Composite Quantile Regression for Alpha Estimation

Zhaoguo Zhan*

September 7, 2024

Abstract

Evaluation of asset pricing models is largely based on the alphas (intercepts) in the linear regression of excess asset returns on risk factors. When regression errors are not normally distributed, the least squares estimator for alphas is inefficient, which further leads to less powerful testing of alphas by the Gibbons, Ross, and Shanken (1989, GRS) test. We use the composite quantile regression to estimate alphas, and show that it provides more accurate alpha estimates under a variety of non-normal distributions. A joint test of alphas using composite quantile regression is also developed, which can reject zero alphas in spanning tests when the GRS test does not.

JEL Classification: C18, C31, G12.

Keywords: Alpha, Composite Quantile Regression, Gibbons-Ross-Shanken Test.

*Email: zzhan@kennesaw.edu. Department of Economics, Finance, and Quantitative Analysis, Coles College of Business, Kennesaw State University, GA 30144, USA. Web: https://sites.google.com/site/zhaoguozhan.

1 Introduction

The influential Gibbons, Ross, and Shanken (1989, GRS hereafter) test of portfolio efficiency has become one of the default tools for evaluating asset pricing models; see, e.g., Cochrane (2005), Fama and French (2015, 2016, 2017, 2018), and Hou, Xue, and Zhang (2015). More recently, Kleibergen and Zhan (2020) extend the GRS test to construct confidence sets of risk premia; see also Kleibergen, Kong, and Zhan (2023). From a methodological perspective, the GRS test, as well as its extension in Kleibergen and Zhan (2020), is a joint test of alphas (intercepts) in a set of linear regression equations, for which the least squares estimator is well known to be efficient under normally distributed regression errors. The GRS test is built on the least squares estimator for alphas, and it is a uniformly most powerful test if the assumed normal distribution holds. Financial asset returns, however, are typically not normally distributed (see, e.g., Affleck-Graves and McDonald (1989)), and their resulting regression errors may not be exactly normal either. The least squares estimator for alphas is thus unlikely to be efficient in empirically relevant settings. Consequently, the estimated alphas by least squares could be imprecise due to non-normality, which further casts doubt on the GRS test that relies on such estimated alphas.

It is worth noting that the alphas, or intercepts, in linear regression equations can be of interest in many different contexts. For instance, in the so-called regression discontinuity designs, the difference in alphas above and below the regression discontinuity cutoff can be interpreted as the local average treatment effect; see, e.g., Cattaneo, Idrobo, and Titiunik (2019). In financial economics, alphas are often interpreted as the abnormal returns with respect to the investment strategy reflected by regressors, so they are expected to be zero under efficient portfolios; see, e.g., Gibbons, Ross, and Shanken (1989). Furthermore, in spanning tests where tested risk factors are regressed on a set of existing factors (see, e.g., Hou, Mo, Xue, and Zhang (2019)), whether the tested factors are considered redundant depends on the values of their alphas. All such alphas have been widely estimated in existing empirical studies by least squares, which may not be efficient especially when financial data are involved.

In light of the above, we propose to use an alternative estimator for alphas, which can outperform the commonly used least squares estimator in a variety of data generating processes. The alternative estimator we use is based on the composite quantile regression (CQR) approach of Zou and Yuan (2008), who show that the CQR estimator for regression betas (slope) can be more efficient than the least squares estimator when regression errors are non-normal. The efficiency gain of CQR results from combining the information from multiple quantiles, and can be noticeably large. While the CQR approach is straightforward to implement, our paper differs from the pioneering work of Zou and Yuan (2008) in several aspects. First, Zou and Yuan (2008) consider one linear regression equation, while our setting involves a system of N equations with $N \geq 1$.

Second, Zou and Yuan (2008)'s interest lies in the regression betas (slope) of the single equation, while we develop a joint test for evaluating alphas (intercepts) of N equations. Third, the analytical result provided in Zou and Yuan (2008) relies on the simplification that all regressors are assumed to be centered, so they have mean equal to zero; in contrast, we do not impose this simplification, since the expected values of regressors, which correspond to factor risk premia in our considered setting, are generally non-zero. Put differently, our contribution lies in extending the CQR approach so that it can be used for conducting inference on alphas in a set of linear equations, while Zou and Yuan (2008)'s focus is on the estimation of betas in a single equation. Despite all these differences, using CQR instead of least squares to improve efficiency under non-normality remains to be the motivation for our paper.

Kai, Li, and Zou (2010) and Huang and Zhan (2022) also adopt the CQR approach. Consistent with Zou and Yuan (2008), both Kai, Li, and Zou (2010) and Huang and Zhan (2022) find that there are efficiency gains to use the CQR estimator instead of the least squares estimator under non-normality. However, Kai, Li, and Zou (2010)'s focus is on nonparametric estimation through local composite quantile regression (LCQR), while Huang and Zhan (2022) explore the boundary points of LCQR around the cutoff of regression discontinuity designs. Therefore, these existing studies do not directly apply to our considered setting, although they do shed light on the superior performance of CQR over least squares. Instead of assuming normally distributed regression errors, Kai, Li, and Zou (2010) and Huang and Zhan (2022) make the weaker assumption that regression errors are symmetrically distributed around zero. We therefore adopt this symmetry assumption, so that the resulting CQR estimator (denoted by $\hat{\alpha}_i^{CQR}$) converges to the alpha of the mean regression. Additionally, we also show that the CQR approach can be extended to allow for asymmetric (skewed) error distributions, and propose a second CQR estimator (denoted by $\tilde{\alpha}_i^{CQR}$) whose validity does not depend on whether error distributions are symmetric or asymmetric. While $\tilde{\alpha}_i^{CQR}$ is naturally appealing, we show that $\hat{\alpha}_i^{CQR}$ can perform better under symmetric errors. Both $\hat{\alpha}_i^{CQR}$ and $\tilde{\alpha}_i^{CQR}$ are, however, based on the CQR approach, so they can outperform the least squares estimator in various non-normal settings.

For ease of exposition, we conduct a simulation study to highlight the difference between the least squares estimator and the CQR estimator $\hat{\alpha}_i^{CQR}$ or $\tilde{\alpha}_i^{CQR}$ for estimating alphas. As expected, we find that the CQR estimator appears to be more efficient than the least squares estimator when the data generating process of the linear regression involves non-normal error distributions. We further simulate the power curves of both the GRS test and the CQR-based test for jointly testing zero alphas, under normal and non-normal distributions. Consistent with the existing literature, the GRS test is found to be more powerful than the CQR-based test under normally distributed errors. In contrast, under non-normally distributed errors, we find the evidence that the power of the CQR-based test can exceed that of the GRS test.

We use the proposed CQR-based approach to study the factor models developed by Fama and French

(2018) and Hou, Xue, and Zhang (2015). The six-factor model of Fama and French (2018) includes market (Mkt-RF), size (SMB), value (HML), profitability (RMW), investment (CMA), and momentum (UMD) factors, while the q-factor model of Hou, Xue, and Zhang (2015) involves only four factors: market (R_MKT), size (R_ME), investment (R_IA), and return on equity (R_ROE).[1] Yet both Zhang (2020) and Hou, Mo, Xue, and Zhang (2019) claim that the q-factor model subsumes the six-factor model in spanning tests, which is consistent with their findings that the GRS test could not reject zero alphas when regressing HML, RMW, CMA, and UMD on the four factors of the q-factor model, i.e., the resulting GRS p-values are just too large to reject the null of zero alphas. We, however, use the CQR-based approach to estimate alphas and conduct the joint test of zero alphas. In contrast with the GRS test, the CQR-based test yields small p-values around 0.05, so we can reject zero alphas at the commonly used 5% or 10% significance level, indicating that the q-factor model does not fully subsume the six-factor model of Fama and French (2018). On the other hand, when we use the q^5 model of Hou, Mo, Xue, and Zhang (2021), which augments the q-factor model with the additional expected growth factor (R_EG), we find that both the GRS test and the CQR-based test do not reject zero alphas when regressing HML, RMW, CMA, and UMD on the q^5 factors. All-in-all, our findings indicate that the additional expected growth factor in the q^5 model is crucial for helping explain the momentum factor used in the Fama and French (2018) six-factor model.

In line with our paper, there exists a sizeable literature that aims to extend the GRS test to non-normal error distributions. Zhou (1993), for example, assumes the class of elliptical distributions, which nests the normal distribution as a special case; see also Harvey and Zhou (1993). Beaulieu, Dufour, and Khalaf (2007) further allow for more general distributions including elliptical and non-elliptical ones. Unlike this branch of literature that relies on non-normal distributional assumptions, the CQR approach does not assume any specific distribution of regression errors. In this regard, the CQR approach is more comparable to the generalized method of moments (GMM) approach, which also provides an asymptotic distribution of alpha estimators without making distributional assumptions; see, e.g., Cochrane (2005) for a textbook discussion. Unlike the GMM estimator that coincides the least squares estimator for alphas in the just-identified linear regression, we show that the CQR approach can provide an appealing alternative under non-normal error distributions.

The rest of the paper is organized as follows. Section 2 presents the CQR approach under symmetric error distributions, which we use to estimate alphas and develop the joint test of alphas. Section 3 further extends the CQR approach to allow for asymmetric (skewed) error distributions. To illustrate our methodology, Section 4 provides a simulation study. The empirical application is discussed in Section 5. Section 6

[1] The Fama and French (2018) factors can be downloaded from Kenneth R. French's online data library. The q factors are available at https://global-q.org.

concludes. Technical details and additional results are relegated to the Appendix.

2 GRS and CQR

Consider a system of N linear regression equations, where X_t is the $L \times 1$ vector of regressors:

$$\begin{cases} Y_{1t} = \alpha_1 + \beta_1' X_t + \epsilon_{1t} \\ Y_{2t} = \alpha_2 + \beta_2' X_t + \epsilon_{2t} \\ \quad \vdots \\ Y_{Nt} = \alpha_N + \beta_N' X_t + \epsilon_{Nt} \end{cases} \tag{1}$$

for $t = 1, ..., T$, and $T > N + L$. Let $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_N)'$ be the $N \times 1$ vector of alphas (intercepts). In line with the GRS test, our objective is to evaluate whether $H_0 : \boldsymbol{\alpha} = \mathbf{0}$ holds. Unlike the GRS test, we do not assume the normal distribution of regression errors in $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \epsilon_{2t}, ..., \epsilon_{Nt})'$.

In the context of asset pricing, the dependent variable Y_{it}, with $i = 1, ..., N$, in (1) can be the excess return on the i-th test asset at time t, while X_t is for risk factor returns. If an asset pricing model using X_t as risk factors is able to fully explain expected returns, then $\boldsymbol{\alpha} = \mathbf{0}$. The GRS test is thus commonly used for testing $H_0 : \boldsymbol{\alpha} = \mathbf{0}$ and evaluating models. Similarly, in spanning tests, the dependent variable Y_{it} can be the i-th tested factor, which is potentially spanned by the L existing factors in X_t. If each Y_{it} is redundant in the sense that it is fully spanned by X_t, then $\boldsymbol{\alpha} = \mathbf{0}$. Thus, the GRS test for testing $H_0 : \boldsymbol{\alpha} = \mathbf{0}$ has also been employed for evaluating Y_{it}.

Since regression errors in $\boldsymbol{\epsilon}_t$ are unobservable, their distributions are typically unknown. The existing literature, however, has made distributional assumptions on $\boldsymbol{\epsilon}_t$, such as normal, t, mixtures of normal distributions; see, e.g., Gibbons, Ross, and Shanken (1989), Zhou (1993), and Harvey and Zhou (1993). In contrast, we do not make such distributional assumptions. In this section, we just impose the weaker condition that the distribution of $\boldsymbol{\epsilon}_t$ is symmetric around zero, as in Kai, Li, and Zou (2010) and Huang and Zhan (2022). This symmetry condition is weaker, because it is nested by, e.g., the normal assumption. For completeness, we further allow for asymmetric (skewed) error distributions in the later Section 3.

2.1 Gibbons-Ross-Shanken

Denote the ordinary least squares estimator for $\boldsymbol{\alpha}$ by $\hat{\boldsymbol{\alpha}}^{LS} = (\hat{\alpha}_1^{LS}, \hat{\alpha}_2^{LS}, ..., \hat{\alpha}_N^{LS})'$, where $\hat{\alpha}_i^{LS}$ with $i = 1, ..., N$ is the intercept estimator by ordinary least squares for the i-th equation in (1). The commonly used GRS statistic for testing $H_0 : \boldsymbol{\alpha} = \mathbf{0}$ reads:

$$\text{GRS-stat} \equiv \frac{T(T-N-L)}{N(T-L-1)}(1+\hat{\mu}_X'\hat{\Omega}_X^{-1}\hat{\mu}_X)^{-1}\hat{\boldsymbol{\alpha}}^{LS\prime}(\hat{\Sigma}^{LS})^{-1}\hat{\boldsymbol{\alpha}}^{LS} \sim F_{N,T-N-L} \tag{2}$$

where $\hat{\mu}_X = \frac{1}{T}\sum_{t=1}^{T} X_t$, $\hat{\Omega}_X = \frac{1}{T}\sum_{t=1}^{T}(X_t - \hat{\mu}_X)(X_t - \hat{\mu}_X)'$, and $\hat{\Sigma}^{LS} = \frac{1}{T-L-1}\sum_{t=1}^{T} \hat{\boldsymbol{\epsilon}}_t\hat{\boldsymbol{\epsilon}}_t'$ with $\hat{\boldsymbol{\epsilon}}_t$ the least squares residual for $\boldsymbol{\epsilon}_t$. The $F_{N,T-N-L}$ distribution of the GRS statistic results from the assumed normal distribution of $\boldsymbol{\epsilon}_t$; see, e.g., Gibbons, Ross, and Shanken (1989).

An asymptotic counterpart of the GRS test can be written as follows, as $T \to \infty$:

$$\chi^2(\hat{\boldsymbol{\alpha}}^{LS}) \equiv T\hat{\boldsymbol{\alpha}}^{LS\prime}\widehat{Var}(\hat{\boldsymbol{\alpha}}^{LS})^{-1}\hat{\boldsymbol{\alpha}}^{LS} \xrightarrow{d} \chi^2_N \tag{3}$$

where $\widehat{Var}(\hat{\boldsymbol{\alpha}}^{LS})$ is the estimated asymptotic variance of $\hat{\boldsymbol{\alpha}}^{LS}$. The variance expression for $\hat{\boldsymbol{\alpha}}^{LS}$ can be derived from the GMM framework; see, e.g., Cochrane (2005) for a more detailed discussion. The χ^2_N distribution in (3) does not require the normality of $\boldsymbol{\epsilon}_t$. Instead, it results from the asymptotic normal distribution of the least squares estimator $\hat{\boldsymbol{\alpha}}^{LS}$.

Both the GRS test in (2) and its asymptotic counterpart in (3) are, however, built on the least squares estimator $\hat{\boldsymbol{\alpha}}^{LS}$. Under non-normal error distributions, $\hat{\boldsymbol{\alpha}}^{LS}$ becomes less efficient, which further affects the power of the tests in (2) and (3).

2.2 CQR for a single equation

Instead of using least squares to estimate each equation in (1), we use the CQR approach proposed by Zou and Yuan (2008). Without loss of generality, we focus on the i-th equation in (1) in this subsection, so α_i is our parameter of interest:

$$Y_{it} = \alpha_i + \beta_i' X_t + \epsilon_{it},\ t = 1, ..., T, \tag{4}$$

where the p.d.f. and c.d.f. of ϵ_{it} are denoted by f_{ϵ_i} and F_{ϵ_i}, respectively.

To construct the CQR estimator for α_i, we start with q quantile positions τ_k with $k = 1$, ..., q, such that $\tau_k = \frac{k}{1+q}$.[2] The objective function for CQR thus reads (see also Zou and Yuan (2008), Kai, Li, and Zou (2010), Huang and Zhan (2022)):

$$(\hat{\alpha}_{i1}, ...\hat{\alpha}_{iq}, \hat{\beta}_i^{CQR}) = \underset{\alpha_{ik},\beta_i}{\operatorname{argmin}} \sum_{k=1}^{q}\sum_{t=1}^{T} \rho_{\tau_k}\left(Y_{it} - \alpha_{ik} - \beta_i' X_t\right) \tag{5}$$

[2]For example, if $q = 5$, then $\tau_1 = 1/6$, $\tau_2 = 2/6$, $\tau_3 = 3/6$, $\tau_4 = 4/6$, $\tau_5 = 5/6$. Kai, Li, and Zou (2010) and Huang and Zhan (2022) show that q is a tuning parameter chosen by researchers, while $q = 5$ is often adequate for many non-normal distributions. We thus also use $q = 5$ for our implementation of CQR.

where $\rho_{\tau_k}(r) = \tau_k r - r\mathbf{1}(r < 0)$ is the so-called check function, and α_{ik} is the intercept at the k-th quantile. Note that if $q = 1$, then (5) reduces to the objective function of the median regression (i.e., quantile regression at the median). By using $q > 1$, we aim to combine the information from multiple quantiles to improve the estimation of α_i.

