

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.

November 6, 2024

Abstract

Quantile crossing has been a challenge for quantile regression, leading to research in how to obtain monotonically increasing quantile estimates. While important contributions, these papers do not provide insight into how enforcing monotonicity influences the estimated coefficients. This paper fills this gap and shows that non-crossing constraints are a type of fused-shrinkage. The proposed estimator has good fit and (fused) variable selection properties: it can reliably identify quantile varying parameters. We investigate the ‘heat-or-eat’ dilemma and show that prepayment has a non-linear impact on households’ consumption choices. In a growth-at-risk application the estimator has the best forecast performance.

Keywords: Fused-shrinkage, Quantile Regression, Non-Crossing Constraints.

*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. The authors would also like to thank Andrew Burlinson in providing guidance with UKHLS. 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

Since the work of Koenker and Bassett (1978), quantile regression has been an active field of research. One aspect that has seen recent developments is estimating quantile curves that do not cross. In essence, quantile crossing occurs when the fitted quantiles do not increase monotonically with the estimated quantile. To address this there have been several proposals: Chernozhukov et al. (2009) suggests sorting the fitted quantiles to obtain proper densities, Mitchell et al. (2024) proposed constructing densities from quantile estimates, while Adrian et al. (2019) and Korobilis (2017) opts to fit densities on the estimated quantiles.

The key point about the above methods is that they are directed only at the fitted density. If the sole interest is fit, then post-processing solutions are appropriate. Nevertheless, one might also be interested in estimating the model parameters.¹ One such way to do this is via sequential estimation (Liu and Wu, 2009). Here the practitioner estimates the median first, and sequentially estimates 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. To address this, Bondell et al. (2010) propose estimating quantiles jointly, with a set of non-crossing constraints. This last method is particularly powerful, as elements of the constraints have been carried over to other quantile estimators (see Yang and Tokdar (2017) for example). But how these constraints impact the estimated model parameters has not been much explored in the literature.

The paper introduces a new estimator with adaptive non-crossing constraints that generalises some pre-existing quantile estimators. This is achieved by implementing a set of non-crossing constraints that can be scaled, which in turn makes the non-crossing constraints tighter or looser. By doing so we are able to study the impact on the estimated coefficients of tightening the non-crossing constraint. We propose adaptive constraints which encompass traditional quantile regression ($\alpha = 0$), to Bondell et al. (2010) (hereinafter BRW) non-crossing quantile regression ($\alpha = 1$), as well as composite quantile regression (hereinafter CQR) of Koenker (1984) and Zou and Yuan (2008) when the parameter is set to very high levels ($\alpha \rightarrow \infty$). We call this new estimator the ‘Generalised Non-crossing Constraint Quantile Regression’ (GNCQR), as we can recover the various types of quantile estimators by varying the hyperparameter.² We use

¹An additional feature of methods that focus on estimating the model parameters is that they can improve the fitted density as well.

²We follow Powell (2020) in the naming convention, who develop the ‘Generalized Quantile Regression’.

this adaptive non-crossing constraint to show that non-crossing constraints are simply a special type of fused shrinkage, i.e., shrinkage on the difference in β parameters across quantiles.³

We conduct comprehensive Monte Carlo exercises based on those used in the original paper by Bondell et al. (2010), that first 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 original BRW estimator. We also consider the variable selection properties of the different estimators and demonstrate that the BRW estimator employs a type of fused shrinkage. The Monte Carlos also reveal how the BRW estimator undershrinks quantile variation while the traditional Fused Lasso, as outlined in Jiang et al. (2013), overshrinks.

We then examine the trade-off between in-sample and out-of sample fits of GNCQR and FLQR in two applications. As expected, as the hyperparameter increases the in-sample fits of both estimators gets worse. Nevertheless, there is a difference between the estimators in how this deterioration in fit occurs: the in-sample fit profile of GNCQR deteriorates more gradually as the hyperparameter is increased. We provide an intuitive explanation for this: as we increase α , we introduce bias in the quantile property and gradually start to introduce the ‘quantile subset’ property⁴.

In our cross-section empirical application, we consider the ‘heat-or-eat’ dilemma at the heart of the recent UK cost-of-living crisis applied to data on the 9 English regions. To do so we follow Burlinson et al. (2022) and focus on the impact of household expenditures on gas and electricity on the portions of fruit and vegetables they consume per week. We identify quantile variation in the majority of English regions, highlighting how (conditional on covariates) the variable of interest shows a clearly nonlinear impact on the types of food consumed by households. In this application we see the advantage of GNCQR: the estimator only imposes additional shrinkage if it yields improvements above BRW. We find that except for 2 regions of England, the optimal hyperparameter of GNCQR is around $\alpha = 1$, which in effect recovers the BRW estimates. This highlights the power of BRW estimator, and why it performs well if one is not specifically interested in optimising the degree of shrinkage in the quantile profile.

In our time series empirical application, we estimate the Growth-at-Risk (GaR) of US GDP, following Adrian et al. (2019). This canonical GaR includes only lagged GDP and a lagged financial conditions index. We show that our proposed method yields the most intuitive

³Jiang et al. (2013) refer to this as interquantile shrinkage.

⁴The quantile subset property is that the observations below the $(\tau - 1)^{th}$ quantile is a subset of the observations that fall below the τ^{th} quantile.

coefficient profiles and also offers the best forecast performance for most cases. In particular, the GNCQR estimator fares much better at distinguishing quantile varying variables than the traditional Fused LASSO. This shows that while the non-crossing constraints are a special type of fused shrinkage profile, when one would like to impose fused shrinkage in their quantile estimation, it is better to impose it via the adaptive non-crossing constraints proposed here.

This paper is structured as follows. In Section 2 we introduce the quantile regression of 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 them into the Fused LASSO constraint of Jiang et al. (2013). The section concludes by describing how the hyperparameter is chosen. Section 3 describes the Monte Carlo experiment before presenting the fit and variable selection performance of the different estimators. This is followed by the empirical application Section 4 using cross-sectional data. We first present results for the quantile coefficient of the payment variable for the different English regions. We then look at how the in-sample and out-of-sample fits change as the hyperparameter is increased, using data relating to London, and discuss the coefficients of all other covariates used in the estimation for London. Next we consider the time series data application that looks at the growth-at-risk in US data. We first examine the in-sample and out-of-sample fits as hyperparameters are changed before discussing the quantile profiles of GDP and NFCI, on 2 forecast horizons $h = 1$ and $h = 4$. We then compare the forecasting ability of the different models in a small pseudo out-of-sample forecast exercise. The paper concludes with the key takeaways of the method in Section 5.

