\| \right\}^2 \right) \tag{35}$$

$$\le c_4 \cdot E\left[\|\mathbf{X}_t\|^4 + |Y_t|\|\mathbf{X}_t\|^3 + |Y_t|^2\|\mathbf{X}_t\|^2 + |Y_t|\|\mathbf{X}_t\|^2 \right] \tag{36}$$

$$< \infty, \tag{37}$$

for some finite constant c_4. The third inequality follows from expansion and noting that $E\|\mathbf{X}_t\|^2 < E\|\mathbf{X}_t\|^3 < E\|\mathbf{X}_t\|^4$ by Hölder's inequality.

It follows that $\hat{\boldsymbol{\Sigma}}_T(\boldsymbol{\beta}) \xrightarrow{p} E[\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)']$, and for some consistent estimate $\hat{\boldsymbol{\beta}}_T$ of $\boldsymbol{\beta}_0$ specifically, $\hat{\boldsymbol{\Sigma}}_T(\hat{\boldsymbol{\beta}}_T) \xrightarrow{p} \boldsymbol{\Sigma}_0$. $\qquad\square$

B Proofs

B.1 Proof of Lemma 1.

Under the assumption that F_t is continuously differentiable we find that

$$\nabla_{\boldsymbol{\beta}_V} E_t \left[\mathbb{1}(Y_t - \mathbf{X}_t'\boldsymbol{\beta}_V < 0)(Y_t - \mathbf{X}_t'\boldsymbol{\beta}_V) \right] = -E_t[\mathbb{1}(Y_t - \mathbf{X}_t'\boldsymbol{\beta}_V < 0)\mathbf{X}_t], \tag{38}$$

as shown in the proof of Lemma 1 in Giacomini and Komunjer (2005).

We find that the respective parts of the gradient of the conditional mean of the loss function $\nabla_{\boldsymbol{\beta}} E_t[S(Y_t, \mathbf{X}_t\boldsymbol{\beta}_{0,Q,i}, \mathbf{X}_t\boldsymbol{\beta}_{0,L,i})]$ are as follows.

$$\nabla_{\boldsymbol{\beta}_{Q,i}} E_t[S(Y_t, \mathbf{X}_t\boldsymbol{\beta}_{Q,i}, \mathbf{X}_t\boldsymbol{\beta}_{L,i})]$$

$$= \nabla_{\boldsymbol{\beta}_{Q,i}} E_t[(\alpha_i - \mathbb{1}(Y_t - \mathbf{X}_t'\boldsymbol{\beta}_{Q,i} < 0))(Y_t - \mathbf{X}_t'\boldsymbol{\beta}_{Q,i})]$$

$$- \frac{1}{\alpha_i} \exp\left(\mathbf{X}_t'\boldsymbol{\beta}_{L,i}\right) \nabla_{\boldsymbol{\beta}_{Q,i}} E_t[\mathbb{1}(Y_t - \mathbf{X}_t'\boldsymbol{\beta}_{Q,i} < 0)(Y_t - \mathbf{X}_t'\boldsymbol{\beta}_{Q,i})] - \exp\left(\mathbf{X}_t'\boldsymbol{\beta}_{L,i}\right) \nabla_{\boldsymbol{\beta}_{Q,i}} \mathbf{X}_t'\boldsymbol{\beta}_{Q,i}$$

$$= E_t\Big(\frac{1}{\alpha_i} \exp(\mathbf{X}_t'\boldsymbol{\beta}_{L,i}) \mathbb{1}(Y_t - \mathbf{X}_t'\boldsymbol{\beta}_{Q,i} < 0) - \exp\left(\mathbf{X}_t'\boldsymbol{\beta}_{L,i}\right) - (\alpha_i - \mathbb{1}(Y_t - \mathbf{X}_t'\boldsymbol{\beta}_{Q,i} < 0)) \Big) \mathbf{X}_t \tag{39}$$

$$= \Big(1 + \frac{1}{\alpha_i} \exp(\mathbf{X}_t'\boldsymbol{\beta}_{L,i}) \Big) E_t\Big(\mathbb{1}(Y_t - \mathbf{X}_t'\boldsymbol{\beta}_{Q,i} < 0) - \alpha_i \Big) \mathbf{X}_t, \tag{40}$$

and,

$$\nabla_{\boldsymbol{\beta}_{L,i}} E_t[S(Y_t, \mathbf{X}_t\boldsymbol{\beta}_{Q,i}, \mathbf{X}_t\boldsymbol{\beta}_{L,i})]$$

$$= \nabla_{\boldsymbol{\beta}_{L,i}} \Big(-\frac{1}{\alpha_i} \exp\left(\mathbf{X}_t'\boldsymbol{\beta}_{L,i}\right) E_t[\mathbb{1}(Y_t - \mathbf{X}_t'\boldsymbol{\beta}_{Q,i} < 0)(Y_t - \mathbf{X}_t'\boldsymbol{\beta}_{Q,i})] \Big)$$

$$+ \nabla_{\boldsymbol{\beta}_{L,i}} \Big(\exp\left(\mathbf{X}_t'\boldsymbol{\beta}_{L,i}\right)(\mathbf{X}_t'\boldsymbol{\beta}_{L,i} - \mathbf{X}_t'\boldsymbol{\beta}_{Q,i}) - \exp\left(\mathbf{X}_t'\boldsymbol{\beta}_{L,i}\right) \Big)$$

$$= E_t\Big(\mathbf{X}_t'\boldsymbol{\beta}_{L,i} - \mathbf{X}_t'\boldsymbol{\beta}_{Q,i} - \frac{1}{\alpha_i} \mathbb{1}(Y_t - \mathbf{X}_t'\boldsymbol{\beta}_{Q,i} < 0)(Y_t - \mathbf{X}_t'\boldsymbol{\beta}_{Q,i}) \Big) \exp\left(\mathbf{X}_t'\boldsymbol{\beta}_{L,i}\right) \mathbf{X}_t. \tag{41}$$