Minimizing (5) yields $(\hat{\alpha}_{i1}, ...\hat{\alpha}_{iq}, \hat{\beta}_i^{CQR})$, and the CQR estimator for α_i is defined as

$$\hat{\alpha}_i^{CQR} = \frac{1}{q}\sum_{k=1}^{q} \hat{\alpha}_{ik}. \tag{6}$$

Theorem 1. Under the regularity conditions provided in the Appendix, as $T \to \infty$:

$$\sqrt{T}(\hat{\alpha}_i^{CQR} - \alpha_i) \xrightarrow{d} \mathbb{N}\left(0, \frac{1}{q^2} e_q'(S_{\epsilon_i}^{-1}\mathbf{\Sigma}S_{\epsilon_i}^{-1})_{11} e_q\right) \tag{7}$$

where e_q is the $q \times 1$ vector of ones, $(S_{\epsilon_i}^{-1}\mathbf{\Sigma}S_{\epsilon_i}^{-1})_{11}$ is the upper-left $q \times q$ submatrix of $S_{\epsilon_i}^{-1}\mathbf{\Sigma}S_{\epsilon_i}^{-1}$, and S_{ϵ_i} and $\mathbf{\Sigma}$ are $(q + L) \times (q + L)$ dimensional matrices whose expressions are provided in the Appendix.

Proof. See the Appendix, where the regularity conditions are also presented. □

Theorem 1 shows that $\hat{\alpha}_i^{CQR}$ is asymptotically normally distributed, and its asymptotic variance depends on the error distribution in the i-th equation. At first glance, the variance expression in Theorem 1 may appear messy, yet it nests two well-known results as we discuss in the remarks below, whose proof is also provided in the Appendix.

2.2.1 Remark 1

When $q = 1$, $\hat{\alpha}_i^{CQR}$ reduces to the intercept estimator of the median regression. Theorem 1 therefore nests the well-known result for the median regression, for which $S_{\epsilon_i}^{-1}\mathbf{\Sigma}S_{\epsilon_i}^{-1}$ reduces to:

$$S_{\epsilon_i}^{-1}\mathbf{\Sigma}S_{\epsilon_i}^{-1} = \frac{1}{4f_{\epsilon_i}^2(0)}\begin{bmatrix} 1 & E(X_t)' \\ E(X_t) & E(X_t X_t') \end{bmatrix}^{-1}, \tag{8}$$

whose upper-left element is the asymptotic variance of the intercept estimator, while the lower-right $L \times L$ submatrix is the asymptotic variance of the slope estimator. See, e.g., Koenker (2005). Under symmetric error distributions, their mean and median coincide, which further implies that $\hat{\alpha}_i^{CQR}$ converges to α_i of the mean regression.

2.2.2 Remark 2

Theorem 1 also nests the asymptotic behavior of $\hat{\beta}_i^{CQR}$ in Zou and Yuan (2008). These authors consider a single linear equation with $E(X_t) = 0$, so $S_{\epsilon_i}^{-1}\mathbf{\Sigma}S_{\epsilon_i}^{-1}$ becomes block-diagonal:

$$S_{\epsilon_i}^{-1}\mathbf{\Sigma}S_{\epsilon_i}^{-1} = \begin{bmatrix} \frac{\tau_{11}}{f_{\epsilon_i}^2(c_{i1})} & \cdots & \frac{\tau_{1q}}{f_{\epsilon_i}(c_{i1})f_{\epsilon_i}(c_{iq})} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ \frac{\tau_{q1}}{f_{\epsilon_i}(c_{iq})f_{\epsilon_i}(c_{i1})} & \cdots & \frac{\tau_{qq}}{f_{\epsilon_i}^2(c_{iq})} & 0 \\ 0 & \cdots & 0 & [E(X_tX_t')]^{-1}\frac{\sum_{k=1}^q\sum_{k'=1}^q \tau_{kk'}}{(\sum_{k=1}^q f_{\epsilon_i}(c_{ik}))^2} \end{bmatrix}, \tag{9}$$

with $c_{ik} = F_{\epsilon_i}^{-1}(\tau_k)$, and $\tau_{kk'} = min(\tau_k, \tau_{k'}) - \tau_k\tau_{k'}$. The lower-right $L \times L$ submatrix of $S_{\epsilon_i}^{-1}\mathbf{\Sigma}S_{\epsilon_i}^{-1}$ is the asymptotic variance of the slope estimator provided by Zou and Yuan (2008) (see their Theorem 2.1):

$$\sqrt{T}(\hat{\beta}_i^{CQR} - \beta_i) \xrightarrow{d} \mathbb{N}\left(0, [E(X_tX_t')]^{-1}\frac{\sum_{k=1}^q\sum_{k'=1}^q \tau_{kk'}}{(\sum_{k=1}^q f_{\epsilon_i}(c_{ik}))^2}\right). \tag{10}$$

Unlike Zou and Yuan (2008), we focus on α_i instead of β_i, so the upper-left $q \times q$ submatrix of $S_{\epsilon_i}^{-1}\mathbf{\Sigma}S_{\epsilon_i}^{-1}$, which is the asymptotic covariance of $(\hat{\alpha}_{i1}, ...\hat{\alpha}_{iq})$, is of interest. Under $E(X_t) = 0$, Theorem 1 reduces to:

$$\sqrt{T}(\hat{\alpha}_i^{CQR} - \alpha_i) \xrightarrow{d} \mathbb{N}\left(0, \frac{1}{q^2}\sum_{k=1}^q\sum_{k'=1}^q \frac{\tau_{kk'}}{f_{\epsilon_i}(c_{ik})f_{\epsilon_i}(c_{ik'})}\right). \tag{11}$$

Since $E(X_t) = 0$ is not imposed in (1), (7) instead of (11) provides the limit behavior of $\hat{\alpha}_i^{CQR}$ in our setting.

2.3 CQR-based testing of alphas

For $N > 1$ equations in (1), we denote the CQR estimator of alphas by $\hat{\boldsymbol{\alpha}}^{CQR} = (\hat{\alpha}_1^{CQR}, \hat{\alpha}_2^{CQR}, ..., \hat{\alpha}_N^{CQR})'$, where each $\hat{\alpha}_i^{CQR}$ for $i = 1, ..., N$ is described in the previous subsection. In particular, Theorem 1 provides the asymptotic normal distribution of $\hat{\alpha}_i^{CQR}$. For joint testing of alphas, we also need the asymptotic covariance of $\hat{\alpha}_i^{CQR}$ and $\hat{\alpha}_j^{CQR}$ across any two equations in (1), as shown in Theorem 2 below.

Theorem 2. Consider $\hat{\alpha}_i^{CQR}$ and $\hat{\alpha}_j^{CQR}$ with $i \neq j$, and $1 \leq i, j \leq N$. Under the regularity conditions provided in the Appendix, as $T \to \infty$:

$$\sqrt{T}\begin{bmatrix} \hat{\alpha}_i^{CQR} - \alpha_i \\ \hat{\alpha}_j^{CQR} - \alpha_j \end{bmatrix} \xrightarrow{d} \mathbb{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{1}{q^2}e_q'(S_{\epsilon_i}^{-1}\mathbf{\Sigma}S_{\epsilon_i}^{-1})_{11}e_q & \frac{1}{q^2}e_q'(S_{\epsilon_i}^{-1}\mathbf{\Sigma}_{\epsilon_i\epsilon_j}S_{\epsilon_j}^{-1})_{11}e_q \\ \frac{1}{q^2}e_q'(S_{\epsilon_i}^{-1}\mathbf{\Sigma}_{\epsilon_i\epsilon_j}S_{\epsilon_j}^{-1})_{11}e_q & \frac{1}{q^2}e_q'(S_{\epsilon_j}^{-1}\mathbf{\Sigma}S_{\epsilon_j}^{-1})_{11}e_q \end{bmatrix}\right) \tag{12}$$

where S_{ϵ_i}, S_{ϵ_j}, $\mathbf{\Sigma}$, and $\mathbf{\Sigma}_{\epsilon_i\epsilon_j}$ are $(q+L)\times(q+L)$ dimensional matrices whose expressions are provided in the Appendix.

Proof. See the Appendix. $\square$

Given that Theorem 2 provides the joint behavior of any two elements in $\hat{\boldsymbol{\alpha}}^{CQR} = (\hat{\alpha}_1^{CQR}, \hat{\alpha}_2^{CQR}, ..., \hat{\alpha}_N^{CQR})'$, we develop a joint test for testing $H_0 : \boldsymbol{\alpha} = \mathbf{0}$, as shown in Theorem 3.

Theorem 3. Under the regularity conditions provided in the Appendix and $H_0 : \boldsymbol{\alpha} = \mathbf{0}$, as $T \to \infty$:

$$CQR(\hat{\boldsymbol{\alpha}}^{CQR}) \equiv T\hat{\boldsymbol{\alpha}}^{CQR'}\widehat{Var}(\hat{\boldsymbol{\alpha}}^{CQR})^{-1}\hat{\boldsymbol{\alpha}}^{CQR} \xrightarrow{d} \chi^2_N \tag{13}$$

where the $N \times N$ dimensional $\widehat{Var}(\hat{\boldsymbol{\alpha}}^{CQR})$ is the estimated covariance matrix, whose expression is provided in the Appendix.

Proof. See the Appendix. $\square$

The χ^2_N distribution in Theorem 3 results from the asymptotic normal distribution of the N-dimensional $\hat{\boldsymbol{\alpha}}^{CQR}$ with $N \geq 1$, because the quadratic form of $\hat{\boldsymbol{\alpha}}^{CQR}$ makes the CQR test statistic. We note that the GRS statistic in (2) and the CQR test statistic in (13) are not directly comparable, since they follow two different distributions: $F_{N,T-N-L}$ and χ^2_N, respectively. Therefore, the GRS statistic multiplied by N will have a magnitude comparable to that of the CQR test statistic, since $N \cdot F_{N,T-N-L} \xrightarrow{d} \chi^2_N$, as $T \to \infty$.

3 CQR for skewed error distributions

In this section, we extend the CQR approach for alpha estimation to allow for skewed error distributions. An alternative alpha estimator denoted by $\tilde{\alpha}_i^{CQR}$, which remains consistent for estimating α_i under skewed errors, is proposed in this section. Correspondingly, Theorems 1, 2, and 3 based on $\hat{\alpha}_i^{CQR}$ in Section 2 are updated to Theorems 4, 5, and 6 based on $\tilde{\alpha}_i^{CQR}$ in this section, respectively.

3.1 Construction of $\tilde{\alpha}_i^{CQR}$

Recall that Equation (5) yields a CQR estimator $\hat{\beta}_i^{CQR}$ for the regression slope β_i, whose consistency and asymptotic normality are established in Zou and Yuan (2008). Given $\alpha_i = E(Y_{it}) - E(X_t)'\beta_i$ in the i-th equation, we can construct the alpha estimator $\tilde{\alpha}_i^{CQR}$ by using $\hat{\beta}_i^{CQR}$:

$$\tilde{\alpha}_i^{CQR} = \hat{\mu}_{Y_i} - \hat{\mu}_X'\hat{\beta}_i^{CQR} \tag{14}$$

with $\hat{\mu}_{Y_i} = \frac{1}{T}\sum_{t=1}^{T} Y_{it}$, and $\hat{\mu}_X = \frac{1}{T}\sum_{t=1}^{T} X_t$. The least squares counterpart of $\tilde{\alpha}_i^{CQR}$ is $\hat{\alpha}_i^{LS}$, which can be similarly written as:

$$\hat{\alpha}_i^{LS} = \hat{\mu}_{Y_i} - \hat{\mu}_X' \hat{\beta}_i^{LS} \tag{15}$$

where $\hat{\beta}_i^{LS}$ is the ordinary least squares estimator for β_i. While $\hat{\beta}_i^{CQR}$ tends to be more efficient than $\hat{\beta}_i^{LS}$ under non-normal regression errors (see Zou and Yuan (2008)), the comparison of (14) and (15) indicates that $\tilde{\alpha}_i^{CQR}$ can also outperform $\hat{\alpha}_i^{LS}$.

Unlike $\hat{\alpha}_i^{CQR}$ in (6), $\tilde{\alpha}_i^{CQR}$ constructed in (14) does not require regression errors to be symmetrically distributed. This is due to the fact that the consistency of $\hat{\beta}_i^{CQR}$ does not require the symmetry condition as shown by Zou and Yuan (2008). Therefore, $\tilde{\alpha}_i^{CQR}$ is valid for conducting inference on alphas under skewed error distributions, for which we provide further details next.

3.2 Testing of alphas by using $\tilde{\alpha}_i^{CQR}$

Based on $\tilde{\alpha}_i^{CQR}$, Theorems 4, 5, and 6 below are the counterparts of Theorems 1, 2, and 3 based on $\hat{\alpha}_i^{CQR}$, respectively. In particular, Theorem 4 provides the limiting distribution of $\tilde{\alpha}_i^{CQR}$. Given that $\hat{\beta}_i^{CQR}$, as well as $\hat{\mu}_{Y_i}$ and $\hat{\mu}_X$, is asymptotically normally distributed under regularity conditions, the limiting distribution of $\tilde{\alpha}_i^{CQR}$ in Theorem 4 can be derived by the delta method.

Theorem 4. Under the regularity conditions provided in the Appendix, as $T \to \infty$:

$$\sqrt{T}(\tilde{\alpha}_i^{CQR} - \alpha_i) \xrightarrow{d} \mathbb{N}\left(0, E(X_t)'(S_{\epsilon_i}^{-1}\mathbf{\Sigma} S_{\epsilon_i}^{-1})_{22} E(X_t) + var(\epsilon_{it})\right) \tag{16}$$

where $(S_{\epsilon_i}^{-1}\mathbf{\Sigma} S_{\epsilon_i}^{-1})_{22}$ is the lower-right $L \times L$ submatrix of $S_{\epsilon_i}^{-1}\mathbf{\Sigma} S_{\epsilon_i}^{-1}$ in Theorem 1.

Proof. See the Appendix. □

Theorem 5 provides the asymptotic covariance of any two elements in $\tilde{\boldsymbol{\alpha}}^{CQR} = (\tilde{\alpha}_1^{CQR}, \tilde{\alpha}_2^{CQR}, ..., \tilde{\alpha}_N^{CQR})'$ for estimating the $N \times 1$ vector of alphas.

Theorem 5. Consider $\tilde{\alpha}_i^{CQR}$ and $\tilde{\alpha}_j^{CQR}$ with $i \neq j$, and $1 \leq i, j \leq N$. Under the regularity conditions provided in the Appendix, as $T \to \infty$, the covariance of $\sqrt{T}(\tilde{\alpha}_i^{CQR} - \alpha_i)$ and $\sqrt{T}(\tilde{\alpha}_j^{CQR} - \alpha_j)$ converges to: $E(X_t)'(S_{\epsilon_i}^{-1}\mathbf{\Sigma}_{\epsilon_i\epsilon_j} S_{\epsilon_j}^{-1})_{22} E(X_t) + cov(\epsilon_{it}, \epsilon_{jt})$, where $(S_{\epsilon_i}^{-1}\mathbf{\Sigma}_{\epsilon_i\epsilon_j} S_{\epsilon_j}^{-1})_{22}$ is the lower-right $L \times L$ submatrix of $S_{\epsilon_i}^{-1}\mathbf{\Sigma}_{\epsilon_i\epsilon_j} S_{\epsilon_j}^{-1}$ in Theorem 2.

Proof. See the Appendix. □

Theorem 6 develops a joint test for testing $H_0 : \boldsymbol{\alpha} = \mathbf{0}$ by using the quadratic form of $\tilde{\boldsymbol{\alpha}}^{CQR}$.

Theorem 6. Under the regularity conditions provided in the Appendix and $H_0 : \boldsymbol{\alpha} = \mathbf{0}$, as $T \to \infty$:

$$CQR(\tilde{\boldsymbol{\alpha}}^{CQR}) \equiv T\tilde{\boldsymbol{\alpha}}^{CQR'}\widehat{Var}(\tilde{\boldsymbol{\alpha}}^{CQR})^{-1}\tilde{\boldsymbol{\alpha}}^{CQR} \xrightarrow{d} \chi^2_N \tag{17}$$

where the $N \times N$ dimensional $\widehat{Var}(\tilde{\boldsymbol{\alpha}}^{CQR})$ is the estimated covariance matrix, whose expression is provided in the Appendix.

Proof. See the Appendix. □

3.3 $\tilde{\alpha}_i^{CQR}$ or $\hat{\alpha}_i^{CQR}$ for spanning tests

As $\tilde{\alpha}_i^{CQR}$ remains consistent under skewed error distributions, it might be tempting to conclude that we should just choose $\tilde{\alpha}_i^{CQR}$ over $\hat{\alpha}_i^{CQR}$ for practical purposes. The comparison of (14) and (15), however, suggests that the performance of $\tilde{\alpha}_i^{CQR}$ depends on the magnitude of the mean of regressors, since $\hat{\beta}_i^{CQR}$ is multiplied by $\hat{\mu}_X$ in (14). When $\hat{\mu}_X$ is tiny, $\tilde{\alpha}_i^{CQR}$ would effectively become similar to $\hat{\alpha}_i^{LS}$. In the limit case that $E(X_t) = 0$, $\tilde{\alpha}_i^{CQR}$ and $\hat{\alpha}_i^{LS}$ are asymptotically equivalent. This provides the reason that $\tilde{\alpha}_i^{CQR}$ is not always a better choice than $\hat{\alpha}_i^{CQR}$ under symmetric error distributions.