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 P estimated quantiles can be used to describe the distribution of a response variable Y conditional on a vector of response variables $X \in \mathbb{R}^K$. Formally, the q^{th} quantile is modeled in regression setting as:

$$Q_q = X^T \beta_{\tau_q}.$$

The collection of the Q quantiles parameters $\beta = \{\beta_{\tau_1}, \beta_{\tau_2}, \dots, \beta_{\tau_Q}\}$ describe the conditional distribution. The goal is to estimate the vector of coefficients $\beta_{\tau_q} \in \mathbb{R}^K$ for all quantiles.⁵ This can be done using quantile regression:

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

$$\rho_q(u) = u(p - I(u < 0))$$

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 quantile regression’s. Furthermore, when setting $\tau = 0.5$ in an expectile regression, one recovers the OLS estimator, i.e., the conditional mean.

Expectile regression does come with some disadvantages. First, of expectiles is their interpretation: 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 are the τ^{th} quantile of some distribution which related to the original distribution of Y as shown in 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 disadvantages, in this paper we will focus on quantile regression and we leave for future research to explore how our findings translate to expectiles.

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

⁵Note that with this notation X includes a constant or intercept.

(Koenker, 2005). Such limits in data availability are frequently encountered 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). 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.

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:

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

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 interest to $\mathcal{D} \in [0, 1]^K$ and focus on the worst case scenario in the data,⁶ which reduces the number of constraints to $(Q - 1)$.⁷ This simplifies computation a great deal and enables non-crossing constraints to be included without too much additional computational cost. Because quantile regression is invariant to monotone transformations, any affine invertible transformation can be applied: to obtain the coefficients pertaining to the 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: $(\gamma_{0,\tau_1}, \dots, \gamma_{K,\tau_1})^T = \beta_{\tau_1}$ and $(\gamma_{0,\tau_q}, \dots, \gamma_{K,\tau_q})^T = \beta_{\tau_q} - \beta_{\tau_{q-1}}$ for $q = 2, \dots, Q$. With this quantile

⁶The 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.

⁷Technically, any domain of interest which has an affine transformation that maps to $\mathcal{D} \in [0, 1]^K$ works.

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:

$$\gamma_{0,\tau_q} \geq \sum_{j=1}^K \gamma_{j,\tau_q}^- \quad (q = 2, \dots, Q) \quad (3)$$

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

Throughout the paper we follow the framework and assumptions laid out 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 provided by Lemma 1 below.

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

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

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

Proof. It is easy to see that the right hand side of the equation is equivalent to the one in Bondell et al. (2010): simply use the min-max transformation, $\frac{a - \min(X_j)}{\max(X_j) - \min(X_j)}$, which equals 1 when $a = \max(X_j)$. Using the same formula one can see that it equals 0 when $a = \min(X_j)$. Using these conditions it is easy to see that equation (4) is equivalent to equation (3). Note that this equivalence only holds if one can scale the domain of equation (4) to be on $\mathcal{D} \in [0, 1]^K$. Further, we assume that $\max(X_j) > \min(X_j)$ for all $j > 0$, i.e., $\text{Var}(X_j) > 0$. \square

The above set of constraints are still a necessary condition for non-crossing because the point will be in \mathbb{R}^{K+1} . However, this also encompasses a sufficient condition, since if equation (4) is satisfied for the worst case, it is immediately satisfied for every observation. The novelty of equation (4) is that it works on the domain of \mathbb{R} , while the original formulation in Bondell et al. (2010), works for $\mathbb{D} \in [0, 1]$ only.

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 $\alpha = 0$. This leads to the following adaptive non-crossing constraints:

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

When $\alpha = 0$, these constraints impose 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 (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. Furthermore, these constraints equal the non-crossing constraints when $\alpha = 1$ as equation (5) 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).

While it is possible to construct alternate constraints for the traditional quantile regression when $\alpha = 0$, equation (5) yields traditional quantile regression β parameters without needing indicator functions in the formulation. This implies that as α increases, it has a gradual influence on the estimated β 's, by making the constraints tighter so that it becomes more difficult to satisfy the constraint in equation (5). When α is large enough, we can see that 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).⁸ This suggests that non-crossing constraints impact the coefficient profiles in a similar fashion to fused shrinkage, leading to Theorem 1.

⁸Note that in the limit there is no unique solution. Because of this, as $\alpha \rightarrow \infty$ it becomes numerically challenging to find the solution, leading to numerical instability.

Theorem 1. *In the Fused LASSO of Jiang et al. (2013),*

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

$$\text{s.t. } k^* \geq \sum_{k=1}^K (\gamma_{k,\tau_q}^+ + \gamma_{k,\tau_q}^-),$$

non-crossing constraints are equivalent to Fused LASSO shrinkage when $k^ \propto \frac{\gamma_{0,\tau_q}}{\alpha}$.*

Proof. Rearrange equation (5) and collect terms under one summation:

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

To see the equivalence more clearly, use Lemma 1 to recast the data to be on $\mathbb{D} \in [-1, 1]$ and assume that the each covariate is mean 0 after rescaling. We note that when we rescale a symmetric distribution, the mean becomes trivially close to 0. If the data is not distributed symmetrically, one can standardise the data before rescaling. Rescaling the data will lead us to:

$$\gamma_{0,\tau_q} \geq \sum_{j=1}^K \alpha (\gamma_{j,\tau_q}^- + \gamma_{j,\tau_q}^+)$$

Note that $\gamma_{j,\tau_q}^j \geq 0$, so the right hand side is always positive which entails that $(\gamma_{j,\tau_q}^- + \gamma_{j,\tau_q}^+) = |\gamma_{j,\tau_q}^- + \gamma_{j,\tau_q}^+|$. Rearranging the terms yields:

$$\frac{\gamma_{0,\tau_q}}{\alpha} \geq \sum_{j=1}^K |\gamma_{j,\tau_q}^+ + \gamma_{j,\tau_q}^-| \tag{6}$$

The right hand side of equation (6) is exactly how Jiang et al. (2013) and Jiang et al. (2014) ensure fused shrinkage. Note that the term $\frac{\gamma_{0,\tau_q}}{\alpha}$ is equivalent to a quantile specific k^* parameter of the usual Fused LASSO as described in these papers. We assume $\alpha > 0$ for this equivalence, as the definition does not hold for $\alpha = 0$. Consequently, this equivalence with k^* is only proportional. \square

While fused shrinkage for quantile regression has previously been suggested in the literature (Jiang et al., 2013, 2014), we are the first to show the equivalence between fused shrinkage and non-crossing constraints.⁹ Because of this equivalence with fused shrinkage, as $\alpha \rightarrow \infty$, the

⁹Jiang et al. (2013) explicitly state that one can include non-crossing constraints *in addition* to fused shrinkage. Given Theorem 1 it is clear that this would entail two different parametrisations of non-crossing constraints.

model converges to the composite quantile regression estimator of Koenker (1984) and Zou and Yuan (2008).

