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.

August 13, 2025

Abstract

Growth-at-Risk is vital for assessing macroeconomic uncertainty but is often suspect to quantile crossing due to sampling variation. The existing literature addresses this through post-processing of the fitted quantiles, but these methods do not modify the estimated coefficients. We propose imposing non-crossing constraints during estimation and demonstrate their equivalence to fused LASSO with quantile-specific shrinkage parameters. By re-examining Growth-at-Risk through a fused shrinkage lens, we obtain improved left-tail forecasts and better identification of determinants that drive quantile variation, reflecting improved understanding of downside growth risks. We show that these improvements have implications for policy tools such as Expected Shortfall and Quantile Local Projections.

Keywords: Interquantile shrinkage, Crossing Quantile Curves, High-Dimensional Econometrics, Growth-at-Risk.

*The authors thank Atanas Christev, Rod McCrorie, Paul Allanson, Isaiah Andrews, István Járási, Katalin Varga, David Kohns, and all the participants of the 2022 and 2023 PhD conference in Crieff 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

Introduced by Adrian et al. (2019), Growth-at-Risk (GaR) has become increasingly popular and is now a key measure of economic vulnerability. The key idea is to view GDP growth through the lens of value-at-risk and use quantile regression of Koenker and Bassett (1978) to uncover nonlinear macro-financial linkages. The importance of GaR was highlighted particularly by the global financial crisis, showing how downside risks, or lower quantiles of the density of GDP growth, evolve with the state of financial markets.

GaR has become a popular tool for policy makers and researchers alike. Figueres and Jarociński (2020) apply the method to Euro Area data and show that the macro-financial linkages driving skewness are not specific to the US. Kohns and Szendrei (2024) use shrinkage methods to fit GDP densities with over 200 potential covariates and find that accounting for multiple sources of risk leads to better density fits of GDP growth. Iseringhausen et al. (2025) use the same dataset and study the skewness of the fitted quantile estimates, which encompasses the aggregate impact of the considered covariates. They show that aggregate skewness is highly pro-cyclical. Mitchell et al. (2024) use quantiles estimated on US GaR and show that multimodality in the fitted densities is just as important as skewness.[1] Finally, Carriero et al. (2025) considered specifications for quantile regression for empirical macroeconomics, including GaR among the candidate models.

While GaR undoubtedly provides important policy insights, it is a macroeconometric approach which requires substantial data. Hence, it suffers substantial finite sample variations with usual data constraints. This data scarcity is particularly a problem for tail estimation (Koenker, 2005), which is the main goal for a GaR model. One big concern when applying quantile regression in data scarce settings is the phenomenon of quantile crossing, whereby the fitted quantiles are not monotonically increasing and therefore quantile profiles can intersect yielding improper densities. The literature has addressed this problem using one of two methods. One approach is to fit some distribution on the estimated quantiles in each time period; see, for example, Adrian et al. (2019) and Korobilis (2017). An alternate approach, advocated by Chernozhukov et al. (2009) and Chernozhukov et al. (2010), is to simply sort the estimated quantiles in each period.[2]

[1] Adrian et al. (2021) and Kohns and Szendrei (2024) also discuss the importance of multimodality stemming from macro-financial linkages.

[2] In addition, there is the novel method of Mitchell et al. (2024) which proposes a nonparametric method to construct densities from sorted quantiles.

All the above approaches have been popular in the macroeconometric literature and they all involve some form of post-processing of the estimated quantiles. While these two-step methods yield proper densities, simply adjusting the fitted quantiles does not quantify corresponding changes in the estimated coefficients, i.e., the coefficients estimated in the first step remain uncorrected. This is particularly a problem as quantile coefficients are used to construct quantile IRFS (Chavleishvili and Manganelli, 2024), or quantile local projections (Ruzicka, 2021). As such, it would be desirable to impose non-crossing during estimation of the quantiles. This is the motivation in Szendrei and Varga (2023), who use the non-crossing constraints of Bondell et al. (2010) to identify variables that drive macro-financial nonlinearity in the Euro Area GaR. They also find that imposing non-crossing constraints improves the forecasted densities of GaR.

While non-crossing constraints are appealing for empirical macroeconomic applications of quantile regression, we do not know how these constraints influence the estimated coefficients. The primary motivation of this paper is to study the impact of such constraints on the model parameters. This is achieved by implementing a set of non-crossing constraints that can be scaled, which in turn renders the non-crossing constraints tighter or looser. Using this adaptive non-crossing constraint we show that non-crossing constraints are equivalent to fused LASSO with quantile specific shrinkage parameters. Then, we propose an estimator using these adaptive non-crossing constraints (Generalised Non-crossing Constraint Quantile Regression), where the hyperparameter regulating interquantile shrinkage is obtained by cross-validation.

We conduct comprehensive Monte Carlo exercises based on Bondell et al. (2010) who originally proposed non-crossing quantile regression. Through these experiments we show that the proposed estimator is capable of providing further improvements in fit compared to the estimator proposed in Bondell et al. (2010); henceforth denoted by BRW estimator. We also consider the variable selection properties of the different estimators and verify that non-crossing constraints are a type of fused shrinkage. The Monte Carlo experiments also reveal that the BRW estimator undershrinks quantile variation while the traditional Fused LASSO (Jiang et al., 2013) overshrinks, while our proposed estimator provides improvements on both.

We then estimate US GaR using the variables outlined in Adrian et al. (2019). Specifically, we investigate the canonical GaR within an interquantile shrinkage framework, i.e. identifying variables that drive nonlinearities. This approach was mentioned in Adrian et al. (2019), but their estimation did not enforce limiting quantile variation during estimation, which in turn impacts their downside risk measures. By contrast, our proposed estimator performs much better in distinguishing quantile varying variables than the traditional Fused LASSO.

Furthermore, conducting a pseudo forecasting exercise we confirm that the estimator with adaptive non-crossing constraints yields the best left tail forecasts. These findings complement Carriero et al. (2025), namely that fused shrinkage is also important for GaR. Our results also indicate that how this shrinkage is induced is important: adaptive non-crossing constraints yield improvements, while the traditional fused LASSO provide inferior forecast performance.

These differences in GaR coefficients have an influence on policy tools. We show that the choice of imposing interquantile shrinkage, and how it is imposed, influences the estimated Expected Shortfall of GDP growth. In particular, Expected Shortfall based on traditional quantile regression is more volatile from period to period, while the proposed estimator produces smoother expected shortfall values over time at the same time capturing periods of financial stress. We also produce quantile local projections of how GDP responds over time to a unit shock in our measure of financial stress. Similar to expected shortfall, the local projections (LPs) of the proposed estimator are smoother than those of quantile regression. These evidences highlight the policy relevance of implementing interquantile shrinkage in GaR.

The remaining paper is structured 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. This is followed by the US Growth-at-Risk application in Section 4. Section 5 concludes with the key takeaways.