Put differently, when error distributions are symmetric, $\hat{\alpha}_i^{CQR}$ can outperform $\tilde{\alpha}_i^{CQR}$, since $\hat{\alpha}_i^{CQR}$ uses the symmetry condition while $\tilde{\alpha}_i^{CQR}$ does not. On the other hand, when error distributions are asymmetric, it is proper to use $\tilde{\alpha}_i^{CQR}$ if the purpose is to consistently estimate alphas of the mean regression.[3]

For spanning tests considered in our later empirical study, however, the purpose is to evaluate whether existing factors span tested factors, for which the GRS test has been commonly adopted. We will thus attach more weight to the test outcome based on $\hat{\alpha}_i^{CQR}$, for the following reasons. Firstly, the symmetry condition, which is also imposed in Kai, Li, and Zou (2010) and Huang and Zhan (2022), is weaker than and nested by the normal distributional assumption of the GRS test. Secondly, skewed regression errors may indicate a model specification problem in spanning tests. For example, if a skewed factor Y_{it} leads to skewed errors in the i-th equation, then it indicates that Y_{it} is not fully spanned by X_t. In this scenario, we would wish to signal that Y_{it} is not spanned by X_t, while accepting $H_0 : \boldsymbol{\alpha} = \mathbf{0}$ would lead to a misleading conclusion. As for the GRS test, we therefore impose the symmetry condition in spanning tests, while acknowledging that the rejection of $H_0 : \boldsymbol{\alpha} = \mathbf{0}$ could occur due to: (i) $\boldsymbol{\alpha} \neq \mathbf{0}$ when $\boldsymbol{\epsilon}_t$ is symmetric; or (ii) $\boldsymbol{\epsilon}_t$ is asymmetric. Both (i) and (ii) indicate the possibility that Y_{it} is not fully spanned by X_t, so rejecting $H_0 : \boldsymbol{\alpha} = \mathbf{0}$ due to either (i) or (ii) is practically meaningful for the purpose of spanning tests.

[3] Under skewed error distributions, however, quantile regression is often chosen over mean regression, since mean regression estimates are sensitive to skewed errors. This is similar to the view that median, not mean, is a preferred measure of central location when data are skewed.

4 Simulation

In this section, we conduct a simulation study to compare the GRS approach with the CQR-based approach for evaluating alphas. Since the difference of these two approaches is driven by the least squares estimator versus the CQR estimator for estimating alphas, we also illustrate the performance of these two types of estimators under a variety of data generating processes.

4.1 $\hat{\alpha}_i^{LS}$ vs. $\hat{\alpha}_i^{CQR}$

We start by comparing the least squares estimator $\hat{\alpha}_i^{LS}$ with the CQR estimator $\hat{\alpha}_i^{CQR}$.

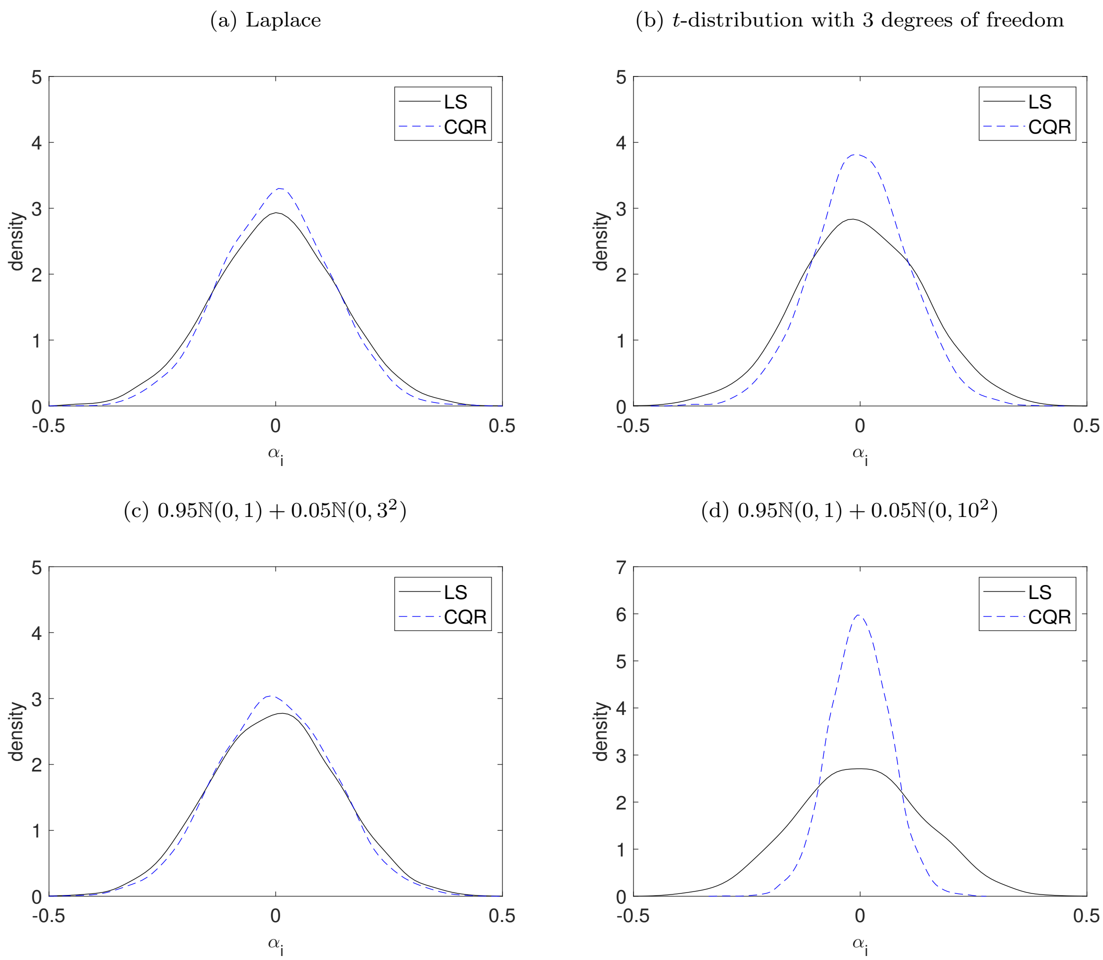
For our data generating processes of (1), we calibrate X_t to the market factor, and $X_t \sim \mathbb{N}(\mu_X, \Omega_X)$, where μ_X and Ω_X are calibrated to the excess market return data we use for the empirical study in the later Section 5. Similarly, the betas are also calibrated to their estimated empirical counterparts. The value of alphas is set to zero under the null hypothesis $H_0 : \boldsymbol{\alpha} = \mathbf{0}$, and is non-zero for power analysis. The error distributions we draw $\boldsymbol{\epsilon}_t$ from play a key role in our simulation study, and we calibrate the variance of $\boldsymbol{\epsilon}_t$ to the residual variance we observe in the time-series regression of the size and book-to-market sorted portfolios on the market factor.

For a single regression equation considered in this subsection, we set $\alpha_i = 0$, and calibrate β_i to the average value of betas when we regress the twenty-five size and book-to-market sorted portfolios on the market factor. The error ϵ_{it} is drawn from four non-normal distributions taken from Kai, Li, and Zou (2010) as well as Huang and Zhan (2022), respectively, whose variance is then re-scaled to mimic the variance of residuals we observe in the regression described above. With the simulated data, we repeatedly conduct the least squares estimation and the CQR estimation of α_i. The reported distributions of $\hat{\alpha}_i^{LS}$ (solid black) and $\hat{\alpha}_i^{CQR}$ (dashed blue) in Figure 1 result from 5000 Monte Carlo replications with the sample size $T = 500$.[4]

The comparison of $\hat{\alpha}_i^{LS}$ vs. $\hat{\alpha}_i^{CQR}$ in Figure 1 for four non-normal distributions shows the efficiency gain of using the CQR estimator over the least squares estimator. Since the true α_i is set to zero, Figure 1 shows that the CQR estimator for α_i (dashed blue) is more concentrated around zero, compared to the least squares estimator (solid black). In addition, Figure 1 shows that the efficiency gain of using $\hat{\alpha}_i^{CQR}$ over $\hat{\alpha}_i^{LS}$ could be small or big, depending on which non-normal distribution is employed in the data generating process. All these findings are consistent with Theorem 1 and the sizeable CQR literature (see, e.g., Zou and Yuan (2008), Kai, Li, and Zou (2010), and Huang and Zhan (2022)), which emphasize the advantage of using the CQR estimator under non-normality.

[4] Figure A1 in the Appendix sets $T = 100$ to show that the CQR estimator performs similarly well under relatively small sample sizes.

Figure 1: **Distributions of $\hat{\alpha}_i^{LS}$ and $\hat{\alpha}_i^{CQR}$ under nonnormalities**



(a) Laplace

(b) t-distribution with 3 degrees of freedom

(c) $0.95\mathbb{N}(0,1) + 0.05\mathbb{N}(0,3^2)$

(d) $0.95\mathbb{N}(0,1) + 0.05\mathbb{N}(0,10^2)$

Notes: The true value of α_i is zero in the data generating process of $Y_{it} = \alpha_i + \beta_i' X_t + \epsilon_{it}$. The solid black line is the simulated density of the least squares estimator $\hat{\alpha}_i^{LS}$, while the dashed blue line is the simulated density of the CQR estimator $\hat{\alpha}_i^{CQR}$. $X_t \sim \mathbb{N}(\mu_X, \Omega_X)$, and μ_X, Ω_X, β_i are all calibrated to data. ϵ_{it} is drawn from non-normal distributions taken from Kai, Li, and Zou (2010) as well as Huang and Zhan (2022): (a) Laplace; (b) t-distribution with 3 degrees of freedom; (c) $0.95\mathbb{N}(0,1) + 0.05\mathbb{N}(0,3^2)$; (d) $0.95\mathbb{N}(0,1) + 0.05\mathbb{N}(0,10^2)$. The variance of ϵ_{it} is then re-scaled to match empirical data. The sample size T is 500, while the number of Monte Carlo replications is 5000.

Table 1: **Comparison of $\hat{\alpha}_i^{LS}$ and $\hat{\alpha}_i^{CQR}$ when $\alpha_i = 0$**

	$\hat{\alpha}_i^{LS}$			$\hat{\alpha}_i^{CQR}$		
	mean	s.d.	$\lvert\hat{\alpha}_i\rvert$	mean	s.d.	$\lvert\hat{\alpha}_i\rvert$
Panel A: $T = 100$						
(a) Laplace	-0.0026	0.3118	0.2484	-0.0045	0.2819	0.2230
(b) t-distribution with 3 degrees of freedom	-0.0012	0.3093	0.2477	-0.0056	0.2541	0.2000
(c) $0.95\mathbb{N}(0,1) + 0.05\mathbb{N}(0,3^2)$	0.0010	0.3151	0.2526	0.0004	0.2984	0.2377
(d) $0.95\mathbb{N}(0,1) + 0.05\mathbb{N}(0,10^2)$	0.0024	0.3131	0.2567	0.0012	0.1820	0.1379
(e) $\mathbb{N}(0,1)$	0.0048	0.3159	0.2515	0.0042	0.3270	0.2604
Panel B: $T = 500$						
(a) Laplace	-0.0002	0.1374	0.1093	0.0003	0.1219	0.0971
(b) t-distribution with 3 degrees of freedom	-0.0003	0.1370	0.1096	-0.0002	0.1053	0.0834
(c) $0.95\mathbb{N}(0,1) + 0.05\mathbb{N}(0,3^2)$	-0.0024	0.1384	0.1111	-0.0021	0.1288	0.1031
(d) $0.95\mathbb{N}(0,1) + 0.05\mathbb{N}(0,10^2)$	-0.0009	0.1397	0.1125	-0.0011	0.0665	0.0528
(e) $\mathbb{N}(0,1)$	-0.0007	0.1395	0.1116	-0.0004	0.1456	0.1165

Notes: For the least squares and CQR estimators $\hat{\alpha}_i^{LS}$ and $\hat{\alpha}_i^{CQR}$, this table reports their mean and standard deviations (s.d.) in the settings of Figure 1 and Figure A1 in the Appendix. $|\hat{\alpha}_i|$, the mean absolute value of estimated alphas, is also reported. The true value of α_i is zero in the data generating process of $Y_{it} = \alpha_i + \beta_i' X_t + \epsilon_{it}$. $X_t \sim \mathbb{N}(\mu_X, \Omega_X)$, and μ_X, Ω_X, β_i are all calibrated to data. ϵ_{it} is drawn from: (a) Laplace; (b) t-distribution with 3 degrees of freedom; (c) $0.95\mathbb{N}(0,1) + 0.05\mathbb{N}(0,3^2)$; (d) $0.95\mathbb{N}(0,1) + 0.05\mathbb{N}(0,10^2)$; (e) $\mathbb{N}(0,1)$. The variance of ϵ_{it} is then re-scaled to match the residual variance observed from data. The sample size T is 100 (Panel A) or 500 (Panel B), while the number of Monte Carlo replications is 5000.

The findings in Figure 1 are further summarized in Table 1, where we report the simulated mean and standard deviations of the least squares and CQR estimators under four non-normal error distributions as Cases (a)(b)(c)(d). To facilitate comparison, we also consider the normal distribution Case (e) in Table 1. In addition, the mean absolute value of estimated alphas denoted by $|\hat{\alpha}_i|$, a metric commonly used for model comparison, is also reported.

For the simulated mean values of $\hat{\alpha}_i^{LS}$ and $\hat{\alpha}_i^{CQR}$, Table 1 shows that they are all around the true value of α_i, which is set to zero in the data generating process. Thus, the tiny mean values of $\hat{\alpha}_i^{LS}$ and $\hat{\alpha}_i^{CQR}$ in Table 1 indicate that both estimators appear to have negligible bias.

For the standard deviations of $\hat{\alpha}_i^{LS}$ and $\hat{\alpha}_i^{CQR}$, however, Table 1 documents their sizeable differences. In particular, under the normal distribution in Table 1 Case (e), the standard deviation of $\hat{\alpha}_i^{LS}$ is slightly smaller than that of $\hat{\alpha}_i^{CQR}$. This finding should not be surprising, since $\hat{\alpha}_i^{LS}$ is the efficient estimator under normally distributed errors. Yet the difference in standard deviations between $\hat{\alpha}_i^{LS}$ and $\hat{\alpha}_i^{CQR}$ in Case (e) appears minor, so using $\hat{\alpha}_i^{CQR}$ in the normal case does not appear to suffer too much efficiency loss. On the other hand, under the four non-normal distributions in Table 1 Cases (a)(b)(c)(d), it is clear that $\hat{\alpha}_i^{CQR}$ is now associated with smaller values for its standard deviation than those of $\hat{\alpha}_i^{LS}$. Most notably in Case (d) with $T = 500$, the standard deviation of $\hat{\alpha}_i^{CQR}$, 0.0665, is only about half as large as that of $\hat{\alpha}_i^{LS}$, 0.1397.

The findings above are also reflected by $|\hat{\alpha}_i|$ reported in Table 1. Since the true α_i is set to zero, we expect that a better estimator for α_i is associated with smaller $|\hat{\alpha}_i|$. For the non-normal Cases (a)(b)(c)(d) in Table 1, the reported values of $|\hat{\alpha}_i|$ show that $\hat{\alpha}_i^{CQR}$ does much better than $\hat{\alpha}_i^{LS}$, especially in Case (d). On the other hand, $\hat{\alpha}_i^{LS}$ only slightly outperforms $\hat{\alpha}_i^{CQR}$ in the normal Case (e).

The existing asset pricing studies widely use the magnitude of estimated alphas for model evaluation and model comparison; see, e.g., Fama and French (2015, 2016, 2017, 2018). The findings presented in Figure 1 and Table 1 therefore highlight the relevance of the CQR approach for asset pricing, since it can potentially provide more accurate alpha estimates.

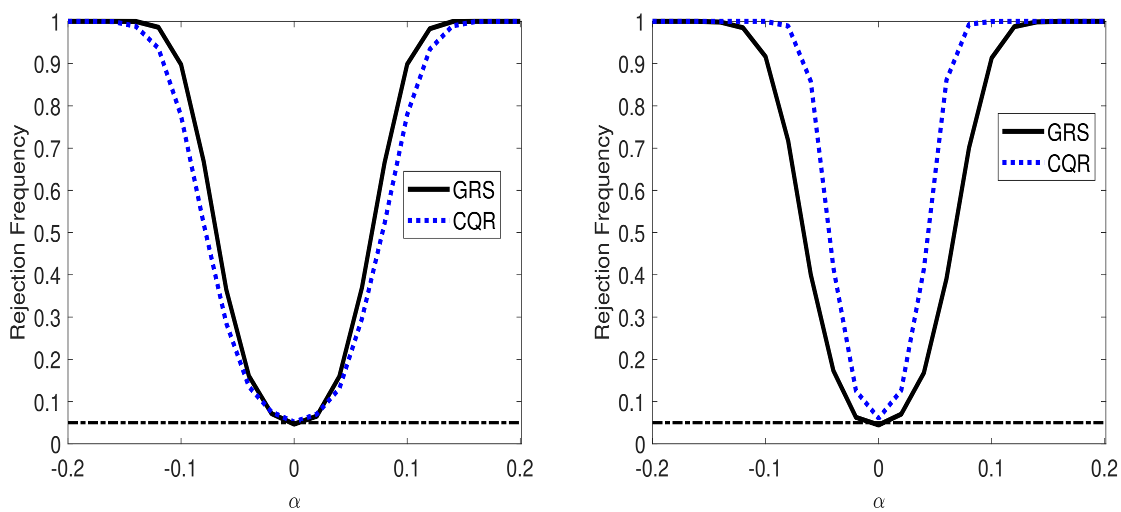
4.2 GRS vs. the CQR-based test

The GRS test is based on the least squares estimator of alphas, which is not as efficient as the CQR estimator under non-normality as we observe in Figure 1. Therefore, it is natural to expect that the CQR-based test for testing zero alphas can exhibit more power than the GRS test, for which we present the simulated power curves in Figure 2.