Because 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).¹⁰ Formally GNCQR is defined as:

$$\begin{aligned} \hat{\beta}_{GNCQR}|\alpha = \underset{\beta}{\operatorname{argmin}} & \sum_{q=1}^Q \sum_{t=1}^T \rho_q(y_t - x_t^T \beta_{\tau_q}) \\ \text{s.t. } & \gamma_{0,\tau_q} + \sum_{k=1}^K \left[\bar{X}_k - \alpha(\bar{X}_k - \min(X_k)) \right] \gamma_{k,\tau_q}^+ \geq \\ & \sum_{k=1}^K \left[\bar{X}_k + \alpha(\max(X_k) - \bar{X}_k) \right] \gamma_{k,\tau_q}^- \end{aligned} \quad (7)$$

Since equation (7) recovers commonly used quantile estimators given specific values of α , the GNCQR estimator inherits the key properties of these estimators. Specifically, when $\alpha = 0$, the estimator is consistent and the coefficients are asymptotically normal (Koenker and Bassett, 1978). When $\alpha = 1$, we can use Theorem 1 of Bondell et al. (2010), which states assuming no model misspecification the non-crossing estimator has the same asymptotic properties as quantile regression, to infer that the GNCQR retains consistency and asymptotic normality. When α is sufficiently large the asymptotic properties of the composite quantile regression apply to the GNCQR. Furthermore, due to the equivalence to Fused LASSO, as shown in Theorem 1, the GNCQR estimator retains the properties of the Fused LASSO for intermediate values of α .

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.¹¹ 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

¹⁰The different types of estimators are defined, with a unified notation, in the appendix.

¹¹The 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.

‘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:

$$\{I(\varepsilon_{\tau_{q-1}} - \xi \leq 0)\} \subseteq \{I(\varepsilon_{\tau_q} \leq 0)\} \quad (8)$$

Here $\{I(\varepsilon_{\tau_q} \leq 0)\}$ is 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 (8) when $\xi = 0$.

As α of equation (7) goes above 1, the ξ term in the quantile subset property (equation (8)) 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, quantile that get too close to each other are treated as if they cross. As such, when $\alpha > 1$, the ξ term in equation (8) 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 (8) trivial to satisfy.

2.4 Hyperparameter selection

A natural candidate for choosing the degree of penalisation is cross-validation. In cross-section applications, this would be the usual k -fold cross-validation or a variant thereof. Of the several types of cross-validation methods available for time series application, we choose the hv -block CV setup of Racine (2000). This block setup has several desirable properties: (1) it has been shown to have 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,¹² 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.¹³

¹²See Cerqueira et al. (2020) for a description and comparison of the different types of cross-validation for time-series data

¹³We use this equivalence in the Monte Carlo setups and applications so that the same cross-validation code can be employed throughout.

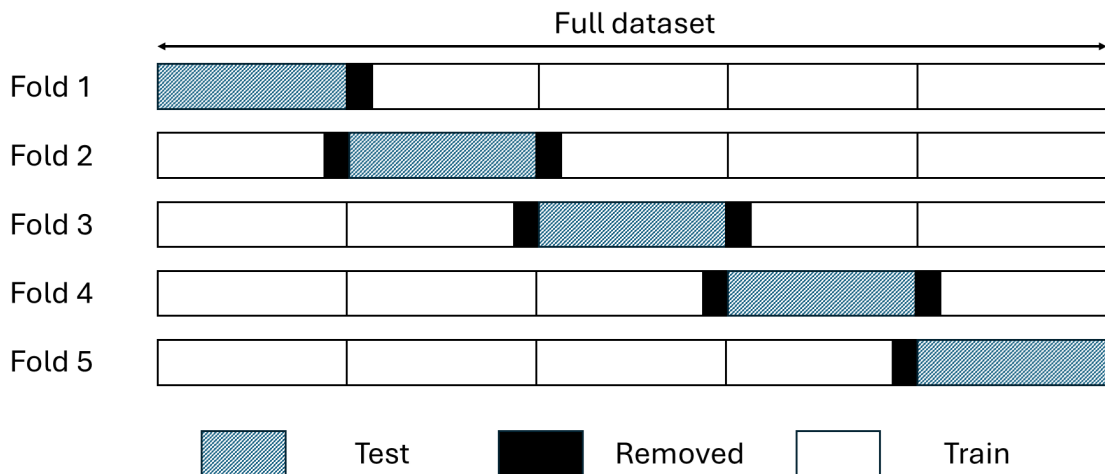


Figure 1: hv-Cross Validation

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 guard 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 will utilise grid-search to obtain the optimal hyperparameter values.

3 Monte Carlo

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), GNCQR, and the Fused LASSO (FLQR).¹⁴

Our Monte Carlo setup takes as its starting point the design used in Bondell et al. (2010). Each Monte Carlo experiment is generated using the location scale heteroskedastic error model of the form:

$$y_t = \beta_0 + \beta^T x_t + (\theta_0 + \vartheta_t \otimes \theta^T x_t) \varepsilon_t, \quad x_{t,k} \sim U(0, 1), \quad \varepsilon_t \sim N(0, 1) \quad (9)$$

An intercept is included in each setup (i.e., $\beta_0 = \theta_0 = 1$). Note the presence of $\vartheta_k \in \{0, 1\}$ 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 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, $\vartheta = \mathbf{1}_k$.

There are four 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 an amendment of 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 $\vartheta = (\mathbf{1}_4^T, \mathbf{1}_4^T \times [I(\varepsilon_t > F_\varepsilon^{-1}(0.9)) + I(\varepsilon_t \leq F_\varepsilon^{-1}(0.1))], \mathbf{0}_2^T)^T$.

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 ‘not ideal’ situations.

To test 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 macroeconomic applications

¹⁴Note 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.

it is not uncommon to apply quantile regression for such small samples.¹⁵ Applying quantile regression to analyse GDP distribution for European economies has been particularly daunting on account of data availability, so we feel that these Monte Carlo results are particularly 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: $\Delta_\tau \in \{0.2, 0.1\}$.¹⁶ 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, which is a claim we can verify by examining the variable selection properties of BRW as the number of quantiles increases.

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

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

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

where *iter* indexes a given Monte Carlo experiment, $n = 500$ is the total number of evaluated Monte Carlo experiments for each DGP, \hat{g}_τ is the estimated quantile and g_τ is the true quantile given equation (9). 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

¹⁵See 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.

¹⁶for $\Delta_\tau = 0.2$, the first quantile is set to 0.1.

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 is an upper limit of quantile variation for GNCQR.

3.2 Results

Table 1 presents the results for the average fit of the different estimators based on the Monte Carlo experiments, while Table 2 shows the variable selection performance. The RIMSE 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).

Results on fit in Table 1 reveal that GNCQR and FLQR provide best fits, even beating BRW estimator, for y_1 , y_2 , and y_3 . However, for y_4 FLQR fails to yield improvements over BRW. While GNCQR also has difficulties in y_4 , it is much closer to the fits of BRW, with both yielding better fits than the traditional QR or FLQR. This is because GNCQR recovers BRW when $\alpha = 1$, so it will 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 accumulates, 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 gives further improvements in fit over BRW, or (at worst) does as well as BRW.