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.[3] 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).

Quantile regression is not the only asymmetric estimator which uses the 'tick-loss' function as a way to model different parts of the distribution. Newey and Powell (1987) introduced the concept of expectile regression, i.e., the 'tick-loss' weight applied to the ℓ_2 norm of the residuals. The advantage of expectile regression is that the objective function is differentiable, unlike the case of quantile regression. Furthermore, when setting $\tau = 0.5$ in an expectile regression, one recovers the OLS estimator, i.e., the conditional mean.

Expectile regression has some disadvantages relative to quantile regression. First, interpretation is challenging: while conditional quantiles can be interpreted as splitting the data into τ and $1 - \tau$ parts, i.e. the τ^{th} conditional quantile of the sample, the same is not true for expectiles. Instead, expectiles represent the τ^{th} quantile of some distribution which is related to the original distribution of Y (Jones, 1994). Second, since expectile regression is based on the ℓ_2 norm of the residuals, outliers in the dependent variable impacts all expectiles, while outliers only exert undue influence on extreme quantiles in the case of quantile regression. As such, "fan-shaped" estimated quantiles are a clear indication of heteroskedasticity, while the same shape for expectiles might simply be due to outliers in the dependent variable (Newey and Powell, 1987; Sobotka and Kneib, 2012). On account of these challenges, in this paper we focus on quantile regression and retain for future research the extension of our findings to expectiles.

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

[3]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) and Chernozhukov et al. (2010). While these two-step methods yield proper densities, correcting the fitted quantiles does not quantify corresponding changes in the estimated coefficients, i.e., 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. 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.

2.2 Non-Crossing Constraints

Non-crossing constraints incorporated into quantile regression are a way to ensure that the estimated quantiles remain monotonically increasing. Imposing them can be done via 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 there being $T \times (Q-1)$ 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,[4] which reduces the number of constraints to $(Q-1)$.[5] This simplifies computation a great deal and enables non-crossing constraints to be included without too much additional

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

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

computational cost. Because quantile regression is invariant to monotone transformations, any affine invertible transformation can be applied: to obtain the coefficients pertaining to the untransformed data, it is enough to apply the inverse transformation to the estimated coefficients (Koenker, 2005).

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) shows that (3) is a necessary as well as a 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 (and 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 when setting $\alpha = 0$ and the Bondell et al. (2010) when $\alpha = 1$.[6] This leads to the following adaptive non-crossing constraints:

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

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

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 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):

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

where $\gamma_j^- \geq 0$ and $\gamma_j^+ \geq 0$ represent the positive and negative parts of γ_j such that $\gamma_j = \gamma_j^+ - \gamma_j^-$. 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. This equivalence requires $\alpha > 0$, as the relationship is undefined when $\alpha = 0$. As $\alpha \to 0$, $k^*_{\tau_q} \to \infty$, and we recover the QR coefficients for sufficiently small α. □

Theorem (1) implies that when α is large enough, the only way to satisfy the constraint is by having $\gamma^-_{j,\tau_q} = \gamma^+_{j,\tau_q} = 0$, which is equivalent to the Composite Quantile Regression setup of Koenker (1984) and Zou and Yuan (2008).[7] Since it is possible to recover the traditional quantile regression estimator, the Bondell et al. (2010) estimator, and the composite quantile regression estimator by simply varying α, we refer to the estimator using the constraints of equation (5) as Generalised Non-Crossing Quantile Regression (GNCQR).[8,9] Formally GNCQR is defined as:

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

2.4 Hyperparameter Properties

It is clear from equation (6) that the choice of α plays a strong role in the estimation. This role is similar to λ in the LASSO, and here we have similar restrictions on α as well.

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

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

[8] In the Appendix, we present the different types of estimators within a unified notation.

[9] Powell (2020) developed the 'Generalized Quantile Regression', and we follow the same naming convention.

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).[10] We leave for future research to study how different forms of misspecification influence crossing incidence in quantile estimation.

The advantage of the GNCQR 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.

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

Assumption (2) of vanishing penalty, 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) of decaying penalty, the bias introduced by fused penalty is of smaller order than the sampling noise. This ensures that the limiting distribution of the GNCQR is the same as that of the classical quantile regression.

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

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

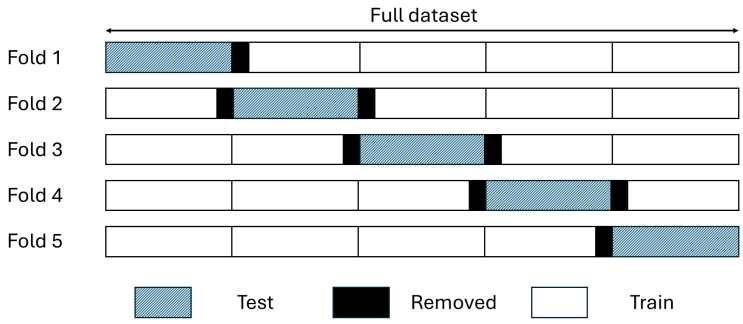

Figure 1: hv-Cross Validation

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,[11] 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. When data are dependent, the information from the test dataset can leak into the training set. This can inadvertently lead to overfitting and hence poor generalization. Removing the data around the test dataset can mitigate 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:*

[11] 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.

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

2. *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 $\hat{\alpha} \xrightarrow{P} \alpha_0$.

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. This theorem also shows that GNCQR is equivalent to having quantile-specific hyperparameters. 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. Specifically, we consider the following estimators: the non-crossing quantile regression of Bondell et al. (2010) (BRW), our proposed GNCQR, and the Fused LASSO (FLQR).[12] We also consider the two-step approach proposed in Chernozhukov et al. (2010), i.e. running regular quantile regression and sorting the fitted quantiles. This allows us to evaluate how far better model selection results translate into improved fit.

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.

[12]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.

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 in macroeconometric applications it is often necessary to apply quantile regression to such small samples.[13] Applying quantile regression to analyse GDP distribution for European economies has been particularly challenging on account of data availability, so we feel that these Monte Carlo results can be useful for policy makers.

We also consider variation in the number of quantiles to be estimated, by generating equidistant quantiles with varying distance between the quantiles: $\Delta_\tau \in \{0.2, 0.1\}$.[14] 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 hyperparameter – 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

[13] 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.

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

(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 ($\times 100$):

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

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.

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 and GNCQR, we use 10-fold cross-validation.

Note that while 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 for GNCQR.

Table 1reports 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 relate to the extensions to the Monte Carlo experiments of Bondell et al. (2010).

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
GNCQR	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	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
GNCQR	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	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
GNCQR	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	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

3.2 Variable Selection results

The results on variable selection presented in Table 1 are particularly revealing. Since the GNCQR 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 the findings of 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, which corroborates Theorem 1, which states 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. GNCQR always ranks second and FLQR 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 produces the best results, with GNCQR a close second. From these results we can see that BRW undershrinks quantile variation, FLQR overshrinks, and GNCQR yields a middle-ground option.