For convenience of our power analysis, we let alphas gradually deviate from zero, and we set $\boldsymbol{\alpha} \propto \mathbf{1}$ when it is nonzero. Put differently, we impose that each α_i deviates from zero in the same manner for $i = 1, ..., N$

Figure 2: **Power comparison of GRS and CQR for testing $H_0 : \boldsymbol{\alpha} = \mathbf{0}$ at the 5% level**

(a) Normal (b) Non-normal



Notes: The solid black line is the simulated power curve of the GRS test for testing $H_0 : \boldsymbol{\alpha} = \mathbf{0}$ at the 5% significance level, while the dashed blue line is the simulated power curve of the CQR-based test. The benchmark 5% line (black dash-dotted) is also provided to illustrate the 5% size at $H_0 : \boldsymbol{\alpha} = \mathbf{0}$. For the data generating process of (1), $X_t \sim \mathbb{N}(\mu_X, \Omega_X)$, and μ_X, Ω_X, β_i are all calibrated to the regression of the twenty-five size and book-to-market sorted portfolios on the market factor. For (a) Normal: $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \epsilon_{2t}, ..., \epsilon_{Nt})'$ is drawn from a joint normal distribution whose covariance is calibrated to the residual covariance. For (b) Non-normal: each element of $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \epsilon_{2t}, ..., \epsilon_{Nt})'$ is drawn from $0.95\mathbb{N}(0, 1) + 0.05\mathbb{N}(0, 10^2)$, and the covariance of $\boldsymbol{\epsilon}_t$ is then re-scaled to match the residual covariance. $\boldsymbol{\alpha}$ is set to zero at the null, and $\boldsymbol{\alpha} \propto \mathbf{1}$ under the alternative, so the scalar alpha value makes the horizontal line in this figure. The number of test assets N is 25, and the sample size T is 500. The power curves result from the average of 5000 Monte Carlo replications.

to simplify the simulation study. To generate the power curves in Figure 2, we set $N = 25$ and $T = 500$, while the level of significance is 5%.

For the joint distribution of $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \epsilon_{2t}, ..., \epsilon_{Nt})'$ in the data generating process, we consider two cases. In the first case for Figure 2(a), we simulate $\boldsymbol{\epsilon}_t$ from a multivariate normal distribution, whose covariance is calibrated to the residual covariance in the time-series regression of the twenty-five size and book-to-market sorted portfolios on the market factor. In the second case for a non-normal distribution in Figure 2(b), each ϵ_{it} is firstly drawn from the $0.95\mathbb{N}(0,1) + 0.05\mathbb{N}(0,10^2)$ distribution we used for Figure 1(d), where the difference between the least squares estimator and the CQR estimator appears sizeable. We then re-scale the covariance of $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \epsilon_{2t}, ..., \epsilon_{Nt})'$ so that it is also calibrated to the residual covariance we observe from data. The power curves of the GRS test and the CQR-based test in Figure 2 are then generated by using the simulated data to test $H_0 : \boldsymbol{\alpha} = \mathbf{0}$ under a sequence of values of $\boldsymbol{\alpha}$.

Figure 2(a) shows that the GRS test is (slightly) more powerful than the CQR-based test under normally distributed errors in the data generating process. Both tests, however, reject $H_0 : \boldsymbol{\alpha} = \mathbf{0}$ with the probability near the nominal 5% level at $\boldsymbol{\alpha} = \mathbf{0}$, so they are size-correct tests. As $\boldsymbol{\alpha}$ moves away from zero, both the GRS test and the CQR-based test increasingly reject $H_0 : \boldsymbol{\alpha} = \mathbf{0}$. The GRS test is known to be a most powerful test under the imposed normal distribution, so its power curve (solid black) is overall above the power curve of the CQR-based test (dashed blue).

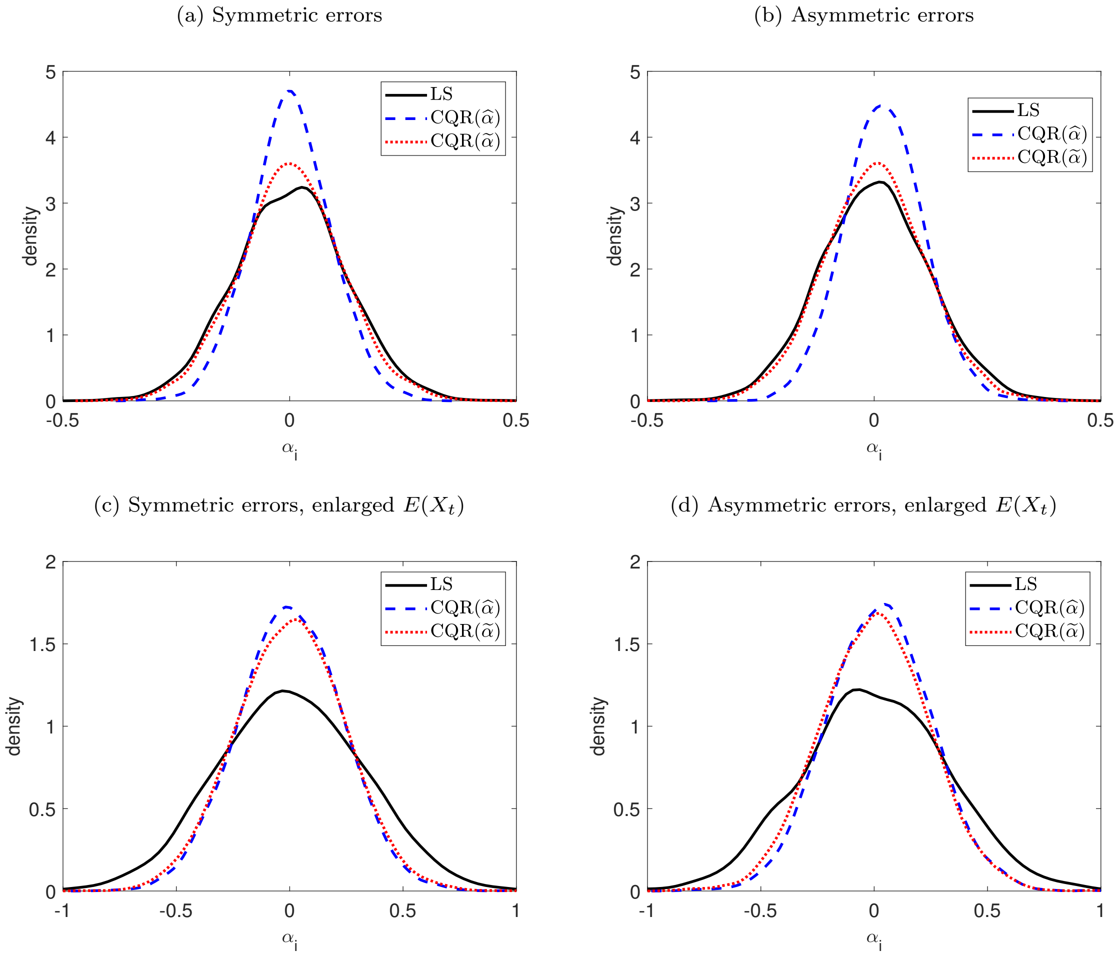
In contrast with Figure 2(a), Figure 2(b) shows that the GRS test is less powerful than the CQR-based test under the non-normal distribution imposed in the data generating process. Both tests, however, still reject $H_0 : \boldsymbol{\alpha} = \mathbf{0}$ with the probability close to the nominal 5% level at $\boldsymbol{\alpha} = \mathbf{0}$, so they remain asymptotically size-correct. Yet the GRS power curve (solid black) is now below the power curve of the CQR-based test (dashed blue) in Figure 2(b).

The comparison of Figure 2(a) and Figure 2(b) suggests that it can be practically useful to use the CQR-based approach. If regression errors are normal, then using the CQR-based approach does not appear to suffer too much power loss. On the other hand, if regression errors are non-normal, then using the CQR-based approach can have relatively larger power gains.

4.3 Empirical residuals for simulating skewed errors and $\tilde{\alpha}_i^{CQR}$

In addition to the well-known distributions previously listed in Table 1, which aim to mimic various error features that could go beyond asset pricing, we further use the empirical distribution of residuals for the simulation study conducted in this subsection. More specifically, we just take the residuals from the spanning regression considered in our later empirical study, and repeatedly draw regression errors from their empirical

Figure 3: **Comparison of $\hat{\alpha}_i^{LS}$, $\hat{\alpha}_i^{CQR}$ and $\tilde{\alpha}_i^{CQR}$ using empirical residuals for simulation**



Notes: The true value of α_i is zero in the data generating process of $Y_{it} = \alpha_i + \beta_i' X_t + \epsilon_{it}$. The solid black line is the simulated density of the least squares estimator $\hat{\alpha}_i^{LS}$. The dashed blue line is the simulated density of the CQR estimator $\hat{\alpha}_i^{CQR}$. The dotted red line is the simulated density of the CQR estimator $\tilde{\alpha}_i^{CQR}$. $X_t \sim \mathbb{N}(\mu_X, \Omega_X)$, and μ_X, Ω_X, β_i for (a)(b) are all calibrated to data, while in (c)(d), μ_X is scaled by 5 to make it larger. ϵ_{it} is drawn from the empirical distribution of the spanning regression residuals considered in the later Section 5. For (a)(c): the empirical distribution is made symmetric by incorporating the positive/negative mirror image of residuals. For (b)(d): the empirical distribution directly results from residuals, so it is asymmetric (skewed). The sample size T is 500, while the number of Monte Carlo replications is 5000.

distribution. Given that the empirical distribution is not perfectly symmetric, it provides a skewed setting which helps illustrate the performance of $\tilde{\alpha}_i^{CQR}$. To facilitate the comparison of $\tilde{\alpha}_i^{CQR}$ with $\hat{\alpha}_i^{CQR}$ as well as $\hat{\alpha}_i^{LS}$ in various scenarios, we also consider a symmetric setting. This is achieved by making the empirical distribution of residuals symmetric through incorporating the positive/negative mirror image of the residuals. Furthermore, in order to emphasize that the magnitude of $E(X_t)$ affects the performance of $\tilde{\alpha}_i^{CQR}$, a setting with enlarged $E(X_t)$ is also simulated. The resulting performances of $\hat{\alpha}_i^{LS}$, $\hat{\alpha}_i^{CQR}$, and $\tilde{\alpha}_i^{CQR}$ are presented in Figure 3 for four scenarios, depending on whether error distributions are symmetric, and whether $E(X_t)$ is enlarged.

Figure 3(a) presents the distributions of $\hat{\alpha}_i^{LS}$, $\hat{\alpha}_i^{CQR}$, and $\tilde{\alpha}_i^{CQR}$ when regression errors are drawn from a symmetric but non-normal distribution. It shows that all three estimators are centered around the true zero alpha, while the two CQR estimators $\hat{\alpha}_i^{CQR}$ and $\tilde{\alpha}_i^{CQR}$ are more concentrated, compared to the least squares estimator $\hat{\alpha}_i^{LS}$. More importantly, Figure 3(a) also shows that $\hat{\alpha}_i^{CQR}$ performs better than $\tilde{\alpha}_i^{CQR}$. All these findings are consistent with the discussions in the previous Section 3, making $\hat{\alpha}_i^{CQR}$ the recommended choice under symmetric but non-normal regression errors.

Figure 3(b) is generated under asymmetric (skewed) regression errors. It shows that $\tilde{\alpha}_i^{CQR}$ remains centered around zero, while $\hat{\alpha}_i^{CQR}$ does not, i.e., its distribution is slightly shifted towards the right of zero. Thus, $\tilde{\alpha}_i^{CQR}$ can outperform $\hat{\alpha}_i^{CQR}$ if regression errors are skewed. Both Figure 3(a) and Figure 3(b), however, show that the difference between $\tilde{\alpha}_i^{CQR}$ and $\hat{\alpha}_i^{LS}$ can be minor. This minor difference occurs when $E(X_t)$ is not large, since the performance of $\tilde{\alpha}_i^{CQR}$ depends on the magnitude of $E(X_t)$ as explained in Section 3. For Figure 3(c) and Figure 3(d), we therefore enlarge $E(X_t)$ (scaled by 5) in the data generating process to compare with Figure 3(a) and Figure 3(b), respectively.

As $E(X_t)$ becomes larger, Figures 3(c)(d) show that the difference between $\tilde{\alpha}_i^{CQR}$ and $\hat{\alpha}_i^{LS}$ is more visible. For instance, in Figure 3(c), $\tilde{\alpha}_i^{CQR}$ and $\hat{\alpha}_i^{CQR}$ almost coincide, both of which clearly outperform $\hat{\alpha}_i^{LS}$. Similar to Figure 1, Figure 3 overall shows that the least squares estimator $\hat{\alpha}_i^{LS}$ may not be efficient, while the CQR approach can provide an alternative choice. While the advantage of using $\tilde{\alpha}_i^{CQR}$ is minor under possibly small $E(X_t)$, $\hat{\alpha}_i^{CQR}$ becomes appealing especially under symmetric errors (or errors with minor skewness).

The estimators presented in Figure 3 lead to different tests for testing $H_0 : \boldsymbol{\alpha} = \mathbf{0}$ with different power, as shown by Figure A2 in the Appendix. Like Figure 2, Figure A2 also suggests that the CQR test using $\hat{\alpha}_i^{CQR}$ can have more power than the GRS test, so we relegate the figure to the Appendix.[5]

[5] Figure A2 in the Appendix is consistent with Theorems 1-6, and it also carries the following messages: (i) The GRS test and its asymptotic χ^2 counterpart are found to perform similarly in large samples; (ii) The CQR test using $\tilde{\alpha}_i^{CQR}$ has slightly more power than the GRS test, unless $E(X_t)$ is large; (iii) The CQR test using $\hat{\alpha}_i^{CQR}$ appears most powerful, yet its size at $H_0 : \boldsymbol{\alpha} = \mathbf{0}$ is a bit distorted under asymmetric errors (see Figure A2(b)). This size distortion is minor, when compared to the large power gains under $\boldsymbol{\alpha} \neq \mathbf{0}$; in addition, it helps signal that the corresponding spanning regression is possibly misspecified, since regression errors are skewed.

5 Application

In this section, we contrast the commonly used GRS test with the proposed CQR-based test in an empirical study. The purpose is to investigate whether using the CQR-based test instead of the GRS test could cause any meaningful difference in practice. To this end, we explore whether the q-factor model of Hou, Xue, and Zhang (2015) subsumes the six-factor model of Fama and French (2018).

Zhang (2020) states that "... *despite having two fewer factors, the Hou-Xue-Zhang q-factor model fully subsumes the Fama–French six-factor model, including UMD.*" Similarly, the abstract of Hou, Mo, Xue, and Zhang (2019) states: "*In spanning tests, the q-factor model largely subsumes the Fama–French five- and six-factor models ...*" These statements, however, are largely based on their GRS test outcomes for evaluating alphas in spanning regressions, which we examine by using the CQR approach.

5.1 Data

The six factors of Fama and French (2018) include: $Mkt\text{-}RF$ (market), SMB (size), HML (value), RMW (profitability), CMA (investment), and UMD (momentum), which nest those in the three-factor model of Fama and French (1993), the four-factor model of Carhart (1997), and the five-factor model of Fama and French (2015). On the other hand, the q-factor model of Hou, Xue, and Zhang (2015) uses only four factors: R_MKT (market), R_ME (size), R_IA (investment), and R_ROE (return on equity). More recently, Hou, Mo, Xue, and Zhang (2021) augment the q-factor model with the additional R_EG (expected growth) to construct the q^5 model. The data of Fama and French (2018) factors we use are downloaded from Kenneth R. French's online data library, while the data of q factors are from https://global-q.org. Overall, we consider eleven factors as presented in Table 2.

Table 2 provides the summary statistics as well as correlation coefficients of the eleven factors over January 1967 to December 2022, so $T = 672$. It is well known that the Fama and French (2018) factors and q factors are often closely related. For example, the two market factors ($Mkt\text{-}RF$ vs. R_MKT) are almost identical, so their correlation coefficient is rounded to 1 in Panel B of Table 2. Similarly, the two size factors (SMB vs. R_ME) also have a large correlation coefficient 0.97, while the two investment factors (CMA vs. R_IA) have the correlation coefficient 0.92. Such large correlations, however, do not appear for HML, RMW, UMD. In particular, UMD stands out as its largest correlation coefficient 0.49 with q factors is much smaller, compared to the counterparts of the other Fama and French (2018) factors with q factors ($Mkt\text{-}RF$: 1; SMB: 0.97; HML: 0.68; RMW: 0.66; CMA: 0.92). Therefore, it is natural to explore whether UMD (and similarly, but to a lesser extent, HML, RMW, CMA) is fully spanned by q factors.