The results on variable selection presented in Table 2 are more revealing. Since the GNCQR can recover both the BRW (when setting $\alpha = 1$) and the QR (as $\alpha \rightarrow 0$), 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 ‘goldilocks’ option.

Coupled with the fit results of Table (1), we find that GNCQR not only provides robustness over FLQR for y_4 , but is also able to give 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.

Unlike the fits, increasing the number of quantiles has a marked impact on variable selection: it lowers TPR and increases TNR for all estimators. 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 tendency 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.

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.

Table 2: True Positive and True Negative Rates for the different Monte Carlo experiments

	y_1		y_2		y_3		y_4	
	TPR	TNR	TPR	TNR	TPR	TNR	TPR	TNR
$\Delta\tau = 0.2$								
T=50								
BRW	0.806	-	0.360	0.639	0.535	0.525	0.305	0.691
GNCQR	0.372	-	0.232	0.769	0.345	0.715	0.243	0.753
QR	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
T=100								
BRW	0.938	-	0.510	0.501	0.694	0.393	0.390	0.582
GNCQR	0.364	-	0.228	0.780	0.430	0.689	0.279	0.715
QR	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
T=200								
BRW	0.988	-	0.659	0.340	0.846	0.262	0.481	0.459
GNCQR	0.381	-	0.225	0.787	0.564	0.672	0.280	0.700
QR	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
$\Delta\tau = 0.1$								
T=50								
BRW	0.618	-	0.246	0.752	0.383	0.669	0.224	0.776
GNCQR	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
FLQR	0.127	-	0.107	0.901	0.154	0.870	0.114	0.888
T=100								
BRW	0.792	-	0.363	0.641	0.531	0.557	0.318	0.687
GNCQR	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
FLQR	0.110	-	0.074	0.927	0.196	0.860	0.078	0.932
T=200								
BRW	0.923	-	0.488	0.517	0.690	0.445	0.409	0.575
GNCQR	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
FLQR	0.139	-	0.061	0.946	0.298	0.832	0.098	0.917

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 to 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 fit and variable selection results together, we can conclude that GNCQR provides good fit while retaining better variable selection properties than FLQR.

4 Empirical Application

4.1 Cross-section data: Heat-or-eat

To showcase the estimators in a cross-section setting, we use data from English regions to examine the ‘heat-or-eat’ dilemma. Cullen et al. (2004) was the first to raise concerns regarding heat-or-eat. The authors find little evidence in US data of excess sensitivity when the shock to fuel prices is anticipated. However, they find that households without substantial financial assets react to unanticipated fuel price changes by adjusting their consumption. This motivated Frank et al. (2006) to study how Low-Income Home Energy Assistance Program impacted households’ consumption decisions; they show that children living in households that received support for energy efficiency were at a lower risk of malnutrition. Similarly, Bhattacharya et al. (2003) find that low-income households were more likely to reduce their caloric intake during cold weather shocks.

For the UK, there are two prominent studies in the area: Beatty et al. (2014) and Burlinson et al. (2022). Similar to Bhattacharya et al. (2003) and Cullen et al. (2004), Beatty et al. (2014) finds that British households reduced their food spending in an effort to keep their homes warm. Burlinson et al. (2022) finds that British individuals with pre-payment meters (hereinafter PPM) are more likely to consume less fruit and vegetables per week, thereby increasing the risk of cancer and cardiovascular problems (Boeing et al., 2012; Murray et al., 2013).

In this section we follow Burlinson et al. (2022) in analysing data on the Portion of Fruit and Vegetable consumed per week (hereinafter PFNV) over January 2019 to May 2021 from Wave 11 of the *Understanding Society* survey (University of Essex, 2023). In contrast to previous studies, we examine the impacts of PPM on fruit and vegetable consumption through a quantile lens. Specifically, we consider the response to the survey question “how many days in a usual week do you eat fruit” and multiply it with the response to the question “on the days one eats fruit, how many portions”. The same is done for vegetables, and the results are summed to yield the individuals’ consumption of fruit and vegetable per week. Note that the response to the frequency of fruit (or vegetable) consumption is one of the following categories: Never, 1-3 days, 4-6 days, every day. Following Burlinson et al. (2022) we use the mid-points (i.e., 0, 2, 5, 7) as the response for each category. This can lead to ‘bunching’ data around the midpoints and a non-smooth density. To alleviate this, we deviate from Burlinson et al. (2022) and focus

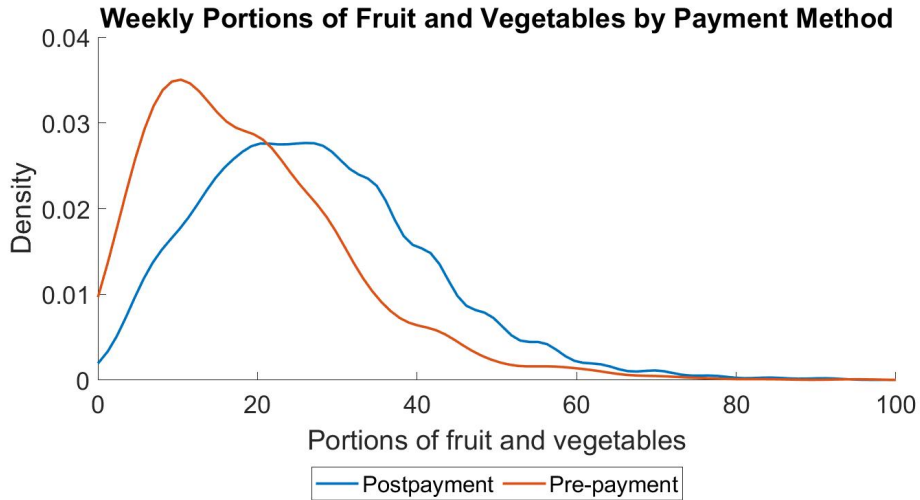


Figure 2: PFNV by payment method

on households, looking at the average PFNV in a given household.

As in Burlinson et al. (2022), our focus lies on the impact of PPM. We follow the authors in the definition for PPM but adjust it to factor in inference on households, i.e., PPM is 1 if any of the individuals in the household pay for their gas or electricity using prepayment method, and 0 otherwise.

We also vary from Burlinson et al. (2022) in that we examine English regions only. Furthermore, we estimate separate models for each English region individually. This has two advantages: (1) it allows the estimated quantile coefficients to be region specific; and (2) it allows the hyperparameter selected to be region-specific. The first point is important because we do not assume that the coefficients of PPM are shared across regions. Furthermore, running region-specific regressions ensures that any nonlinearities that stem from regional differences do not impact our estimates. The cost of running region specific regressions is that the sample size is smaller. The second point is also important, as it is difficult to ascertain ex ante whether the values of the hyperparameter are shared or not.

