Fused LASSO as Non-Crossing Quantile Regression*

Tibor Szendrei†
Department of Economics, Heriot-Watt University, UK.
National Institute of Economic and Social Research, UK.

Arnab Bhattacharjee
Department of Economics, Heriot-Watt University, UK.
National Institute of Economic and Social Research, UK.

Mark E. Schaffer
Department of Economics, Heriot-Watt University, UK.

February 3, 2026

Abstract

This paper establishes a formal equivalence between non-crossing constraints in quantile regression and Fused LASSO regularisation with quantile-specific hyperparameters. This equivalence implies that imposing non-crossing constraints implicitly induces interquantile shrinkage, positioning the estimator of Bondell et al. (2010) as one point on a bias-variance trade-off spectrum. We propose an adaptive framework that nests unconstrained quantile regression, non-crossing quantile regression, and composite quantile regression as special cases. Monte Carlo experiments demonstrate that cross-validation reliably selects regularisation intensities that outperform both unconstrained estimation and fully constrained alternatives. An empirical application to Fama-French factor models augmented with a market downturn indicator finds evidence of quantile-varying factor loadings, consistent with the directional predictability documented by Linton and Whang (2007).

Keywords: Interquantile Shrinkage, Crossing Quantile Curves, High-Dimensional Econometrics, Fama-French Factor Model.

*The authors thank Paul Allanson, Isaiah Andrews, Atanas Christev, István Járási, David Kohns, Rod McCrorie, Katalin Varga, and participants of the 2022 and 2023 PhD conference in Crieff, and the 2025 Money, Macro and Finance Society's Annual Conference for their feedback. Tibor Szendrei thanks the ESRC for PhD studentship as well as Heriot-Watt University for institutional support. The usual disclaimer applies.

†Corresponding author: t.szendrei@niesr.ac.uk

1 Introduction

Moving beyond the mean to understand the full distribution of economic and financial outcomes has become an increasingly important in theory and applications. Quantile regression, originally proposed by Koenker and Bassett (1978), is an important approach to achieve this. The method allows one to model different quantiles of the conditional distribution of an outcome variable analogous to the role of the conditional mean in regular regression models. Quantile regression has emerged as an important tool in applied econometrics, with many applications in both economics and finance; see Adrian et al. (2019) and Engle and Manganelli (2004), among others.

Quantile regression has several advantages which make it appealing in finance applications. It places emphasis on extreme outcomes, it is robust to outliers in the dependent variable and the estimates remain consistent even in the presence of heteroskedasticity (Koenker, 2005). However, there are also some key challenges, particularly that the estimated quantiles curves may cross, violating the monotonicity property that must hold for any valid distribution. While the probability of crossing is asymptotically negligible if the model is correctly specified, it frequently occurs in finite samples either when the conditional distribution has limited variation across quantiles or when the model is misspecified (Koenker, 2005; Koenker and Xiao, 2006).

The literature suggests two primary approaches to address quantile crossing. The first approach was proposed by Chernozhukov et al. (2009, 2010) and involves post-estimation refinement, whereby rearranging the fitted quantiles at each point is a valid way to recover to underlying distribution. This approach is computationally simple and asymptotically valid, but has one key limitation: it only corrects the fitted values, without adjusting the corresponding coefficients. This is a drawback because for many questions it is the quantile coefficients (and not the fitted quantiles) which are of interest. To address this, Bondell et al. (2010) proposed incorporating non-crossing constraints directly into the estimation procedure. By imposing these constraints, this method produces coefficient estimates that inherently yield non-crossing fitted quantiles.

While imposing non-crossing constraints is attractive, previous literature has treated these constraints as a tool for ensuring distributional coherence without fully exploring their impact on the estimated coefficients. Understanding the impact on the coefficients is particularly important since imposing these constraints is equivalent to implicitly assuming that the model is correctly specified. By exploring how non-crossing constraints influence the estimated coeffi-

cients, we can also gain insights into the consequences of imposing them on misspecified models. This motivates the fundamental question of the paper: what precisely is the relationship between non-crossing constraints and the estimated coefficients?

We provide a precise answer to this question by establishing a formal equivalence between non-crossing constraints and Fused-LASSO regularisation. Our main theoretical result shows that non-crossing constraints of Bondell et al. (2010) are mathematically equivalent to a specific form of Fused-LASSO shrinkage with quantile-specific regularisation. In essence, imposing non-crossing constraints leads to interquantile shrinkage, i.e., they penalise variation in the coefficients across the quantile dimension. This reframes non-crossing quantile regression within the broader regularisation literature, connecting it to the literature on penalised quantile methods (Jiang et al., 2013, 2014).

The consequence of this equivalence is that the non-crossing solution should be treated as one point on the bias-variance trade-off spectrum. The implication is that the degree of non-crossing can be optimised, just as one would tune the hyperparameter in LASSO regression. Simply imposing these constraints without any optimisation, we either over- or under- shrink quantile variation, introducing excessive bias into the estimation procedure.

We complement our theoretical analysis with an extensive Monte Carlo simulation following the setup of Bondell et al. (2010). The results reveal a clear bias-variance trade-off: increasing the regularisation parameter (α in our setting) reduces the true positive rate, i.e., detecting quantile variation, while improving the true negative rate, i.e., detecting quantile constant covariates. Our Monte Carlo experiments highlight that using cross-validation to tune the degree of crossing leads to fitted quantiles that are closer to the data generating process than traditional quantile regression and the Bondell et al. (2010) estimator. In our Monte Carlo experiments we also compare tuning the degree of non-crossing constraints with the traditional Fused-LASSO specification of Jiang et al. (2013). Through our theorem we know that this comparison is equivalent to testing global shrinkage against quantile specific shrinkage. We find that the traditional Fused LASSO overshrinks.

The empirical application of this paper examines the Fama-French factor model, connecting our methodological contribution to the growing literature on asymmetric risk pricing. Since Sharpe (1964) and Lintner (1975), factor models have been the workhorse of empirical asset pricing, yet they have been looked at almost exclusively through the conditional mean lens. This implicitly assumes that the factor-return relationship is characterised by location shift effects only, i.e., the factors have no impact on the shape of the returns distribution. This assumption

has been increasingly question in the literature (Linton and Whang, 2007; Han et al., 2016). Recent work by Bollerslev et al. (2022) provides a theoretical foundation for why market beta might load differently depending on the state of the market and the individual asset return. They propose decomposing the traditional beta into four "semibetas" that capture the different dimensions of risk. This semibeta decomposition can be incorporated within a quantile Fama-French factor setup with the addition of a market downturn indicator. In this way we can use quantile regression to split the sample along the dimension of returns, and use the market downturn indicator to split the sample in the market dimension.

We apply our Fused LASSO quantile regression framework to this adjusted five factor model – the Fama-French five factor model augmented with market downturn indicator – using monthly returns for the 50 largest S&P500 stocks over the period 2000 to 2024. This specification allows us to test for quantile variation in factor loadings while controlling for established risk factors. Our results strongly support the semibeta hypothesis: the market downturn coefficient exhibits a pronounced downward-sloping quantile profile for nearly all stocks, indicating that stocks are more sensitive to market declines at lower quantiles of their own return distribution.

Beyond the market downturn factor, our approach to replicating the semibeta framework allows testing whether the other factors have an influence on the shape of the return distribution. We find that the CMA (Conservative Minus Aggressive) factor displays an upward-sloping quantile profile for many stocks, suggesting that firms with aggressive investment strategies face wider conditional return distributions. In contrast, the SMB (Small Minus Big), HML (High Minus Low), and RMW (Robust Minus Weak) factors exhibit relatively sparse quantile variation patterns, consistent with these factors operating primarily as location shifters rather than shape shifters of the conditional return distribution. These findings align with the risk premium function estimates of Bollerslev et al. (2025), who find a flat premium function for SMB and RMW.

A consequence of our shrinkage approach is that it leads to lower crossing incidence. In our empirical application, the quantile-specific Fused LASSO estimator substantially reduces the degree of quantile crossing relative to the unconstrained quantile regression while also preserving quantile variation for key factors. In contrast, the traditional Fused-LASSO formulation (with global shrinkage) leads to more quantile crossing than the quantile-specific Fused LASSO while purging much more quantile variation.

The rest of the paper is organised as follows. Section 2 introduces quantile regression (Koenker and Bassett, 1978) along with the non-crossing constraint of Bondell et al. (2010),

before providing the adaptive non-crossing constraints that vary with a hyperparameter α. Using these new constraints, we show that one can rewrite non-crossing constraints as the Fused LASSO constraint of Jiang et al. (2013). This is followed by a discussion of how the hyperparameter can be chosen. Section 3 describes the Monte Carlo experiment and presents fit and variable selection performance of the different estimators. Section 4 applies the proposed methodology to quantile Fama-French factor models, documenting patterns of quantile variation in factor loadings and draws their implications for asymmetric risk pricing.

2 Methodology

2.1 Quantile Regression

The first building block of the proposed methodology is the quantile regression framework of Koenker and Bassett (1978). Quantile regression is a weighted version of the least absolute deviation regression, and yields lines of best fit that explain different parts of the distribution. The collection of Q estimated quantiles can be used to describe the distribution of a response variable Y conditional on a vector of independent variables (regressors) $X \in \mathbb{R}^K$. Formally, the τ_q-th quantile is modelled in regression setting as:

$$\mathcal{Q}(\tau_q) = X^T \beta_{\tau_q}.$$

The collection of the Q quantiles parameters $\beta = \{\beta_{\tau_1}, \beta_{\tau_2}, ..., \beta_{\tau_Q}\}$ describe the conditional distribution. The goal is to estimate the vector of coefficients $\beta_{\tau_q} \in \mathbb{R}^{K+1}$ for all quantiles.[1] This can be done using quantile regression:

$$\begin{aligned} \hat{\beta} &= \underset{\beta}{argmin} \sum_{q=1}^{Q} \sum_{t=1}^{T} \rho_{\tau_q}(y_t - x_t^T \beta_{\tau_q}) \\ \rho_{\tau_q}(u) &= u(\tau_q - I(u < 0)) \end{aligned} \tag{1}$$

where the second equation is the 'tick-loss' function (Koenker and Bassett, 1978).

Equation (1) provides an estimate for the parameters with which a description of the conditional distribution is obtained, but it is possible that these fitted quantiles cross. Quantile crossing is a violation of monotonocity assumption and often occurs on account of data scarcity

[1] Note that with this notation X includes an intercept.

or model misspecification (Koenker, 2005). Limits in data availability are frequently encounted in practice, particularly in time-series settings. For forecasting applications, the two main methods for addressing quantile crossing are: (1) use the fitted quantiles to fit some distribution for each time period as in Adrian et al. (2019) or Korobilis (2017); or (2) sort the estimated quantiles in each period as proposed by Chernozhukov et al. (2009, 2010). While these two-step methods yield proper densities, correcting the fitted quantiles does not quantify corresponding changes in the estimated coefficients; the coefficients estimated in the first step remain uncorrected. This prompted Bondell et al. (2010) to propose an estimator which yields no crossing for the estimated quantiles in-sample.[2]

2.2 Non-Crossing Constraints

Non-crossing constraints incorporated into quantile regression are a way to ensure that the estimated quantiles remain monotonically increasing. This implies inequality constraints:

$$\begin{aligned} \hat{\beta} &= \underset{\beta}{argmin} \sum_{q=1}^{Q} \sum_{t=1}^{T} \rho_q(y_t - x_t^T \beta_{\tau_q}) \\ s.t.\ & x^T \beta_{\tau_q} \geq x^T \beta_{\tau_{q-1}} \end{aligned} \tag{2}$$

While conceptually simple, the number of constraints in equation (2) can be rather large on account of the $T \times (Q-1)$ inequality constraints. To address this, Bondell et al. (2010) restrict the domain of the covariates to $\mathcal{D} \in [0,1]^K$ and focus on the worst case scenario in the data,[3] which reduces the number of constraints to $(Q-1)$.[4] This simplifies computation considerably and enables non-crossing constraints to be included without much additional computational cost. Because quantile regression is invariant to monotone transformations, any affine invertible transformation can be applied: to obtain the coefficients pertaining to the un-transformed data, it is enough to apply the inverse transformation to the estimated coefficients (Koenker, 2005).

[2]We note that there are other means to estimate non-crossing quantiles like Liu and Wu (2009) who estimate the median first, and sequentially estimate further quantiles, conditional on the previously estimated quantiles not being crossed. While this method will yield non-crossing quantiles, the choice of the first estimated quantile can be arbitrary. As such, in this paper we focus exclusively on non-crossing constraints as proposed in Bondell et al. (2010) since elements of their constraints have been carried over to other quantile estimators; see Yang and Tokdar (2017) for example.

[3]The situation where the negative difference coefficients (γ^-_{j,τ_q}) corresponding variables values are 1 and the positive difference coefficients (γ^+_{j,τ_q}) corresponding variables equal 0.

[4]Any domain of interest which has an affine transformation that maps to $\mathcal{D} \in [0,1]^K$ will work.

The method of Bondell et al. (2010) recasts the parameters in terms of quantile differences: $(\gamma_{0,\tau_1}, ..., \gamma_{K,\tau_1})^T = \beta_{\tau_1}$ and $(\gamma_{0,\tau_q}, ..., \gamma_{K,\tau_q})^T = \beta_{\tau_q} - \beta_{\tau_{q-1}}$ for $q = 2, ..., Q$. With this quantile difference reparametrisation, the constraint in equation (2) becomes $x^T \gamma_{\tau_q} \geq 0$. Note that we can without further assumptions decompose the j^{th} difference as $\gamma_{j,\tau_q} = \gamma^+_{j,\tau_q} - \gamma^-_{j,\tau_q}$, where γ^+_{j,τ_q} is its positive and $-\gamma^-_{j,\tau_q}$ its negative part. For each γ_{j,τ_q} both parts are non-negative and only one part is allowed to be non-zero. With this reparameterisation, along with the restriction of the data to $\mathcal{D} \in [0,1]^K$, the non-crossing constraint can be redefined as:

$$\gamma_{0,\tau_q} \geq \sum_{j=1}^{K} \gamma^-_{j,\tau_q} \ (q = 2, ..., Q) \tag{3}$$

A non-crossing constraint is therefore equivalent to ensuring that the sum of negative shifts do not push the quantile below the change in intercept, which acts as a pure location shifter. Bondell et al. (2010) show that (3) is a necessary as well as sufficient condition for non-crossing quantiles.