The skewness and kurtosis reported in Panel A of Table 2 provide the evidence that factor returns can

Table 2: **Summary statistics and correlation coefficients of factors during 1967:01 – 2022:12**

Panel A: Summary statistics											
	Mkt-RF	SMB	HML	RMW	CMA	UMD	R_MKT	R_ME	R_IA	R_ROE	R_EG
mean	0.56	0.21	0.31	0.30	0.32	0.62	0.56	0.27	0.40	0.53	0.78
s.d.	4.59	3.05	3.04	2.26	2.08	4.27	4.59	3.05	2.04	2.60	2.05
skewness	-0.49	0.37	0.13	-0.29	0.36	-1.28	-0.49	0.60	0.36	-0.80	-0.02
(p-val)	(.00)	(.00)	(.17)	(.00)	(.00)	(.00)	(.00)	(.00)	(.00)	(.00)	(.81)
kurtosis	4.61	6.13	5.17	13.89	4.25	12.62	4.61	7.92	4.75	8.31	6.73
(p-val)	(.00)	(.00)	(.00)	(.00)	(.00)	(.00)	(.00)	(.00)	(.00)	(.00)	(.00)
ks-stat	0.054	0.049	0.069	0.089	0.041	0.107	0.054	0.050	0.033	0.072	0.051
(p-val)	(.04)	(.08)	(.00)	(.00)	(.21)	(.00)	(.04)	(.07)	(.44)	(.00)	(.06)
ck-stat	0.092	0.278	0.857	0.842	0.529	1.641	0.049	0.176	0.140	0.284	0.300
(p-val)	(.20)	(.47)	(.00)	(.00)	(.04)	(.00)	(.93)	(.26)	(.88)	(.42)	(.44)
Panel B: Correlation coefficients											
Mkt-RF	1.00										
SMB	0.28	1.00									
HML	-0.22	-0.03	1.00								
RMW	-0.18	-0.36	0.11	1.00							
CMA	-0.38	-0.09	0.69	0.00	1.00						
UMD	-0.18	-0.09	-0.21	0.08	-0.01	1.00					
R_MKT	1.00	0.28	-0.22	-0.18	-0.38	-0.18	1.00				
R_ME	0.27	0.97	0.02	-0.36	-0.04	-0.05	0.27	1.00			
R_IA	-0.35	-0.14	0.68	0.08	0.92	0.02	-0.35	-0.10	1.00		
R_ROE	-0.21	-0.39	-0.13	0.66	-0.05	0.49	-0.21	-0.32	0.05	1.00	
R_EG	-0.43	-0.44	0.04	0.38	0.20	0.36	-0.43	-0.39	0.19	0.53	1.00

Notes: Panel A reports the mean, standard deviation (s.d.), skewness, and kurtosis of factor returns. The p-value for skewness results from testing the null of zero, while the p-value of kurtosis results from testing the null of 3 or less. In addition, Panel A reports the ks-stat and its associated p-value from the Kolmogorov-Smirnov normality test. A large ks-stat leads to a small p-value to reject the null of a normal distribution. Panel A also reports the conditional Kolmogorov test statistic (ck-stat) of Andrews (1997) and its associated p-value: for each of the Fama and French (2018) factors, the null is that its distribution conditional on the four factors in the q-factor model is normal; on the other hand, for each of the q factors, the null is that its distribution conditional on the six factors in the Fama and French (2018) model is normal. The p-value associated with ck-stat results from 10000 bootstrap replications, since ck-stat is not nuisance parameter-free. Fama and French (2018) use six factors: Mkt-RF, SMB, HML, RMW, CMA, UMD. The q-factor model of Hou, Xue, and Zhang (2015) uses four factors: R_MKT, R_ME, R_IA, R_ROE, while Hou, Mo, Xue, and Zhang (2021) add R_EG to the q-factor model. Panel B reports the correlation coefficients of factor returns. The sample contains monthly data starting from January 1967 to December 2022 with $T = 672$, which are available from Kenneth R. French's online data library and https://global-q.org.

be substantially non-normal. For a normal distribution, the skewness is 0 while the kurtosis is 3. Overall, the reported values for skewness and kurtosis are quite different from the normal benchmarks, as reflected by their associated small p-values.

It is worth noting that the momentum factor denoted by UMD also stands out in Panel A of Table 2, where we report the Kolmogorov-Smirnov normality test statistic (ks-stat), as well as its associated p-value. Specifically, we standardize each factor, and then compare its empirical distribution after standardization with the standard normal distribution. The reported Kolmogorov-Smirnov normality test statistic just reflects the difference in c.d.f. of these two distributions. A large Kolmogorov-Smirnov test statistic, together with a small p-value, indicates that the distribution of factor returns is substantially different from a normal distribution. Out of the eleven factors listed in Table 2, we can thus reject the null of a normal distribution for nine factors at the 10% level, for six factors at the 5% level, and for four factors at the 1% level. Among all these factors, UMD has the largest Kolmogorov-Smirnov test statistic, and thus the smallest p-value, so its distribution is substantially different from a normal distribution.

5.1.1 Conditional Kolmogorov test

A subtle point is worth emphasizing: we are motivated by possibly non-normal regression errors, not factors themselves. In other words, non-normal factor distributions do not necessarily imply that their regression errors are also non-normal.[6] Instead of the distribution of factor returns themselves, the distribution of their regression errors in spanning tests is more relevant for the purpose of this paper. Therefore, we use the conditional Kolmogorov test proposed by Andrews (1997) to test whether the error distribution in spanning regressions is normal. The test outcome is presented in Panel A of Table 2, where the test statistic ck-stat and its associated p-value are reported.

Specifically, for each of the six Fama and French (2018) factors such as UMD, we consider the null hypothesis that its distribution conditional on the four factors in the q-factor model is normal. On the other hand, for each of the q factors, the null is that its distribution conditional on the six factors in the Fama and French (2018) model is normal. The conditional Kolmogorov test is conducted, whose test statistic depends on nuisance parameters. Thus, its reported p-value results from bootstrap replications; see Andrews (1997).

The conditional Kolmogorov test outcome in Panel A of Table 2 shows that UMD and similarly, HML, RMW, CMA, lead to tiny p-values. These tiny p-values thus cast doubt on whether the error distribution should be considered as normal as for the GRS test, when regressing UMD, as well as HML, RMW, CMA, on the four factors of the q-factor model.

[6] For example, if Y_{it} is just a noisy version of Y_{it}^*, i.e., $Y_{it} = Y_{it}^* + \epsilon_{it}$, then it is possible that ϵ_{it} is normally distributed, while Y_{it} and Y_{it}^* are both allowed to be non-normal and skewed.

5.2 Testing of zero alphas in spanning regressions of the q-factor model

Next, we regress the momentum factor denoted by UMD on the four factors in the q-factor model:

$$UMD = \alpha_{UMD} + \beta_{MKT} R\_MKT + \beta_{ME} R\_ME + \beta_{IA} R\_IA + \beta_{ROE} R\_ROE + \epsilon \tag{18}$$

Our interest is on testing whether $\alpha_{UMD} = 0$, for which both the GRS test and the CQR-based test can be applied with $N = 1$.

As shown in Panel A of Table 3, the GRS test does not reject $H_0 : \alpha_{UMD} = 0$, since its p-value 0.18 is well above the commonly used significance levels such as 5% or 10%. Therefore, the GRS test could not rule out the possibility that UMD is spanned by the four factors in the q-factor model, which is consistent with Zhang (2020). In contrast with the GRS test, Panel A of Table 3 shows that the CQR test using $\hat{\boldsymbol{\alpha}}^{CQR}$ can easily reject $H_0 : \alpha_{UMD} = 0$. Since the CQR p-value is 0.02, we can reject the null of zero alpha at the commonly used 5% level. Therefore, a researcher who uses the CQR-based test would conclude that UMD is not spanned by the four factors in the q-factor model, while the GRS test leads to the opposite conclusion.

In a similar fashion as in (18), we regress CMA, RMW, and HML on the four factors in the q-factor model, so α_{CMA}, α_{RMW}, and α_{HML} are the resulting intercepts. We then jointly test whether these alphas, together with α_{UMD}, are equal to zero. This leads to three joint null hypotheses in Panel A of Table 3: $H_0 : \alpha_{CMA} = \alpha_{UMD} = 0$; $H_0 : \alpha_{RMW} = \alpha_{CMA} = \alpha_{UMD} = 0$; $H_0 : \alpha_{HML} = \alpha_{RMW} = \alpha_{CMA} = \alpha_{UMD} = 0$. For these joint hypotheses, we report the test outcomes by both the GRS test and the CQR-based test, with $N = 2$, $N = 3$, and $N = 4$, respectively.

Unlike the large GRS p-values in Panel A of Table 3, the CQR-based test yields much smaller p-values. These CQR p-values are all below 10%, so we can reject the null of zero alphas at the 10% level for all the considered hypotheses. Similarly, at the 5% level, we can reject three out of four hypotheses listed in Panel A of Table 3. These findings thus cast doubt on the claim that the q-factor model of Hou, Xue, and Zhang (2015) fully subsumes the six-factor model of Fama and French (2018). Put differently, the GRS test and the CQR-based test lead to opposite statistical conclusions in Panel A of Table 3.

The seemingly contradictory performance of GRS and CQR in Panel A of Table 3 is mainly due to their different alpha estimates. As shown by the mean absolute value of estimated alphas $A|\alpha_i|$, the CQR approach yields larger values of $A|\alpha_i|$ than those by GRS. In particular, the magnitude of $A|\alpha_i|$ by CQR is about 50% larger than the GRS counterpart when we focus solely on α_{UMD}, as reported in the first row of Table 3 Panel A. This finding is consistent with Table 2, where the distribution of UMD conditional on the four q factors is substantially non-normal as indicated by the conditional Kolmogorov test. Under non-normal regression errors, least squares and CQR could lead to substantially different alpha estimates.

Table 3: **Spanning tests by regressing Fama-French factors on q factors**

	GRS			CQR($\hat{\boldsymbol{\alpha}}^{CQR}$)		
	GRS-stat	p-value	$A\|\alpha_i\|$	CQR-stat	p-value	$A\|\alpha_i\|$
Panel A: HML, RMW, CMA, UMD on q						
$H_0 : \alpha_{UMD} = 0$	1.78	0.18	0.203	5.61	0.02	0.307
$H_0 : \alpha_{CMA} = \alpha_{UMD} = 0$	1.03	0.36	0.111	6.56	0.04	0.167
$H_0 : \alpha_{RMW} = \alpha_{CMA} = \alpha_{UMD} = 0$	1.10	0.35	0.087	8.39	0.04	0.129
$H_0 : \alpha_{HML} = \alpha_{RMW} = \alpha_{CMA} = \alpha_{UMD} = 0$	0.83	0.51	0.066	9.05	0.06	0.115
Panel B: HML, RMW, CMA, UMD on q^5						
$H_0 : \alpha_{UMD} = 0$	0.07	0.79	0.043	0.01	0.92	0.013
$H_0 : \alpha_{CMA} = \alpha_{UMD} = 0$	0.69	0.50	0.042	0.37	0.83	0.016
$H_0 : \alpha_{RMW} = \alpha_{CMA} = \alpha_{UMD} = 0$	0.54	0.66	0.043	1.56	0.67	0.033
$H_0 : \alpha_{HML} = \alpha_{RMW} = \alpha_{CMA} = \alpha_{UMD} = 0$	0.41	0.80	0.036	2.09	0.72	0.040

Notes: For GRS and CQR, this table reports their test statistics, p-values, and the mean absolute value of estimated alphas $A|\alpha_i|$. The null hypothesis is that the alphas of Fama-French factors (HML, RMW, CMA, UMD) are zero, when regressing them on q factors for Panel A, or q^5 factors for Panel B. We consider the single $H_0 : \alpha_{UMD} = 0$, as well as three joint cases: $H_0 : \alpha_{CMA} = \alpha_{UMD} = 0$; $H_0 : \alpha_{RMW} = \alpha_{CMA} = \alpha_{UMD} = 0$; $H_0 : \alpha_{HML} = \alpha_{RMW} = \alpha_{CMA} = \alpha_{UMD} = 0$. The q-factor model uses R_MKT, R_ME, R_IA, R_ROE, while the q^5 model adds R_EG to the q-factor model. The sample is from January 1967 to December 2022 with $T = 672$ as in Table 2.

5.3 Testing of zero alphas in spanning regressions of the q^5 model

If the four factors in the q-factor model of Hou, Xue, and Zhang (2015) do not fully span the momentum factor, how about the q^5 model of Hou, Mo, Xue, and Zhang (2021)? We therefore regress the momentum factor denoted by UMD on the five factors of the q^5 model to examine the intercept α_{UMD}:

$$UMD = \alpha_{UMD} + \beta_{MKT} R\_MKT + \beta_{ME} R\_ME + \beta_{IA} R\_IA + \beta_{ROE} R\_ROE + \beta_{EG} R\_EG + \epsilon \tag{19}$$

Likewise, the CMA, RMW, HML factors are similarly considered, leading to α_{CMA}, α_{RMW}, and α_{HML} as intercepts. Test outcomes on the alphas of such factors using the q^5 model are thus presented in Panel B of Table 3, to contrast Panel A of Table 3 for the q-factor model.

For both the GRS test and the CQR-based test, Panel B of Table 3 shows that they lead to large p-values. Therefore, we could not reject the null of zero alphas, no matter whether GRS or CQR is adopted. In other words, neither the GRS test nor the CQR-based test has the power to rule out the possibility that the Fama and French (2018) six-factor model is subsumed by the q^5 model in our studied sample. Similar findings can be seen from Table A1 in the Appendix, where we use a different number of quantiles for sensitivity analysis.

To sum up, the findings in Table 3 suggest that the additional expected growth factor (R_EG) in the q^5 model is crucial for helping explain the momentum factor (UMD). Without R_EG, the CQR-based test rejects zero alpha for the momentum factor in Panel A of Table 3 for the q-factor model. With R_EG, the CQR-based test does not reject zero alpha for the momentum factor in Panel B of Table 3 for the q^5 model. Given that Table 2 shows a sizeable correlation (0.36) of R_EG and UMD, the findings documented in Table 3 should not be surprising.

The existing asset pricing literature largely relies on the GRS approach to estimate alphas and evaluate models. Based on the GRS test outcomes in Table 3, a researcher would thus draw a conclusion similar to those in Zhang (2020) and Hou, Mo, Xue, and Zhang (2019): the q-factor model is sufficient to span the Fama and French (2018) factors such as UMD, so it does not appear necessary to have the additional expected growth factor. Yet the proposed CQR approach conveys the different message that the q-factor model needs the expected growth factor to better explain the Fama and French (2018) factors. This clearly shows the value of having the alternative CQR approach to accompany the GRS test, so that researchers can assess their empirical findings with a second thought.

Lastly, we note that using the asymptotic counterpart of the GRS test does not alter our findings in Table 3; see Table A2 in the Appendix. Since our sample size $T = 672$ is already large, using p-values resulting from χ^2_N for the asymptotic counterpart of the GRS test does not cause any substantial difference to the p-values we report for GRS in Table 3. Therefore, the power improvement of CQR over GRS we observe in Panel A of Table 3 is not caused by using $F_{N,T-N-L}$ or χ^2_N distributions. Table A2 in the Appendix also contains the findings based on the CQR test using $\tilde{\boldsymbol{\alpha}}^{CQR}$. Since the mean of factors is generally small as reported in Table 2, the CQR test using $\tilde{\boldsymbol{\alpha}}^{CQR}$ leads to similar, but mostly smaller p-values, compared to those from the asymptotic counterpart of the GRS test in Table A2. All these findings are thus consistent with the power comparison of tests in Figure A2. One might criticize the symmetry condition associated with the CQR-based test conducted in Table 3. Yet the purpose of Table 3 is to contrast the CQR-based test with the GRS test, whose normality assumption also implies the symmetry condition. When error distributions are asymmetric (skewed), the corresponding spanning regressions are likely misspecified, so rejecting zero alphas in this scenario helps signal that the tested factors are not fully spanned.

5.4 Further discussions

Asset pricing models, especially their linear simplifications, are approximations to reality. From this perspective, it is not surprising that the null of zero alphas can be rejected by a test with power, when accompanied with informative data. Failure to reject zero alphas could occur, when the test itself lacks power, or the employed samples are not sufficiently informative. Therefore, we opt not to overly interpret the test outcomes of the GRS test or the CQR-based test, regardless of whether these outcomes are in favor of or against researchers' prior thoughts.

Nevertheless, given that a variety of models have been proposed in the asset pricing literature, it is important to have econometric tools that can be used for evaluating the alphas of such models. For this purpose, the CQR-based approach is proposed in this paper, since it is designed for non-normality in empirical studies, while the least squares estimator for alphas and the resulting GRS test are not.

6 Conclusion

We propose the CQR-based approach to complement the popular GRS test for evaluating asset pricing models. When regression errors are normally distributed, the least squares estimator is efficient for estimating alphas in linear equations, and the resulting GRS test is a powerful test for jointly testing zero alphas. On the other hand, in empirically relevant settings where regression errors are possibly non-normal, the CQR estimator can be more efficient than the least squares estimator for estimating alphas, and consequently, the CQR-based test can have more power than the GRS test. In our empirical study as well as simulation experiment, we find the evidence that the CQR-based approach outperforms the GRS test in many cases of interest.

References

Affleck-Graves, J. and McDonald, B. (1989). Nonnormalities and tests of asset pricing theories. *Journal of Finance*, **44** (4), 889–908.

Andrews, D. (1997). A conditional Kolmogorov test. *Econometrica*, **65** (5), 1097–1128.

Beaulieu, M.-C., Dufour, J.-M. and Khalaf, L. (2007). Multivariate tests of mean–variance efficiency with possibly non-Gaussian errors: An exact simulation-based approach. *Journal of Business & Economic Statistics*, **25** (4), 398–410.

Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance*, **52** (1), 57–82.

Cattaneo, M. D., Idrobo, N. and Titiunik, R. (2019). A practical introduction to regression discontinuity designs: Foundations. *Cambridge Elements: Quantitative and Computational Methods for Social Science, Cambridge University Press*.

Cochrane, J. (2005). *Asset Pricing: Revised Edition*. Princeton University Press.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, **33** (1), 3–56.

— and — (2015). A five-factor asset pricing model. *Journal of Financial Economics*, **116** (1), 1–22.

— and — (2016). Dissecting anomalies with a five-factor model. *Review of Financial Studies*, **29** (1), 69–103.

— and — (2017). International tests of a five-factor asset pricing model. *Journal of Financial Economics*, **123** (3), 441–463.

— and — (2018). Choosing factors. *Journal of Financial Economics*, **128** (2), 234–252.

Gibbons, M. R., Ross, S. A. and Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica*, **57** (5), 1121–1152.

Harvey, C. R. and Zhou, G. (1993). International asset pricing with alternative distributional specifications. *Journal of Empirical Finance*, **1** (1), 107–131.

Hou, K., Mo, H., Xue, C. and Zhang, L. (2019). Which factors? *Review of Finance*, **23** (1), 1–35.