As will be seen in the model fit results of Table 2, we find that GNCQR not only provides robustness over FLQR for y_4, but is also able to provide fits close to FLQR while retaining better model selection properties. In particular, GNCQR is able to yield much better TPR than FLQR 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 GNCQR, increasing the sample size increases TPR but lowers TNR for all DGPs and all Δ_τ. However, for FLQR we see similar pattern when $\Delta_\tau = 0.2$ but not when $\Delta_\tau = 0.1$. In particular, for FLQR for y_4, the TPR decreases as the sample size increases when $\Delta_\tau = 0.1$. This is particularly troubling since FLQR has the worst TPR of all the estimators.

3.3 Goodness-of-Fit results

Results on model fit in Table 2 reveal that GNCQR and FLQR provide the best fits, even beating the BRW estimator, for y_1, y_2, and y_3.[15] However, for y_4, FLQR fails to yield improvements over BRW. While GNCQR 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. This is because GNCQR recovers BRW when $\alpha = 1$, so it does not do much worse than BRW. Hence, GNCQR is more robust to DGPs that have quantile-specific sparsity than the simple FLQR.

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, 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 GNCQR 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.

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

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
GNCQR	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	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
GNCQR	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	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
GNCQR	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	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
GNCQR	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	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
GNCQR	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	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
GNCQR	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	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
GNCQR	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	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
GNCQR	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	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

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 GNCQR 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 GNCQR provides the best linear approximation of the quantiles while regularising some of the quantile variation in the coefficients.

In summary, GNCQR and FLQR provides better fit than BRW for y_1, y_2, and y_3, while for y_4 BRW performs best but GNCQR is nearly as good and both are preferred over the FLQR. This robustness of GNCQR 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 overshrinks. Taking the goodness-of-fit and variable selection results together, we conclude that GNCQR provides good fit while retaining better variable selection properties than FLQR.

4 Growth-at-Risk results

The canonical GaR uses US quarterly GDP growth in conjunction with the NFCI to obtain estimates of downside risk. Using quantile regression, GaR estimates can be obtained by estimating the following model:

$$y_{t+h} = x_t' \beta_{\tau_q} + \varepsilon_{t+h}, \tag{8}$$

for $t = 1, \cdots, T-h$, where h refers to the forecast horizon. x_t includes a constant (intercept), current quarterly GDP growth (annualised), and the average of the NFCI for the given quarter. Note that we focus on forecasting specific quantiles of the distribution of y_{t+h} conditional on x_t. Our data are quarterly and cover the period 1973Q1 to 2023Q1. This sample includes the original sample of Adrian et al. (2019), augmented by the COVID-19 crisis. We consider one- and four-quarter ahead forecast horizons ($h = 1, 4$). We follow Adrian et al. (2019) and forecast average growth rates over the specific forecast horizons.

To evaluate the performance of the different estimators, we first conduct a small psuedo forecasting exercise similar to Carriero et al. (2025). This is followed by examining the coefficient profiles of the different models to see whether we can corroborate the finding of Adrian et al. (2019), i.e. that lagged GDP growth acts as a location shifter while NFCI is the key driver of nonlinearities in GDP growth. We briefly consider the in-sample and out-of-sample weighted residuals for several values of α to judge the bias-variance trade-off. We then turn to the policy implications of selecting models that incorporate interquantile shrinkage.

4.1 Out-of-Sample Performance

To evaluate out-of-sample performance of the estimators we conduct a pseudo forecasting exercise similar to the one in Carriero et al. (2025). For the one- and four-quarter ahead forecast horizon we compute the out-of-sample GDP densities over an expanding window, where the initial in-sample period uses the first 50 observations of the sample. This means that there are in total $150 - h$ forecast windows to evaluate.

4.1.1 Quantile Crossing Incidence

Before evaluating the out-of-sample performance, we note that $\alpha = 1$ only ensures non-crossing quantiles in-sample but not out-of-sample. However, setting $\alpha > 1$, the non-crossing constraints are tighter which leads to shrinking the fitted quantiles towards parallel lines. This in turn leads to less out-of-sample crossing as well. Since the FLQR also shrinks towards parallel fitted quantiles, a key question is whether the FLQR also leads to less out-of-sample crossing. To evaluate this, we compare the out-of-sample crossing incidence for all the estimators (QR, BRW, GNCQR, FLQR). The crossing incidence is calculated by comparing the fitted quantiles with the sorted quantiles following the procedure of Chernozhukov et al. (2010):

$$CrossI = \frac{1}{Q(T-50)} \sum_{t=50}^{T} \sum_{q=1}^{Q} \mathbf{I}[\hat{\mathcal{Q}}(\tau_q, t+h|\mathcal{F}_t) \neq \hat{\mathcal{Q}}_{sort}(\tau_q, t+h|\mathcal{F}_t)], \tag{9}$$

where $\hat{\mathcal{Q}}(\tau_q, t+h|\mathcal{F}_t)$ is the forecasted quantile at time t and forecast horizon h and $\hat{\mathcal{Q}}_{sort}(\tau_q, t+h|\mathcal{F}_t)$ is the sorted forecast quantile. The crossing incidence measures the proportion of quantiles that need sorting after estimation to obtain a valid CDF. The lower the $CrossI$ value, the less quantiles need to be rearranged after estimation. The out-of-sample crossing incidence results are presented in Table 3.

Table 3: Out-of-Sample Crossing Incidence for the different estimators

	$h=1$	$h=4$
QR	6.32%	5.62%
BRW	1.37%	1.47%
$GNCQR$	0.91%	0.00%
$FLQR$	3.19%	1.79%

The first thing to note is that more crossing occurs at the shorter forecast horizon than the longer one for all estimators. Second, the out-of-sample crossing incidence is the highest for the traditional quantile regression method. This indicates that the rearrangement algorithm of Chernozhukov et al. (2010) would have the most gains for the traditional quantile estimator. Note that, as mentioned by Bondell et al. (2010), post-processing methods like the rearrangement method have the potential to improve fitted quantiles without changing the estimated coefficients. Third, GNCQR yields the lowest out-of-sample crossing incidence, even yielding proper densities for $h = 4$. Finally, the FLQR estimator provides worse out-of-sample crossing than the BRW estimator. Taken together these results indicate that the GNCQR provides the best potential to yield proper forecasted densities.

4.1.2 Forecast Performance

Given that the sorting procedure is widely used in forecasting settings, we evaluate forecast performance of sorted and unsorted quantiles. Both sorted and unsorted quantiles are considered because of the results in Table 3 showing that sorting is likely to improve the performance of the QR and FLQR more than the BRW and GNCQR. By examining both sorted and unsorted forecast performance we can check whether correctly identifying quantile varying variables yields gains beyond sorting quantiles.