The summary statistics of the key variables, PFNV and PPM, for the different English regions are presented in Table 3 along with the optimal hyperparameters for FLQR and GNCQR. We can see that there is sizeable variation for optimal hyperparameter choice especially for FLQR. In contrast, the optimal hyperparameter for GNCQR hovers around 1 (i.e., the BRW value) for most of the regions, except for East Midlands and Yorkshire and the Humber, where the need for more fused shrinkage is identified. The fact that most hyperparameters of GNCQR

Table 3: Summary statistics and hyperparameters for the different English regions

Region	Obs.	$PFNV_{mean}$	$PFNV_{SD}$	PPM_{mean}	PPM_{SD}	α_{GNCQR}	α_{FLQR}	β_{OLS}^{PPM}
North East	581	24.212	13.717	0.133	0.339	1.630	8.111	-5.583
North West	1549	24.715	14.368	0.122	0.327	1.520	53.367	-6.290
Yorkshire & H.	1293	24.866	14.790	0.111	0.314	1000.000	24.771	-4.247
East Midlands	1141	26.118	14.536	0.085	0.279	16.298	657.933	-3.616
West Midlands	1244	25.498	13.846	0.108	0.310	1.072	10.723	-4.332
East of England	1379	26.731	13.615	0.081	0.273	1.000	11.498	-3.726
London	1687	27.184	14.879	0.166	0.372	0.890	24.771	-5.717
South East	1892	28.449	14.446	0.075	0.264	1.233	10.000	-7.291
South West	1328	27.749	13.877	0.084	0.277	0.690	305.386	-5.605

hover around the BRW value demonstrates the usefulness of the BRW estimator.

In table (3) we also show the OLS results for the PPM coefficient. Note that the impacts of prepayment on PFNV is larger than in Burlinson et al. (2022) (who find a coefficient of -2.78), but this is simply on account of our analysis focusing on household PFNV as opposed to individuals. The table reveals that, just like with the optimal hyperparameter, there is a large degree of variation across the different English regions with the impacts of PPM ranging from -3.62 to -7.29. This underscores the regional heterogeneity across not just the UK, but in England as well, which necessitates modelling strategies that allow for such difference.

The PFNV by PPM for England is presented in Figure 2. We can see that beyond an impact on the mean, the impact of payment method on the density is sizeable. This can raise concerns about the impact on the mean: Does PPM have an influence on the location, or is the mean effect influenced by the inherent skewness of the distributions? Nevertheless, the shape of the PFNV distribution can simply be on account of confounding factors and as such we include additional socio-economic controls.¹⁷

We focus on 10 estimated quantiles, i.e., $\Delta_\tau = 0.1$. To obtain the optimal α parameters, we create a grid of 100 equally spaced points between $[0, 1)$. To this list of candidate solutions we append a list of 100 points between $[0, 3]$ as choices of exponent of base 10. Compared to the grid in the Monte Carlo section, this grid is smaller but has the same number of candidate points between 0 and 1000.

¹⁷Namely, we include the following additional variables in our empirical model: the log of gross household income (\loginc) and its square (\logincsq), household size ($HHsize$), the number of children in the household ($nkids$), whether there is someone in the household with a university degree ($HHeduc$), whether there is an unemployed individual in the household ($HHUnemp$), whether there is a person with disability living in the household ($HHDisab$), and whether there is a retired person in the household ($HHRetire$). We also include a rural identifier ($Rural$) for all regions except London.

The quantile profiles of the PPM variable for the different estimators is shown in Figure 3. The first thing to note is that there is evidence for quantile variation in PPM for the majority of the regions. In particular, GNCQR identifies no quantile variation only in Yorkshire and the Humber and the East Midlands, while FLQR identifies no quantile variation in East Midlands and the South West. Of regions that do portray quantile variation, we see a clear downwards sloping quantile profile. Given that a household cannot consume less than 0 portions of fruit and vegetable a week, it is not surprising that there is a downwards sloping quantile profile, simply driven by the fact that the median can ‘move’ more to the left than the lower tails of the distribution. An interesting insight is that this decreasing tendency in the quantile profile continues into the right tail for a lot of the regions. This highlights that even households who consume enough fruit and vegetables, i.e. households who likely know of the positive health impacts of consuming fresh fruits and vegetables, are impacted by the means with which they pay for energy. Given that the PPM variable identifies households most at risk of having to choose between heating and eating, these results highlight the nonlinear impact of the ‘heat-or-eat’ dilemma.

Looking at the PPM coefficients for London and the North West, we see that FLQR has a tendency of shrinking away quantile variation at the tails. This is particularly problematic as quantile regression is often used to reveal how covariates drive the tails of a particular distribution. If the estimator of choice is prone to shrink away quantile variation at the tails, we risk missing key drivers of nonlinearity.

Next, we consider the findings for London in more detail. London is particularly interesting because it is the region with the highest proportion of PPM. Figure 4 shows the in-sample and out-of-sample average quantile score for different α values and compares the profiles with the Fused LASSO estimator. The figures show that the equivalent hyperparameters of the different estimators do not lead to the same values, leading to different profiles. Furthermore, the figures reveal that GNCQR shows a more gradual change in in-sample fit. For London, the out-of-sample fits of GNCQR have a more pronounced minimum. Another feature of GNCQR is that its in-sample and out-of-sample profiles are less ‘jagged’ than that of FLQR. These indicate that GNCQR shrinks quantile variation of the correct variable. These tendencies are not unique to London, and are present in all the regions investigated.¹⁸

We can see from Figure 4 that there is a clear bias-variance trade-off given that in-sample performance gets worse as we introduce bias, but out-of-sample performance (initially) improves

¹⁸The figures for the other regions are presented in the appendix.