—, —, — and — (2021). An augmented q-factor model with expected growth. *Review of Finance*, **25** (1), 1–41.

—, Xue, C. and Zhang, L. (2015). Digesting anomalies: An investment approach. *Review of Financial Studies*, **28** (3), 650–705.

Huang, X. and Zhan, Z. (2022). Local composite quantile regression for regression discontinuity. *Journal of Business & Economic Statistics*, **40** (4), 1863–1875.

Kai, B., Li, R. and Zou, H. (2010). Local composite quantile regression smoothing: an efficient and safe alternative to local polynomial regression. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **72** (1), 49–69.

Kleibergen, F., Kong, L. and Zhan, Z. (2023). Identification robust testing of risk premia in finite samples. *Journal of Financial Econometrics*, **21** (2), 263–297.

— and Zhan, Z. (2020). Robust inference for consumption-based asset pricing. *Journal of Finance*, **75** (1), 507–550.

Koenker, R. (2005). *Quantile Regression*. Cambridge University Press.

Zhang, L. (2020). q-factors and Investment CAPM. In *Oxford Research Encyclopedia of Economics and Finance*.

Zhou, G. (1993). Asset-pricing tests under alternative distributions. *Journal of Finance*, **48** (5), 1927–1942.

ZOU, H. and YUAN, M. (2008). Composite quantile regression and the oracle model selection theory. *Annals of Statistics*, **36** (3), 1108–1126.

Appendix

A. Regularity conditions

Conditions 1-2 below are similar to the regularity conditions provided in Zou and Yuan (2008), except that we do not require $E(X_t)$ to be zero, i.e., regressors are not required to be centered.

Condition 1:

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix}1 & X_t' \\ X_t & X_tX_t'\end{bmatrix}=\begin{bmatrix}1 & E(X_t)' \\ E(X_t) & E(X_tX_t')\end{bmatrix},$$

and the $(L+1)\times(L+1)$ matrix in the limit is positive definite.

Condition 2: The p.d.f. and c.d.f. of ϵ_{it} denoted by f_{ϵ_i} and F_{ϵ_i} exist, and satisfy:

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\int_0^{u_0+X_t'\mathbf{u}}\sqrt{T}(F_{\epsilon_i}(a+v/\sqrt{T})-F_{\epsilon_i}(a))dv=\frac{f_{\epsilon_i}(a)}{2}(u_0,\mathbf{u}')\begin{bmatrix}1 & E(X_t)' \\ E(X_t) & E(X_tX_t')\end{bmatrix}(u_0,\mathbf{u}')'$$

where $\mathbf{u}$ is the L-dimensional vector, u_0 is a scalar, and $f_{\epsilon_i}(a)$ is a positive p.d.f. value at a.

Condition 3 below is imposed in Kai, Li, and Zou (2010); see also Huang and Zhan (2022). The symmetric error distribution condition is to ensure that the CQR estimator $\hat{\alpha}_i^{CQR}$ converges to the intercept of the mean regression. It is required for Theorems 1, 2, 3, but not for Theorems 4, 5, 6.

Condition 3: The regression error ϵ_{it} has a symmetric distribution.

Condition 4 below is imposed for Theorem 3 and Theorem 6, so that the covariance matrix can be consistently estimated.

Condition 4: There exist consistent estimators for f_{ϵ_i}, F_{ϵ_i}, and the joint c.d.f. $F_{\epsilon_i\epsilon_j}$ around the q quantile positions.

Condition 5 below is imposed for Theorems 4, 5, 6, so that the sample means are asymptotically normally distributed.

Condition 5: A central limit theorem applies to the sample means of Y_{it} and X_t, so that

$$\sqrt{T}\begin{bmatrix}\hat{\mu}_{Y_i}-E(Y_{it}) \\ \hat{\mu}_X-E(X_t)\end{bmatrix}\xrightarrow{d}\mathbb{N}\left(\begin{bmatrix}0 \\ 0\end{bmatrix},\begin{bmatrix}var(Y_{it}) & cov(Y_{it},X_t)' \\ cov(Y_{it},X_t) & var(X_t)\end{bmatrix}\right)$$

with $\hat{\mu}_{Y_i}=\frac{1}{T}\sum_{t=1}^{T}Y_{it}$, and $\hat{\mu}_X=\frac{1}{T}\sum_{t=1}^{T}X_t$.

B. Notation

S_{ϵ_i} and S_{ϵ_j} are $(q+L) \times (q+L)$ dimensional matrices. They are defined in the same fashion, with:

$$S_{\epsilon_i} = \begin{bmatrix} f_{\epsilon_i}(c_{i1}) & & & f_{\epsilon_i}(c_{i1})E(X_t)' \\ & \ddots & & \vdots \\ & & f_{\epsilon_i}(c_{iq}) & f_{\epsilon_i}(c_{iq})E(X_t)' \\ f_{\epsilon_i}(c_{i1})E(X_t) & ... & f_{\epsilon_i}(c_{iq})E(X_t) & \sum_{k=1}^{q} f_{\epsilon_i}(c_{ik})E(X_t X_t') \end{bmatrix},$$

where the upper-left $q \times q$ submatrix of S_{ϵ_i} is a diagonal matrix, and $c_{ik} = F_{\epsilon_i}^{-1}(\tau_k)$. S_{ϵ_j} is similarly defined as above by replacing i with j.

$\mathbf{\Sigma}$ is a $(q+L) \times (q+L)$ dimensional matrix:

$$\mathbf{\Sigma} = \begin{bmatrix} \tau_{11} & ... & \tau_{1q} & E(X_t)' \sum_{k'=1}^{q} \tau_{1k'} \\ \vdots & \ddots & \vdots & \vdots \\ \tau_{q1} & ... & \tau_{qq} & E(X_t)' \sum_{k'=1}^{q} \tau_{qk'} \\ E(X_t) \sum_{k'=1}^{q} \tau_{1k'} & ... & E(X_t) \sum_{k'=1}^{q} \tau_{qk'} & E(X_t X_t') \sum_{k=1}^{q} \sum_{k'=1}^{q} \tau_{kk'} \end{bmatrix},$$

with $\tau_{kk'} = min(\tau_k, \tau_{k'}) - \tau_k \tau_{k'}$, so $\tau_{kk} = \tau_k(1-\tau_k)$.

Similarly, $\mathbf{\Sigma}_{\epsilon_i \epsilon_j}$ is a $(q+L) \times (q+L)$ dimensional matrix:

$$\mathbf{\Sigma}_{\epsilon_i \epsilon_j} = \begin{bmatrix} \tau_{ij,11} & ... & \tau_{ij,1q} & E(X_t)' \sum_{k'=1}^{q} \tau_{ij,1k'} \\ \vdots & \ddots & \vdots & \vdots \\ \tau_{ij,q1} & ... & \tau_{ij,qq} & E(X_t)' \sum_{k'=1}^{q} \tau_{ij,qk'} \\ E(X_t) \sum_{k=1}^{q} \tau_{ij,k1} & ... & E(X_t) \sum_{k=1}^{q} \tau_{ij,kq} & E(X_t X_t') \sum_{k=1}^{q} \sum_{k'=1}^{q} \tau_{ij,kk'} \end{bmatrix},$$

with $\tau_{ij,kk'} = F_{\epsilon_i \epsilon_j}(c_{ik}, c_{jk'}) - \tau_k \tau_{k'}$, and $F_{\epsilon_i \epsilon_j}$ is the joint c.d.f. of ϵ_{it} and ϵ_{jt}.

C. Proof of Theorem 1

The CQR objective function is provided by:

$$(\hat{\alpha}_{i1}, ...\hat{\alpha}_{iq}, \hat{\beta}_i^{CQR}) = \underset{\alpha_{ik}, \beta_i}{\text{argmin}} \sum_{k=1}^{q} \sum_{t=1}^{T} \rho_{\tau_k} \left(Y_{it} - \alpha_{ik} - \beta_i' X_t \right).$$

Define: $\sqrt{T}(\hat{\beta}_i^{CQR} - \beta_i) = \mathbf{U}_T$ and $\sqrt{T}(\hat{\alpha}_{ik} - \alpha_{ik}) = U_{T,k}$, then $(U_{T,1}, ..., U_{T,q}, \mathbf{U}_T)$ is the minimizer of

$$L_T = \sum_{k=1}^{q}\sum_{t=1}^{T}\left(\rho_{\tau_k}\left(\epsilon_{it} - c_{ik} - \frac{U_k + X_t'\mathbf{U}}{\sqrt{T}}\right) - \rho_{\tau_k}\left(\epsilon_{it} - c_{ik}\right)\right)$$

with $c_{ik} = F_{\epsilon_i}^{-1}(\tau_k)$.

Note that the identity

$$|r - s| - |r| = -s(1(r > 0) - 1(r < 0)) + 2\int_0^s [1(r \le t) - 1(r \le 0)]dt$$

implies

$$\rho_\tau(r - s) - \rho_\tau(r) = s(1(r < 0) - \tau) + \int_0^s [1(r \le t) - 1(r \le 0)]dt.$$

We can thus rewrite L_T as:

$$\begin{aligned} L_T &= \sum_{k=1}^{q}\sum_{t=1}^{T}\frac{U_k + X_t'\mathbf{U}}{\sqrt{T}}(1(\epsilon_{it} < c_{ik}) - \tau_k) + \sum_{k=1}^{q}\sum_{t=1}^{T}\int_0^{(U_k + X_t'\mathbf{U})/\sqrt{T}} [1(\epsilon_{it} \le c_{ik} + v) - 1(\epsilon_{it} \le c_{ik})]dv \\ &= \sum_{k=1}^{q} Z_{T,k}U_k + \mathbf{Z}_T'\mathbf{U} + \sum_{k=1}^{q} B_T^{(k)} \end{aligned}$$

with

$$\begin{aligned} Z_{T,k} &\equiv \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(1(\epsilon_{it} < c_{ik}) - \tau_k) \\ \mathbf{Z}_T &\equiv \frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_t[\sum_{k=1}^{q}(1(\epsilon_{it} < c_{ik}) - \tau_k)] \\ B_T^{(k)} &= \sum_{t=1}^{T}\int_0^{(U_k + X_t'\mathbf{U})/\sqrt{T}} [1(\epsilon_{it} \le c_{ik} + v) - 1(\epsilon_{it} \le c_{ik})]dv. \end{aligned}$$

For $B_T^{(k)}$:

$$\begin{aligned} E(B_T^{(k)}) &= \sum_{t=1}^{T}\int_0^{(U_k + X_t'\mathbf{U})/\sqrt{T}} [F_{\epsilon_i}(c_{ik} + v) - F_{\epsilon_i}(c_{ik})]dv \\ &= \frac{1}{T}\sum_{t=1}^{T}\int_0^{(U_k + X_t'\mathbf{U})} \sqrt{T}[F_{\epsilon_i}(c_{ik} + \frac{v}{\sqrt{T}}) - F_{\epsilon_i}(c_{ik})]dv \\ &\to \frac{f_{\epsilon_i}(c_{ik})}{2}(U_k, \mathbf{U}')\begin{bmatrix} 1 & E(X_t)' \\ E(X_t) & E(X_tX_t') \end{bmatrix}(U_k, \mathbf{U}')' \end{aligned}$$

and its variance converges to zero (see Zou and Yuan (2008)). Thus, we can rewrite L_T as:

$$\begin{aligned} L_T &\xrightarrow{d} \sum_{k=1}^{q} Z_{T,k}U_k + \mathbf{Z}'_T\mathbf{U} + \sum_{k=1}^{q} \frac{f_{\epsilon_i}(c_{ik})}{2}(U_k, \mathbf{U}') \begin{bmatrix} 1 & E(X_t)' \\ E(X_t) & E(X_tX'_t) \end{bmatrix} (U_k, \mathbf{U}')' \\ &= (Z_{T,1}, ..., Z_{T,q}, \mathbf{Z}'_T)(U_1, ..., U_q, \mathbf{U}')' + \frac{1}{2}(U_1, ..., U_q, \mathbf{U}')S_{\epsilon_i}(U_1, ..., U_q, \mathbf{U}')' \end{aligned}$$

with

$$S_{\epsilon_i} = \begin{bmatrix} f_{\epsilon_i}(c_{i1}) & & & f_{\epsilon_i}(c_{i1})E(X_t)' \\ & \ddots & & \vdots \\ & & f_{\epsilon_i}(c_{iq}) & f_{\epsilon_i}(c_{iq})E(X_t)' \\ f_{\epsilon_i}(c_{i1})E(X_t) & ... & f_{\epsilon_i}(c_{iq})E(X_t) & \sum_{k=1}^{q} f_{\epsilon_i}(c_{ik})E(X_tX'_t) \end{bmatrix},$$

where the upper-left $q \times q$ submatrix of S_{ϵ_i} is a diagonal matrix.

$(U_{T,1}, ..., U_{T,q}, \mathbf{U}_T)$ results from minimizing L_T above:

$$(U_{T,1}, ..., U_{T,q}, \mathbf{U}'_T)' = -S_{\epsilon_i}^{-1}(Z_{T,1}, ..., Z_{T,q}, \mathbf{Z}'_T)' + o_p(1)$$

For $(Z_{T,1}, ..., Z_{T,q}, \mathbf{Z}'_T)'$:

$$(Z_{T,1}, ..., Z_{T,q}, \mathbf{Z}'_T)' \xrightarrow{d} \mathbb{N}(0, \mathbf{\Sigma})$$

where

$$\mathbf{\Sigma} = \begin{bmatrix} \tau_{11} & ... & \tau_{1q} & E(X_t)' \sum_{k'=1}^{q} \tau_{1k'} \\ \vdots & \ddots & \vdots & \vdots \\ \tau_{q1} & ... & \tau_{qq} & E(X_t)' \sum_{k'=1}^{q} \tau_{qk'} \\ E(X_t) \sum_{k'=1}^{q} \tau_{1k'} & ... & E(X_t) \sum_{k'=1}^{q} \tau_{qk'} & E(X_tX'_t) \sum_{k=1}^{q} \sum_{k'=1}^{q} \tau_{kk'} \end{bmatrix},$$

with $\tau_{kk'} = min(\tau_k, \tau_{k'}) - \tau_k\tau_{k'}$, so $\tau_{kk} = \tau_k(1 - \tau_k)$.

Thus, we have:

$$(U_{T,1}, ..., U_{T,q}, \mathbf{U}'_T)' = \sqrt{T} \begin{bmatrix} \hat{\alpha}_{i1} - \alpha_{i1} \\ \vdots \\ \hat{\alpha}_{iq} - \alpha_{iq} \\ \hat{\beta}_i^{CQR} - \beta_i \end{bmatrix} \xrightarrow{d} \mathbb{N}\left(\mathbf{0}, S_{\epsilon_i}^{-1}\mathbf{\Sigma}S_{\epsilon_i}^{-1}\right).$$

Given the definition $\hat{\alpha}_i^{CQR} = \frac{1}{q}\sum_{k=1}^{q} \hat{\alpha}_{ik}$, the expression above leads to:

$$\sqrt{T}(\hat{\alpha}_i^{CQR} - \alpha_i) \xrightarrow{d} \mathbb{N}\left(0, \frac{1}{q^2} e_q' (S_{\epsilon_i}^{-1} \mathbf{\Sigma} S_{\epsilon_i}^{-1})_{11} e_q\right),$$

where $\alpha_i = \frac{1}{q}\sum_{k=1}^{q} \alpha_{ik}$ results from the equally distributed q quantile positions and the symmetric error distribution.

Special Case 1: If $q = 1$: $\tau_k = 1/2$, $c_{ik} = 0$, $\tau_{kk'} = 1/4$, and

$$S_{\epsilon_i} = \begin{bmatrix} f_{\epsilon_i}(0) & f_{\epsilon_i}(0)E(X_t)' \\ f_{\epsilon_i}(0)E(X_t) & f_{\epsilon_i}(0)E(X_t X_t') \end{bmatrix}, \ \mathbf{\Sigma} = \begin{bmatrix} \frac{1}{4} & \frac{1}{4}E(X_t)' \\ \frac{1}{4}E(X_t) & \frac{1}{4}E(X_t X_t') \end{bmatrix}.$$

Thus,

$$S_{\epsilon_i}^{-1} \mathbf{\Sigma} S_{\epsilon_i}^{-1} = \frac{1}{4 f_{\epsilon_i}^2(0)} \begin{bmatrix} 1 & E(X_t)' \\ E(X_t) & E(X_t X_t') \end{bmatrix}^{-1}.$$

Special Case 2: If $E(X_t) = 0$, then S_{ϵ_i} and $\mathbf{\Sigma}$ are block-diagonal:

$$S_{\epsilon_i} = \begin{bmatrix} f_{\epsilon_i}(c_{i1}) & & & 0 \\ & \ddots & & \vdots \\ & & f_{\epsilon_i}(c_{iq}) & 0 \\ 0 & ... & 0 & \sum_{k=1}^{q} f_{\epsilon_i}(c_{ik}) E(X_t X_t') \end{bmatrix},$$

and

$$\mathbf{\Sigma} = \begin{bmatrix} \tau_{11} & ... & \tau_{1q} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ \tau_{q1} & ... & \tau_{qq} & 0 \\ 0 & ... & 0 & E(X_t X_t') \sum_{k=1}^{q} \sum_{k'=1}^{q} \tau_{kk'} \end{bmatrix}.$$

Thus,

$$S_{\epsilon_i}^{-1} \mathbf{\Sigma} S_{\epsilon_i}^{-1} = \begin{bmatrix} \frac{\tau_{11}}{f_{\epsilon_i}^2(c_{i1})} & ... & \frac{\tau_{iq}}{f_{\epsilon_i}(c_{i1}) f_{\epsilon_i}(c_{iq})} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ \frac{\tau_{q1}}{f_{\epsilon_i}(c_{iq}) f_{\epsilon_i}(c_{i1})} & ... & \frac{\tau_{qq}}{f_{\epsilon_i}^2(c_{iq})} & 0 \\ 0 & ... & 0 & [E(X_t X_t')]^{-1} \frac{\sum_{k=1}^{q} \sum_{k'=1}^{q} \tau_{kk'}}{(\sum_{k=1}^{q} f_{\epsilon_i}(c_{ik}))^2} \end{bmatrix}$$

Theorem 2.1 of Zou and Yuan (2008) is on the lower-right $L \times L$ submatrix of $S_{\epsilon_i}^{-1}\mathbf{\Sigma}S_{\epsilon_i}^{-1}$:

$$\sqrt{T}(\hat{\beta}_i^{CQR} - \beta_i) \xrightarrow{d} \mathbb{N}\left(0, [E(X_t X_t')]^{-1}\frac{\sum_{k=1}^q \sum_{k'=1}^q \tau_{kk'}}{(\sum_{k=1}^q f_{\epsilon_i}(c_{ik}))^2}\right)$$

while we focus on the upper-left $q \times q$ submatrix of of $S_{\epsilon_i}^{-1}\mathbf{\Sigma}S_{\epsilon_i}^{-1}$.