2.3 Adaptive Non-Crossing Constraints

We follow the framework and assumptions in Bondell et al. (2010). To derive adaptive non-crossing constraints we first need to recast the non-crossing constraints of equation (2) in a way that does not restrict the domain of interest to $\mathcal{D} \in [0,1]^K$. This is provided by Lemma 1 below.

Lemma 1. *Given that a non-crossing constraint implies that the sum of positive shifters is larger than the sum of negative shifters, in the worst case scenario, these constraints can be reformulated as:*

$$\gamma_{0,\tau_q} + \sum_{j=1}^{K} min(X_j)\gamma^+_{j,\tau_q} \geq \sum_{j=1}^{K} max(X_j)\gamma^-_{j,\tau_q} \tag{4}$$

where X_j is the j^{th} variable in the design matrix.

Proof. Recall the standard non-crossing constraint formulation from Bondell et al. (2010) needs that in the worst case scenario the sum of positive shifters needs to be larger than the sum of negative shifters:

$$\gamma_{0,\tau_q} + \sum_{j=1}^{K} (Z_j = 0) \cdot \gamma^+_{j,\tau_q} \geq \sum_{j=1}^{K} (Z_j = 1) \cdot \gamma^-_{j,\tau_q}$$

where $Z_j \in [0,1]$ is a transformed variable of X_j. Assume that $\max(X_j) > \min(X_j)$ for all $j \in \{1,2,...,K\}$, i.e., $Var(X_j) > 0$ for all j. Consider a design matrix variable X_j with domain $[\min(X_j), \max(X_j)]$. X_j can be normalised using the min-max transformation: $Z_{t,j} = \frac{X_{t,j} - \min(X_j)}{\max(X_j) - \min(X_j)}$. We can express $X_j = min(X_j) + Z_j \cdot [max(X_j) - min(X_j)]$. Instead of expressing the constraint in terms of the transformed variable Z_j, we can write it in terms of the untransformed X_j:

$$\begin{aligned}
&\gamma_{0,\tau_q} + \sum_{j=1}^{K} [\min(X_j) + (Z_j = 0) \cdot (\max(X_j) - \min(X_j))] \gamma^{+}_{j,\tau_q} \\
&\geq \sum_{j=1}^{K} [\min(X_j) + (Z_j = 1) \cdot (\max(X_j) - \min(X_j))] \gamma^{-}_{j,\tau_q}
\end{aligned}$$

This equation can be simplified to recover:

$$\gamma_{0,\tau_q} + \sum_{j=1}^{K} \min(X_j) \gamma^{+}_{j,\tau_q} \geq \sum_{j=1}^{K} \max(X_j) \gamma^{-}_{j,\tau_q}$$

Thus there is an equivalence between this equation and the one in Bondell et al. (2010). □

This new constraint is a sufficient condition for non-crossing, since if equation (4) is satisfied for the worst case (i.e.; when $Z_j = 0$ for variables that shift the quantile curve up and $Z_j = 1$ for variables that shift the quantile curve down), it is immediately satisfied for every observation. The novelty of equation (4) is that it works on the domain of $\mathbb{D} \in \mathbb{R}$, while the original formulation in Bondell et al. (2010), works for $\mathbb{D} \in [0,1]$ only.

To study the impact of non-crossing constraints on the estimated coefficients, it is important to be able to tighten or loosen these constraints. As such, the next step is to formulate a set of constraints that can become adaptive as a hyperparameter α varies. An intuitively appealing formulation is one that yields the traditional quantile regression estimator by setting $\alpha = 0$ and the Bondell et al. (2010) non-crossing estimator when $\alpha = 1$.[5] This leads to the following adaptive non-crossing constraints:

$$\gamma_{0,\tau_q} + \sum_{j=1}^{K} \Big[\bar{X}_j - \alpha(\bar{X}_j - min(X_j)) \Big] \gamma^{+}_{j,\tau_q} \geq \sum_{j=1}^{K} \Big[\bar{X}_j + \alpha(max(X_j) - \bar{X}_j) \Big] \gamma^{-}_{j,\tau_q} \tag{5}$$

When $\alpha = 0$, the constraint simplifies to imposing non-crossing quantiles evaluated at the average value of the covariates. Quantile monotonicity at this value is satisfied even by the

[5] Any scalar $\alpha > 0$ would work. We set $\alpha = 1$ for simplicity and to remain consistent with Bondell et al. (2010).

traditional quantile regression estimator: $\bar{X}^T\beta$ yields the empirical quantiles of Y, which are monotonically increasing by definition (Koenker, 2005; Koenker and Xiao, 2006). As such imposing equation (5) and setting $\alpha = 0$ will yield the same β coefficients as the traditional quantile regression without constraints.

The constraints in equation (5) equal the non-crossing constraints when $\alpha = 1$ as the equation becomes equation (4). As such using equation (5) as a constraint allows us to recover both the traditional quantile regression estimator of Koenker and Bassett (1978) as well as the non-crossing quantile regression estimator of Bondell et al. (2010).

There are other ways to construct adaptive non-crossing constraints that yield traditional quantile regression as well as Bondell et al. (2010) estimator, specifically through the use of indicator functions that activate the non-crossing constraint. The advantage of equation (5) is that it produces a gradual transition from quantile estimates to non-crossing estimates as α increases. This allows us to study the impacts these constraints have on the estimated coefficients.

Theorem 1. *Non-crossing constraints are a type of Fused LASSO, with quantile-specific hyperparameters:* $k^*_{\tau_q} = \frac{\gamma_{0,\tau_q}}{\alpha}$.

Proof. For simplicity, assume $Q = 2$. We begin with the non-crossing constraint given by equation (5):

$$\gamma_0 \geq \sum_{j=1}^{K} \Big[[\bar{X}_j + \alpha(max(X_j) - \bar{X}_j)]\gamma_j^- - [\bar{X}_j - \alpha(\bar{X}_j - min(X_j))]\gamma_j^+ \Big]$$

By Lemma 1, we can rescale the data to $\mathbb{D} \in [-1, 1]$. When the data are properly standardized and rescaled, we can assume $\bar{X}_j = 0$ for each covariate j. For a symmetric distribution, this standardization is trivial; for asymmetric distributions, one can normalise the data prior to rescaling. After normalising and rescaling, $\bar{X}_j \approx 0$, $max(X_j) = 1$, and $min(X_j) = -1$. Then, the constraint simplifies to:

$$\gamma_0 \geq \sum_{j=1}^{K} \alpha(\gamma_j^- + \gamma_j^+)$$

Since $\gamma_j^- \geq 0$ and $\gamma_j^+ \geq 0$ by definition, we have $(\gamma_j^- + \gamma_j^+) = |\gamma_j^- + \gamma_j^+|$. Rearranging terms yields:

$$\frac{\gamma_0}{\alpha} \geq \sum_{j=1}^{K} |\gamma_j^+ + \gamma_j^-|$$

This corresponds exactly to the fused shrinkage formulation in Jiang et al. (2013) with $k^* = \frac{\gamma_0}{\alpha}$.

For $Q > 2$, all γ parameters become quantile-specific, leading to quantile-specific hyperparameters: $k^*_{\tau_q} = \frac{\gamma_{0,\tau_q}}{\alpha}$. As such, non-crossing constraints lead to fused LASSO shrinkage of the parameters, with quantile specific hyperparameters. □

Remark 1. *The relationship between α and the constrained space parallels the role of the hyperparameter in LASSO regression. In LASSO, setting the hyperparameter to 0 recovers the unconstrained estimator, while setting it to a positive value leads to shrinkage. Similarly, Theorem 1 establishes the equivalence between the (non-crossing) constrained space and fused shrinkage for $\alpha > 0$. The boundary case $\alpha = 0$ lies outside this equivalence as this corresponds to the unconstrained space where no fused shrinkage is imposed. One can set $\alpha = 0$, but this will simply recover the unconstrained estimator: traditional quantile regression.*

Remark 2. *There is no a priori reason to privilege $\alpha = 1$ over other values on the positive real line. The optimal α represents a bias-variance trade-off: larger α reduces variance from quantile crossing at the cost of introducing bias from shrinking quantile variation.*

The estimator that uses the adaptive non-crossing constraints, given a value of α can be formally written as:

$$\begin{aligned} \hat{\beta}_{FLQR-QS}|\alpha = \underset{\beta}{argmin} \sum_{q=1}^{Q}\sum_{t=1}^{T} \rho_q(y_t - x_t^T\beta_{\tau_q}) \\ s.t.\ \gamma_{0,\tau_q} + \sum_{k=1}^{K}\Big[\bar{X_k} - \alpha(\bar{X_k} - min(X_k))\Big]\gamma^+_{k,\tau_q} \geq \\ \sum_{k=1}^{K}\Big[\bar{X_k} + \alpha(max(X_k) - \bar{X_k})\Big]\gamma^-_{k,\tau_q} \end{aligned} \tag{6}$$

Corollary 1. *The estimator in equation (6) nests the following estimators as special cases:*

(i) $\alpha = 0$: Unconstrained quantile regression (Koenker and Bassett, 1978)

(ii) $\alpha = 1$: Non-crossing quantile regression (Bondell et al., 2010)

(iii) $\alpha = \infty$: Composite quantile regression (Koenker, 1984; Zou and Yuan, 2008)

Proof. For case (i) setting $\alpha = 0$ gives a constraint of $\bar{X}\beta_{\tau_q} \geq \bar{X}\beta_{\tau_{q-1}}$, which is automatically satisfied by any unconstrained QR solution. Case (ii) follows directly from setting $\alpha = 1$ in equation (5) and recovering equation (4). Case (iii) follows from Theorem 1: as $\alpha \to \infty$, $k^*_{\tau_q} \to 0$, forcing $\gamma_{j,\tau_q} = 0$ for all j and q.[6] □

[6]Note that in the limit there is no unique solution. Because of this, as $\alpha \to \infty$ we encounter numerical instability and it becomes numerically challenging to find the solution.

2.4 Hyperparameter Properties

It is clear from equation (6) that the choice of α plays a central role in the estimation.

Assumption 1. *The tuning parameter is a scalar and assumes non-negative values, $\alpha \geq 0$, for all sample sizes T.*

Assumption (1) of non-negativity excludes the case of negative shrinkage. One might be tempted to restrict potential value of $\alpha \geq 1$, i.e. always enforcing non-crossing quantiles in estimation. We caution against this, because the estimated conditional quantiles are merely an approximation of the true (and unknown) DGP (Koenker and Xiao, 2006). In particular, analytical results in Bondell et al. (2010) ensure that the limiting distribution of the parameters of the non-crossing and regular quantile regression are identical, which only applies when crossing occurs on account of data sparsity (i.e. when $T^{1/2} min_q(\tau_q - \tau_{q-1}) \to \infty$ as $T \to \infty$). In a situation where we have model misspecification, enforcing non-crossing constraints could lead to excessive interquantile shrinkage as a consequence of Theorem (1).[7]

The advantage of the estimator is that it leads to quantile specific shrinkage parameters, while only needing to set the scalar α. In essence, the problem becomes one of model selection, but instead of selecting variables from X, the interest here is on interquantile shrinkage, i.e., selecting variables whose impact on Y captured by β varies by quantile.

We now consider the asymptotic behaviour of the tuning parameter. Let α_T denote the tuning parameter as a function of sample size T.

Assumption 2. *As the sample size $T \to \infty$, the tuning parameter vanishes:* $\lim_{T\to\infty} \alpha_T = 0$.

Assumption (2) ensures that the fused shrinkage penalty becomes asymptotically negligible, allowing the estimator to recover the true quantile parameters. For this assumption we rely on the theorem of Bondell et al. (2010). Without this assumption, the estimator would remain biased even for large T.

Assumption 3. *The penalty decays faster than sample size:* $\sqrt{T}\alpha_T \to 0$.

With assumption (3), the bias introduced by fused penalty is of smaller order than the sampling noise. This ensures that the limiting distribution of the estimator is the same as that of the classical quantile regression.

[7]The CaViaR setting in Szendrei (2025) is an example where traditional QR yields lower coefficient bias than non-crossing constrained estimator.

2.5 Hyperparameter Selection

A natural candidate for choosing the degree of penalisation is cross-validation. In cross-section applications, a natural choice is the usual k-fold cross-validation or variations thereof. Of the several types of cross-validation methods available for the time-series context, we choose the hv-block CV setup of Racine (2000). This block setup has several desirable properties: (1) it has been shown to provide a good trade-off between bias and variance in various applications; (2) the required number of computations does not increase with the number of observations to the degree it would with leave-one-out cross-validation,[8] and (3) when the "h" in the hv-block is set equal to 0, we recover the standard k-fold cross-validation setup, so the method can be used for both cross-section and time-series data.

Figure 1 visualises this block setup and which data are removed from the dataset. The reason to remove the data points around the test datasets in time-series data is to ensure that no data leakage occurs when evaluating the performance of the model. This can lead to overfitting and hence poor generalization. Removing the data around the test dataset mitigates against this.

Since the hyperparameter α is a scalar, one can find the optimal hyperparameter using simple grid search. Hence, the curse of dimensionality that often limits the applicability of grid search is not present here. Crucially, the grid search is 'embarrassingly parallel' and this can be utilised to speed up computation times (Bergstra and Bengio, 2012). As such, throughout the paper we use grid search to obtain the optimal hyperparameter values. When the grid search yields the same testing error for two candidate values, we choose the hyperparameter that is smaller.

We rely on the CV model selection results of Yang (2007) and Wager (2020) to ensure that the correct hyperparameter is selected. This is summarised in the following assumption:

Assumption 4. *Let* $\hat{R}_T(\alpha) = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{Q}\sum_{q=1}^{Q}\rho_{\tau_q}(y_t - x_T^T\beta_{-t,\tau_q}(\alpha))$ *be the average error over leave-one-out (or K-fold) CV estimates for a given* α*. Suppose:*

(i) $\hat{R}_T(\alpha)$ *converges uniformly in probability to* $R(\alpha)$ *with* $\alpha \in [0, A]$

(ii) The true error $R(\alpha_0)$ *has a unique minimiser* $\alpha_0 \in [0, A]$

Then $\hat{\alpha} = argmin_{0\leq\alpha\leq A}\hat{R}_T(\alpha)$ *satisfies* $\hat{\alpha} \xrightarrow{P} \alpha_0$.

[8] See Cerqueira et al. (2020) for a description and comparison of the different types of cross-validation for time-series data and Arlot and Celisse (2010) for a survey of CV procedures for model selection.