If $\boldsymbol{\beta}_0$ is a solution to (6) then $\nabla_{\boldsymbol{\beta}} E_t[S(Y_t, \mathbf{X}_t\boldsymbol{\beta}_{Q,i}, \mathbf{X}_t\boldsymbol{\beta}_{L,i})]|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} = 0$ a.s.-P, such that

$$E_t \left[\begin{matrix} \Big(1 + \frac{1}{\alpha_i} \exp(\mathbf{X}_t'\boldsymbol{\beta}_{0,L,i}) \Big) \Big(\mathbb{1}(Y_t - \mathbf{X}_t'\boldsymbol{\beta}_{0,Q,i} < 0) - \alpha_i \Big) \mathbf{X}_t \\ \Big(\mathbf{X}_t'\boldsymbol{\beta}_{0,L,i} - \mathbf{X}_t'\boldsymbol{\beta}_{0,Q,i} - \frac{1}{\alpha_i} \mathbb{1}(Y_t - \mathbf{X}_t'\boldsymbol{\beta}_{0,Q,i} < 0)(Y_t - \mathbf{X}_t'\boldsymbol{\beta}_{0,Q,i}) \Big) \exp(\mathbf{X}_t'\boldsymbol{\beta}_{0,L,i}) \mathbf{X}_t \end{matrix} \right] = \mathbf{0} \text{ a.s.-}P, \tag{42}$$

because the discontinuities occur at the events $Y_t = \mathbf{X}_t'\boldsymbol{\beta}_{0,Q,i}$, which have zero probability due to Y_t being a continuous random variable.

This is equivalent to

$$E_t \left[\begin{array}{c} \mathbb{1}(Y_t - \mathbf{X}_t'\boldsymbol{\beta}_{0,Q,i} < 0) - \alpha_i \\ \mathbf{X}_t'\boldsymbol{\beta}_{0,L,i} - \mathbf{X}_t'\boldsymbol{\beta}_{0,Q,i} - \frac{1}{\alpha_i}\mathbb{1}(Y_t - \mathbf{X}_t'\boldsymbol{\beta}_{0,Q,i} < 0)(Y_t - \mathbf{X}_t'\boldsymbol{\beta}_{0,Q,i}) \end{array} \right] = \mathbf{0} \text{ a.s.-}P, \qquad (43)$$

since $\mathbf{X}_t$ and $\exp(\mathbf{X}_t'\boldsymbol{\beta}_{0,L,i})$ are $\mathcal{F}_t$-measurable.

We obtain the first order conditions for $(\boldsymbol{\beta}_{0,Q,i}', \boldsymbol{\beta}_{0,U,i}')'$ straightforwardly by recognizing that $U_t(\alpha_i) = E_t[Y_t|Y_t > Q_t(\alpha_i)] = -E_t[-Y_t| -Y_t < -Q_t(\alpha_i)]$ and that $-Q_t(\alpha_i)$ is the $(1 - \alpha_i)$-quantile of $-Y_t$. Consider $\alpha_j = 1 - \alpha_i$, and let $\tilde{F}_t^{-1}(\alpha_j) = -\mathbf{X}_t\tilde{\boldsymbol{\beta}}_{Q,j}$ and define linear models $E_t[-Y_t| -Y_t < \tilde{F}_t^{-1}(\alpha_j)] = -\mathbf{X}_t\tilde{\boldsymbol{\beta}}_{L,j}$, with $\tilde{F}_t^{-1}$ denoting the inverse distribution function of $-Y_t$. We then find the equalities $\boldsymbol{\beta}_{U,i} = \tilde{\boldsymbol{\beta}}_{L,j}$ and $\boldsymbol{\beta}_{Q,i} = \tilde{\boldsymbol{\beta}}_{Q,j}$, such that we find the first order conditions for $(\boldsymbol{\beta}_{0,Q,i}', \boldsymbol{\beta}_{0,U,i}')'$, given by

$$E_t \left[\begin{array}{c} \mathbb{1}(Y_t - \mathbf{X}_t'\boldsymbol{\beta}_{0,Q,i} < 0) - \alpha_i \\ \mathbf{X}_t'\boldsymbol{\beta}_{0,U,i} - \mathbf{X}_t'\boldsymbol{\beta}_{0,Q,i} - \frac{1}{1-\alpha_i}\mathbb{1}(Y_t - \mathbf{X}_t'\boldsymbol{\beta}_{0,Q,i} > 0)(Y_t - \mathbf{X}_t'\boldsymbol{\beta}_{0,Q,i}) \end{array} \right] = \mathbf{0} \text{ a.s.-}P, \qquad (44)$$

by substituting $\boldsymbol{\beta}_{U,i} = \tilde{\boldsymbol{\beta}}_{L,j}$, $\boldsymbol{\beta}_{Q,i} = \tilde{\boldsymbol{\beta}}_{Q,j}$, and $\alpha_i = 1 - \alpha_j$ in the first order conditions of $(\tilde{\boldsymbol{\beta}}_{0,L,j}', \tilde{\boldsymbol{\beta}}_{0,L,j}')'$. $\qquad \square$

B.2   Proof of Lemma 2.