D. Proof of Theorem 2

Given Theorem 1, we only need to derive the covariance expression in Theorem 2. The proof of Theorem 1 shows that, for the i-th equation:

$$\sqrt{T}\begin{bmatrix} \hat{\alpha}_{i1} - \alpha_{i1} \\ \vdots \\ \hat{\alpha}_{iq} - \alpha_{iq} \\ \hat{\beta}_i^{CQR} - \beta_i \end{bmatrix} = -S_{\epsilon_i}^{-1}(Z_{T,1}, ..., Z_{T,q}, \mathbf{Z}_T')' + o_p(1)$$

with $Z_{T,k} \equiv \frac{1}{\sqrt{T}}\sum_{t=1}^T(1(\epsilon_{it} < c_{ik}) - \tau_k)$, $\mathbf{Z}_T \equiv \frac{1}{\sqrt{T}}\sum_{t=1}^T X_t[\sum_{k=1}^q(1(\epsilon_{it} < c_{ik}) - \tau_k)]$.

For the j-th equation, we thus similarly have:

$$\sqrt{T}\begin{bmatrix} \hat{\alpha}_{j1} - \alpha_{j1} \\ \vdots \\ \hat{\alpha}_{jq} - \alpha_{jq} \\ \hat{\beta}_j^{CQR} - \beta_j \end{bmatrix} = -S_{\epsilon_j}^{-1}(\mathcal{Z}_{T,1}, ..., \mathcal{Z}_{T,q}, \boldsymbol{\mathcal{Z}}_T')' + o_p(1)$$

with $\mathcal{Z}_{T,k} \equiv \frac{1}{\sqrt{T}}\sum_{t=1}^T(1(\epsilon_{jt} < c_{jk}) - \tau_k)$, $\boldsymbol{\mathcal{Z}}_T \equiv \frac{1}{\sqrt{T}}\sum_{t=1}^T X_t[\sum_{k=1}^q(1(\epsilon_{jt} < c_{jk}) - \tau_k)]$.

The asymptotic covariance of $(Z_{T,1}, ..., Z_{T,q}, \mathbf{Z}_T')'$ and $(\mathcal{Z}_{T,1}, ..., \mathcal{Z}_{T,q}, \boldsymbol{\mathcal{Z}}_T')'$ is:

$$\mathbf{\Sigma}_{\epsilon_i\epsilon_j} = \begin{bmatrix} \tau_{ij,11} & \cdots & \tau_{ij,1q} & E(X_t)'\sum_{k'=1}^q \tau_{ij,1k'} \\ \vdots & \ddots & \vdots & \vdots \\ \tau_{ij,q1} & \cdots & \tau_{ij,qq} & E(X_t)'\sum_{k'=1}^q \tau_{ij,qk'} \\ E(X_t)\sum_{k=1}^q \tau_{ij,k1} & \cdots & E(X_t)\sum_{k=1}^q \tau_{ij,kq} & E(X_t X_t')\sum_{k=1}^q\sum_{k'=1}^q \tau_{ij,kk'} \end{bmatrix},$$

with $\tau_{ij,kk'} = F_{\epsilon_i\epsilon_j}(c_{ik}, c_{jk'}) - \tau_k\tau_{k'}$, and $F_{\epsilon_i\epsilon_j}$ is the joint c.d.f. of ϵ_{it} and ϵ_{jt}.

Thus, the covariance of $-S_{\epsilon_i}^{-1}(Z_{T,1}, ..., Z_{T,q}, \mathbf{Z}_T')'$ and $-S_{\epsilon_j}^{-1}(\mathcal{Z}_{T,1}, ..., \mathcal{Z}_{T,q}, \boldsymbol{\mathcal{Z}}_T')'$ is $S_{\epsilon_i}^{-1}\mathbf{\Sigma}_{\epsilon_i\epsilon_j}S_{\epsilon_j}^{-1}$, whose upper-left $q \times q$ matrix is for the covariance of the estimated alphas: $\hat{\alpha}_{i1}, ..., \hat{\alpha}_{iq}, \hat{\alpha}_{j1}, ..., \hat{\alpha}_{jq}$. The covariance $\frac{1}{q^2} e_q' (S_{\epsilon_i}^{-1} \boldsymbol{\Sigma}_{\epsilon_i \epsilon_j} S_{\epsilon_j}^{-1})_{11} e_q$ in Theorem 2 thus results from using the definition of the CQR estimator: $\hat{\alpha}_i^{CQR} = \frac{1}{q} \sum_{k=1}^{q} \hat{\alpha}_{ik}$, and similarly, $\hat{\alpha}_j^{CQR} = \frac{1}{q} \sum_{k=1}^{q} \hat{\alpha}_{jk}$.

Since $\hat{\alpha}_i^{CQR}$ and $\hat{\alpha}_j^{CQR}$ are driven by their error terms ϵ_{it} and ϵ_{jt}, respectively, Theorem 2 involves the joint distribution of these two error terms. It is worth noting that we do not assume the same error distribution across equations, i.e., ϵ_{it} and ϵ_{jt} are allowed to follow different distributions.

Special Case 3: It is straightforward to verify that $\tau_{ij,kk'}$ reduces to zero if $\epsilon_{it}, \epsilon_{jt}$ are independent. In this case, $\boldsymbol{\Sigma}_{\epsilon_i \epsilon_j}$ reduces to a zero matrix, so the off-diagonal covariance in (12) equals zero, implying that $\hat{\alpha}_i^{CQR}$ and $\hat{\alpha}_j^{CQR}$ are asymptotically independent. On the other hand, $\tau_{ij,kk'}$ reduces to $\tau_{kk'}$ if $i = j$, so $\boldsymbol{\Sigma}_{\epsilon_i \epsilon_j}$ reduces to $\boldsymbol{\Sigma}$, and the off-diagonal covariance in (12) coincides with the variance on the diagonal.

Special Case 4: If we impose $E(X_t) = 0$, then (12) becomes:

$$\sqrt{T} \begin{bmatrix} \hat{\alpha}_i^{CQR} - \alpha_i \\ \hat{\alpha}_j^{CQR} - \alpha_j \end{bmatrix} \xrightarrow{d} \mathbb{N} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{1}{q^2} \sum_{k=1}^{q} \sum_{k'=1}^{q} \frac{\tau_{kk'}}{f_{\epsilon_i}(c_{ik}) f_{\epsilon_i}(c_{ik'})} & \frac{1}{q^2} \sum_{k=1}^{q} \sum_{k'=1}^{q} \frac{\tau_{ij,kk'}}{f_{\epsilon_i}(c_{ik}) f_{\epsilon_j}(c_{jk'})} \\ \frac{1}{q^2} \sum_{k=1}^{q} \sum_{k'=1}^{q} \frac{\tau_{ij,kk'}}{f_{\epsilon_i}(c_{ik}) f_{\epsilon_j}(c_{jk'})} & \frac{1}{q^2} \sum_{k=1}^{q} \sum_{k'=1}^{q} \frac{\tau_{kk'}}{f_{\epsilon_j}(c_{jk}) f_{\epsilon_j}(c_{jk'})} \end{bmatrix} \right),$$

which clearly shows that the distributions of two error terms affect the joint behavior of $\hat{\alpha}_i^{CQR}$ and $\hat{\alpha}_j^{CQR}$.

E. Proof of Theorem 3

Given that Theorem 2 establishes the joint normal distribution of alpha estimators, we only need to show the covariance estimator $\widehat{Var}(\hat{\boldsymbol{\alpha}}^{CQR})$ used in Theorem 3 is consistent:

$$\widehat{Var}(\hat{\boldsymbol{\alpha}}^{CQR}) = \begin{bmatrix} \frac{1}{q^2} e_q' (\hat{S}_{\epsilon_1}^{-1} \hat{\boldsymbol{\Sigma}} \hat{S}_{\epsilon_1}^{-1})_{11} e_q & \dots & \frac{1}{q^2} e_q' (\hat{S}_{\epsilon_1}^{-1} \hat{\boldsymbol{\Sigma}}_{\epsilon_1 \epsilon_N} \hat{S}_{\epsilon_N}^{-1})_{11} e_q \\ \vdots & \ddots & \vdots \\ \frac{1}{q^2} e_q' (\hat{S}_{\epsilon_N}^{-1} \hat{\boldsymbol{\Sigma}}_{\epsilon_N \epsilon_1} \hat{S}_{\epsilon_1}^{-1})_{11} e_q & \dots & \frac{1}{q^2} e_q' (\hat{S}_{\epsilon_N}^{-1} \hat{\boldsymbol{\Sigma}} \hat{S}_{\epsilon_N}^{-1})_{11} e_q \end{bmatrix}.$$

We use $\hat{S}_{\epsilon_i}$ (and similarly $\hat{S}_{\epsilon_j}$), $\hat{\boldsymbol{\Sigma}}$, and $\hat{\boldsymbol{\Sigma}}_{\epsilon_i \epsilon_j}$ as follows:

$$\hat{S}_{\epsilon_i} = \begin{bmatrix} \hat{f}_{\epsilon_i}(\hat{c}_{i1}) & & & \hat{f}_{\epsilon_i}(\hat{c}_{i1}) \frac{1}{T} \sum_{t=1}^{T} X_t' \\ & \ddots & & \vdots \\ & & \hat{f}_{\epsilon_i}(\hat{c}_{iq}) & \hat{f}_{\epsilon_i}(\hat{c}_{iq}) \frac{1}{T} \sum_{t=1}^{T} X_t' \\ \hat{f}_{\epsilon_i}(\hat{c}_{i1}) \frac{1}{T} \sum_{t=1}^{T} X_t & \dots & \hat{f}_{\epsilon_i}(\hat{c}_{iq}) \frac{1}{T} \sum_{t=1}^{T} X_t & \sum_{k=1}^{q} \hat{f}_{\epsilon_i}(\hat{c}_{ik}) \frac{1}{T} \sum_{t=1}^{T} X_t X_t' \end{bmatrix},$$

$$
\hat{\boldsymbol{\Sigma}} = \begin{bmatrix} \tau_{11} & \cdots & \tau_{1q} & \frac{1}{T}\sum_{t=1}^{T} X_t' \sum_{k'=1}^{q} \tau_{1k'} \\ \vdots & \ddots & \vdots & \vdots \\ \tau_{q1} & \cdots & \tau_{qq} & \frac{1}{T}\sum_{t=1}^{T} X_t' \sum_{k'=1}^{q} \tau_{qk'} \\ \frac{1}{T}\sum_{t=1}^{T} X_t \sum_{k'=1}^{q} \tau_{1k'} & \cdots & \frac{1}{T}\sum_{t=1}^{T} X_t \sum_{k'=1}^{q} \tau_{qk'} & \frac{1}{T}\sum_{t=1}^{T} X_t X_t' \sum_{k=1}^{q}\sum_{k'=1}^{q} \tau_{kk'} \end{bmatrix},
$$

and

$$
\hat{\boldsymbol{\Sigma}}_{\epsilon_i\epsilon_j} = \begin{bmatrix} \hat{\tau}_{ij,11} & \cdots & \hat{\tau}_{ij,1q} & \frac{1}{T}\sum_{t=1}^{T} X_t' \sum_{k'=1}^{q} \hat{\tau}_{ij,1k'} \\ \vdots & \ddots & \vdots & \vdots \\ \hat{\tau}_{ij,q1} & \cdots & \hat{\tau}_{ij,qq} & \frac{1}{T}\sum_{t=1}^{T} X_t' \sum_{k'=1}^{q} \hat{\tau}_{ij,qk'} \\ \frac{1}{T}\sum_{t=1}^{T} X_t \sum_{k=1}^{q} \hat{\tau}_{ij,k1} & \cdots & \frac{1}{T}\sum_{t=1}^{T} X_t \sum_{k=1}^{q} \hat{\tau}_{ij,kq} & \frac{1}{T}\sum_{t=1}^{T} X_t X_t' \sum_{k=1}^{q}\sum_{k'=1}^{q} \hat{\tau}_{ij,kk'} \end{bmatrix},
$$

with $\hat{c}_{ik} = \hat{F}_{\epsilon_i}^{-1}(\tau_k)$, and $\hat{F}_{\epsilon_i}$, $\hat{f}_{\epsilon_i}$ are the estimated c.d.f. and p.d.f. based on the CQR residual estimate $\hat{\epsilon}_{it}^{CQR}$ of ϵ_{it}, $\hat{\tau}_{ij,kk'} = P(\hat{\epsilon}_{it}^{CQR} \leq \hat{c}_{ik}$ and $\hat{\epsilon}_{jt}^{CQR} \leq \hat{c}_{jk'}) - \tau_k\tau_{k'}$.

Since the CQR estimators of α_i and β_i are $\sqrt{T}$-consistent, the resulting residual $\hat{\epsilon}_{it}^{CQR} = \epsilon_{it} + O_p(T^{-1/2})$. The p.d.f. f_{ϵ_i} and c.d.f. F_{ϵ_i} can thus be consistently estimated by $\hat{f}_{\epsilon_i}$ and $\hat{F}_{\epsilon_i}$ using residuals, which further leads to the consistency of $\hat{S}_{\epsilon_i}$ and $\hat{\boldsymbol{\Sigma}}_{\epsilon_i\epsilon_j}$. The consistency of $\hat{\boldsymbol{\Sigma}}$ directly results from the first regularity condition.

F. Proof of Theorem 4

$\tilde{\alpha}_i^{CQR} = \hat{\mu}_{Y_i} - \hat{\mu}_X'\hat{\beta}_i^{CQR}$ is a smooth function of $\hat{\mu}_{Y_i}$, $\hat{\mu}_X$, and $\hat{\beta}_i^{CQR}$, where $\hat{\mu}_{Y_i}$, $\hat{\mu}_X$, and $\hat{\beta}_i^{CQR}$ are all asymptotically normally distributed. In particular, the asymptotic variance of $\hat{\beta}_i^{CQR}$ is $(S_{\epsilon_i}^{-1}\boldsymbol{\Sigma}S_{\epsilon_i}^{-1})_{22}$.

Note that $\hat{\mu}_{Y_i}$, $\hat{\mu}_X$ converge to $E(Y_{it})$ and $E(X_t)$ respectively, and $\hat{\beta}_i^{CQR}$ converges to β_i (see Zou and Yuan (2008)). This implies that:

$$
\tilde{\alpha}_i^{CQR} = \hat{\mu}_{Y_i} - \hat{\mu}_X'\hat{\beta}_i^{CQR} \xrightarrow{p} E(Y_{it}) - E(X_t)'\beta_i = \alpha_i
$$

In addition, $\frac{\partial \alpha_i}{\partial E(Y_{it})} = 1$, $\frac{\partial \alpha_i}{\partial E(X_t)} = -\beta_i$, and $\frac{\partial \alpha_i}{\partial \beta_i} = -E(X_t)$. By the delta method, the asymptotic variance of $\tilde{\alpha}_i^{CQR}$ equals:

$$(1, -\beta_i', -E(X_t)')V_{\hat{\mu}_{Y_i}, \hat{\mu}_X, \hat{\beta}_i^{CQR}}(1, -\beta_i', -E(X_t)')'$$

where $V_{\hat{\mu}_{Y_i}, \hat{\mu}_X, \hat{\beta}_i^{CQR}}$ stands for the joint covariance of $\hat{\mu}_{Y_i}$, $\hat{\mu}_X$, $\hat{\beta}_i^{CQR}$, which reads:

$$V_{\hat{\mu}_{Y_i}, \hat{\mu}_X, \hat{\beta}_i^{CQR}} = \begin{bmatrix} var(Y_{it}) & cov(Y_{it}, X_t)' & 0 \\ cov(Y_{it}, X_t) & var(X_t) & 0 \\ 0 & 0 & (S_{\epsilon_i}^{-1}\mathbf{\Sigma}S_{\epsilon_i}^{-1})_{22} \end{bmatrix}.$$

Completing the calculation above leads to $E(X_t)'(S_{\epsilon_i}^{-1}\mathbf{\Sigma}S_{\epsilon_i}^{-1})_{22}E(X_t) + var(\epsilon_{it})$ for the asymptotic variance expression of $\tilde{\alpha}_i^{CQR}$.