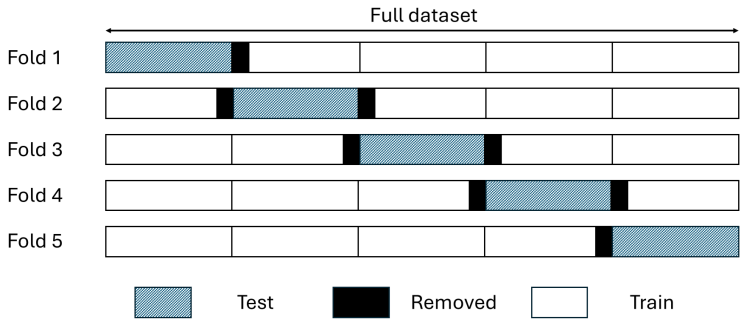

Figure 1: hv-Cross Validation

For further discussion on hyperparameter selection, please refer to the appendix.

3 Monte Carlo experiments

3.1 Setup

Theorem 1 demonstrated the connection between Fused LASSO and non-crossing constraints. One implication of the theorem is that the traditional non-crossing constraints are simply a point on the hyperparameter space and should be treated as something to be optimised. In this section we explore the implications of this using Monte Carlo evidence, and evaluate variable selection properties of the various estimators as well as examine their ability to recover the true quantiles. To this end we will compare the performance of the estimator for three values of α: $\alpha = 0$ which is the traditional quantile estimator (QR), $\alpha = 1$ which is the estimator of Bondell et al. (2010) (BRW), and $\alpha = \alpha_{opt}$ which is estimator using the optimal hyperparameter as obtained with cross validation (FLQR-QS).

Theorem 1 also shows that the equivalence needs quantile-specific hyperparameters. While global hyperparameters could recover non-crossing quantiles, it is likely to achieve this through overshrinking quantile differences. To investigate this, we will also consider the Fused LASSO with global hyperparameter (FLQR-G).[9]

Since the method of Chernozhukov et al. (2009) achieves non-crossing quantiles as well, we will also consider it in our Monte Carlo simulations. These sorted quantiles will be obtained by running the regular quantile regression and sorting the fitted quantiles (QR (sort)). We include

[9]Note that we consider the Fused LASSO and not the Fused Adaptive LASSO in Jiang et al. (2013). We focus on the non-adaptive version as this allows us to examine the value added impact of quantile-specific hyperparameters.

this estimator to evaluate how better variable selection can translate to improved model fit beyond simply sorting the fitted quantiles. We note that because this procedure sorts the fitted quantiles, it will apply only to the model fit results and not the variable selection results.

We consider as starting point the Monte Carlo setup of Bondell et al. (2010). Each Monte Carlo experiment is generated using the location scale heteroskedastic error model of the form:

$$y_t = \beta_0 + \beta^T x_t + (\theta_0 + \vartheta_t \odot \theta^T x_t)\varepsilon_t,\ x_{t,k} \sim U(0,1),\ \varepsilon_t \sim N(0,1) \tag{7}$$

An intercept is included in each setup (i.e., $\beta_0 = \theta_0 = 1$). Note the k^{th} element of $\vartheta_t \in \{0, 1\}$ which regulates whether the given variable has quantile variation at the given quantile. This term is included to allow 'quantile varying sparsity', i.e. cases where a certain variable enters only parts of the distribution (Kohns and Szendrei, 2021). A simple way to implement quantile-specific sparsity is by setting ϑ_t as an indicator function, where it takes the value of 1 only for cases when ε_t is below (or above) a specific quantile. For cases where there is full quantile variation, $\vartheta_t = \mathbf{1}_k$. The t subscript is needed since the presence of quantile variation will be dependent on the magnitude of ε_t.

We consider four different data generating processes (DGPs), each with 500 generated datasets. The first three (y_1, y_2, and y_3) are identical to Examples 1-3 of Bondell et al. (2010). The fourth DGP (y_4) is a variation on Example 2, with some variables only varying at the tails. Specifically:

- y_1: $k = 4$, with the parameters $\beta = \mathbf{1}_k$, $\theta = 0.1 * \mathbf{1}_k$, and $\vartheta = \mathbf{1}_k$.
- y_2: $k = 10$, with the parameters $\beta = (\mathbf{1}_4^T, \mathbf{0}_6^T)^T$, $\theta = (0.1 * \mathbf{1}_4^T, \mathbf{0}_6^T)^T$, and $\vartheta = \mathbf{1}_k$.
- y_3: $k = 7$, with the parameters $\beta = \mathbf{1}_k$, $\theta = (\mathbf{1}_3^T, \mathbf{0}_4^T)^T$, and $\vartheta = \mathbf{1}_k$.
- y_4: $k = 10$, with the parameters $\beta = (\mathbf{1}_4^T, \mathbf{0}_6^T)^T$, $\theta = (0.1 * \mathbf{1}_8^T, \mathbf{0}_2^T)^T$, and $\vartheta = (\mathbf{1}_4^T, \mathbf{1}_4^T \times [I(\epsilon_t \geq F_\varepsilon^{-1}(0.9)) + I(\epsilon_t \leq F_\varepsilon^{-1}(0.1))], \mathbf{0}_2^T)^T$.

Note that for y_4, variable selection and fused shrinkage will both be necessary. The models considered will only allow for fused shrinkage, and as such this DGP is only included to judge the performance of the estimators in non-ideal situations.

To evaluate the performance of the different estimators three sample sizes are considered for each DGP: $T \in \{50, 100, 200\}$. Sample sizes 100 and 200 were also considered in Bondell

et al. (2010), but 50 was not. We include this small sample setting because for rolling-window applications it is often necessary to apply quantile regression to such small samples.[10]

We also consider variation in the number of quantiles to be estimated, by generating equidistant quantiles with varying distance between the quantiles: $\Delta_\tau \in \{0.2, 0.1\}$.[11] Note that by increasing the number of quantiles, the number of estimated parameters varies. The choice to vary the estimated quantiles was driven by the fact that Theorem 1 shows how γ_0 acts as a quantile-specific slackness – a claim that can be verified by examining the variable selection properties of BRW as the number of quantiles increases.

Two measures are used to judge the variable selection performance of the estimators: True Positive Rate (TPR) and True Negative Rate (TNR). Both measures take a value between 0 (worst performance possible) and 1 (best performance possible). TPR measures the degree to which the estimator is able to capture the quantile varying coefficients, while the TNR measures the ability of the estimator to identify where no quantile variation occurs. Considering these measures together allows one to conclude whether a given estimator over-shrinks or under-shrinks. When calculating these measures, we only consider the difference in β coefficients of the variables (without the intercept).

To measure model fit to data, we follow Bondell et al. (2010) and report the average root mean integrated square error ($\times 100$):

$$RMISE = \Big[\frac{1}{n}\sum_{iter=1}^{n}\Big(\hat{g}_\tau(x_{iter}) - g_\tau(x_{iter})\Big)^2\Big]^{1/2}$$

where *iter* indexes a given Monte Carlo experiment, $n = 500$ is the total number of evaluated Monte Carlo experiments for each DGP, $\hat{g}_\tau$ is the estimated quantile and g_τ is the true quantile given by equation (7). We also report the standard error of the RMISE for all the estimators.

Grid search is done separately for values of α below and above 1. We create a grid of 100 equally-spaced points between $[0, 1)$, and use grid search to obtain the optimal α parameter. To consider large α options, we append to this list of candidate solutions 200 points between $[0, 6]$ as exponents of base 10. Setting the hyperparameter to 10^6 yields solutions that are close to the composite quantile regression solution. For FLQR-G and FLQR-QS, we use 10-fold

[10] It is also common to run quantile regression on such small samples for macro-financial applications. See Szendrei and Varga (2023) for an application with around 50 observations, Figueres and Jarociński (2020) for an application with around 100 observations, and Adrian et al. (2019) for an application with around 200 observations.

[11] For $\Delta_\tau = 0.2$, the first quantile is set to 0.1.

cross-validation.

Note that although Theorem 1 implies that changing the α parameter will have a similar impact on the β's as the λ of a Fused LASSO estimator, the same α and λ values will not lead to the same β coefficients. This is because the quantile-specific difference in constants constitutes an upper limit of quantile variation that is scaled by α for FLQR-QS.

Table 1 reports the variable selection performance based on the Monte Carlo experiments, while Table 2 presents the results for the average goodness-of-fit of the different estimators. The RMISE set of results for y_1, y_2, and y_3 when T=100 (and T=200 for y_3) and $\Delta\tau = 0.2$ is the closest setup to Bondell et al. (2010). The results for these setups for BRW and QR are almost identical and as such our new findings are extensions to the Monte Carlo experiments of Bondell et al. (2010).

We also investigate the impact of varying α more explicitly after discussing variable selection and goodness of fit. Specifically we will show how the TPR, TNR, and average quantile score (QS) behave as we increase the hyperparameter from 0 to 10. The average quantile score is calculated by averaging the average tick-loss value for each quantile, i.e. the average of quantile regression objective function.

3.2 Variable Selection results

The results on variable selection presented in Table 1 are particularly revealing. Since the FLQR-QS can recover both the BRW (when setting $\alpha = 1$) and the QR (as $\alpha \to 0$), we can compare the performance of these estimators to see the impact α has on the coefficient profiles. Given Theorem 1, we expect the TPR to decrease and the TNR to increase as α increases. We find that the TPR of BRW is below 1 and the TNR is above 0 for all DGPs, consistent with the implication of Theorem 1 that non-crossing constraints are a special type of fused shrinkage. This is not simply a feature of the Monte Carlo design, as the QR yields a TPR of 1 and a TNR of 0 in all cases. Note that for y_1, all variables are quantile-varying and as such TNR does not exist for this case.

Among the estimators that have some fused shrinkage, BRW yields the highest TPR for all DGPs for all cases considered. FLQR-QS always ranks second and FLQR-G has the worst performance for TPR. However, this superior performance in TPR for BRW is coupled with the worst performance when it comes to TNR. For TNR, FLQR-G produces the best results,

Table 1: True Positive and True Negative Rates for the different Monte Carlo experiments

	y_1	y_2		y_3		y_4		y_1	y_2		y_3		y_4	
	TPR	TPR	TNR	TPR	TNR	TPR	TNR	TPR	TPR	TNR	TPR	TNR	TPR	TNR
	$\Delta\tau = 0.2$							$\Delta\tau = 0.1$						
T-50														
BRW	0.806	0.360	0.639	0.535	0.525	0.305	0.691	0.618	0.246	0.752	0.383	0.669	0.224	0.776
FLQR-QS	0.372	0.232	0.769	0.345	0.715	0.243	0.753	0.304	0.195	0.812	0.271	0.788	0.173	0.825
QR	1.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	1.000	0.000	1.000	0.000	1.000	0.000
FLQR-G	0.215	0.136	0.868	0.246	0.790	0.194	0.819	0.127	0.107	0.901	0.154	0.870	0.114	0.888
T=100														
BRW	0.938	0.510	0.501	0.694	0.393	0.390	0.582	0.792	0.363	0.641	0.531	0.557	0.318	0.687
FLQR-QS	0.364	0.228	0.780	0.430	0.689	0.279	0.715	0.294	0.183	0.828	0.349	0.767	0.179	0.833
QR	1.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	1.000	0.000	1.000	0.000	1.000	0.000
FLQR-G	0.202	0.122	0.885	0.299	0.776	0.184	0.834	0.110	0.074	0.927	0.196	0.860	0.078	0.932
T=200														
BRW	0.988	0.659	0.340	0.846	0.262	0.481	0.459	0.923	0.488	0.517	0.690	0.445	0.409	0.575
FLQR-QS	0.381	0.225	0.787	0.564	0.672	0.280	0.700	0.325	0.166	0.846	0.465	0.765	0.203	0.797
QR	1.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	1.000	0.000	1.000	0.000	1.000	0.000
FLQR-G	0.215	0.130	0.883	0.413	0.746	0.209	0.821	0.139	0.061	0.946	0.298	0.832	0.098	0.917

with FLQR-QS a close second. From these results we can see that BRW undershrinks quantile variation, FLQR-G overshrinks, and FLQR-QS yields a middle-ground option.

As will be seen in the model fit results of Table 2, we find that FLQR-QS not only provides robustness over FLQR-G for y_4, but is also able to provide fits close to FLQR-G while retaining better model selection properties. In particular, FLQR-QS is able to yield much better TPR than FLQR-G without substantially compromising its ability to identify the true negative differences.

Increasing the number of quantiles has a marked impact on variable selection: it lowers TPR and increases TNR for all estimators (except the QR). This further corroborates Theorem 1. Increasing sample size also influences the TPR and TNR of all estimators. For BRW and FLQR-QS, increasing the sample size increases TPR but lowers TNR for all DGPs and all Δ_τ. However, for FLQR-G we see similar pattern when $\Delta_\tau = 0.2$ but not when $\Delta_\tau = 0.1$. In particular, for FLQR-G for y_4, the TPR decreases as the sample size increases when $\Delta_\tau = 0.1$. This is particularly troubling since FLQR-G has the worst TPR of all the estimators.

3.3 Goodness-of-Fit results

Results on model fit in Table 2 reveal that FLQR-QS and FLQR-G provide the best fits, even beating the BRW estimator, for y_1, y_2, and y_3.[12] However, for y_4, FLQR-G fails to yield

[12]Since the quantile specific QR fits are unaffected by the number of quantiles being estimated, we do not repeat their values for $\Delta_\tau = 0.1$.

improvements over BRW. While FLQR-QS also faces challenges with model fit in y_4, it is much closer to the fits of BRW, with both estimators yielding better fits than the traditional QR or FLQR-G. This is because FLQR-QS recovers BRW when $\alpha = 1$, so it does not do much worse than BRW. Hence, FLQR-QS is more robust to DGPs that have quantile-specific sparsity than the simple FLQR-G.

Unsurprisingly, y_4 produces the worst fit for all estimators, but as more data becomes available, the performance of all the estimators initially improves. The key difference lies with FLQR-G, where the fits do not improve as much as the other estimators when the sample size increases from $T = 100$ to $T = 200$. This also highlights that to yield improvements in the traditional LASSO setting for such DGPs one needs to explicitly penalise the level of β_τ coefficients too; see also Jiang et al. (2014). Overall, increasing sample size improves the fit for all estimators, while increasing the fineness of the grid of quantiles yields no significant differences. The key takeaway is that FLQR-QS either provides improvements in fit over BRW, or (at worst) does as well as BRW.