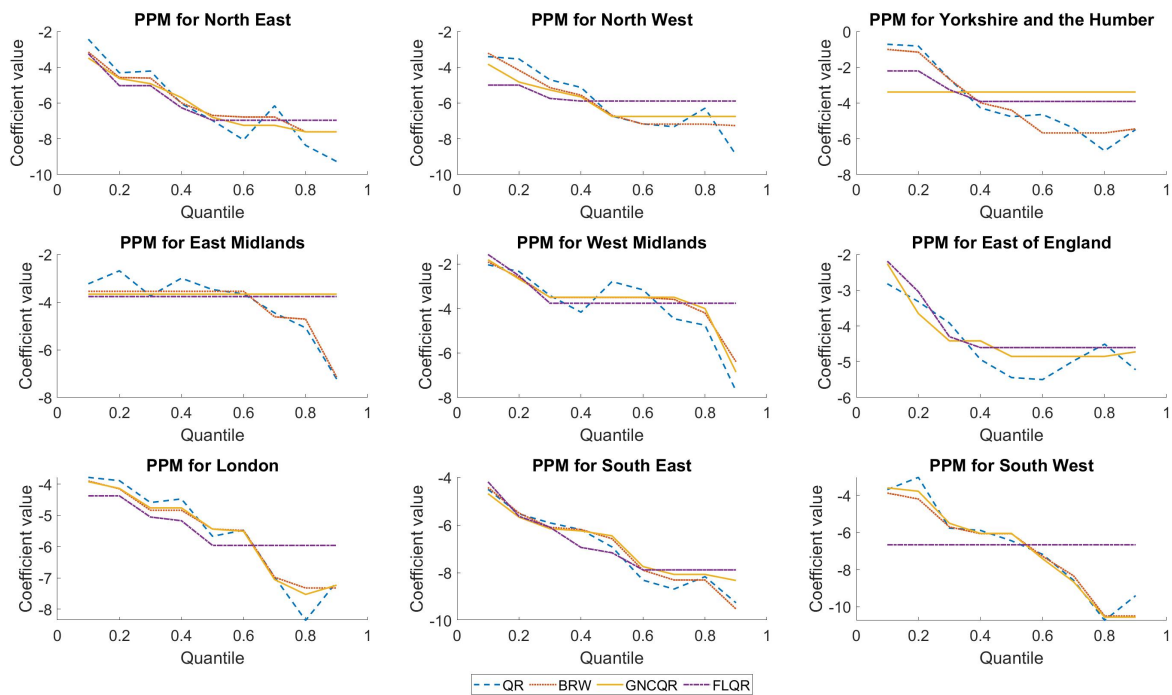


Figure 3: Coefficient profile of PPM across the different quantiles

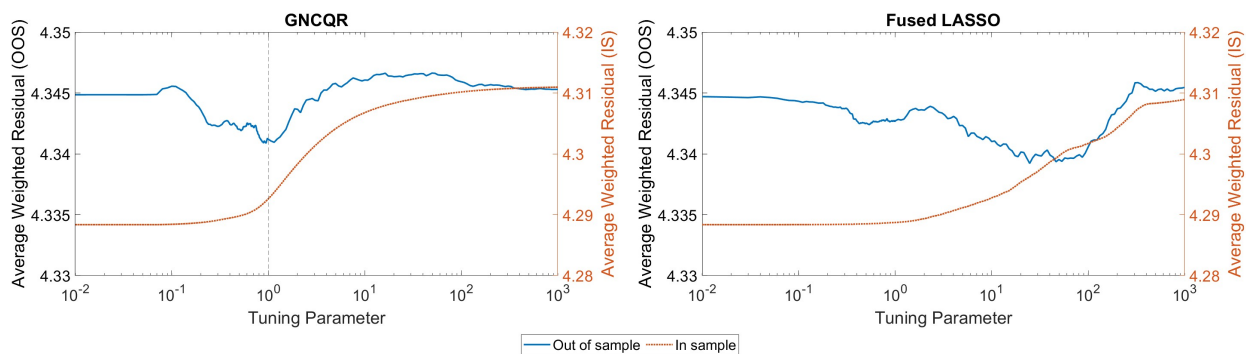


Figure 4: In-sample and Out-of-sample fits for GNCQR and Fused LASSO for London

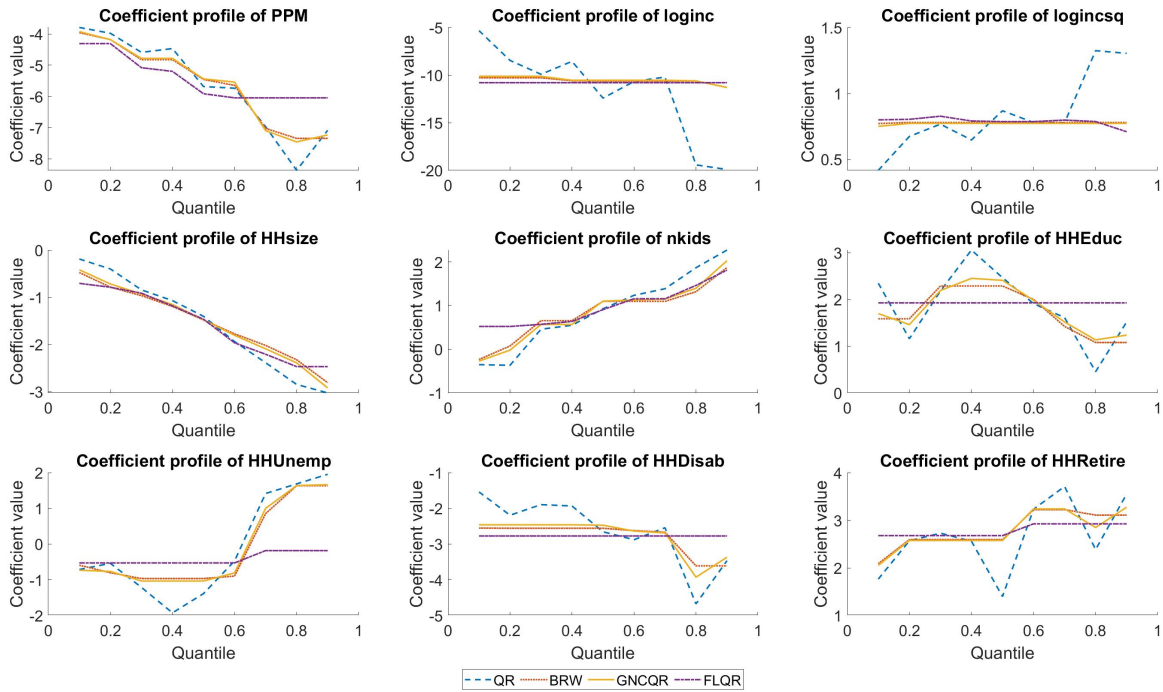


Figure 5: Coefficient profiles for all covariates for London

through the reduction of variance. The smoother profile of GNCQR is likely because the bias we introduce is by moving from the quantile property of QR to the ‘quantile subset’ property of BRW. Once we fully satisfy the quantile subset property, increasing α is equivalent to penalising quantile closeness on top of quantile crossing.

Figure 5 shows the coefficient profiles of all the covariates for London. Note first that FLQR is very judicious in shrinking away quantile variation: it shrinks most quantile variation in 6 of the 9 variables. Of the remaining variables it shrinks away variation at the tails in 2 variables. This corroborates the results of Table 2 that FLQR overshrinks. The CV method identified a hyperparameter value of 0.890 for GNCQR which is very close to 1 which would yield BRW. As such, it is not surprising that GNCQR and BRW quantile profiles are quite close. Nevertheless, there are differences and due to α being less than 1, GNCQR portrays more quantile variation than BRW.