G. Proof of Theorem 5

$\tilde{\alpha}_i^{CQR} = \hat{\mu}_{Y_i} - \hat{\mu}_X'\hat{\beta}_i^{CQR} = \alpha_i + \bar{\epsilon}_i - \hat{\mu}_X'(\hat{\beta}_i^{CQR} - \beta_i)$, with $\bar{\epsilon}_i = \frac{1}{T}\sum_{t=1}^T \epsilon_{it}$. Similarly, $\tilde{\alpha}_j^{CQR} = \hat{\mu}_{Y_j} - \hat{\mu}_X'\hat{\beta}_j^{CQR} = \alpha_j + \bar{\epsilon}_j - \hat{\mu}_X'(\hat{\beta}_j^{CQR} - \beta_j)$, with $\bar{\epsilon}_j = \frac{1}{T}\sum_{t=1}^T \epsilon_{jt}$.

The proof of Theorem 2 above implies that the asymptotic covariance of $\hat{\beta}_i^{CQR} - \beta_i$ and $\hat{\beta}_j^{CQR} - \beta_j$ is $(S_{\epsilon_i}^{-1}\mathbf{\Sigma}_{\epsilon_i\epsilon_j}S_{\epsilon_j}^{-1})_{22}$, which further leads to the covariance expression $E(X_t)'(S_{\epsilon_i}^{-1}\mathbf{\Sigma}_{\epsilon_i\epsilon_j}S_{\epsilon_j}^{-1})_{22}E(X_t) + cov(\epsilon_{it}, \epsilon_{jt})$ for $\tilde{\alpha}_i^{CQR} - \alpha_i$, and $\tilde{\alpha}_j^{CQR} - \alpha_j$.

If $i = j$, the covariance expression $E(X_t)'(S_{\epsilon_i}^{-1}\mathbf{\Sigma}_{\epsilon_i\epsilon_j}S_{\epsilon_j}^{-1})_{22}E(X_t) + cov(\epsilon_{it}, \epsilon_{jt})$ reduces to the variance expression $E(X_t)'(S_{\epsilon_i}^{-1}\mathbf{\Sigma}S_{\epsilon_i}^{-1})_{22}E(X_t) + var(\epsilon_{it})$ in Theorem 4.

H. Proof of Theorem 6

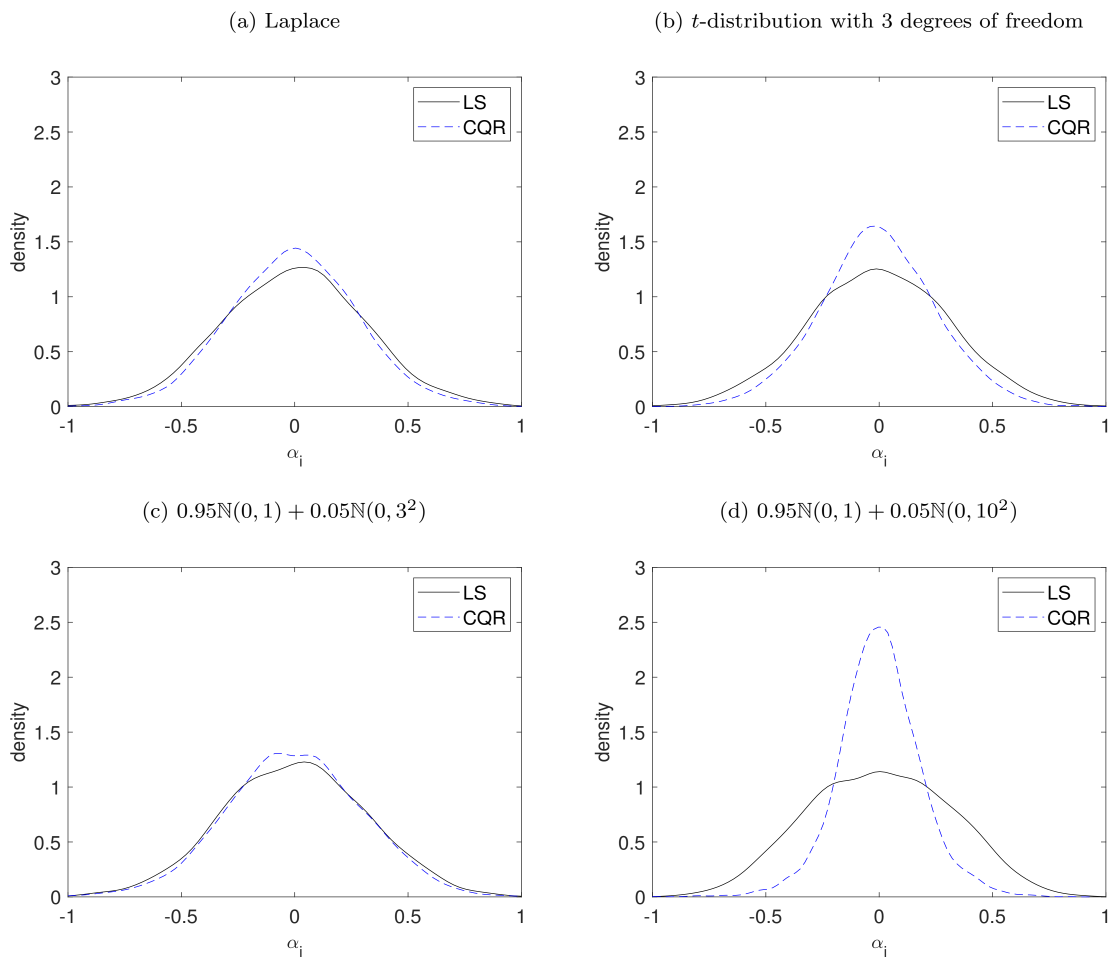
Given the joint normal distribution of alpha estimators, we only need to show the covariance estimator $\widehat{Var}(\tilde{\boldsymbol{\alpha}}^{CQR})$ is consistent:

$$\widehat{Var}(\tilde{\boldsymbol{\alpha}}^{CQR}) = \begin{bmatrix} \hat{\mu}_X'(\hat{S}_{\epsilon_1}^{-1}\hat{\mathbf{\Sigma}}\hat{S}_{\epsilon_1}^{-1})_{22}\hat{\mu}_X + \widehat{Var}(\epsilon_1) & \dots & \hat{\mu}_X'(\hat{S}_{\epsilon_1}^{-1}\hat{\mathbf{\Sigma}}_{\epsilon_1\epsilon_N}\hat{S}_{\epsilon_N}^{-1})_{22}\hat{\mu}_X + \widehat{Cov}(\epsilon_1, \epsilon_N) \\ \vdots & \ddots & \vdots \\ \hat{\mu}_X'(\hat{S}_{\epsilon_N}^{-1}\hat{\mathbf{\Sigma}}_{\epsilon_N\epsilon_1}\hat{S}_{\epsilon_1}^{-1})_{22}\hat{\mu}_X + \widehat{Cov}(\epsilon_N, \epsilon_1) & \dots & \hat{\mu}_X'(\hat{S}_{\epsilon_N}^{-1}\hat{\mathbf{\Sigma}}\hat{S}_{\epsilon_N}^{-1})_{22}\hat{\mu}_X + \widehat{Var}(\epsilon_N) \end{bmatrix}.$$

Consistency of $\hat{\mu}_X$, $\hat{S}_{\epsilon_i}$, $\hat{\mathbf{\Sigma}}_{\epsilon_i\epsilon_j}$, and $\hat{\mathbf{\Sigma}}$ is as in the proof for Theorem 3. Consistency of $\widehat{Var}(\epsilon_i)$ and $\widehat{Cov}(\epsilon_i, \epsilon_j)$ results from using the CQR residual $\tilde{\epsilon}_{it}^{CQR} = Y_{it} - \tilde{\alpha}_i^{CQR} - X_t'\hat{\beta}_i^{CQR} = \epsilon_{it} + O_p(T^{-1/2})$.

I. Additional numerical and empirical results

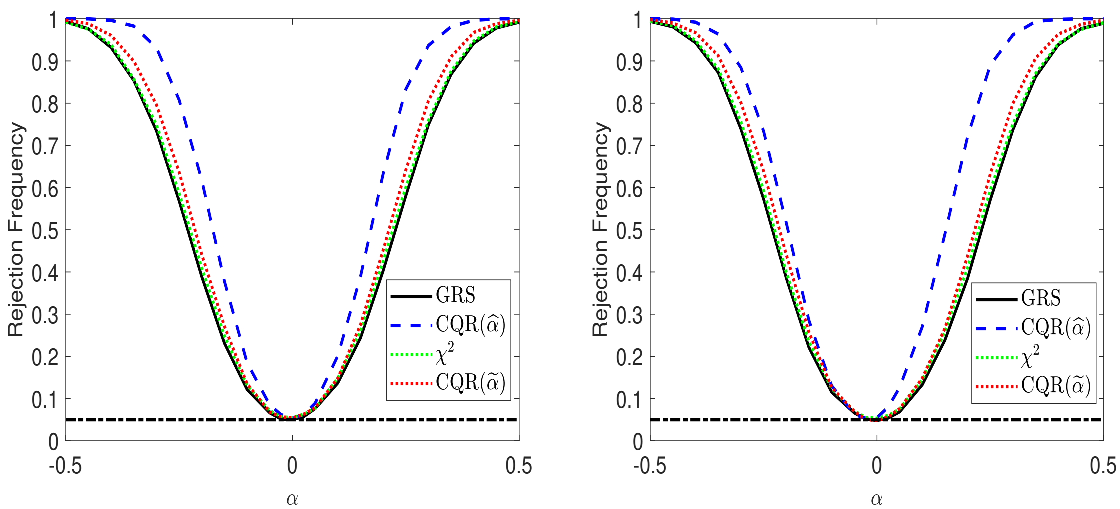
Figure A1: Distributions of $\hat{\alpha}_i^{LS}$ and $\hat{\alpha}_i^{CQR}$ under nonnormalities, $T = 100$



Notes: See also Figure 1 in the main text. The true value of α_i is zero in the data generating process of $Y_{it} = \alpha_i + \beta_i' X_t + \epsilon_{it}$. The solid black line is the simulated density of the least squares estimator $\hat{\alpha}_i^{LS}$, while the dashed blue line is the simulated density of the CQR estimator $\hat{\alpha}_i^{CQR}$. $X_t \sim \mathbb{N}(\mu_X, \Omega_X)$, and μ_X, Ω_X, β_i are all calibrated to data. ϵ_{it} is drawn from non-normal distributions taken from Kai, Li, and Zou (2010) as well as Huang and Zhan (2022): (a) Laplace; (b) t-distribution with 3 degrees of freedom; (c) $0.95\mathbb{N}(0,1) + 0.05\mathbb{N}(0,3^2)$; (d) $0.95\mathbb{N}(0,1) + 0.05\mathbb{N}(0,10^2)$. The variance of ϵ_{it} is then re-scaled to match empirical data. The sample size T is 100, while the number of Monte Carlo replications is 5000.

Figure A2: Power comparison of four tests for testing $H_0 : \boldsymbol{\alpha} = \mathbf{0}$ at the 5% level

(a) Symmetric (b) Asymmetric



Notes: The solid black line is the simulated power curve of the GRS test for testing $H_0 : \boldsymbol{\alpha} = \mathbf{0}$ at the 5% significance level. The dashed blue line is the simulated power curve of the CQR-based test using $\hat{\boldsymbol{\alpha}}^{CQR}$. The dotted green line is the simulated power curve of the asymptotic χ^2 test using $\hat{\boldsymbol{\alpha}}^{LS}$. The dotted red line is the simulated power curve of the CQR-based test using $\tilde{\boldsymbol{\alpha}}^{CQR}$. The benchmark 5% line (black dash-dotted) is also provided to illustrate the 5% size at $H_0 : \boldsymbol{\alpha} = \mathbf{0}$. For the data generating process of (1), $X_t \sim \mathbb{N}(\mu_X, \Omega_X)$, and μ_X, Ω_X, β_i are all calibrated to data. For (a) Symmetric, and (b) Asymmetric, they correspond to the settings in Figure 3. $\boldsymbol{\alpha}$ is set to zero at the null, and $\boldsymbol{\alpha} \propto \mathbf{1}$ under the alternative, so the scalar alpha value makes the horizontal line in this figure. The number of test assets N is 2, and the sample size T is 500. The power curves result from the average of 5000 Monte Carlo replications.

Table A1: Spanning tests by regressing Fama-French factors on q factors: sensitivity analysis

	GRS			CQR($\hat{\boldsymbol{\alpha}}^{CQR}$)		
	GRS-stat	p-value	$A\lvert\alpha_i\rvert$	CQR-stat	p-value	$A\lvert\alpha_i\rvert$
Panel A: HML, RMW, CMA, UMD on q						
$H_0 : \alpha_{UMD} = 0$	1.78	0.18	0.203	6.79	0.01	0.333
$H_0 : \alpha_{CMA} = \alpha_{UMD} = 0$	1.03	0.36	0.111	7.84	0.02	0.180
$H_0 : \alpha_{RMW} = \alpha_{CMA} = \alpha_{UMD} = 0$	1.10	0.35	0.087	9.08	0.03	0.134
$H_0 : \alpha_{HML} = \alpha_{RMW} = \alpha_{CMA} = \alpha_{UMD} = 0$	0.83	0.51	0.066	9.89	0.04	0.121
Panel B: HML, RMW, CMA, UMD on q^5						
$H_0 : \alpha_{UMD} = 0$	0.07	0.79	0.043	0.08	0.77	0.039
$H_0 : \alpha_{CMA} = \alpha_{UMD} = 0$	0.69	0.50	0.042	0.62	0.73	0.032
$H_0 : \alpha_{RMW} = \alpha_{CMA} = \alpha_{UMD} = 0$	0.54	0.66	0.043	1.21	0.75	0.036
$H_0 : \alpha_{HML} = \alpha_{RMW} = \alpha_{CMA} = \alpha_{UMD} = 0$	0.41	0.80	0.036	2.17	0.70	0.046

Notes: This table corresponds to Table 3 in the main text. While Table 3 sets $q = 5$ for CQR, this table sets $q = 3$ (three quantiles). For GRS and CQR, this table reports their test statistics, p-values, and the mean absolute value of estimated alphas $A|\alpha_i|$. The null hypothesis is that the alphas of Fama-French factors (HML, RMW, CMA, UMD) are zero, when regressing them on q factors for Panel A, or q^5 factors for Panel B. We consider the single $H_0 : \alpha_{UMD} = 0$, as well as three joint cases: $H_0 : \alpha_{CMA} = \alpha_{UMD} = 0$; $H_0 : \alpha_{RMW} = \alpha_{CMA} = \alpha_{UMD} = 0$; $H_0 : \alpha_{HML} = \alpha_{RMW} = \alpha_{CMA} = \alpha_{UMD} = 0$. The q-factor model uses R_MKT, R_ME, R_IA, R_ROE, while the q^5 model adds R_EG to the q-factor model. The sample is from January 1967 to December 2022 with $T = 672$ as in Table 2.

Table A2: Spanning tests by regressing Fama-French factors on q factors: alternative tests

	$\chi^2(\hat{\alpha}^{LS})$			$\mathrm{CQR}(\tilde{\boldsymbol{\alpha}}^{CQR})$		
	GRS-stat	p-value	$A\lvert\alpha_i\rvert$	CQR-stat	p-value	$A\lvert\alpha_i\rvert$
Panel A: HML, RMW, CMA, UMD on q						
$H_0 : \alpha_{UMD} = 0$	1.23	0.27	0.203	1.40	0.24	0.176
$H_0 : \alpha_{CMA} = \alpha_{UMD} = 0$	1.44	0.49	0.111	2.61	0.27	0.107
$H_0 : \alpha_{RMW} = \alpha_{CMA} = \alpha_{UMD} = 0$	2.98	0.39	0.087	5.93	0.12	0.100
$H_0 : \alpha_{HML} = \alpha_{RMW} = \alpha_{CMA} = \alpha_{UMD} = 0$	3.02	0.56	0.066	6.03	0.20	0.090
Panel B: HML, RMW, CMA, UMD on q^5						
$H_0 : \alpha_{UMD} = 0$	0.05	0.82	0.043	0.51	0.47	0.111
$H_0 : \alpha_{CMA} = \alpha_{UMD} = 0$	1.34	0.51	0.042	0.76	0.69	0.064
$H_0 : \alpha_{RMW} = \alpha_{CMA} = \alpha_{UMD} = 0$	1.53	0.67	0.043	2.24	0.52	0.076
$H_0 : \alpha_{HML} = \alpha_{RMW} = \alpha_{CMA} = \alpha_{UMD} = 0$	1.57	0.81	0.036	2.27	0.69	0.069

Notes: This table corresponds to Table 3 in the main text, while using two alternative tests. Instead of the GRS test and the CQR test using $\hat{\boldsymbol{\alpha}}^{CQR}$ in Table 3, this table uses the asymptotic counterpart of the GRS test (denoted by $\chi^2(\hat{\boldsymbol{\alpha}}^{LS})$) with White standard errors, and the CQR test using $\tilde{\boldsymbol{\alpha}}^{CQR}$. For these two alternative tests, this table reports their test statistics, p-values, and the mean absolute value of estimated alphas $A|\alpha_i|$. The null hypothesis is that the alphas of Fama-French factors (HML, RMW, CMA, UMD) are zero, when regressing them on q factors for Panel A, or q^5 factors for Panel B. We consider the single $H_0 : \alpha_{UMD} = 0$, as well as three joint cases: $H_0 : \alpha_{CMA} = \alpha_{UMD} = 0$; $H_0 : \alpha_{RMW} = \alpha_{CMA} = \alpha_{UMD} = 0$; $H_0 : \alpha_{HML} = \alpha_{RMW} = \alpha_{CMA} = \alpha_{UMD} = 0$. The q-factor model uses R_MKT, R_ME, R_IA, R_ROE, while the q^5 model adds R_EG to the q-factor model. The sample is from January 1967 to December 2022 with $T = 672$ as in Table 2.