Considering the rearrangement method proposed by Chernozhukov et al. (2010) reveals that sorting the fitted quantiles for the quantile regression estimator yields modest improvements. These improvements increase as the number of estimated quantiles increases especially for smaller sample sizes. This finding is intuitive, since increasing the number of quantiles enhances the chance of quantile crossing. However, the improvements in fit from rearrangement become smaller as the sample size increases.

Comparing the rearrangement method with the alternative estimators shows that although rearrangement helps improve fit, the alternative estimators still perform better than the sorted QR. This highlights how better interquantile variable selection translates into superior fit beyond simple quantile sorting.

Note that in all the Monte Carlo runs, we have estimated a correctly specified model. In such situations the BRW estimator will (almost) always yield improvements over the QR. When the estimated model is misspecified, imposing strict non-crossing constraints can lead to worse coefficient bias than the QR. The fact that the FLQR-QS allows for some quantile crossing when $\alpha < 1$ means that the estimator is more robust to misspecification than the BRW. In essence, when $0 < \alpha_{opt} < 1$ the FLQR-QS provides the best linear approximation of the quantiles while regularising some of the quantile variation in the coefficients.

Table 2: RIMSE of different models for the different Monte Carlo experiments

	T=50										T=100										T=200									
τ	0.1		0.3		0.5		0.7		0.9		0.1		0.3		0.5		0.7		0.9		0.1		0.3		0.5		0.7		0.9	
	bias	std. err	bias	std. err	bias	std. err	bias	std. err	bias	std. err	bias	std. err	bias	std. err	bias	std. err	bias	std. err	bias	std. err	bias	std. err	bias	std. err	bias	std. err	bias	std. err	bias	std. err
$\Delta\tau = 0.2$																														
y_1																														
BRW	55.3	0.76	44.3	0.65	42.8	0.62	44.8	0.66	56.5	0.80	41.7	0.58	32.9	0.47	30.3	0.45	32.0	0.49	41.1	0.61	30.4	0.45	23.5	0.33	22.1	0.32	23.2	0.33	30.4	0.42
FLQR-QS	46.5	0.72	41.7	0.61	40.8	0.61	41.9	0.63	48.2	0.78	33.6	0.54	29.2	0.44	27.8	0.42	29.3	0.45	33.6	0.57	23.9	0.38	20.9	0.30	20.0	0.29	20.5	0.30	24.3	0.39
QR	60.7	0.84	46.8	0.68	45.3	0.65	48.0	0.71	62.4	0.93	43.9	0.63	33.9	0.50	31.2	0.47	33.0	0.49	43.5	0.66	30.8	0.46	23.8	0.33	22.4	0.33	23.5	0.33	31.0	0.44
QR (sort)	59.9	0.82	46.4	0.67	45.0	0.64	47.6	0.69	61.6	0.91	43.7	0.62	33.9	0.49	31.2	0.47	32.9	0.49	43.3	0.65	30.8	0.46	23.8	0.33	22.4	0.33	23.5	0.33	30.9	0.44
FLQR-G	46.6	0.72	41.8	0.61	40.6	0.61	42.1	0.65	48.2	0.77	33.2	0.55	29.5	0.45	27.9	0.44	29.3	0.45	34.2	0.59	24.1	0.39	21.0	0.29	20.1	0.30	20.6	0.30	24.9	0.41
y_2																														
BRW	73.4	0.69	63.0	0.58	61.2	0.57	63.1	0.62	73.2	0.72	55.0	0.51	45.9	0.44	44.6	0.43	46.4	0.42	54.4	0.52	40.8	0.39	33.1	0.32	31.7	0.30	33.2	0.32	40.7	0.37
FLQR-QS	69.4	0.75	61.9	0.60	60.1	0.58	62.0	0.63	68.3	0.76	48.7	0.53	43.8	0.42	43.3	0.42	44.0	0.43	48.5	0.53	34.3	0.38	30.6	0.31	29.9	0.29	30.8	0.30	33.8	0.37
QR	94.5	0.84	72.2	0.70	67.6	0.67	72.3	0.73	94.8	0.90	67.1	0.65	51.9	0.47	49.3	0.47	52.2	0.49	67.9	0.66	47.6	0.47	36.5	0.35	34.6	0.33	36.7	0.36	47.2	0.45
QR (sort)	90.2	0.81	70.7	0.66	66.2	0.63	70.5	0.68	90.7	0.86	65.8	0.64	51.3	0.46	48.6	0.46	51.5	0.47	66.7	0.64	47.3	0.46	36.3	0.35	34.5	0.32	36.6	0.35	47.0	0.44
FLQR-G	68.5	0.74	61.8	0.61	59.9	0.59	61.7	0.62	69.3	0.78	47.9	0.54	43.6	0.43	43.3	0.42	44.0	0.43	49.5	0.55	33.8	0.38	30.6	0.32	30.1	0.30	30.9	0.31	34.7	0.41
y_3																														
BRW	136.8	1.62	110.4	1.28	108.3	1.28	114.2	1.36	137.6	1.70	101.3	1.19	79.7	0.94	76.2	0.83	80.8	0.91	100.0	1.19	75.7	0.83	58.8	0.65	54.8	0.62	57.6	0.65	74.0	0.83
FLQR-QS	132.5	1.71	109.4	1.30	106.9	1.27	112.5	1.32	133.9	1.69	98.6	1.22	78.0	0.92	74.7	0.83	80.0	0.92	98.6	1.16	75.2	0.83	57.2	0.63	52.9	0.62	56.9	0.65	74.4	0.80
QR	167.7	2.00	124.1	1.38	118.2	1.39	127.8	1.49	165.7	1.95	116.8	1.37	87.8	1.01	82.7	0.91	89.3	0.97	115.2	1.37	83.1	0.96	62.8	0.71	58.1	0.65	61.8	0.70	81.1	0.91
QR (sort)	163.3	1.92	121.8	1.32	116.0	1.34	124.9	1.43	161.4	1.90	115.7	1.35	87.1	0.98	82.0	0.89	88.5	0.95	113.9	1.34	82.9	0.95	62.6	0.70	58.0	0.65	61.6	0.70	80.9	0.91
FLQR-G	133.7	1.74	109.7	1.30	106.9	1.29	113.5	1.37	136.4	1.67	99.9	1.18	78.3	0.92	75.8	0.86	81.7	0.92	100.1	1.13	77.9	0.85	57.4	0.64	53.8	0.62	58.6	0.66	74.5	0.86
y_4																														
BRW	225.1	2.40	85.9	1.61	77.3	1.27	85.5	1.66	224.8	2.29	194.4	2.28	53.1	0.66	48.4	0.55	52.9	0.63	197.5	2.22	176.9	2.32	37.4	0.40	33.6	0.36	37.4	0.38	180.5	2.34
FLQR-QS	224.7	3.06	85.5	1.66	77.8	1.36	85.4	1.78	222.9	2.97	196.5	3.18	52.2	0.68	48.6	0.59	51.9	0.66	199.8	3.06	181.6	3.28	36.1	0.40	34.1	0.37	36.0	0.39	186.0	3.33
QR	279.7	1.74	100.0	2.34	72.5	1.09	98.9	2.32	282.4	1.78	231.0	1.42	54.5	0.81	48.7	0.51	54.7	0.75	232.2	1.44	199.5	1.50	36.7	0.41	34.2	0.35	36.8	0.37	201.6	1.52
QR (sort)	269.2	1.71	99.7	2.22	75.8	1.14	98.8	2.20	271.2	1.70	225.7	1.40	55.5	0.77	49.3	0.50	55.9	0.72	226.9	1.44	196.8	1.50	37.9	0.40	34.6	0.34	37.9	0.36	199.0	1.51
FLQR-G	228.3	3.40	85.7	1.66	79.0	1.38	85.8	1.84	228.6	2.91	200.9	3.55	51.9	0.67	49.2	0.61	51.2	0.65	205.7	3.06	195.3	3.39	35.4	0.40	34.0	0.37	35.1	0.38	192.2	3.11
$\Delta\tau = 0.1$																														
y_1																														
BRW	54.3	0.76	44.0	0.64	42.8	0.62	44.8	0.66	55.6	0.78	40.6	0.57	32.5	0.46	29.8	0.44	31.6	0.49	40.0	0.59	29.7	0.44	23.3	0.33	21.8	0.33	22.9	0.32	29.7	0.42
FLQR-QS	46.2	0.72	41.6	0.61	40.7	0.61	41.7	0.64	47.6	0.76	33.5	0.53	29.2	0.45	27.7	0.42	28.9	0.44	32.7	0.54	24.0	0.37	20.9	0.29	19.9	0.29	20.3	0.29	23.7	0.37
QR (sort)	59.1	0.81	46.1	0.67	44.8	0.63	47.2	0.68	60.7	0.89	43.4	0.62	33.7	0.49	31.0	0.47	32.7	0.48	43.0	0.65	30.7	0.46	23.8	0.33	22.3	0.33	23.5	0.33	30.9	0.44
FLQR-G	46.4	0.72	41.5	0.61	40.7	0.60	42.0	0.64	48.1	0.81	32.5	0.53	29.0	0.44	27.8	0.42	29.1	0.44	33.2	0.56	24.2	0.39	20.9	0.29	20.0	0.29	20.3	0.30	24.0	0.38
y_2																														
BRW	73.2	0.69	63.1	0.59	61.4	0.58	63.2	0.61	72.3	0.71	54.3	0.50	45.9	0.43	44.6	0.43	46.1	0.43	53.6	0.52	40.0	0.38	32.9	0.32	31.6	0.30	33.0	0.32	39.9	0.37
FLQR-QS	69.9	0.74	62.4	0.61	60.6	0.59	62.5	0.63	69.1	0.78	48.8	0.53	43.8	0.42	43.2	0.43	43.9	0.43	48.6	0.54	33.6	0.36	30.3	0.30	29.7	0.29	30.5	0.30	33.3	0.35
QR (sort)	88.1	0.80	69.8	0.65	65.7	0.62	69.7	0.68	88.8	0.84	64.6	0.62	50.7	0.45	48.1	0.45	50.8	0.47	65.5	0.62	46.8	0.46	36.0	0.34	34.2	0.32	36.3	0.35	46.4	0.44
FLQR-G	69.6	0.76	62.3	0.63	60.3	0.60	62.4	0.63	69.6	0.79	48.0	0.53	43.9	0.43	43.1	0.42	43.9	0.43	48.8	0.55	33.1	0.35	30.3	0.31	29.6	0.29	30.4	0.30	33.3	0.38
y_3																														
BRW	135.2	1.62	110.6	1.26	108.0	1.25	113.6	1.36	137.3	1.67	100.1	1.18	79.6	0.91	76.2	0.82	80.5	0.89	98.0	1.19	73.9	0.82	57.9	0.64	54.1	0.62	57.3	0.67	73.0	0.81
FLQR-QS	130.9	1.67	109.7	1.25	106.5	1.27	112.1	1.32	133.3	1.65	97.5	1.20	77.8	0.90	74.1	0.82	78.2	0.90	97.3	1.14	74.4	0.80	56.8	0.63	52.7	0.61	56.2	0.65	73.1	0.80
QR (sort)	160.1	1.89	120.4	1.31	115.0	1.32	123.4	1.40	158.8	1.87	114.1	1.33	86.3	0.97	81.3	0.88	87.7	0.93	112.4	1.32	82.3	0.94	62.3	0.70	57.7	0.64	61.3	0.69	80.4	0.90
FLQR-G	132.6	1.73	110.9	1.29	106.7	1.30	113.5	1.34	135.2	1.64	100.6	1.19	78.6	0.96	74.6	0.84	79.4	0.90	98.0	1.14	78.2	0.83	57.4	0.64	53.4	0.62	57.8	0.65	74.4	0.85
y_4																														
BRW	224.3	2.52	84.2	1.63	75.5	1.26	84.1	1.67	223.2	2.36	193.5	2.37	50.9	0.64	47.0	0.53	51.1	0.64	196.9	2.30	175.4	2.43	35.6	0.41	32.5	0.34	35.4	0.36	179.4	2.45
FLQR-QS	222.0	3.33	82.6	1.64	75.4	1.29	82.9	1.78	217.1	3.22	190.6	3.53	49.1	0.63	47.1	0.56	49.0	0.62	194.5	3.33	179.8	3.51	33.7	0.39	32.3	0.35	33.4	0.37	182.9	3.54
QR (sort)	263.1	1.70	98.3	2.19	74.0	1.12	97.6	2.19	265.2	1.70	222.3	1.40	54.8	0.78	48.5	0.49	55.1	0.73	223.6	1.44	194.9	1.49	37.2	0.41	34.1	0.34	37.2	0.36	197.2	1.50
FLQR-G	226.2	3.60	84.5	1.84	75.5	1.30	82.7	1.80	222.7	3.27	194.0	3.75	48.7	0.60	47.3	0.57	48.6	0.62	199.9	3.36	190.5	3.64	33.1	0.37	32.1	0.35	33.0	0.36	190.0	3.41

In summary, FLQR-QS and FLQR-G provides better fit than BRW for y_1, y_2, and y_3, while for y_4 BRW performs best but FLQR-QS is nearly as good, and both are preferred over the FLQR-G. This robustness of FLQR-QS is attractive in quantile applications as it is difficult to know *ex ante* whether quantile specific sparsity is present in the DGP. The variable selection results demonstrate that BRW undershrinks while FLQR-G overshrinks. Taking the goodness-of-fit and variable selection results together, we conclude that FLQR-QS provides good fit while retaining better variable selection properties than FLQR-G.

3.4 Consequences of varying α

Figure 2 shows how the TPR, TNR and average QS change as the hyperparameter α is changed for FLQR-QS when T=100 and $\Delta\tau = 0.1$. In this figure we also show where the non-crossing solution is ($\alpha = 1$). We note that the QS in these figures is not the RMISE measure used in goodness-of-fit, since it is not calculated with the true quantiles in mind. Instead it is the value of the quantile regression objective function averaged across the quantiles. Just like for the other Monte Carlo experiments, to generate this figure, 500 Monte Carlo iterations were done, and the averages (across the iterations) is shown. Note that y_1 has no TNR line on account of all variables being quantile varying.

The figures show clearly that as the hyperparameter α is increased, the TPR rate decreases, and the TNR rate increases as expected for any estimator with shrinkage included. The left side of the figure simply verifies that the FLQR-QS estimator behaves as it is expected from reguralised estimators, i.e., as the hyperparameter is increased, the TPR decreases and TNR increases as variation in coefficients becomes more heavily penalised.