We give the proof for $(\boldsymbol{\beta}_{0,Q,i}', \boldsymbol{\beta}_{0,L,i}')'$. The proof for the full vector $\boldsymbol{\beta}_0$ follow straightforwardly from similar steps by noting the uniqueness of $\boldsymbol{\beta}_{0,Q,i}'$, for all $i = 1, \ldots, d_Q$, and the linearity of the moment conditions in the other parameters. At the true parameter vector $(\boldsymbol{\beta}_{0,Q,i}', \boldsymbol{\beta}_{0,L,i}')'$ we have the following equalities by assumption (a):

$$E_t[\mathbb{1}(Y_t - \mathbf{X}_t'\boldsymbol{\beta}_{0,Q,i} < 0)] = F_t(\mathbf{X}_t'\boldsymbol{\beta}_{0,Q,i}) = \alpha_i, \qquad (45)$$

$$\frac{1}{\alpha_i}E_t[\mathbb{1}(Y_t - \mathbf{X}_t'\boldsymbol{\beta}_{0,Q,i} < 0)Y_t] = \mathbf{X}_t'\boldsymbol{\beta}_{0,L,i}. \qquad (46)$$

It is thus obvious that (43) holds at $\boldsymbol{\beta}_0$. Moreover, by assumption (b), the optimal $\boldsymbol{\beta}_{0,Q,i}$ is unique. Finally, it is easily shown that the optimal $\boldsymbol{\beta}_{0,L,i}$ is unique for a given optimal $\boldsymbol{\beta}_{0,Q,i}$ under assumption (c), because $\boldsymbol{\beta}_{0,L,i} = \left(E[\mathbf{X}_t\mathbf{X}_t']\right)^{-1}E\left[\mathbf{X}_t\left\{\mathbf{X}_t'\boldsymbol{\beta}_{0,Q,i} + \frac{1}{\alpha_i}\mathbb{1}(Y_t - \mathbf{X}_t'\boldsymbol{\beta}_{0,Q,i} < 0)(Y_t - \mathbf{X}_t'\boldsymbol{\beta}_{0,Q,i})\right\}\right]$.

We conclude that $\boldsymbol{\beta}_0$ is the unique vector for which the moment condition holds. $\qquad \square$

B.3 Proof of Theorem 1.

Similar to steps in the proof of A.1 we can show by assumption (d) that the sequences $\{\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)\}$ and $\{\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)'\}$ are $\mathcal{F}_t$-measurable, stationary, and ergodic. Moreover, by assumption (g) $\hat{\mathbf{W}}_T \overset{p}{\to} \mathbf{W}_0$, such that we can apply Theorem 2.6 of Newey and McFadden (1994) to show consistency.

Assumption 2.6(i) is satisfied, because the conditions of Lemma 2 hold and $\mathbf{W}_0$ is positive definite under assumption (g), such that $E[\mathbf{g_0}(\boldsymbol{\beta})] = \mathbf{0} \iff \mathbf{W}_0^{-1} E[\mathbf{g_0}(\boldsymbol{\beta})] = \mathbf{0}$ for $\boldsymbol{\beta}_0$ uniquely.

Assumption 2.6(ii) coincides with assumption (e). Assumption 2.6(iii) holds because the discontinuities in $\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)$ occur a.s.-P at the events $Y_t = \mathbf{X}_t'\boldsymbol{\beta}_{Q,i}$, for all $i = 1, \ldots, d$, which have zero mass by Y_t being a continuous random variable under assumption (a).

To show assumption 2.6(iv) holds, notice that

$$\|\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)\| \le c_1 \left(1 + \|\mathbf{X}_t\| + |Y_t|\right) \|\mathbf{X}_t\| < \infty \text{ a.s.-}P, \tag{47}$$

for some finite constant c_1, such that $E[\sup_{\boldsymbol{\beta} \in \mathbf{B}} \|\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)\|] \le c_1 E[(1 + \|\mathbf{X}_t\| + |Y_t|) \|\mathbf{X}_t\|] < c_1 (E\|\mathbf{X}_t\|^2 + E[|Y_t|\|\mathbf{X}_t\|]) < \infty$ by assumption (f), and where the second inequality follows from $E\|\mathbf{X}_t\| < E\|\mathbf{X}_t\|^2$ by Hölder's inequality.

This concludes the proof. □

B.4 Proof of Theorem 2.

We verify that the assumptions of Theorem 7.2 of Newey and McFadden (1994) hold. Firstly, by assumption (e) $\mathbf{B}$ is a compact set, such that in combination with (14), for all $T \ge 1$, $\mathbf{g}_T(\hat{\boldsymbol{\beta}}_T)'\boldsymbol{\Sigma}_T(\hat{\boldsymbol{\beta}}_T)^{-1}\mathbf{g}_T(\hat{\boldsymbol{\beta}}_T) = \inf_{\boldsymbol{\beta} \in \mathbf{B}} \mathbf{g}_T(\boldsymbol{\beta})'\boldsymbol{\Sigma}_T(\boldsymbol{\beta})^{-1}\mathbf{g}_T(\boldsymbol{\beta})$. Moreover, Theorem 1 establishes that $\hat{\boldsymbol{\beta}}_T \overset{p}{\to} \boldsymbol{\beta}_0$, and assumption (g) imposes $\hat{\mathbf{W}}_T \overset{p}{\to} \mathbf{W}_0$, with $\mathbf{W}_0$ positive definite.

The first assumption of Theorem 7.2 is satisfied as shown by Lemma 2. Assumption 7.2(iii) holds by assumption (i).