Based on the results, we can conclude that the households that have prepayment meters are more likely to consume less portions of fruit and vegetables per week. In particular we see that this impact is present across the distribution, with the right tail of the distribution seeing a larger drop in fruit and vegetable consumption. This suggests that households with prepayment meters reduce their fruit and vegetable consumption as much as they can to ensure

that they can heat their homes.

To tackle this two policies are proposed: 1) variable rates, and; 2) nutritional assistance programs. Variable rates can help tackle the “heat-or-eat” dilemma by charging a lower rate for the first few cubic metres of gas for PPM customer during the winter months. In this way, households that have prepayment meters are not forced to cut back as much on their fruit and vegetable consumption during the winter months. The second policy of nutritional assistance programs could directly tackle the adverse effects of PPMs on food consumption by ensuring access to healthy foods for these households during the winter months. Expanding programs such as free school lunches and food vouchers can help mitigate some of the impact of the “heat-or-eat” dilemma by providing a vital safety net.

4.2 Time Series data: Growth-at-Risk

As a time series application we analyse Growth-at-Risk (GaR) (Adrian et al., 2019) for the US. The key idea is to think about GDP growth through the lens of value-at-risk. The canonical GaR uses US quarterly GDP growth in conjunction with the NFCI to obtain estimates of downside risk. The need for GaR was highlighted by the global financial crisis which showed how downside risks, or lower quantiles of the density of GDP growth, evolve with the state of financial market.

GaR has become a popular tool for policy makers and researchers alike. Figueres and Jarociński (2020) and Szendrei and Varga (2023) 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 covariates. They find that accounting for multiple sources of risk leads to better density fits of GDP growth. Iseringhausen et al. (2023) use the same dataset but take a different approach and instead look at 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. As such, testing the performance of the GNCQR on GaR fits nicely in the literature.

Using quantile regression, GaR estimates can be obtained by estimating the following model:

$$y_{t+h} = x_t' \beta_p + \epsilon_{t+h} \quad (10)$$

for $t = 1, \dots, T - h$, 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. The quarterly data cover 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 ($h = 1, 4$). Using the quantile setup, forecasts from each chosen quantile are denoted as $y_{T+h|T}^p$. Forecasts are computed on an expanding window basis where the initial in-sample period uses the first 50 observations of the sample, which makes for $(150 - h)$ forecast windows. We estimate a grid of 19 equidistant quantiles.¹⁹

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 and as such we will run the model on two sample periods: Pre-Covid and Full sample.

Figures 6 and 7 show the in-sample and out-of-sample average quantile score for different α values and compares the profiles with the Fused LASSO estimator. First, the figures show that the equivalent hyperparameters of the different estimators do not lead to the same values, leading to different profiles. In particular, GNCQR shows a more gradual change in in-sample fit. Furthermore, the out-of-sample fits of GNCQR have a more pronounced minimum. Another feature of GNCQR is that its in-sample and out-of-sample profiles are less ‘jagged’ than that of FLQR. This is true for all forecast horizons, as well as pre- and post-COVID periods. These indicate that GNCQR shrinks quantile variation of the correct variable.

The quantile profiles of the estimated β coefficients are presented in Figures 8 and 9, for pre-COVID and full samples. 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. Almost all estimators showcase a general increasing tendency 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. This tendency sees FLQR have 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 is more likely to give a zero coefficient. Given the argument of Adrian et al. (2019), a zero coefficient for the upper quantiles of NFCI makes more sense 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, looking at the NFCI coefficients quantile profile, GNCQR only has negative slope for the pre-covid sample

¹⁹Compared to the cross-section application, here we estimate more quantiles, namely we will set $\Delta_\tau = 0.05$.