The right side of the figure shows that as we increase the degree of shrinkage, the quantile score gradually increases. This is also expected behaviour for regularised estimators: as the degree of shrinkage is increased the in-sample fit degrades on account of introducing bias. Although bias is increased, as TNR increases the variance of the estimator decreases, as unnecessary variation is purged from the model. As such this figure simply confirms the presence of the bias-variance trade-off.

Note that the non-crossing solution of $\alpha = 1$ is just a point on this bias-variance trade-off continuum. While, it is true that this point guarantees non-crossing quantiles in-sample, it does not have any properties that make it a dominant solution in the bias-variance trade-off. While this particular parameterisation guarantees non-crossing quantiles in-sample, our results

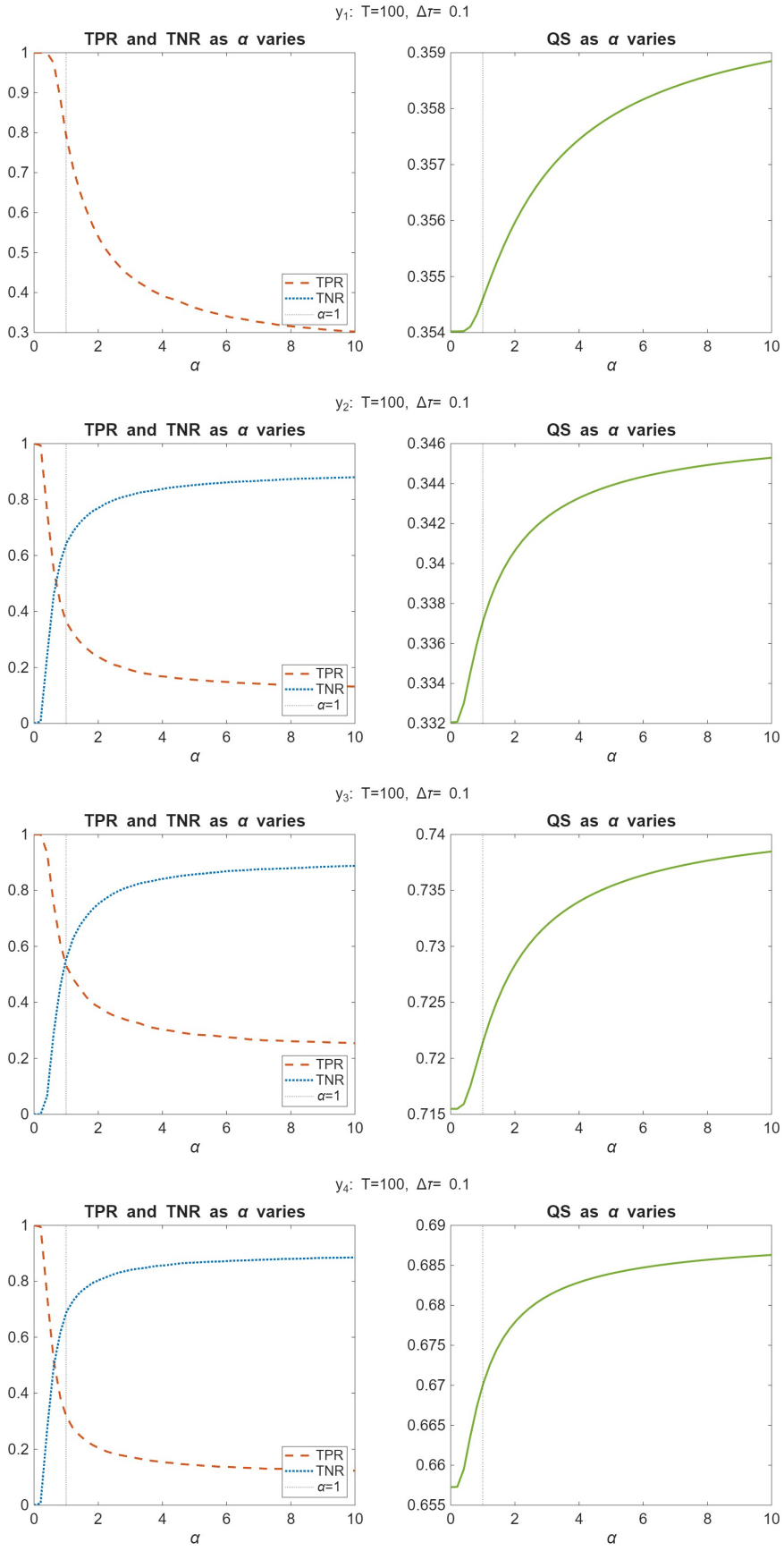

Figure 2: Average TPR, TNR and Quantile Score (QS) for various α parameters.

demonstrate that it does not constitute a uniformly optimal solution. This can be seen from tables (1) and (2), where $\alpha = \alpha_{opt}$ (the rows labelled FLQR-QS) and $\alpha = 1$ (the rows labelled BRW) had different outcomes. In essence, when opting to use non-crossing constraints one chooses to impose a degree of shrinkage implicitly rather than being optimised with respect to out-of-sample fit or variable selection criteria. Consequently, non-crossing constraints should be treated as one option within the regularisation toolkit (albeit a useful option) rather than as a default estimation strategy.

4 Quantile Fama-French Factor Models

Since the Capital Asset Pricing Model of Sharpe (1964) and Lintner (1975), there has been an effort to model how risk is priced on the market. A seminal contribution in this direction is the Fama-French factor model, which decomposes asset returns into specific explanatory factors. The three factor model explains excess returns as a function of market risk (MKT), size (SMB: Small Minus Big), and value (HML: High Minus Low) (Fama and French, 1993). In Fama and French (2015) two additional factors are added, namely profitability (RMW: Robust Minus Weak), and investment (CMA: Conservative Minus Aggressive).

These factors have become a mainstay in financial econometrics, and their application has permeated the entire finance literature. However, examples of the combination of factor models with quantile regression are sparse, and factor models have been almost exclusively implemented using ordinary least squares methods. This implicitly assumes that factor-return relationships are constant across the return distribution. Yet there is evidence that dynamics in the return are quantile-specific. For instance, in Linton and Whang (2007), quantilograms are developed to test the serial dependence at specific quantiles, and whether past extreme events help predict future lower performance. Applied to the S&P 500, the authors find directional predictability varies across the quantiles. If return dynamics are quantile-specific, then the explanatory models using factors should also accommodate quantile-varying relationships.

Han et al. (2016) apply a cross-quantilogram methodology to dependence between two time series across their respective quantile levels. They find that high variance predict left-tail returns more strongly than it predicts median returns. This asymmetry, where the variance forecasts crashes in asset prices better, would be invisible in a mean regression framework. This further motivates utilising quantile regression to estimate quantile specific factor loadings.

Applying quantile regression to the Fama-French factors is not novel by itself. For example,

Allen and Powell (2011) applied quantile regression to Dow Jones stocks. They find that factors do not necessarily have a constant relationship with returns across quantiles. Similarly, de la O González and Jareño (2019) find a large degree of quantile variation in the US stock market, with stocks varying in factor loadings across quantiles and business cycle phases. They also find that to capture tail behaviour, one needs more comprehensive factor models, especially during market downturns. These findings establish an empirical regularity motivates further investigation: why should factor loadings vary across quantiles, and what does such variation imply for asset pricing?

In a recent important contribution, Bollerslev et al. (2022) addressed this gap. They decompose the traditional market beta into four "semibetas" based on the signs of both asset and market returns: β^P (Market+, Stock+), β^N (Market−, Stocks−), β^+ (Market+, Stocks−), β^- (Market−, Stocks+). Using this decomposition Bollerslev et al. (2022) analyse asymmetries in risk pricing. They find that stocks with higher β^N earn significantly higher subsequent returns, while stocks with high β^P earn no significant premium. This highlights an asymmetry not only along the market dimension, but also along the quantile dimension. One can recast the semibeta framework in a quantile regression of the following form:

$$\begin{aligned} r_{i,t} - r_{f,t} =& \beta_i^{Intercept} + \beta_i^{MKT}(r_{m,t} - r_{f,t}) + \beta_i^{DOWN}(r_{m,t} - r_{f,t})I(r_{m,t} - r_{f,t} < 0) \\ &+ \beta_i^{SMB} SMB_t + \beta_i^{HML} HML_t + \beta_i^{RMW} RMW_t + \beta_i^{CMA} CMA_t + \epsilon_{i,t} \end{aligned} \tag{8}$$

where $I(\cdot)$ is an indicator function. This specification motivates a piecewise linear relationship with the market where the factor loading equal β_i^{MKT} when the market excess return is positive and $\beta_i^{MKT} + \beta_i^{DOWN}$ when the market is down. The coefficient β_i^{DOWN} isolates the incremental sensitivity to negative market realisations. Note that, in our setting, the above equation will be estimated by quantile regression, meaning every factor loading will have quantile specific coefficients. In essence, one can re-construct the semibetas in the following way: β^P is upper quantiles of β^{MKT}, β^N is lower quantiles of β^{DOWN}, β^+ is lower quantiles of β^{MKT}, β^- is lower quantiles of β^{MKT}. In essence, we use the indicator function to partition the market returns into two halves along the market dimension, while quantile regression will partition the space of returns into two halves along the (conditional) asset return dimension. The combination of these two partitions creates a structure analogous to the four semibeta quadrants of Bollerslev et al. (2022).

The advantage of using quantile regression with the indicator function is that one can include

the other Fama-French factors to isolate the impact of size, value, profitability, and investment from the market coefficients. In this way we can have "corrected" semi-betas where we isolate the pure impact of market downturns on the conditional distribution of the assets returns. If the downside market coefficient remained important even controlling for these established factors, we can conclude that asymmetric market exposure represents a distinct dimension of risk.

In using the indicator function to split the market into positive and negative impacts, we follow Pettengill et al. (1995) who introduced this formulation to resolve the apparent failure of unconditional beta-return tests. They showed that the cross-sectional relationship between beta and realised returns is significantly positive in up markets but significantly negative in down markets. It is important to note that the indicator function separating up and down markets without quantile regression conflates the impact of β_i^{DOWN} on upside and downside stock returns. The logic of the semibetas predicts that β^{DOWN} will be bigger for the lower quantiles and likely small (or even negative) for upper quantiles (i.e., a downward sloping quantile profile). Quantile regression can separate these effects conveniently, completing the two-dimensional partitioning that the semibeta framework motivates.

The semibeta logic can be recast as a fused variable selection problem: we expect a predominantly negative slope for the β_i^{DOWN}. In essence, if there is a difference across the quantiles for this covariate then there is a difference in the conditional returns between β^N and β^-. As such, if quantile variation exists for β_i^{DOWN}, then the market returns distribution reacts to market downturns. The finding of Bollerslev et al. (2022) of no significant β^P, would then translate to less (or no) quantile variation in β_i^{MKT}. These can be seen by examining the difference between the upper and lower quantiles of the estimated β_i coefficients. To this end we will run a quantile regression exercise on monthly returns for the 50 biggest stocks of the S&P 500 from 2000 January to 2024 December.[13] The list of these 50 stocks are reported in Table 3.

For each stock we run quantile regression with 19 equidistant quantiles (every 5^{th} quantile). When evaluating quantile variation we set the upper quantile to be $\tau_U = 0.9$, and the lower quantile to be $\tau_L = 0.1$. Then, we estimate the traditional quantile regression (QR), the Fused-LASSO with quantile specific shrinkage parameter (FLQR-QS), and the Fused-LASSO with global shrinkage parameter (FLQR-G). Because the BRW solution is a special case of FLQR-QS, we opt to not include that estimator here, and instead focus on quantile specific versus

[13]To select the 50 stocks, we downloaded all stock return information and took the average market cap over the time period for those stocks with complete data. To proxy market cap we multiplied closing price with volume.

Table 3: Selected companies/stocks

Ticker	Company Name	Ticker	Company Name
AAPL	APPLE INC	IBM	INTL BUSINESS MACHINES CORP
ADBE	ADOBE INC	INTC	INTEL CORP
AMAT	APPLIED MATERIALS INC	JNJ	JOHNSON + JOHNSON
AMD	ADVANCED MICRO DEVICES	JPM	JPMORGAN CHASE + CO
AMGN	AMGEN INC	KO	COCA COLA CO/THE
AMZN	AMAZON.COM INC	LLY	ELI LILLY + CO
BA	BOEING CO/THE	MCD	MCDONALD S CORP
BAC	BANK OF AMERICA CORP	MRK	MERCK + CO. INC.
BKNG	BOOKING HOLDINGS INC	MS	MORGAN STANLEY
BRK-B	BERKSHIRE HATHAWAY INC CL B	MSFT	MICROSOFT CORP
C	CITIGROUP INC	MU	MICRON TECHNOLOGY INC
CAT	CATERPILLAR INC	NVDA	NVIDIA CORP
CMCSA	COMCAST CORP CLASS A	ORCL	ORACLE CORP
COP	CONOCOPHILLIPS	PEP	PEPSICO INC
COST	COSTCO WHOLESALE CORP	PFE	PFIZER INC
CSCO	CISCO SYSTEMS INC	PG	PROCTER + GAMBLE CO/THE
CVX	CHEVRON CORP	QCOM	QUALCOMM INC
DIS	WALT DISNEY CO/THE	SLB	SLB LTD
EBAY	EBAY INC	T	AT+T INC
F	FORD MOTOR CO	TXN	TEXAS INSTRUMENTS INC
FCX	FREEPORT MCMORAN INC	UNH	UNITEDHEALTH GROUP INC
GE	GENERAL ELECTRIC	VZ	VERIZON COMMUNICATIONS INC
GILD	GILEAD SCIENCES INC	WFC	WELLS FARGO + CO
GS	GOLDMAN SACHS GROUP INC	WMT	WALMART INC
HD	HOME DEPOT INC	XOM	EXXON MOBIL CORP

global shrinkage. For the Fused-LASSO estimators we implement 10 fold cross-validation with 300 candidate hyperparameters ranging from 0 to 1000 for each firm.

4.1 Quantile varying factors

Figure 3 clearly shows which factors have quantile varying effects. In this figure we see the difference between the selected quantiles for all the factors for all 50 firms. We show the three estimators in the three blocks. We also colour code the value of quantile variation with white cells indicating no quantile variation.

Examining the degree of fused-shrinkage, we can see that FLQR-G exhibits a tendency to shrink the most. This is in line with the findings in the Monte Carlo exercise, where FLQR-G "over-regularised". At the opposite extreme, the traditional QR suffers from spurious quantile variation. Importantly, quantile specific shrinkage provides a data-driven way to allow shrinkage intensity to vary across the quantiles.