We prove 7.2(iv) using the CLT for martingale difference sequences (MDS) in Corollary 5.24 in White (2001). The proof for iid data follows similar steps (using Theorem 5.2 of White (2001)), and is not presented here. Notice that $\{\mathbf{g}(\boldsymbol{\beta}_0; Y_t, \mathbf{X}_t), \mathcal{F}_t\}$ is a MDS under Lemma 1. Here we use the slightly uncommon notation of $\mathcal{F}_t$ instead of $\mathcal{F}_{t-1}$, but this is a labelling choice. By assumption (d) and the MDS property $E[\bar{\mathbf{g}}_T(\boldsymbol{\beta}_0)] = E[\mathbf{g}(\boldsymbol{\beta}_0; Y_t, \mathbf{X}_t)] = \mathbf{0}$ and $\text{var}[\sqrt{T}\bar{\mathbf{g}}_T(\boldsymbol{\beta}_0; Y_t, \mathbf{X}_t)] = \text{var}[\mathbf{g}(\boldsymbol{\beta}_0; Y_t, \mathbf{X}_t)] = E[\mathbf{g}(\boldsymbol{\beta}_0; Y_t, \mathbf{X}_t)\mathbf{g}(\boldsymbol{\beta}_0; Y_t, \mathbf{X}_t)'] = \boldsymbol{\Sigma}_0$. Under as-

sumption (l) $\boldsymbol{\Sigma}_0$ is positive definite. Under assumptions (a), (d), and (j) Theorem A.1 shows that $\hat{\boldsymbol{\Sigma}}_T(\boldsymbol{\beta}_0) = \frac{1}{T} \sum_{t=1}^T \mathbf{g}(\boldsymbol{\beta}_0; Y_t, \mathbf{X}_t) \mathbf{g}(\boldsymbol{\beta}_0; Y_t, \mathbf{X}_t)' \overset{p}{\to} \boldsymbol{\Sigma}_0$.

Consider $\sqrt{T} \boldsymbol{\lambda}' \boldsymbol{\Sigma}_0^{-1/2} \bar{\mathbf{g}}_T(\boldsymbol{\beta}_0; Y_t, \mathbf{X}_t)$, for any real vector $\boldsymbol{\lambda}$, such that $\boldsymbol{\lambda}' \boldsymbol{\lambda} = 1$, and let $\mathcal{Z}_t = \boldsymbol{\lambda}' \boldsymbol{\Sigma}_0^{-1/2} \mathbf{g}_T(\boldsymbol{\beta}_0; Y_t, \mathbf{X}_t)$. It is easily shown that $\{\mathcal{Z}_t, \mathcal{F}_t\}$ is an MDS. Moreover, $\{\mathcal{Z}_t\}$ is stationary and ergodic by Theorem 3.35 in White (1996), such that its distribution, which we denote by F, does not change with t. Moreover, by stationarity $\mathrm{var}[\sqrt{T} \boldsymbol{\lambda}' \boldsymbol{\Sigma}_0^{-1/2} \bar{\mathbf{g}}_T(\boldsymbol{\beta}_0; Y_t, \mathbf{X}_t)] = \mathrm{var}[\mathcal{Z}_t] = \boldsymbol{\lambda}' \boldsymbol{\Sigma}_0^{-1/2} \boldsymbol{\Sigma}_0 \boldsymbol{\Sigma}_0^{-1/2} \boldsymbol{\lambda} = 1$. In similar steps to the proof of Theorem A.1 it can be shown that $\{\mathcal{Z}_t^2\}$ has finite absolute expected values under assumption (j). Then it follows from Theorem 3.34 and Theorem 2.24 in White (1996) that $\frac{1}{T} \sum_{t=1}^T \mathcal{Z}_t^2 - 1 \overset{p}{\to} 0$.

We now verify that the Lindeberg condition of Theorem 5.24 is satisfied by following similar steps to those in the proof of Theorem 5.25 in White (2001). Since $\{\mathcal{Z}_t\}$ is stationary the Lindeberg condition reduces to $\lim_{T \to \infty} \int_{z^2 > \epsilon T} z^2 dF(z) = 0$. Let $r_T(z) = z^2 \mathbb{1}[z^2 \le \epsilon T]$, such that $\{r_T\}$ is an increasing sequence of functions that converges to $r(z) = z^2$. Using the monotone convergence theorem we can interchange limit and integral such that

$$\lim_{T \to \infty} \int_{z^2 \le \epsilon T} z^2 dF(z) = \lim_{T \to \infty} \int_{-\infty}^{\infty} r_T(z)^2 dF(z)$$
$$= \int_{-\infty}^{\infty} \lim_{T \to \infty} r_T(z)^2 dF(z)$$
$$= \int_{-\infty}^{\infty} z^2 dF(z) = 1.$$

Hence,

$$\lim_{T \to \infty} \int_{z^2 > \epsilon T} z^2 dF(z) = \lim_{T \to \infty} \left[1 - \int_{-\infty}^{\infty} r_T(z)^2 dF(z) \right] = 0.$$

Then it follows from Theorem 5.24 that $\sqrt{T} \boldsymbol{\lambda}' \boldsymbol{\Sigma}_0^{-1/2} \bar{\mathbf{g}}_T(\boldsymbol{\beta}_0; Y_t, \mathbf{X}_t) \overset{d}{\to} \mathcal{N}(0, 1)$. Using the Cramér-Wold device we find that $\sqrt{T} \bar{\mathbf{g}}_T(\boldsymbol{\beta}_0) \overset{d}{\to} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_0)$.