We use several measures to evaluate forecast performance. First, the quantile score (QS) is used with the tick-loss weighted residual for a given forecast observation (Giacomini and Komunjer, 2005):

$$QS_{\tau_q,t+h} = (y_{t+h} - \hat{\mathcal{Q}}(\tau_q, t+h|\mathcal{F}_t))(\tau_q - \mathbf{I}[|y_{t+h} \leq \hat{\mathcal{Q}}(\tau_q, t+h|\mathcal{F}_t)|]), \tag{10}$$

where $\hat{\mathcal{Q}}(\tau_q, t+h|\mathcal{F}_t)$ is the forecasted quantile and y_{t+h} is the value of GDP at time $t+h$. Note that, potentially y_{t+h} can be unobserved, and the above QS device is simply meant to compare estimators by obtaining profiles across different quantiles and future values.

To integrate the above construction into an overall picture of density forecast performance we follow Knotek and Zaman (2019) and Carriero et al. (2025) in using diffusion indices to judge relative forecast performance of the different estimators. Specifically, our model comparison is based on the traditional quantile estimator (QR) as the baseline model. Let $rQS_{i,\tau_q,h}$ be the ratio of the Quantile Score for estimator i, at quantile τ_q, for forecast horizon $h = 1, 4$. Given the ratios of QS we calculate the case specific index as follow:

$$f_{i,\tau_q,h} = \begin{cases} 1 & \text{if } rQS \geq 1 + s \\ -1 & \text{if } rQS \leq 1 - s \\ 0 & \text{otherwise} \end{cases}, \tag{11}$$

where s is the chosen sensitivity in relative performance. To obtain the diffusion index we take the average of $f_{i,\tau_q,h}$ over τ_q, and h: $DI_i = N^{-1} \sum_{\forall \tau_q,h} f_{i,\tau_q,h}$. By construction DI_i lies between 1 and -1, with lower values indicating better performance relative to the QR forecasts. Values closer to zero indicate no substantial difference between the estimator and traditional quantile regression.

Note that the diffusion index can be sensitive to the chosen value of s. Setting s to smaller values makes the performance index more sensitive to changes relative to the QR, while setting s to larger values will only consider large improvements in forecast performance. As such, there is ambiguity in the literature regarding an appropriate choice of sensitivity index. Carriero et al. (2025) chose $s = 0.05$ while Knotek and Zaman (2019) set $s = 0.1$. Rather than pick one specific value for s we compare various sensitivity values: $s = 0.005, 0.01, 0.05$, and present results for the diffusion indices for the various s values for the unsorted and sorted cases. The results for the diffusion index are shown in Table 4. Note that in this table the reference model is always the respective QR model, i.e. for unsorted quantiles the reference is QS derived from the unsorted QR fits while for the sorted quantiles the reference is the QS derived from the sorted QR fits.

Table 4: Diffusion Index for the estimators at different sensitivities

	Unsorted quantiles			Sorted quantiles		
	BRW	GNCQR	FLQR	BRW	GNCQR	FLQR
$s = 0.005$	-28.95%	-36.84%	15.79%	-15.79%	-10.53%	31.58%
$s = 0.01$	-13.16%	-31.58%	5.26%	0.00%	-7.89%	21.05%
$s = 0.05$	-7.89%	-5.26%	0.00%	0.00%	0.00%	2.63%

Table 4 reveals that the GNCQR and BRW obtains gains over the QR estimator regardless of whether the sorted or unsorted QS is used. It is also apparent that much of the improvements are smaller given that the DI becomes less negative the larger s is. For the unsorted quantiles, the GNCQR and BRW yield improvements over the QR even when setting $s = 0.05$. The same cannot be said for the sorted quantiles, where s needs to be set to 0.01 to reveal any tangible difference in performance between the sorted QR quantiles and the sorted GNCQR quantiles. Furthermore, the sorted BRW quantiles perform on par with the sorted QR. From Table 3 we also noted that the GNCQR improves the least from sorting given it has the lowest crossing incidence. Therefore, it is reassuring that even in the sorted quantiles, the GNCQR provides gains over sorted quantiles corroborating the fit results from our Monte Carlo exercise. We also find that the FLQR yields no improvements in quantile score compared to QR regardless of what s is set and whether we sort the quantiles. In essence, these findings add to Carriero et al. (2025), namely that fused shrinkage is also important for growth-at-risk. Importantly, our results indicate that perhaps just as important is how the interquantile shrinkage is induced: adaptive non-crossing constraints yield improvements, while the traditional fused LASSO seem to lead to worse forecasting performance.

While the diffusion index of the QS gives a broad view of how the estimator performs, it does not inform us about where in the distribution the estimator does better (or worse). To evaluate this, we also construct quantile weighted CRPS (qwCRPS) scores of Gneiting and Ranjan (2011) for all the estimators for the different forecast horizons. To calculate these measures, we take the QS and apply quantile specific weights:

$$qwCRPS_{t+h} = \int_0^1 w_{\tau_q} QS_{t+h,\tau_q} d\tau_q, \tag{12}$$

where w_{τ_q} denotes a weighting scheme to evaluate specific parts of the forecast density. Through different weighting schemes, we can evaluate differences at different parts of the distribution. We consider four different weighting schemes: (a) $w^1_{\tau_q} = \frac{1}{Q}$ places equal weight on all quantiles (denoted as CRPS), which is equivalent to taking the average of the weighted residuals; (b) $w^2_{\tau_q} = \tau_q(1-\tau_q)$ places more weight on central quantiles; (c) $w^3_{\tau_q} = (1-\tau_q)^2$ places more weight on the left tail; and (d) $w^4_{\tau_q} = {\tau_q}^2$ places more weight on the right tail.

The results using the different weighting schemes are presented in Table (5). The unsorted results show that GNCQR yields the lowest weighted residual for all forecast horizons, for all weighting schemes except the right tail. These results indicate that the gains in forecast

Table 5: Forecast results of the different estimators

	Unsorted				Sorted			
	CRPS	Centre	Left	Right	CRPS	Centre	Left	Right
h=1								
QR	0.916	0.173	0.280	0.463	0.909	0.172	0.277	0.459
BRW	0.914	0.172	0.279	0.462	0.911	0.173	0.278	0.461
GNCQR	0.910	0.172	0.274	0.464	0.909	0.172	0.274	0.463
FLQR	0.918	0.173	0.279	0.466	0.914	0.173	0.277	0.463
h=4								
QR	0.559	0.108	0.168	0.284	0.554	0.107	0.167	0.280
BRW	0.554	0.107	0.166	0.281	0.553	0.107	0.166	0.280
GNCQR	0.552	0.107	0.165	0.280	0.552	0.107	0.165	0.280
FLQR	0.560	0.108	0.167	0.285	0.557	0.108	0.166	0.284

performance in GNCQR largely accrue from better left tail and centre forecasts. Given that the growth-at-risk framework was designed to capture downside (left tail) risk, the inability to yield better right tail forecasts for the short forecast horizon is less of a concern. These results also show that BRW is a strong candidate to obtain density estimates. In particular, it is often only beaten by GNCQR. This highlights the usefulness of BRW in density estimation and for most applications should suffice as a first candidate to estimate.