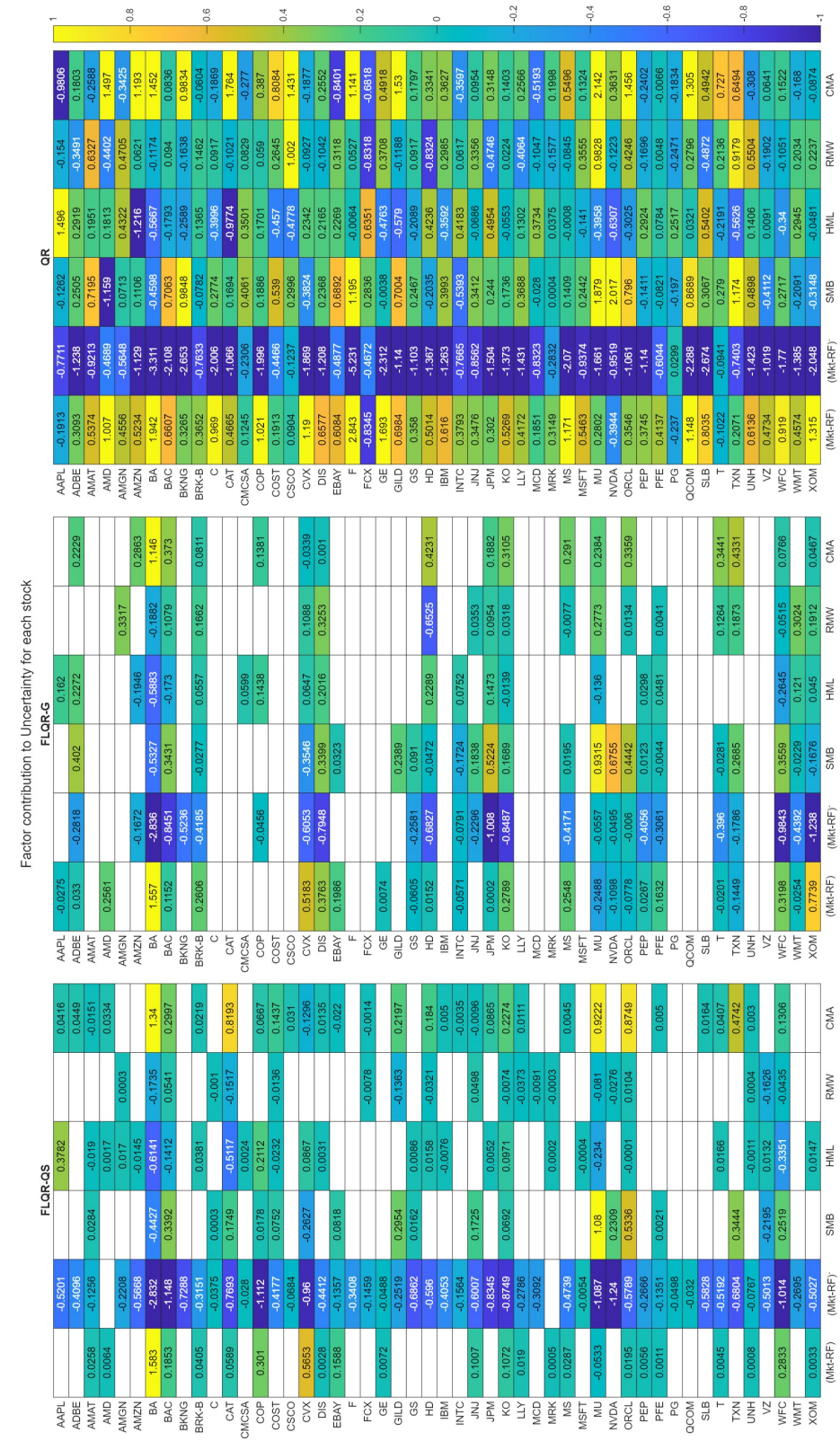

Figure 3: Degree of quantile variation for different factors

Looking at the selected factors we see that β^{DOWN} is negatively sloped, i.e. $\beta(\tau_U) < \beta(\tau_L)$. Furthermore, we see that this is near universal across stocks, the only exceptions being AMD and MRK for FLQR-QS. This pattern is in line with the semibeta logic of Bollerslev et al. (2022): the negative slope implied that when markets decline, the stocks react to these downturns in such a way that the shape of the distribution changes. Since the sign of the market downturn covariate is always negative, this downward sloping profile leads to the spread of the conditional returns distribution widening *ceteris paribus*. This is what generates significant risk premium for β^N in Bollerslev et al. (2022). The fact that the market downturn factor is retained for almost all stocks in the FLQR-QS setup, while other factors are attenuated, suggests that this finding is empirically robust rather than estimation noise.

Interestingly, the second most important factor from a quantile variation perspective is CMA, which is predominantly sloped upwards (if there is quantile variation present). This is likely on account of firms with more aggressive investment strategies (high CMA exposure) taking risks to exploit growth opportunities. Since these firms' valuations depend on discount rates and future cash flow expectations, there is a greater degree of uncertainty regarding their returns. This manifests itself in wider conditional returns distribution driven by an upward sloping CMA profile. Taken together with the downturn market factor, the CMA factor captures additional uncertainty in the return distribution that is driven by the firm's investment strategy: conservative firms (low investment) likely provide some degree of downside protection, which would explain why this factor is not quantile varying for every stock considered. This finding is in line with Bollerslev et al. (2025) who find that the CMA risk premium is negative or zero for the left tail of the factor and positive in the right tail.

Turning to the remaining Fama-French factors, the FLQR-QS panel reveals that SMB and HML display relatively sparse quantile variation profiles. For SMB, this sparsity aligns with the findings of Bollerslev et al. (2025), who find that the risk premium function for size is approximately "flat" across the factor distribution. This suggests that size-related risk has limited impact on the shape of the conditional return distribution. While small-cap exposure could shift the location of returns but it does not systematically widen or narrow the returns distribution.

The HML factor exhibits greater heterogeneity than SMB, both in terms of which stocks display quantile variation and the direction of the slope. This is consistent with de la O González and Jareño (2019) who find time-varying value factor loadings across market conditions. Furthermore, Bollerslev et al. (2025) find that the granular estimates reveal a more nuanced picture

with asymmetric pricing in the tails for HML. The mixed pattern in our results – some stocks showing positive slopes, others negative, and many showing no variation – reflects these previous findings.

The RMW factor has the most sparse profile in both FLQR-QS and FLQR-G. Majority of stocks show no quantile variation for this factor. Bollerslev et al. (2025) also find "flat" patterns for this factor. This suggests that profitability operates more as a level shifter, i.e., firms with robust profitability earn higher average returns, with the premium being stable across the return distribution. Unlike the downside market factor, which captures asymmetric tail behaviour, or the CMA factor, which reflects investment-driven uncertainty, profitability appears to price risk in a more traditional mean-shifting manner.

Beyond the factor-level findings, the FLQR-QR also shows notable heterogeneity in the extent to which quantile variation is present in the individual stocks. Financial sector firms (BAC, JPM, WFC, GS) tend to have quantile variation in multiple factors. This suggests that the conditional return distribution for these stocks are shaped by several dimensions of risk. Financial services firms inherently have higher leverage, which amplifies the sensitivity of their stock return distribution to various factors.

In contrast, consumer staples firms (KO, PG, JNJ, PEP) have sparser quantile variation profiles. This is to be expected, as these firms have stable cash flows and inelastic demand, which insulate them from distributional asymmetries driven by size, value, or investment factors.

Technology firms show more heterogeneity. Downside market factor is present for the majority of the firms but the remaining factors evidence more idiosyncratic selection. For instance NVDA and ORCL have quantile variation in multiple factors, which reflects their growth oriented profiles and sensitivity to discount rate fluctuations. By contrast MSFT and AAPL display comparatively muted factor variation beyond the market downturn factor. These firm-specific patterns reflect genuine economic heterogeneity in how systematic risks propagate through the conditional return distribution.

4.2 Quantile coefficients

The quantile coefficient profiles averaged across the 50 stocks are shown in Figure 4. This figure shows how the different estimators provide different quantile profiles on average for the different factors. The first thing to note is the difference in the profiles between the different estimators. The traditional QR estimator has very "jagged" quantile profiles especially for HML and RMW.

This is on account of QR implicitly assuming that every factor is quantile varying.

The average quantile profiles are as expected given the findings in Figure 3: the market factor slopes upwards, the market downturn factor slopes downwards, and the CMA factor slopes upwards. The SMB, HML, and RMW factors are very different for the different estimators. These three factors have much shallower profiles, as can be seen when comparing the scale of the y-axis, which highlights that market exposure dominates the shape of the return distribution while these factors contribute minimally. This gives further credence to the location shift interpretation for RMW.

For these three factors FLQR-QS has the least steep profile among the estimators. FLQR-QS having shallower profiles for these factors than the QR is not surprising, as many firms have limited uncertainty contribution stemming from these factors. This in turn makes the average slope less steep. Interestingly, the same is not the case for the FLQR-G, which has steeper profiles than the FLQR-QS, especially for RMW and SMB, even though it shrunk more aggressively.

The quantile coefficient profiles for 10 select stocks are shown in Figure 5. These stocks are selected on the basis that they had the largest market downturn profiles based on the FLQR-QS. In line with the average profiles, we can see that traditional QR has the most quantile variation among the estimators. We can also see that the FLQR-G is more likely to produce "step-function" behaviour. The global penalty forces adjacent quantiles to be almost identical unless overwhelming evidence supports a break, which results in a piecewise constant profile for many stocks. This prevents overfitting but it imposes a threshold-model like structure. This is most clearly visible for the CMA and HML factors.

KO displays a remarkably flat profile for nearly all factors in the FLQR-QS model. The main visible variation is in the market downturn factor. This is visual confirmation of the

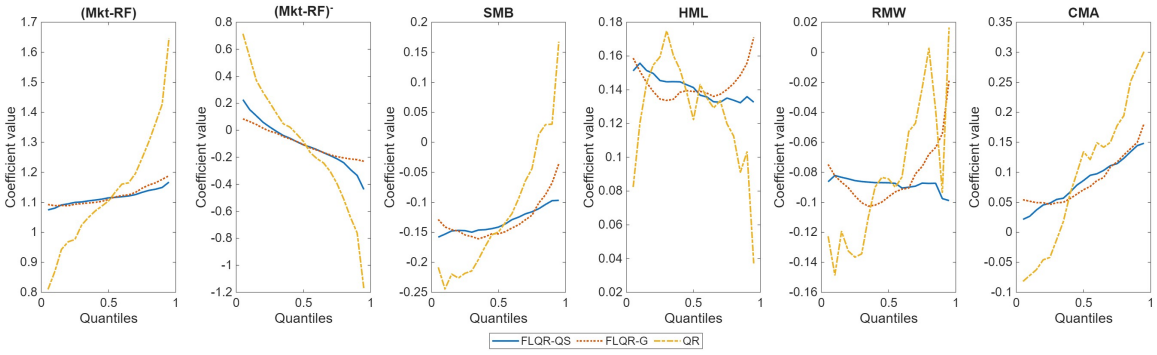

Figure 4: Quantile coefficients (averaged across firms

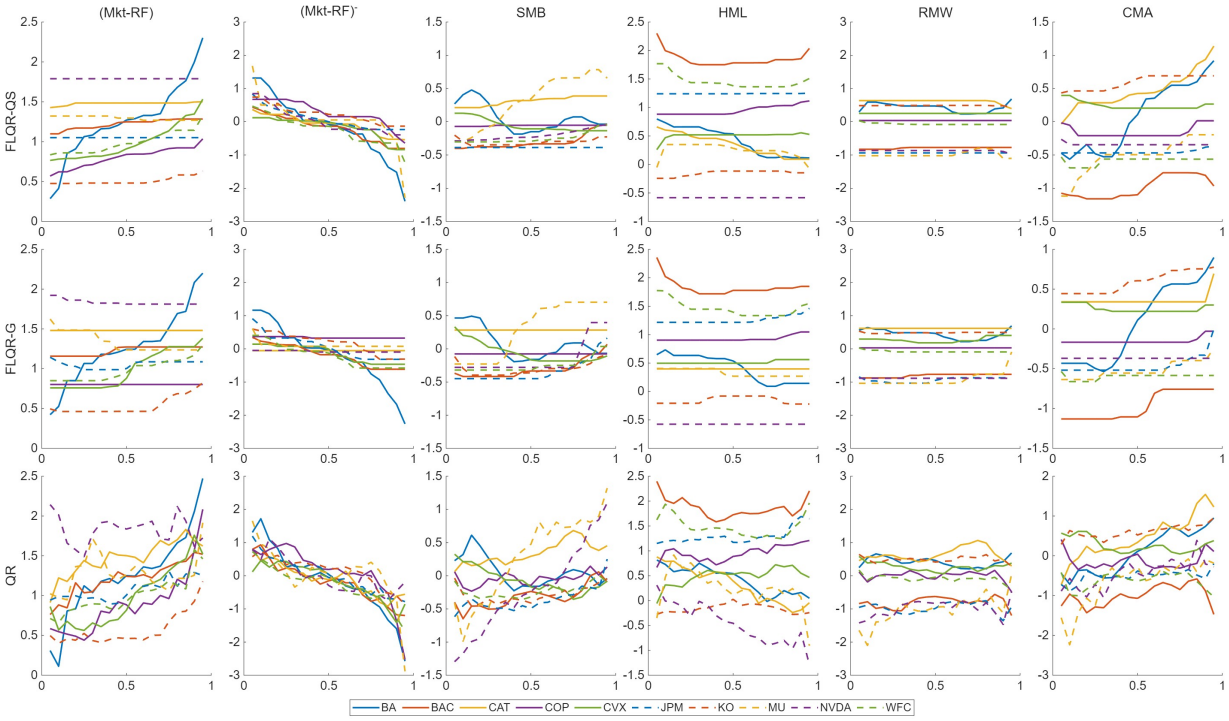

Figure 5: Quantile coefficients (of select stocks)

earlier discussion regarding consumer staples firms, namely that for such firms SMB, HML, and RMW are more likely to act as pure location shifters.

We note that the near-universal downward slope for market downturn with the FLQR-QS suggests the semibeta effect is a systematic phenomenon across firm types rather than being driven by specific sectors. The FLQR-QS is particularly interesting estimator as it holds many of the other factors constant, which leads to a smooth downward sloping profile for all the shown firms. While the FLQR-G also shows downward sloping tendency for these firms, it tends to shrink quantile variation at the tails.