Assumption (k) coincides with 7.2(v). To our knowledge no general theory exists to provide more primitive conditions for stationary ergodic processes where $\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)$ is unbound and non-Lipschitz continuous. For strong mixing martingale difference sequences the stochastic equicontinuity of the last $(k+1)d$ elements of $\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)$ could be shown using the results for Lipschitz continuous functions in Hansen (1996). If the process is m-dependent or iid, results of Andrews (1994) apply to all elements of $\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)$, and stochastic equicontinuity follows from the elements of $\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)$ being VC classes of functions.

We prove 7.2(ii) for the subvector $(\boldsymbol{\beta}_{Q,i}', \boldsymbol{\beta}_{L,i}')'$. The proof for the full vector $\boldsymbol{\beta}$ follows straightforwardly from repeating the steps outlined in this proof by noting that parameters relating to α_i do not enter the moment conditions for parameters relating to α_j, for all $i \neq j$, and by rewriting $g_{I,i,j}(\cdot)$ as

$$g_{I,i,j}(\boldsymbol{\beta}_{Q,i}, \boldsymbol{\beta}_{Q,j}, \boldsymbol{\beta}_{I,i,j}, Y_t, \mathbf{X}_t) \equiv \mathbf{X}_t'\boldsymbol{\beta}_{U,i} - \frac{1}{\alpha_j - \alpha_i}\Big[-\big(\alpha_i \mathbf{X}_t'\boldsymbol{\beta}_{Q,i} + \mathbb{1}(Y_t - \mathbf{X}_t'\boldsymbol{\beta}_{Q,i} < 0)(Y_t - \mathbf{X}_t'\boldsymbol{\beta}_{Q,i})\big)$$
$$+ \big(\alpha_j \mathbf{X}_t'\boldsymbol{\beta}_{Q,j} + \mathbb{1}(Y_t - \mathbf{X}_t'\boldsymbol{\beta}_{Q,j} < 0)(Y_t - \mathbf{X}_t'\boldsymbol{\beta}_{Q,j})\big)\Big]. \quad (48)$$

Notice that $\mathbf{g}_0(\boldsymbol{\beta}) = E[\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)] = E\Big[E_t[\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)]\Big]$ by the law of iterated expectations. We will see that the derivative of the function $E_t[\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)]$ exists and is dominated by a function that is not dependent on $\boldsymbol{\beta}$, everywhere on $\mathbf{B}$, such that it follows that $\nabla_{\boldsymbol{\beta}}\mathbf{g}_0(\boldsymbol{\beta}) = E\Big[\nabla_{\boldsymbol{\beta}}E_t[\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)]\Big]$. We have

$$E_t[\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)] = \begin{bmatrix} (F_t(\mathbf{X}_t'\boldsymbol{\beta}_{Q,1}) - \alpha)\mathbf{X}_t \\ \big(\mathbf{X}_t'\boldsymbol{\beta}_{L,1} - \mathbf{X}_t'\boldsymbol{\beta}_{Q,1} - \frac{1}{\alpha}E_t[\mathbb{1}(Y_t - \mathbf{X}_t'\boldsymbol{\beta}_{Q,1} < 0)(Y_t - \mathbf{X}_t'\boldsymbol{\beta}_{Q,1})]\big)\mathbf{X}_t \end{bmatrix}. \quad (49)$$

Then, using the result in (38), and the fact that $E_t[\mathbb{1}(Y_t < c)] = F_t(c)$ by assumption (a),

$$\nabla_{\boldsymbol{\beta}}E_t[\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)] = \begin{bmatrix} f_t(\mathbf{X}_t'\boldsymbol{\beta}_{Q,1})\mathbf{X}_t\mathbf{X}_t' & \mathbf{0} \\ (\frac{1}{\alpha}F_t(\mathbf{X}_t'\boldsymbol{\beta}_{Q,1}) - 1)\mathbf{X}_t\mathbf{X}_t' & \mathbf{X}_t\mathbf{X}_t' \end{bmatrix}. \quad (50)$$

From assumption (a) it follows that $F_t(\mathbf{X}_t'\boldsymbol{\beta}_{0,Q,1}) = \alpha_1$, such that

$$\nabla_{\boldsymbol{\beta}}E_t[\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)]|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} = \begin{bmatrix} f_t(\mathbf{X}_t'\boldsymbol{\beta}_{0,Q,1})\mathbf{X}_t\mathbf{X}_t' & 0 \\ 0 & \mathbf{X}_t\mathbf{X}_t' \end{bmatrix}. \quad (51)$$

As f_t is bounded by assumption (h), the matrix $\nabla_{\boldsymbol{\beta}}E_t[\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)]$ is uniformly bounded on $\mathbf{B}$, by assumption (j). Therefore, a function $\xi(\mathbf{X}_t)$ exists that dominates $\nabla_{\boldsymbol{\beta}}E_t[\mathbf{g}(\boldsymbol{\beta}; Y_t, \mathbf{X}_t)]$ for all $\boldsymbol{\beta} \in \mathbf{B}$. We can therefore differentiate under the integral and find

$$\mathbf{G}_0 = \nabla_{\boldsymbol{\beta}}\mathbf{g}_0(\boldsymbol{\beta}_0) = E\begin{bmatrix} f_t(\mathbf{X}_t'\boldsymbol{\beta}_{0,Q,1})\mathbf{X}_t\mathbf{X}_t' & 0 \\ 0 & \mathbf{X}_t\mathbf{X}_t' \end{bmatrix}. \quad (52)$$

Notice that $\mathbf{G}_0\mathbf{W}_0\mathbf{G}_0$ is nonsingular, because $\mathbf{G}_0$ is nonsingular by its block-diagonality and assumptions (c) and (m), and $\mathbf{W}_0$ is positive semi-definite by (g).