While sorting the fitted quantiles leads to improved forecast performance for almost all estimators, GNCQR retains its strong forecast performance. Furthermore, GNCQR has better unsorted left tail forecasts than the sorted QR for both forecast horizons. For the one-year ahead forecast horizon, GNCQR has better or on par forecast performance relative to the sorted QR across every part of the distribution. This is in line with the results of the Monte Carlo exercise and highlights that correctly identifying which variables drive quantile variation leads to performance gains beyond simply sorting the fitted quantiles.

The improvement in left tail forecasts is particularly important as it means that GNCQR is better at capturing low-probability high-impact negative events. This is particularly useful for policymakers and financial institutions. For example, it allows for more precise calibration of macroprudential policies, such as the Countercyclical Capital Buffer, that aim to safeguard financial stability during periods of economic stress. With more accurate GaR measures, one can better tune the regulatory tools available to policymakers.

Overall, our proposed method provides gains in forecasting. This highlights how even in small (covariate) dimensional settings, one can obtain better forecast performance through correctly identifying which variables are quantile varying.

4.2 Coefficients

While quantile regression is robust to outliers, extreme values in the covariates can still cause issues. Since the GaR is essentially a quantile autoregressive model of Koenker and Xiao (2006), the extreme GDP observations of the COVID period enter the independent variables through the lag. This could potentially have undue influence on the coefficients. To explore this potential issue, we run the model on two sample periods: pre-Covid and full sample.

The quantile profiles of the estimated β coefficients are presented in Figures 2 and 3, for pre-Covid and full sample. β_1 is the coefficient for GDP growth and β_2 corresponds to the NFCI. A key observation is that the coefficient for the NFCI has a distinct quantile-varying profile for both samples and forecast horizons. The estimators show a general increasing trend across the quantiles. Nevertheless, there are some difference between the different estimators. First, FLQR is likely to shrink quantile variation at the tails, particularly the lower tail. Hence, FLQR has a less pronounced impact of NFCI on the lowest quantiles than the other estimators. Second, the traditional QR estimator has a positive coefficient for the upper quantiles of NFCI, while GNCQR gives an estimate close to zero. Adrian et al. (2019) argue that a zero coefficient for the upper quantiles of NFCI is more meaningful than positive coefficients. Note that none of the estimators have shrinkage imposed on the β, i.e. there is no explicit variable selection, yet GNCQR obtains $\hat{\beta}$'s much closer to 0 at the upper quantile of NFCI. Third, GNCQR has less 'jagged' quantile profile. In fact, for the NFCI coefficient quantile profile, GNCQR only has negative slope for the pre-Covid sample for the $h = 1$ forecast horizon. For all the other cases, GNCQR shows a gentle upward sloping profile.

While there is significant quantile variation for the coefficient of NFCI, the same cannot be said for the coefficient of GDP growth. The GNCQR yields virtually no quantile variation in almost all cases, with only pre-COVID of $h = 1$ showing some variation. The other estimators also show less quantile variation for GDP growth than the NFCI, but they do not fully smooth out spurious quantile variation. Interestingly, FLQR also does not shrink away the quantile variation in GDP growth.

These coefficient profiles demonstrate that GNCQR is capable of identifying quantile variation better than FLQR. Furthermore, for the variables that have quantile variation, GNCQR shrinks away the "jagged" edges resulting in smoother quantile profiles. This is likely the reason for the GNCQR's superior forecast performance: the GNCQR is able to identify quantile varying variables and shrink away any unnecessary variation in these variable profiles.