4.3 Hyperparameters for FLQR-QS

We now look at the distribution of optimal hyperparameters for FLQR-QS for the 50 firms. This is shown in Figure (6). We opt to only focus on the FLQR-QS because of its equivalence with the QR and the BRW estimators. On this distribution we note the $\alpha = 1$ point which is the non-crossing solution.

We can see that the distribution tapers off substantially below $\log_{10}(\alpha) = -1$. This means that very few stocks have optimal hyperparameters approaching the unregularised QR limit. This is in line with a key finding from the Monte Carlo experiments, namely that regularisation

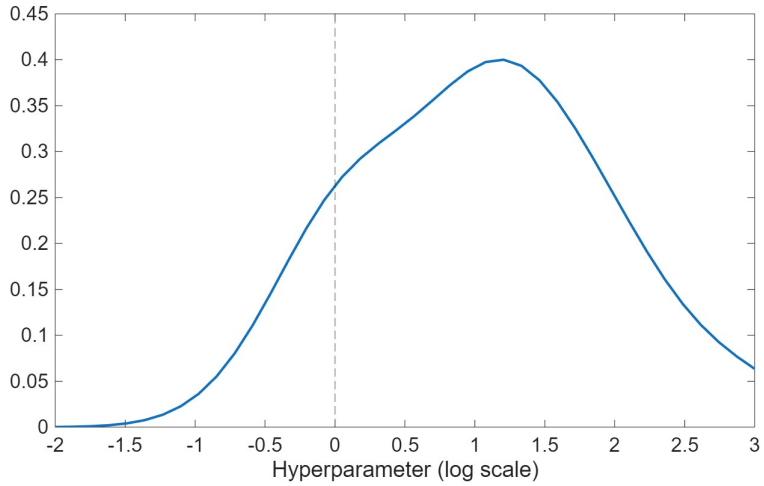

Figure 6: Distribution of optimal hyperparameters for FLQR-QS

is effective.

The distribution spans an approximately 100-fold range in the hyperparameter space. This large scale indicates that different stocks require vastly different degrees of shrinkage. This justifies our stock-by-stock approach. Furthermore, the distribution has a mild skew to the right. As such there are more firms that have factor-return relationships that are "flatter". Then, cross-validation effectively identifies which firms need aggressive shrinkage towards location-shift effects for some factors.

We note the fact that the mode of the distribution is above the $\log_{10}(\alpha) = 1$ point, which is the non-crossing solution. This gives further support to the Monte Carlo finding that the non-crossing solution, while appealing, is not necessarily the optimal point on the bias-variance trade-off spectrum. When picking $\log_{10}(\alpha) = 1$, we are more likely to undershrink some quantile profiles, leading to spurious quantile variation.

4.4 Crossing incidence

Figure 7 shows the crossing incidence for each firm for the different estimators. This crossing incidence was calculated by taking the fitted values for every firm in each time period, and sorting it using the Chernozhukov et al. (2010) procedure. Then we count the instance where the unsorted and sorted quantiles do not match. Taking the average (across time and quantiles) of these crossing incidences gives us a measure for the degree of quantile crossing.

From the figure we can see that QR has the most quantile crossing for all firms. This is

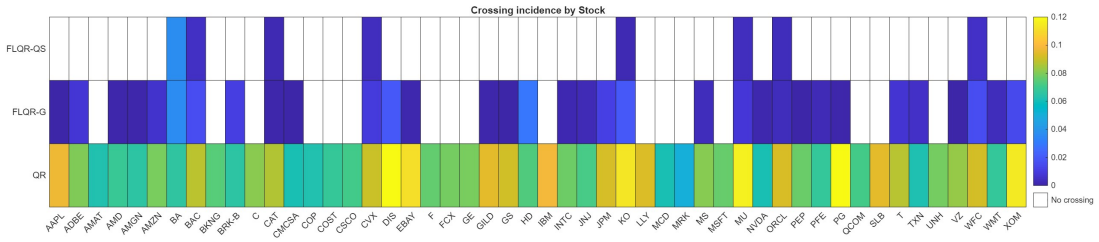

Figure 7: Crossing incidence of different estimators for different stocks

not that surprising, since quantile regression inherently assumes that every factor has quantile variation. This increased variability leads to a higher chance of quantile crossing.

FLQR-G has less crossing overall than QR and also achieves non-crossing quantiles for some of the firms. This finding is important since Figure 3 showed that FLQR-G had the least amount of quantile variation in the factors. However this shrinkage did not translate into non-crossing quantiles. This suggests that where quantile variation is permitted matters more than how much total variation exists across the quantiles. The "step-behaviour" of FLQR-G can create discontinuities in the quantile profile that lead quantile crossing even when total variation is the lowest.

In contrast, FLQR-QS has the least amount of crossing overall and the bulk of the firms have non-crossing quantiles. This shows that the FLQR-QS is not merely reducing variance, it is imposing a smoothness structure that leads to distributional monotonicity as implied by Theorem 1. Interestingly, quantile crossing is largely present for firms that have quantile variation in all the covariates (with the exception of CVX). This implies that for these firms there is some additional explanatory factor (or alternative functional form for the model) that explains some of the remaining quantile variation.

It is worth noting the difference in magnitude for quantile crossing for the different estimators. Traditional QR has 10-12% quantile crossing prevalence for some stocks. FLQR-G improves on this with crossing incidence ranging between 2-6%. FLQR-QS yields non-crossing for most firms, and when it does have quantile crossing it only crosses in 2% of the cases.

4.5 Rolling window estimates

Finally, we test the degree of quantile versus time variation in the sample period. To do this we run a rolling window estimation with 5 years of data in each period (60 observations per time period). For each estimator we run 19 equidistant quantiles (every 5^{th} quantile). For FLQR-G

and FLQR-QS we deviate from the cross validation grid to limit the number of calculations needed to run. Specifically, we keep the 300 candidate hyperparameters for the first period, but in every period afterwards we take the optimal hyperparameter for the firm in time $t-1$, and take the 25 candidate points that are below this optimal hyperparameter and the 25 candidate points above it. In this way we limit the number of calculations that have to be run but we still optimise the degree of shrinkage in every period. Given that the market downturn variable had the most quantile variation, we only show the time varying coefficients for this variable in Figure 8.

The figure reveals that beyond quantile variation, time variation also plays an important role in characterising the conditional return distribution. This can be seen during large market events (the 2008 Global Financial Crisis and the 2020 COVID-19 crash), where the degree of quantile variation increases. The three estimators show very different behaviours in how these dynamics are captured. QR displays substantial variation along both the quantile and time dimensions, though as established earlier, much of this variation likely reflects estimation noise. FLQR-G exhibits the opposite: the pronounced vertical striping indicates that the global shrinkage parameter favours time variation over quantile variation, effectively collapsing the quantile dimension. This creates an artificial appearance of discrete regime switches rather than the smooth quantile gradients that characterise distributional asymmetry. FLQR-QS occupies a middle ground, preserving meaningful variation along both dimensions while filtering out excessive noise in the quantile dimension.

Looking at the FLQR-QS panel, we can see that the characteristic downward-sloping quantile profile for market downturn is consistent across the entire 20-year sample. Even during calmer periods, this downward sloping profile is present. This further strengthens the semibeta mechanism documented by Bollerslev et al. (2022).

The figure also reveals that during crisis episodes, the quantile profile becomes more pronounced: in 2008–2009 and 2020, the upper quantiles become more deeply negative while lower quantiles remain closer to zero. This is indicative that crises periods widen the spread between β^N and β^-. This amplification of asymmetry leads to the conditional return distribution becoming more skewed during stress periods. Interestingly, the two crisis episodes have different temporal dynamics. The 2020 COVID-19 episode shows a sharp intensification of the quantile pattern followed by relatively quick reversion to normal patterns, whereas 2008–2009 displays a more prolonged period of elevated asymmetry. This finding aligns with the "V-shaped" recovery narrative of the pandemic recession versus the protracted deleveraging that followed the

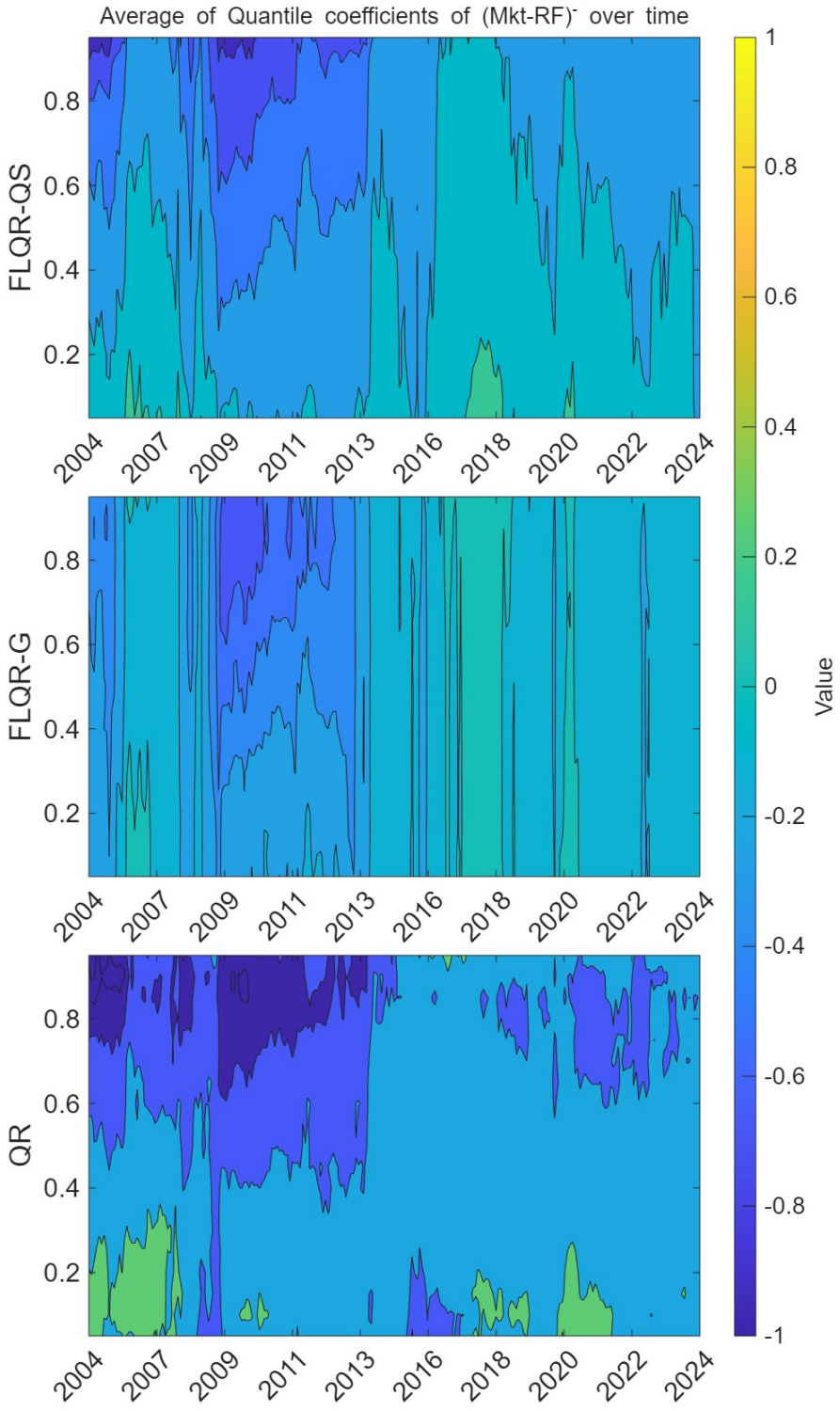

Figure 8: Rolling window quantile coefficients for market downturn

Global Financial Crisis. This suggests that the distributional dynamics of crash risk can differ meaningfully across crisis types.

In summary, our main results confirm the findings of Bollerslev et al. (2022), however our quantile regression approach offers several methodological advantages. First, our quantile regression framework operates within a single frequency, while Bollerslev et al. (2022)'s monthly semibetas are constructed from daily returns. This frequency mismatch introduces additional data requirements and computational complexities. Second, our framework provides a unified approach that incorporates all Fama-French factors simultaneously alongside the market downturn indicator. This enables us to isolate the "pure" asymmetric market exposure, while controlling for size, value, profitability, and investment effects. In contrast, the semibeta approach of Bollerslev et al. (2022) focuses exclusively on the market beta. Third, our proposed fused-shrinkage framework helps in identifying which Fama-French factors drive quantile variation for the different firms' returns. We show that simply extending the semibeta framework to the quantile realm is not enough, as traditional quantile regression assumes all Fama-French factors vary by quantile. By imposing the non-crossing guided shrinkage, we are able to verify the finding of Bollerslev et al. (2022). Finally, our approach brings Value-at-Risk estimation closer to the factor model literature: quantile regression directly estimates conditional quantiles of the return distribution, which correspond to VaR at the specified confidence level. This creates a natural bridge between asset pricing and risk management that is aligned to, but not directly available in the factor model framework.

5 Conclusion

This paper develops an adaptive non-crossing constraint framework that, by varying the tightness of the constraint, encompasses as special cases the traditional quantile regression estimator, the Bondell et al. (2010) non-crossing estimator, and the composite quantile regression,. By developing non-crossing constraints that can be thus tightened, we are able to study the properties of these constraints on the estimated slope coefficients and forecasts. Doing so reveals that non-crossing constraints are simply a type of Fused LASSO with quantile-specific shrinkage parameters. This equivalence has an important implication: imposing non-crossing constraints is not merely a tool for ensuring distributional coherence, but rather a specific choice on the bias-variance trade-off spectrum. This in turn emphasizes that the degree of non-crossing regularisation can and should be optimised via cross-validation, rather than imposed as a fixed

constraint.

Through a Monte Carlo experiment, we verify that imposing non-crossing constraints is equivalent to introducing fused shrinkage. We also show how our proposed Fused LASSO quantile regression with quantile-specific shrinkage (FLQR-QS) estimator provides model fits that are either better or nearly as good as those of the BRW estimator (Bondell et al., 2010). Considering the variable selection properties of the different estimators, we find that traditional Fused LASSO with global shrinkage (FLQR-G) estimator overshrinks while BRW estimator undershrinks relative to the FLQR-QS.

The empirical application of the methods to quantile Fama-French factor models yields findings consistent with the semibeta framework of Bollerslev et al. (2022), but with less intense data requirements and more efficient use of data. The market downturn coefficient has a smooth downward-sloping quantile profile for nearly all stocks. This means that market declines have a differential impact across the conditional return distribution. In contrast, the SMB and RMW factors display flat quantile profiles for many of the stocks considered. These patterns emerge clearly under our quantile-specific shrinkage approach.