This concludes the proof. □

B.5 Proof of Theorem 3.

Theorem A.1 establishes that $\hat{\boldsymbol{\Sigma}}_T(\hat{\boldsymbol{\beta}}_T) \overset{p}{\to} \boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_0$ is positive definite, such that $\boldsymbol{\Sigma}_0^{-1}$ is positive definite as well. Notice that we are specifically concerned with the upper-left block of $\hat{\mathbf{G}}_{T,\tau}$. The moment conditions of $\mathbf{X}_t\mathbf{X}_t'$ as given in (f) and (j) are sufficient for the boundedness conditions in the proof of Lemma 2 in Giacomini and Komunjer (2005). We can therefore apply the steps in their proof to arrive at what needs to be shown. □

C Additional tables

[Table 3 about here.]

[Table 4 about here.]

Table 1: Descriptive statistics of the IQER estimator in the simulation experiment

	$\beta_{L,1,1}$	$\beta_{L,1,2}$	$\beta_{L,2.5,1}$	$\beta_{L,2.5,2}$	$\beta_{L,5,1}$	$\beta_{L,5,2}$	$\beta_{L,10,1}$	$\beta_{L,10,2}$
$T = 50$								
True	-2.59	-0.40	-2.27	-0.32	-2.00	-0.25	-1.70	-0.18
Bias	0.51	0.23	0.29	0.15	0.16	0.09	0.09	0.05
Std. Dev.	0.49	0.57	0.45	0.50	0.37	0.44	0.31	0.37
Median	-2.02	-0.20	-1.94	-0.18	-1.82	-0.16	-1.60	-0.12
90%-10%	1.17	1.37	1.10	1.20	0.93	1.04	0.76	0.88
$T = 100$								
Bias	0.27	0.15	0.14	0.08	0.08	0.04	0.04	0.02
Std. Dev.	0.42	0.45	0.32	0.35	0.26	0.29	0.20	0.23
Median	-2.27	-0.28	-2.11	-0.24	-1.92	-0.20	-1.65	-0.15
90%-10%	1.02	1.09	0.80	0.84	0.64	0.70	0.51	0.56
$T = 1000$								
Bias	0.03	0.02	0.02	0.01	0.01	0.01	0.01	0.00
Std. Dev.	0.14	0.15	0.10	0.11	0.08	0.08	0.06	0.07
Median	-2.55	-0.37	-2.25	-0.30	-1.99	-0.24	-1.70	-0.17
90%-10%	0.36	0.38	0.26	0.27	0.20	0.21	0.16	0.17

	$\beta_{I,10,20,1}$	$\beta_{I,10,20,2}$	$\beta_{I,45,55,1}$	$\beta_{I,45,55,2}$	$\beta_{I,10,90,1}$	$\beta_{I,10,90,1}$	$\beta_{I,1,99,1}$	$\beta_{I,1,99,2}$
$T = 50$								
True	-1.01	0.00	0.00	0.25	0.00	0.25	0.00	0.25
Bias	0.00	0.03	0.00	0.01	0.00	0.00	0.00	0.00
Std. Dev.	0.25	0.32	0.20	0.25	0.16	0.19	0.16	0.18
Median	-1.00	0.01	0.00	0.25	0.00	0.25	0.00	0.25
90%-10%	0.61	0.76	0.49	0.59	0.42	0.46	0.40	0.45
$T = 100$								
Bias	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Std. Dev.	0.16	0.18	0.13	0.16	0.11	0.12	0.11	0.12
Median	-1.01	0.01	0.00	0.25	0.00	0.25	0.00	0.25
90%-10%	0.39	0.44	0.33	0.38	0.28	0.30	0.27	0.29
$T = 1000$								
Bias	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Std. Dev.	0.05	0.05	0.04	0.04	0.03	0.03	0.03	0.03
Median	-1.01	0.00	0.00	0.25	0.00	0.25	0.00	0.25
90%-10%	0.12	0.13	0.10	0.11	0.08	0.09	0.08	0.09

Note: This table presents descriptive statistics for the the LQE regression (L) and IQE regression (I) estimators of the mean-volatility model described in (21) based on 10,000 simulations. We consider LQEs below $\alpha = 0.01, 0.025, 0.05, 0.10$, and IQEs between 0.10 and 0.20, 0.45 and 0.55, 0.10 and 0.90, and 0.01 and 0.99. We indicate the quantile level(s) (in percentages) with the second subscript for the LQE estimator, and with the second and third subscript for the IQE estimator. The table provides results for a normally distributed regressor and error, and $\phi_X = 0.85$ and $\phi_\varepsilon = 0$, such that the regressor X_t has considerable persistence and Y_t is conditionally heteroskedastic and persistent. We consider sample sizes $T = 50, 100, 1000$. The first row gives the true values of the parameters. The subsequent panels show the bias, standard deviation, median, and the difference between the 90% and 10% percentiles.