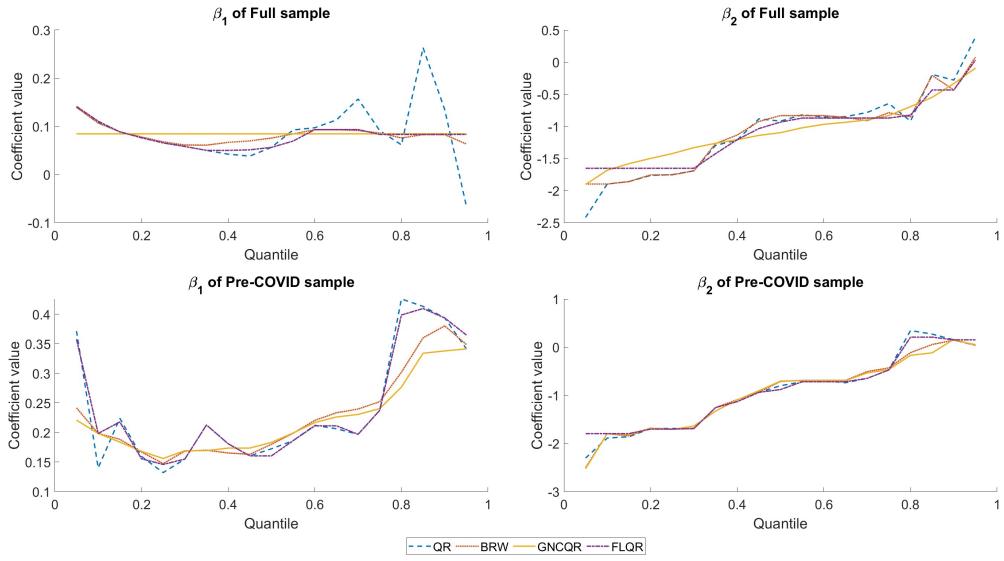

Figure 2: β coefficients for the different estimators at $h = 1$

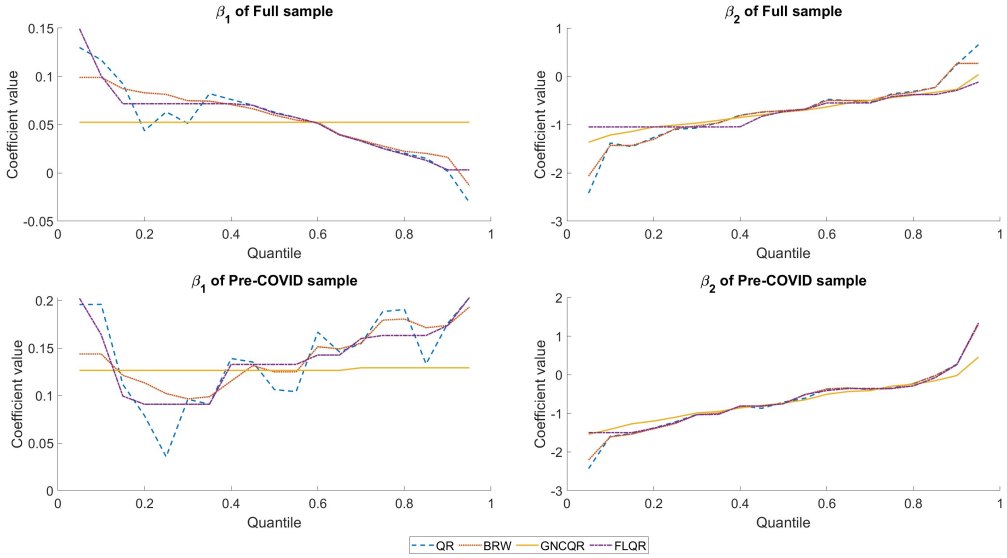

Figure 3: β coefficients for the different estimators at $h = 4$

These results verify the findings of Adrian et al. (2019), namely that macro-financial linkages are important for downside risk of GDP growth. Notably, the GNCQR completely shrinks the quantile variation in the GDP coefficient, implying that, while past GDP is important in determining the location of the GDP distribution, most variation in the shape of the distribution is due to changes in the financial conditions. Further, we add to the finding of Carriero et al. (2025): beyond shrinkage, interquantile shrinkage has benefits for empirical macroeconomics. This is particularly important in estimating the coefficient profiles, since post-processing sorting applies only to the fitted quantiles and not the coefficient profiles.

4.3 Bias-variance trade-off

Figures (4) and (5) plot the in-sample and out-of-sample average quantile scores for different α values and compares the profiles with the fused LASSO estimator with the same hyperparameter value.

The figures show, as expected, that the equivalent hyperparameters of the different estimators do not lead to the same values. This in turn leads to different profiles as the hyperparameter is varied. GNCQR shows a more gradual change in in-sample fit for all horizons and sub-samples. Furthermore, the out-of-sample fits of the GNCQR have a more pronounced minimum. Another feature of the GNCQR is that the in-sample and out-of-sample profiles are less "jagged" than that of the FLQR. This is true for all forecast horizons, as well as pre- and post-Covid. These findings are important for model selection: it is easier to optimise the tuning parameter selection when the minimum is obtained gradually. Furthermore, these properties indicate that the GNCQR shrinks quantile variation of the correct variable gradually.

4.4 Policy implications

The previous section highlighted the importance of fused shrinkage for forecasting and estimating coefficient profiles that have less spurious quantile variation. In this section we highlight how these differences in the estimators could impact policy, focusing on the left-tail.

4.4.1 Expected Shortfall

The first policy tool we investigate is expected shortfall (ES) (Adrian et al., 2019), which measures the total probability mass that the conditional distribution assigns to the left tail:

$$ES_{t+h} = \frac{1}{\pi} \int_0^{\pi} \hat{F}^{-1}_{y_{t+h|\mathcal{F}_t}}(\tau|\mathcal{F}_t)d\tau \tag{13}$$

To construct ES from the estimated conditional quantiles we follow the methodology of Mitchell et al. (2024), using a nonparametric method to construct densities from estimated quantiles. We construct the ES for the 5^{th} quantile using this nonparametric approach. The constructed ES are shown in Figure 6. While the same procedure can be used to estimate Expected Longrise, we opt to only focus on ES, since the variables used in the GaR model are more likely to capture downside risk. Indeed, this paper focuses almost entirely on the policy problem of identifying downside risk in GDP growth.

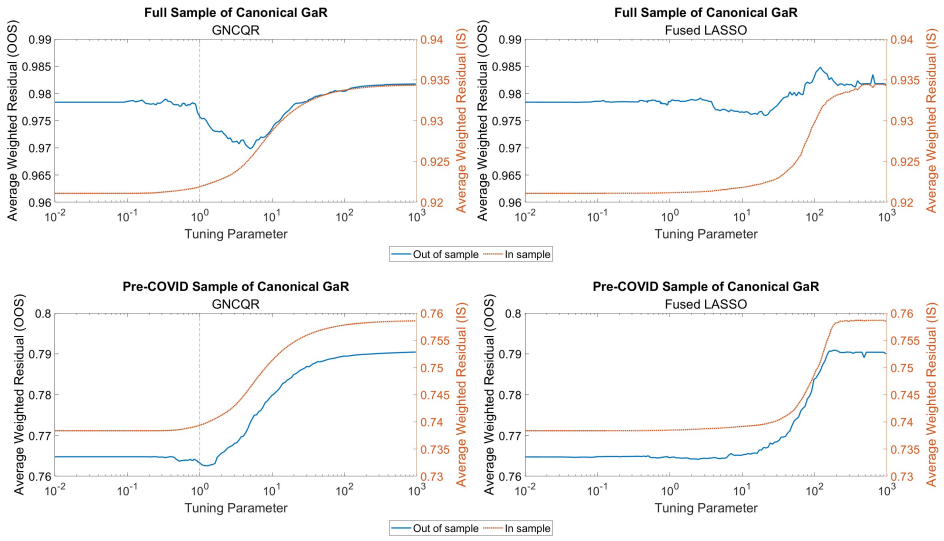

Figure 4: In-sample vs. Out-of-sample fit for various hyperparameter values ($h = 1$)

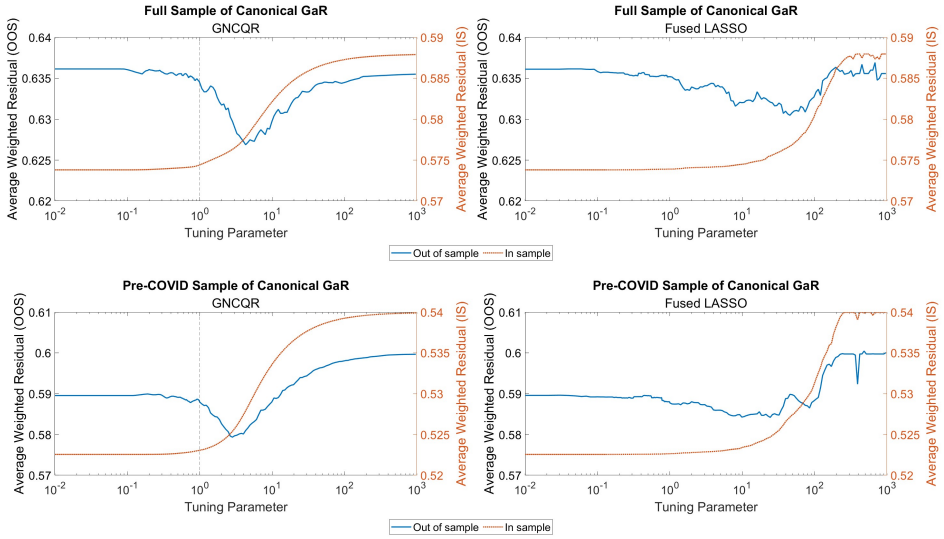

Figure 5: In-sample vs. Out-of-sample fit for various hyperparameter values ($h = 4$)

Figure 6 reveals that there is a stark difference between the estimators for ES, especially at the one year horizon ($h = 4$). These ES estimates are critical for policy-makers as they give an overview for the evolution of downside risk.