In summary, the non-crossing constraints of Equation (5) bestow upon the quantile estimator additional attractive properties: (a) it can distinguish variables with quantile variation from those without, and shrinks the correct variable; and (b) it renders the estimated quantile profiles less 'jagged', i.e., it removes sudden reversions in the difference in the β coefficient.

The implications of the above findings reach beyond the estimator proposed here and suggest several avenues for further research. In particular, because of the equivalence between non-crossing and fused-shrinkage, one can extend the methodology to obtain non-crossing Bayesian quantile regression estimators like Lancaster and Jae Jun (2010). To our knowledge, currently the most popular way to obtain non-crossing Bayesian quantiles involve post-processing methods such as Rodrigues and Fan (2017). Implementing Theorem 1 in the Bayesian realm would have further advantages of better coefficient estimates and forecasts and, importantly, the ability to incorporate prior information or beliefs held by investors and policymakers.

The formulation of the FLQR-QS is simple because here the hyperparameter is simply a scalar. It is probable that one can gain further improvements in estimation by allowing the hyperparameter to have higher dimensions. It is worth noting that making the hyperparameter variable specific has been shown to offer better performance in other contexts (Zou, 2006). Doing so may require an improved hyperparameter tuning procedure, since grid-search is computation-intensive even with a single parameter. We leave this extension to future research.

Another avenue of research is shrinking both the level and difference coefficients, similar to Jiang et al. (2014). However, following Szendrei and Varga (2023), shrinkage in levels would require a separate dedicated hyperparameter. This way the two types of shrinkage could be incorporated yielding the possibility to identify quantile varying sparsity, as in Kohns and Szendrei (2021) and Szendrei and Varga (2023), while being principled about the degree of fused shrinkage one needs to impose. We leave this work to future research as well.

Finally, since the bias introduced by non-crossing constraints is similar to biased bootstrap methods, such as data tilting and data sharpening, proposed for regression under monotone or convexity constraints (Hall and Presnell, 1999; Choi et al., 2000), one can use the degree of bias to test model adequacy. Developing tests of model adequacy using bias induced by the quantile non-crossing constraints is another promising avenue for future research.

References

Adrian, T., N. Boyarchenko, and D. Giannone (2019). Vulnerable growth. *American Economic Review 109*(4), 1263–89.

Allen, D. E. and S. R. Powell (2011). Asset pricing, the fama—french factor model and the implications of quantile-regression analysis. In *Financial econometrics modeling: Market microstructure, factor models and financial risk measures*, pp. 176–193. Springer.

Arlot, S. and A. Celisse (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys 4*(none), 40 – 79.

Bates, S., T. Hastie, and R. Tibshirani (2024). Cross-validation: what does it estimate and how well does it do it? *Journal of the American Statistical Association 119*(546), 1434–1445.

Bergstra, J. and Y. Bengio (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research 13*(2).

Bollerslev, T., A. J. Patton, and R. Quaedvlieg (2022). Realized semibetas: Disentangling "good" and "bad" downside risks. *Journal of Financial Economics 144*(1), 227–246.

Bollerslev, T., A. J. Patton, and R. Quaedvlieg (2025). Granular betas and risk premium functions. *Journal of Econometrics*, 106034.

Bondell, H. D., B. J. Reich, and H. Wang (2010). Noncrossing quantile regression curve estimation. *Biometrika 97*(4), 825–838.

Cerqueira, V., L. Torgo, and I. Mozetič (2020). Evaluating time series forecasting models: An empirical study on performance estimation methods. *Machine Learning 109*, 1997–2028.

Chernozhukov, V., I. Fernandez-Val, and A. Galichon (2009). Improving point and interval estimators of monotone functions by rearrangement. *Biometrika 96*(3), 559–575.

Chernozhukov, V., I. Fernández-Val, and A. Galichon (2010). Quantile and probability curves without crossing. *Econometrica 78*(3), 1093–1125.

Choi, E., P. Hall, and V. Rousson (2000). Data sharpening methods for bias reduction in nonparametric regression. *The Annals of Statistics 28*(5), 1339–1355.

de la O González, M. and F. Jareño (2019). Testing extensions of fama & french models: A quantile regression approach. *The Quarterly Review of Economics and Finance 71*, 188–204.

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

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

Fama, E. F. and K. R. French (2015). A five-factor asset pricing model. *Journal of financial economics 116*(1), 1–22.

Figueres, J. M. and M. Jarociński (2020). Vulnerable growth in the euro area: Measuring the financial conditions. *Economics Letters 191*, 109126.

Hall, P. and B. Presnell (1999). Intentionally biased bootstrap methods. *Journal of the Royal Statistical Society Series B: Statistical Methodology 61*(1), 143–158.

Han, H., O. Linton, T. Oka, and Y.-J. Whang (2016). The cross-quantilogram: Measuring quantile dependence and testing directional predictability between time series. *Journal of Econometrics 193*(1), 251–270.

Jiang, L., H. D. Bondell, and H. J. Wang (2014). Interquantile shrinkage and variable selection in quantile regression. *Computational Statistics & Data Analysis 69*, 208–219.

Jiang, L., H. J. Wang, and H. D. Bondell (2013). Interquantile shrinkage in regression models. *Journal of Computational and Graphical Statistics 22*(4), 970–986.

Koenker, R. (1984). A note on L-estimates for linear models. *Statistics & Probability Letters 2*(6), 323–325.

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

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

Koenker, R. and Z. Xiao (2006). Quantile autoregression. *Journal of the American Statistical Association 101*(475), 980–990.

Kohns, D. and T. Szendrei (2021). Decoupling shrinkage and selection for the bayesian quantile regression. *arXiv preprint arXiv:2107.08498*.

Korobilis, D. (2017). Quantile regression forecasts of inflation under model uncertainty. *International Journal of Forecasting 33*(1), 11–20.

Lancaster, T. and S. Jae Jun (2010). Bayesian quantile regression methods. *Journal of Applied Econometrics 25*(2), 287–307.

Lintner, J. (1975). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. In *Stochastic optimization models in finance*, pp. 131–155. Elsevier.

Linton, O. and Y.-J. Whang (2007). The quantilogram: With an application to evaluating directional predictability. *Journal of Econometrics 141*(1), 250–282.

Liu, Y. and Y. Wu (2009). Stepwise multiple quantile regression estimation using non-crossing constraints. *Statistics and its Interface 2*(3), 299–310.

Pettengill, G. N., S. Sundaram, and I. Mathur (1995). The conditional relation between beta and returns. *Journal of Financial and quantitative Analysis 30*(1), 101–116.

Racine, J. (2000). Consistent cross-validatory model-selection for dependent data: hv-block cross-validation. *Journal of Econometrics 99*(1), 39–61.

Rodrigues, T. and Y. Fan (2017). Regression adjustment for noncrossing bayesian quantile regression. *Journal of Computational and Graphical Statistics 26*(2), 275–284.

Shao, J. (1997). An asymptotic theory for linear model selection. *Statistica Sinica 7*(2), 221–242.

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

Stone, C. J. (1977). Consistent nonparametric regression. *The Annals of Statistics 5*(4), 595–620.

Szendrei, T. (2025). Crossing penalised caviar. *arXiv preprint arXiv:2501.10564*.

Szendrei, T. and K. Varga (2023). Revisiting vulnerable growth in the euro area: Identifying the role of financial conditions in the distribution. *Economics Letters 223*, 110990.

Wager, S. (2020). Cross-validation, risk estimation, and model selection: Comment on a paper by rosset and tibshirani. *Journal of the American Statistical Association 115*(529), 157–160.

Yang, Y. (2005). Can the strengths of AIC and BIC be shared? A conflict between model indentification and regression estimation. *Biometrika 92*(4), 937–950.

Yang, Y. (2007). Consistency of cross validation for comparing regression procedures. *The Annals of Statistics 35*(6), 2450 – 2473.

Yang, Y. and S. T. Tokdar (2017). Joint estimation of quantile planes over arbitrary predictor spaces. *Journal of the American Statistical Association 112*(519), 1107–1120.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association 101*(476), 1418–1429.

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

A Estimators

- Quantile regression (Koenker and Bassett, 1978)

$$\hat{\beta}_{QR} = \underset{\beta}{argmin} \sum_{q=1}^{Q} \sum_{t=1}^{T} \rho_q(y_t - x_t^T \beta_{\tau_q})$$

- Composite quantile regression (Koenker, 1984; Zou and Yuan, 2008)

$$\hat{\beta}_{CQR} = \underset{\beta}{argmin} \sum_{q=1}^{Q} \sum_{t=1}^{T} \rho_q(y_t - x_t^T \beta)$$

- Non-crossing qunatile regression (Bondell et al., 2010). Here, $z \in [0,1]$, i.e. it is the x rescaled.

$$\hat{\beta}_{BRW} = \underset{\beta}{argmin} \sum_{q=1}^{Q} \sum_{t=1}^{T} \rho_q(y_t - z_t^T \beta_{\tau_q})$$
$$s.t.\ \gamma_{0,\tau_p} \geq \sum_{k=1}^{K} \gamma^{-}_{k,\tau_q}$$

- Fused shrinkage quantile regression (Jiang et al., 2013)

$$\hat{\beta}_{FLQR-G} = \underset{\beta}{argmin} \sum_{q=1}^{Q} \sum_{t=1}^{T} \rho_q(y_t - x_t^T \beta_{\tau_q})$$
$$s.t.\ k^* \geq \sum_{q=2}^{Q} \sum_{k=1}^{K} (\gamma^{+}_{k,\tau_q} + \gamma^{-}_{k,\tau_q})$$

- Non-crossing based fused shrinkage (proposed here)

$$\hat{\beta}_{FLQR-QS} = \underset{\beta}{argmin} \sum_{q=1}^{Q} \sum_{t=1}^{T} \rho_q(y_t - x_t^T \beta_{\tau_q})$$
$$s.t.\ \gamma_{0,\tau_p} + \sum_{k=1}^{K} \Big[\bar{X}_k - \alpha(\bar{X}_k - min(X_k))\Big]\gamma^{+}_{k,\tau_q} \geq \sum_{k=1}^{K} \Big[\bar{X}_k + \alpha(max(X_k) - \bar{X}_k)\Big]\gamma^{-}_{k,\tau_q}$$

B Bias-variance trade-off

GNCQR allows us to gradually enforce non-crossing constraints by inducing fused shrinkage. This naturally entails that there is some type of bias-variance trade-off. In essence, as we increase α we increase the amount of non-crossing we want to impose. By introducing non-crossing constraints we impose a bias on the quantile property (i.e., we no longer obtain the minimum of the tick-loss weighted ℓ_1 residuals) and instead enforce monotonically increasing quantiles.[14] This implicitly enforces that the data below τ_q must be a subset of the data below τ_{q+1}. As we move from $\alpha = 0$ to $\alpha = 1$, we gradually move from the quantile property to the "quantile subset" property. Once $\alpha > 1$, GNCQR starts to become more restrictive in how the quantile subset property is satisfied. In particular, it will start to penalise quantile closeness as well as quantile crossing. Formally this can be shown as follows:

$$\{I(\varepsilon_{\tau_{q-1}} - \xi \leq 0)\} \subseteq \{I(\varepsilon_{\tau_q} \leq 0)\} \tag{9}$$

Here $\{I(\varepsilon_{\tau_q} \leq 0)\}$ locates the elements of the quantile residual that are negative for the fitted τ_q^{th} quantile, i.e. the observations below the fitted quantile. The quantile subset property necessitates that the observations with negative residuals of the τ_{q-1} quantile must be a subset of the observations with negative residuals of the τ_q quantile. This is achieved in Equation (9) when $\xi = 0$.

As α of Equation (6) goes above 1, the ξ term in the quantile subset property (Equation (9)) increases as well. Increasing the ξ parameter means that even some observations that are above the τ_{q-1} fitted quantile form the set that has to be below τ_q quantile. In this way, quantiles that get too close to each other are treated as if they cross. As such, when $\alpha > 1$, the ξ term in Equation (9) becomes positive, which in turn leads to penalising quantile closeness in addition to quantile crossing. Conversely, when $\alpha < 1$, the ξ term is negative and the quantile subset property allows for some degree of quantile crossing. When $\alpha = 0$, the ξ term is some large number, making Equation (9) trivial to satisfy.

[14] The bias thus introduced by non-crossing constraints is similar to biased bootstrap methods (data tilting, data sharpening, etc.) proposed for regression under monotone or convexity constraints.

C Hyperparameter selection methods

The choice of cross-validation technique can lead to differences in results. In particular, there is a documented trade-off between model identification and minimising predictive risk (Yang, 2005). Broadly speaking, leave-one-out cross-validation is asymptotically equivalent to AIC (Stone, 1977), while the various block cross-validation methods are closer to the BIC (Shao, 1997) if the size of the training sample (relative to the validation sample) goes to zero as $T \to \infty$. This underlies the key point of Yang (2005) that one cannot simultaneously achieve model consistency (in terms of model selection) and efficiency (in terms of achieving lowest error variance). Note also that by choosing the hv-block CV of Racine (2000), we are implicitly expressing a preference in favour of model selection. If the interest lies exclusively upon data fit, it may be beneficial to opt for a leave-one-out CV.

Given the results of Stone (1977) and Shao (1997), one can also rely on information criteria for hyperparameter selection. The equations for AIC and BIC for quantile regression are reported in Jiang et al. (2014). Utilising the information criteria has the advantage of only needing one estimation (per hyperparameter), rather than one per block. This can speed up computation especially for larger sample sizes.

Cross-validation has been shown to have some drawbacks in some applications (Bates et al., 2024). Specifically, it has been shown that cross-validation estimates the average prediction error of models fit on other unseen training sets drawn from the same population, rather than the prediction error of the model fit on the specific training set. On account of this, confidence intervals for prediction error may have lower than nominal coverage. While these results are undoubtedly important, they are less of a concern for model selection. In essence, all candidate hyperparameters are biased in the same way, which has little influence on their relative rank. This implies that the optimal hyperparameter choice remains valid. For a more rigorous mathematical treatment of the consistency of cross-validation for model selection, see Theorem 2 of Yang (2007) and Proposition 2 of Wager (2020).[15]

[15]While our analysis assumes that the relative ranking of hyperparameters remains valid, further investigation might reveal contexts in which Bates et al. (2024) nested cross-validation method improves hyperparameter tuning.