Table 2: Coverage rates of the IQER estimator in the simulation experiment

	$\beta_{L,1,1}$	$\beta_{L,1,2}$	$\beta_{L,2.5,1}$	$\beta_{L,2.5,2}$	$\beta_{L,5,1}$	$\beta_{L,5,2}$	$\beta_{L,10,1}$	$\beta_{L,10,2}$
$T = 50$	0	0	18	9	57	46	76	70
$T = 100$	1	1	57	46	76	69	86	82
$T = 250$	57	44	80	73	87	82	92	89
$T = 500$	76	67	87	82	91	88	93	91
$T = 1000$	86	80	91	88	93	91	94	92
$T = 2500$	90	87	93	91	94	93	94	94
	$\beta_{I,10,20,1}$	$\beta_{I,10,20,2}$	$\beta_{I,45,55,1}$	$\beta_{I,45,55,2}$	$\beta_{I,10,90,1}$	$\beta_{I,10,90,1}$	$\beta_{I,1,99,1}$	$\beta_{I,1,99,2}$
$T = 50$	88	87	88	90	94	94	94	93
$T = 100$	92	92	91	92	94	94	94	94
$T = 250$	94	93	93	94	95	95	95	95
$T = 500$	94	94	94	94	95	95	95	95
$T = 1000$	94	95	95	95	95	95	95	95
$T = 2500$	95	95	95	95	95	94	95	95

Note: This table presents coverage rates of the 95% confidence intervals of the LQE regression (L) and IQE regression (I) estimators of the mean-volatility model described in (21) based on 10,000 simulations. We consider LQEs below $\alpha = 0.01, 0.025, 0.05, 0.10$, and IQEs between 0.10 and 0.20, 0.45 and 0.55, 0.10 and 0.90, and 0.01 and 0.99. We indicate the quantile level(s) (in percentages) with the second subscript for the LQE estimator, and with the second and third subscript for the IQE estimator. The table provides results for a normally distributed regressor and error, and $\phi_X = 0.85$ and $\phi_\varepsilon = 0$, such that the regressor X_t has considerable persistence and Y_t is conditionally heteroskedastic and persistent. We consider sample sizes $T = 50, 100, 250, 500, 1000, 2500$. Each panel gives the coverage rate in percentages that the true parameter falls within the 95% confidence interval based on the asymptotic covariance matrix.

Table 3: Descriptive statistics of the GMM quantile estimator in the simulation experiment

	$\beta_{Q,1,1}$	$\beta_{Q,1,2}$	$\beta_{Q,2.5,1}$	$\beta_{Q,2.5,2}$	$\beta_{Q,5,1}$	$\beta_{Q,5,2}$
$T = 50$						
True	-2.26	-0.31	-1.90	-0.23	-1.60	-0.15
Bias	0.19	0.15	0.02	0.07	0.01	0.04
Std. Dev.	0.49	0.57	0.43	0.49	0.33	0.38
Median	-2.02	-0.20	-1.84	-0.18	-1.57	-0.12
90%-10%	1.17	1.37	1.05	1.18	0.83	0.91
$T = 100$						
Bias	-0.06	0.07	-0.01	0.04	0.00	0.02
Std. Dev.	0.42	0.45	0.28	0.29	0.22	0.23
Median	-2.27	-0.29	-1.89	-0.21	-1.59	-0.14
90%-10%	1.02	1.10	0.70	0.70	0.56	0.58
$T = 1000$						
Bias	0.00	0.01	0.00	0.01	0.00	0.00
Std. Dev.	0.12	0.11	0.08	0.08	0.07	0.06
Median	-2.25	-0.31	-1.90	-0.22	-1.60	-0.15
90%-10%	0.30	0.27	0.21	0.19	0.17	0.15
	$\beta_{Q,10,1}$	$\beta_{Q,10,2}$	$\beta_{Q,20,1}$	$\beta_{Q,20,2}$	$\beta_{Q,45,1}$	$\beta_{Q,45,2}$
$T = 50$						
True	-1.24	-0.06	-0.82	0.05	-0.12	0.22
Bias	0.01	0.02	0.01	0.01	0.00	0.00
Std. Dev.	0.27	0.31	0.23	0.25	0.20	0.21
Median	-1.23	-0.05	-0.80	0.06	-0.12	0.22
90%-10%	0.68	0.74	0.56	0.61	0.50	0.52
$T = 100$						
Bias	0.00	0.01	0.00	0.01	0.00	0.00
Std. Dev.	0.18	0.18	0.15	0.15	0.13	0.13
Median	-1.23	-0.06	-0.81	0.05	-0.12	0.22
90%-10%	0.45	0.46	0.38	0.38	0.34	0.34
$T = 1000$						
Bias	0.00	0.00	0.00	0.00	0.00	0.00
Std. Dev.	0.05	0.05	0.05	0.04	0.04	0.03
Median	-1.24	-0.06	-0.82	0.05	-0.12	0.22
90%-10%	0.14	0.12	0.12	0.10	0.10	0.09

Note: This table presents descriptive statistics for the GMM quantile estimator of the mean-volatility model described in (21) based on 10,000 simulations. We consider quantile levels $\alpha = 0.01, 0.025, 0.05, 0.10, 0.20, 0.45$, and indicate the quantile level with the second subscript (in percentages). The table provides results for a normally distributed regressor and error, and $\phi_X = 0.85$ and $\phi_\varepsilon = 0$, such that the regressor X_t has considerable persistence and Y_t is conditionally heteroskedastic and persistent. We consider sample sizes $T = 50, 100, 1000$. The first row gives the true values of the parameters. The subsequent panels show the bias, standard deviation, median, and the difference between the 90% and 10% percentiles.