Focusing on the difference between the estimators, we find that the QR estimator produces the most volatile ES compared to the other estimators, especially at the start of the estimation period. Compared to QR, the BRW estimator produces smoother ES over time, while still dipping during extreme financial stress: particularly the global financial crisis and the 1970s energy crisis.

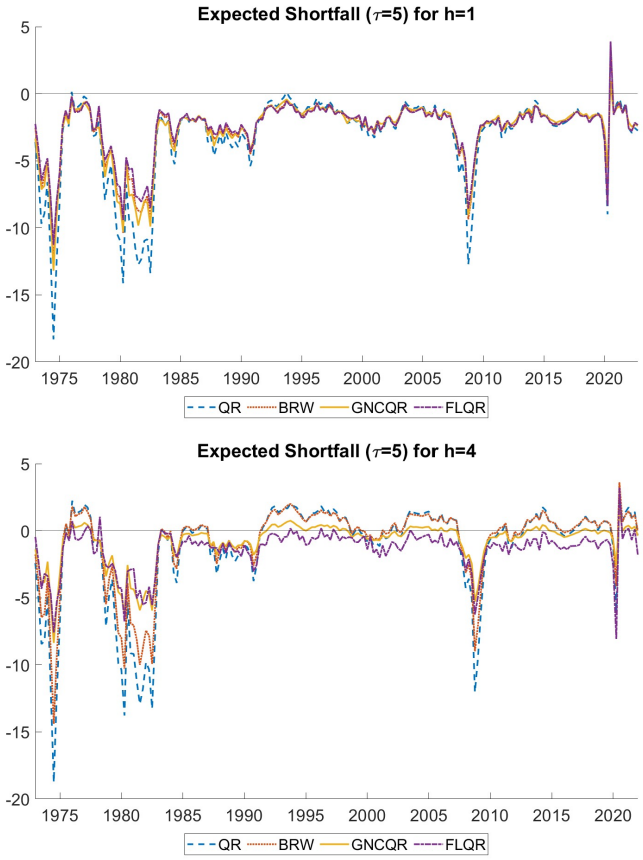

Figure 6: Expected Shortfall for the different estimators

FLQR deviates the most from the QR and BRW, producing more optimistic ES during financial stress episodes, but more pessimistic ES during more tranquil periods. This pattern is particularly prominent for the one year forecast horizon. Notably, FLQR shows consistently more negative ES estimates at this horizon during relatively stable market periods (e.g., 1990-2005), suggesting a potential pessimistic bias of the FLQR.

Similar to the variable selection results, the ES based on GNCQR coefficients yields a middle-ground: between the QR's ES and the FLQR's ES. In particular, for the one-year forecast horizon the GNCQR ES has less negative values than QR during financial stress episodes. On the other hand, during tranquil periods the GNCQR method yields close to zero or positive ES values, which are greater than those for the FLQR. This in turn leads to smoother ES dynamics than QR. Smooth measures of tail-risk that still capture episodes of stress are important for policy makers as abrupt and potentially unnecessary changes in regulatory tools impose costs on the financial sector. These advantages of GNCQR are likely on account of its ability to shrink away spurious quantile variation, which helps in correctly locating variables that drive downside risk.

The difference between QR and the other estimators, particularly during the early sample period, highlights the value of identifying variables that drive quantile variation. GNCQR can effectively moderate the extreme volatility of traditional QR while maintaining sensitivity to genuine downside risks. We argue that this leads to more reliable and practical risk assessments for financial stability purposes.

4.4.2 Tail local projection

Another important policy tool for GaR is local projection (LP) (Loria et al., 2025; Ruzicka, 2021). Ruzicka (2021) notes that there is an identification assumption embedded in Adrian et al. (2019), namely that NFCI has no contemporaneous effects on GDP growth distribution. This assumption was used by Ruzicka (2021), Wojciechowski (2024), and Chavleishvili and Manganelli (2024) to compute quantile IRFs and quantile local projections. Importantly, these methods rely on the estimated coefficients, which are not affected by sorting, to construct the quantile projections through time. We estimate quantile local projections using the four estimators for the 5^{th}, 10^{th}, and 25^{th} quantile. This follows Ruzicka (2021) but departs from Adrian et al. (2019) on two fronts: a) we estimate future growth directly, rather than averaged over the forecast horizon; and b) we include an additional set of lags for GDP and NFCI. The quantile LPs of the response of GDP to a unit shock in NFCI are shown in Figure 7.

Across all three quantiles, there is a general pattern of initial decline followed by recovery, but the timing and magnitude of the dynamics vary by estimator. The QR consistently produces the most volatile projections, while FLQR shows the quickest recovery. BRW and GNCQR typically fall between these extremes, with GNCQR often showing slightly less negative projections than BRW, particularly in the medium run ($h = 4$ to $h = 8$).

At the 5^{th} quantile, we see the largest differences across the estimators. The QR estimator exhibits high volatility over time, with deeper initial drops reaching below -4 around $h = 1$ and $h = 2$ and again around $h = 6$ and $h = 7$. This volatility of the QR projections is noted in the literature, and was the primary motivation for Ruzicka (2021) to propose smoothing the projections over the LP horizons. By contrast, the FLQR estimator shows a remarkably different pattern, returning to near-zero values much more quickly after $h = 4$, reflecting more conservative estimates of tail risk. Figures 2 and 3 showed the coefficient profiles for FLQR being prone to shrinking quantile variation at the tails. As such the quick recovery of FLQR is a consequence of this overshrinking character.

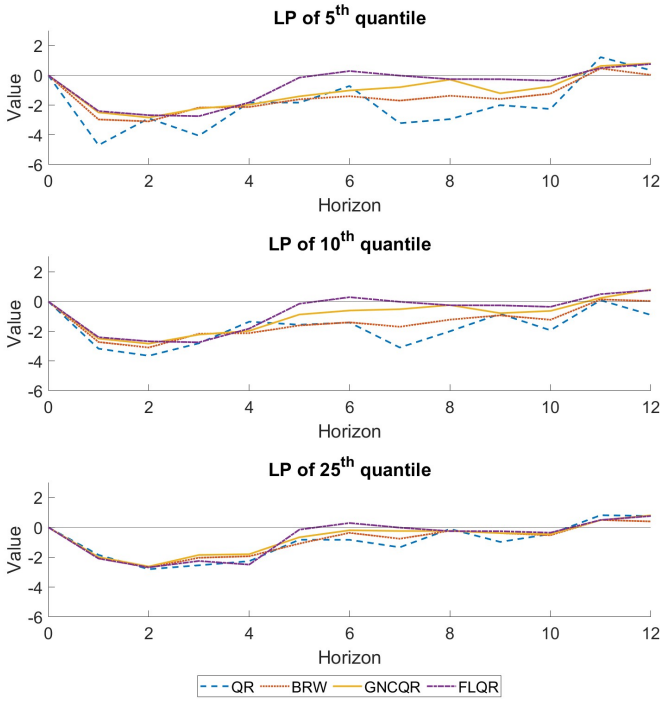

Figure 7: Local projections for the different estimators at different quantiles

GNCQR and BRW produce smoother projections than the QR but do not return to zero as quickly as the FLQR. This finding is particularly interesting and suggests that potentially the smoothed quantile LP framework of Ruzicka (2021) induced fused shrinkage at each horizon. We leave for future research to explore the link between smoothing quantile variation and smoothing quantile projections over time.