Table 4: Coverage rates of the GMM quantile estimator in the simulation experiment

	$\beta_{Q,1,1}$	$\beta_{Q,1,2}$	$\beta_{Q,2.5,1}$	$\beta_{Q,2.5,2}$	$\beta_{Q,5,1}$	$\beta_{Q,5,2}$
$T = 50$	- (30)	- (20)	- (78)	- (71)	- (87)	- (88)
$T = 100$	- (53)	- (43)	- (89)	- (88)	- (91)	- (92)
$T = 250$	80 (89)	79 (86)	87 (92)	82 (93)	89 (93)	86 (95)
$T = 500$	89 (91)	87 (91)	92 (93)	89 (94)	94 (94)	92 (95)
$T = 1000$	93 (92)	91 (93)	95 (94)	94 (94)	96 (95)	95 (95)
$T = 2500$	96 (94)	94 (92)	97 (95)	96 (95)	97 (95)	97 (95)
	$\beta_{Q,10,1}$	$\beta_{Q,10,2}$	$\beta_{Q,20,1}$	$\beta_{Q,20,2}$	$\beta_{Q,45,1}$	$\beta_{Q,45,2}$
$T = 50$	- (91)	- (94)	- (95)	- (97)	- (97)	- (99)
$T = 100$	- (93)	- (95)	- (95)	- (97)	- (96)	- (97)
$T = 250$	92 (94)	90 (95)	93 (95)	91 (96)	92 (95)	91 (97)
$T = 500$	96 (95)	94 (95)	95 (95)	95 (95)	94 (95)	94 (96)
$T = 1000$	96 (95)	96 (95)	96 (95)	96 (95)	95 (95)	95 (96)
$T = 2500$	97 (95)	97 (95)	96 (95)	97 (95)	95 (95)	95 (95)

Note: This table presents coverage rates of the 95% confidence intervals of the GMM quantile estimator of the mean-volatility model described in (21) based on 10,000 simulations. We consider quantile levels $\alpha = 0.01, 0.025, 0.05, 0.10, 0.20, 0.45$, and indicate the quantile level with the second subscript (in percentages). The table provides results for a normally distributed regressor and error, and $\phi_X = 0.85$ and $\phi_\varepsilon = 0$, such that the regressor X_t has considerable persistence and Y_t is conditionally heteroskedastic and persistent. We consider sample sizes $T = 50, 100, 250, 500, 1000, 2500$. Each panel gives the coverage rate in percentages that the true parameter falls within the 95% confidence interval based on the asymptotic covariance matrix. The coverage rates based on the approximated asymptotic covariance matrix of Theorem 3 are given outside brackets, with $\tau = 0.05$ - the best performing tuning parameter for this DGP. We do not provide results for $T = 50, 100$ because the approximation is poor for samples of this size. Coverage rates inside brackets are based on the true standard errors.
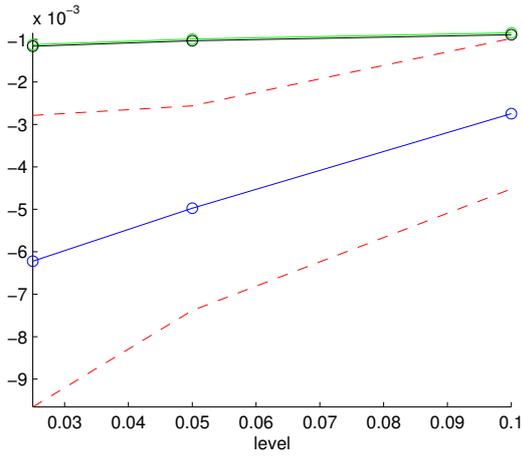
(a) intercept
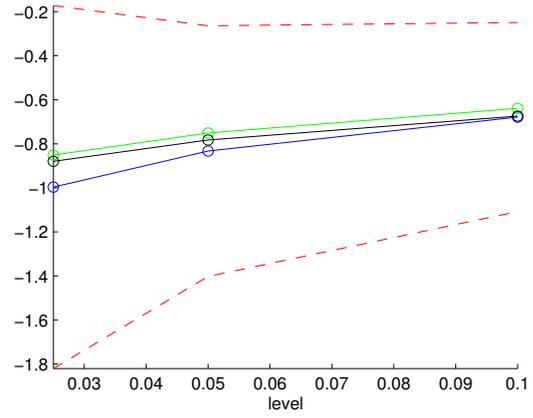
(b) lagged realized volatility
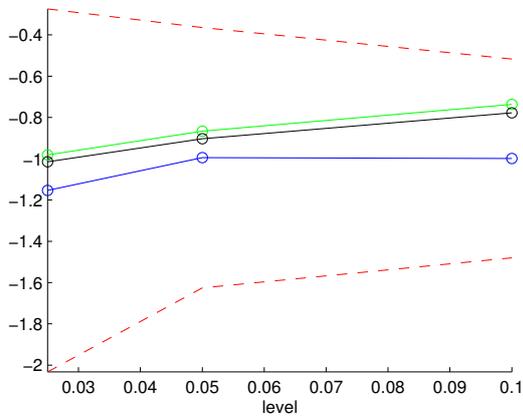
(c) lagged week average realized volatility
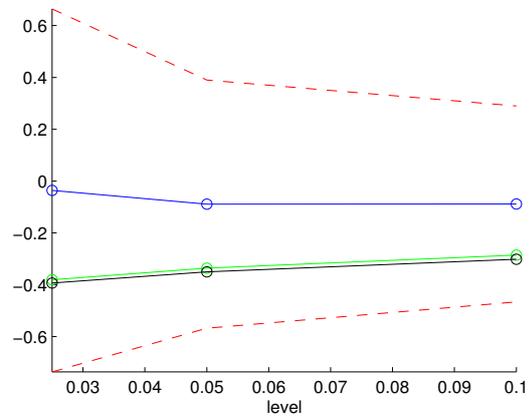
(d) lagged month average realized volatility

Figure 1: S&P 500 ETF return LQE coefficients. IQER coefficients are given in blue, and their 95% confidence bounds in red. LQE coefficients obtained from the two-stage procedures are presented in green and black for the standard normal normal distribution and the EDF, respectively.