The differences between estimators decrease as we move up the quantiles, with the 25^{th} quantile showing limited difference between the estimated profiles especially in the short run. After $h = 4$, the estimators begin to diverge again, though less than that at the lower quantiles.

These differences highlight the importance of estimator choice in GaR analyses, especially when focusing on extreme quantiles that are of particular interest to policymakers concerned with tail risks to economic growth. The varying persistence and magnitude of negative projections across estimators could lead to significantly different assessments of downside risks and potentially different policy measures.

5 Conclusion

This paper develops an adaptive non-crossing constraint, encompassing as special cases the traditional quantile regression estimator, the Bondell et al. (2010) non-crossing estimator, as well as the composite quantile regression, by varying the tightness of the constraint. By developing non-crossing constraints that can be thus tightened, we study the properties of these constraints on the estimated β parameters and forecasts. Doing so reveals that non-crossing constraints are simply a type of Fused LASSO with quantile specific shrinkage parameters.

Through a Monte Carlo experiment, we verify that imposing non-crossing constraints is equivalent to introducing fused shrinkage. We also show how our proposed Generalised Non-Crossing Quantile Regression (GNCQR) 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 (FLQR) estimator overshrinks while BRW estimator undershrinks relative to the GNCQR.

Taking an interquantile shrinkage lens to US Growth-at-Risk, we find that our proposed GNCQR estimator outperforms both traditional quantile regression and regular fused LASSO approaches, particularly in forecasting left-tail risks. The GNCQR effectively identifies variables that drive quantile variation while shrinking away spurious variation, resulting in smoother coefficient profiles and improved forecast performance. Notably, our results confirm the findings of Adrian et al. (2019) that financial conditions constitute key drivers of nonlinearities in GDP distribution, while past GDP growth acts primarily as a location shifter.

The policy implications of better estimated GaR coefficients is also explored. Expected Shortfall estimates derived from GNCQR exhibit less volatility while remaining sensitive to prominent financial stress episodes, offering policymakers a more reliable assessment of downside risks. Similarly, quantile local projections based on GNCQR provide balanced medium-term forecasts that avoid both the excessive volatility of traditional quantile regression and the potentially over-optimistic recovery paths suggested by standard fused LASSO. These improvements in policy-relevant metrics underscore the importance of appropriate interquantile shrinkage in macroeconomic risk assessment.

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 has clear advantages of better coefficient estimates and forecasts and importantly, the ability to incorporate prior information or beliefs held by policymakers. These developments are retained for future research.

The formulation of the GNCQR is very simple: 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 also for future research.

Finally, since the bias introduced by non-crossing constraints is similar to biased bootstrap methods proposed for regression under monotone or convexity constraints, 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 a promising avenue for future research.

References

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

Adrian, T., N. Boyarchenko, and D. Giannone (2021). Multimodality in macrofinancial dynamics. *International Economic Review 62*(2), 861–886.

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).

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

Carriero, A., T. E. Clark, and M. Marcellino (2025). Specification choices in quantile regression for empirical macroeconomics. *Journal of Applied Econometrics 40*(1), 57–73.

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.

Chavleishvili, S. and S. Manganelli (2024). Forecasting and stress testing with quantile vector autoregression. *Journal of Applied Econometrics 39*(1), 66–85.

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.

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

Giacomini, R. and I. Komunjer (2005). Evaluation and combination of conditional quantile forecasts. *Journal of Business & Economic Statistics 23*(4), 416–431.

Gneiting, T. and R. Ranjan (2011). Comparing density forecasts using threshold-and quantile-weighted scoring rules. *Journal of Business & Economic Statistics 29*(3), 411–422.

Iseringhausen, M., I. Petrella, and K. Theodoridis (2025). Aggregate skewness and the business cycle. *Review of Economics & Statistics Forthcoming*.

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.

Jones, M. C. (1994). Expectiles and m-quantiles are quantiles. *Statistics & Probability Letters 20*(2), 149–153.

Knotek, E. S. and S. Zaman (2019). Financial nowcasts and their usefulness in macroeconomic forecasting. *International Journal of Forecasting 35*(4), 1708–1724.

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*.

Kohns, D. and T. Szendrei (2024). Horseshoe prior bayesian quantile regression. *Journal of the Royal Statistical Society Series C: Applied Statistics 73*(1), 193–220.

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.

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

Loria, F., C. Matthes, and D. Zhang (2025). Assessing macroeconomic tail risk. *The Economic Journal 135*(665), 264–284.

Mitchell, J., A. Poon, and D. Zhu (2024). Constructing density forecasts from quantile regressions: Multimodality in macrofinancial dynamics. *Journal of Applied Econometrics 39*(5), 790–812.

Newey, W. K. and J. L. Powell (1987). Asymmetric least squares estimation and testing. *Econometrica 55*(4), 819–847.

Powell, D. (2020). Quantile treatment effects in the presence of covariates. *Review of Economics & Statistics 102*(5), 994–1005.

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.

Ruzicka, J. (2021). Quantile local projections: Identification, smooth estimation, and inference. *Universidad Carlos III de Madrid*.

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

Sobotka, F. and T. Kneib (2012). Geoadditive expectile regression. *Computational Statistics & Data Analysis 56*(4), 755–767.

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.

Wojciechowski, R. (2024). A structural approach to growth-at-risk. *arXiv preprint arXiv:2410.04431*.

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)

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

- Generalised Non-Crossing Quantile Regression (proposed here)

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

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.[16] 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:

$$\{I(\varepsilon_{\tau_{q-1}} - \xi \leq 0)\} \subseteq \{I(\varepsilon_{\tau_q} \leq 0)\} \tag{14}$$

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 (14) when $\xi = 0$.

As α of Equation (6) goes above 1, the ξ term in the quantile subset property (Equation (14)) 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 (14) 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 (14) trivial to satisfy.

[16] 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. The degree of bias can be used to test model adequacy, which is a topic retained for future research.

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. We retain consideration of the GNCQR using other CV methods for future research.

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 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).[17]

[17]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.