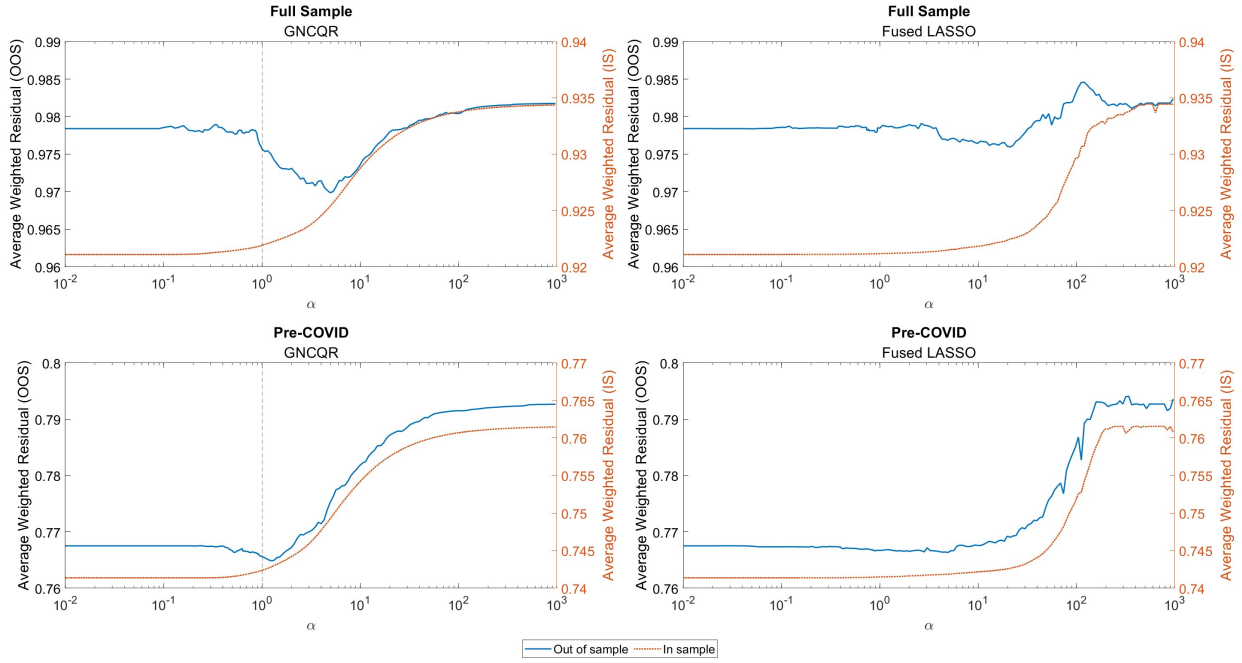


Figure 6: In-sample and Out-of-sample fits for GNCQR and Fused LASSO at $h = 1$

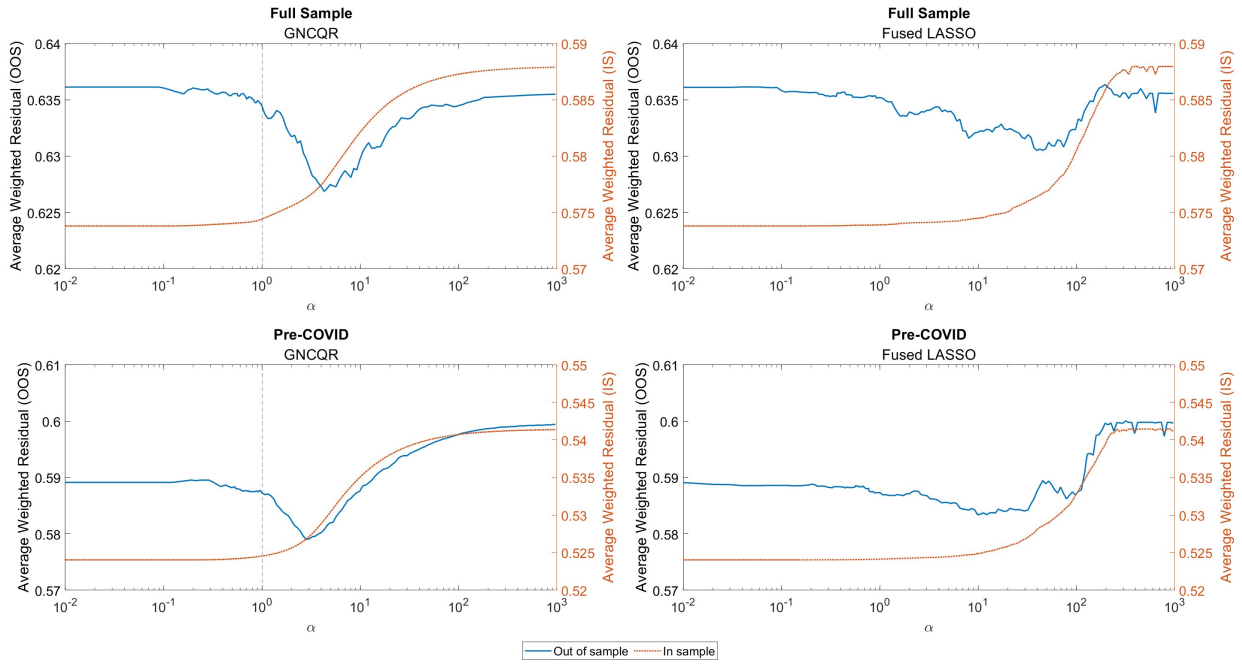


Figure 7: In-sample and Out-of-sample fits for GNCQR and Fused LASSO at $h = 4$

for the $h = 1$ forecast horizon. For all the other cases, GNCQR portrays a gentle upwards 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.

The coefficient profile figures demonstrate that GNCQR is capable of identifying quantile variation better than simple FLQR. Furthermore, for the variables that have quantile variation, GNCQR shrinks away the “jagged” edges resulting in smoother quantile profiles.

To check the forecast performance of the different estimators, the quantile weighted CRPS (qwCRPS) of Gneiting and Ranjan (2011) is chosen as a scoring rule. To calculate this measure, we first take the Quantile Score (QS), which is the weighted residual for a given forecast observation, $\hat{y}_{t+h,p}$. Using the QS, the qwCRPS is calculated as:

$$qwCRPS_{t+h} = \int_0^1 w_q QS_{t+h,q} dq, \quad (11)$$

where w_q denotes a weighting scheme to evaluate specific parts of the forecast density. We choose this measure as the scoring rule because through different weighting schemes we can evaluate differences at different parts of the distribution. We consider 3 different weighting schemes: $w_q^1 = \frac{1}{Q}$ places equal weight on all quantiles (denoted as CRPS), which is equivalent to taking the average of the weighted residuals; $w_q^2 = q(1 - q)$ places more weight on central quantiles; and $w_q^3 = (1 - q)^2$ places more weight on the left tail. The results using the different weighting schemes are presented in Table (4).

Note that in a forecasting setting, having $\alpha = 1$ only ensures non-crossing quantiles in-sample. Nevertheless, for GNCQR, when $\alpha > 1$, we start imposing non-crossing quantiles out-of-sample as well. As such for time-series settings, when $\alpha > 1$, we are more likely to have non-crossing forecasted quantiles. We also sort the fitted quantiles following Chernozhukov et al. (2009) to see whether GNCQR yields improvements beyond simple sorting.

Looking at the unsorted results of Table 4, we can see that GNCQR yields the lowest weighted residual for all forecast horizons, for all weighting schemes. Given the evidence from the previous section, we can say that performance of GNCQR is largely driven by its ability

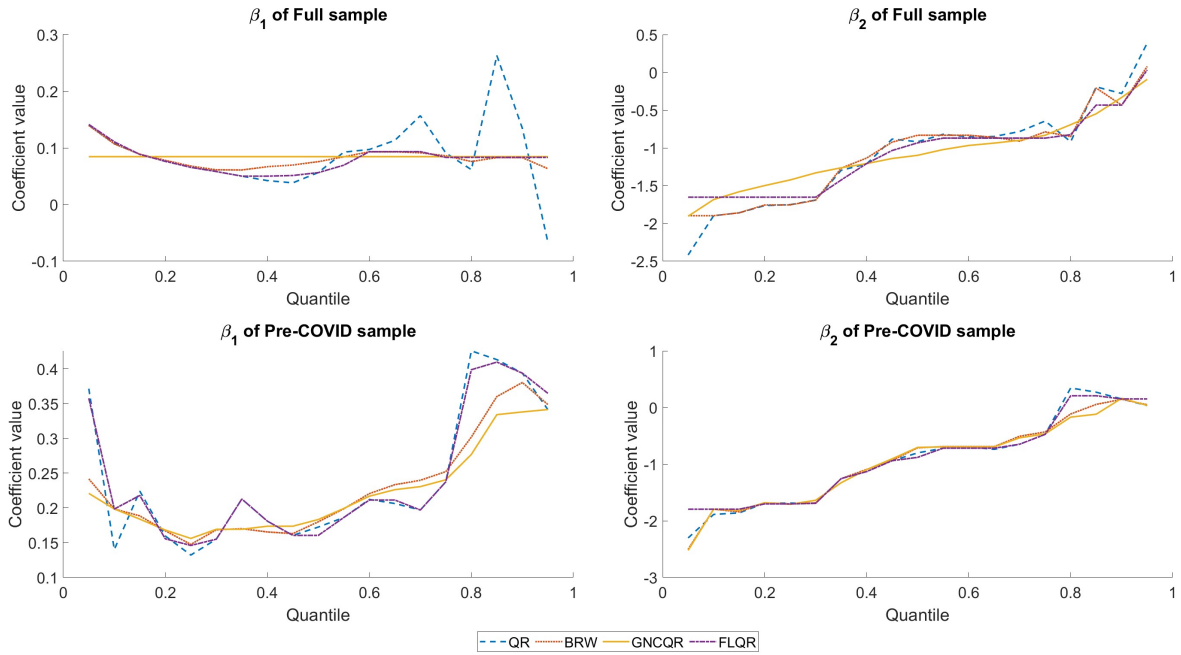


Figure 8: β coefficients for the different estimators at $h=1$

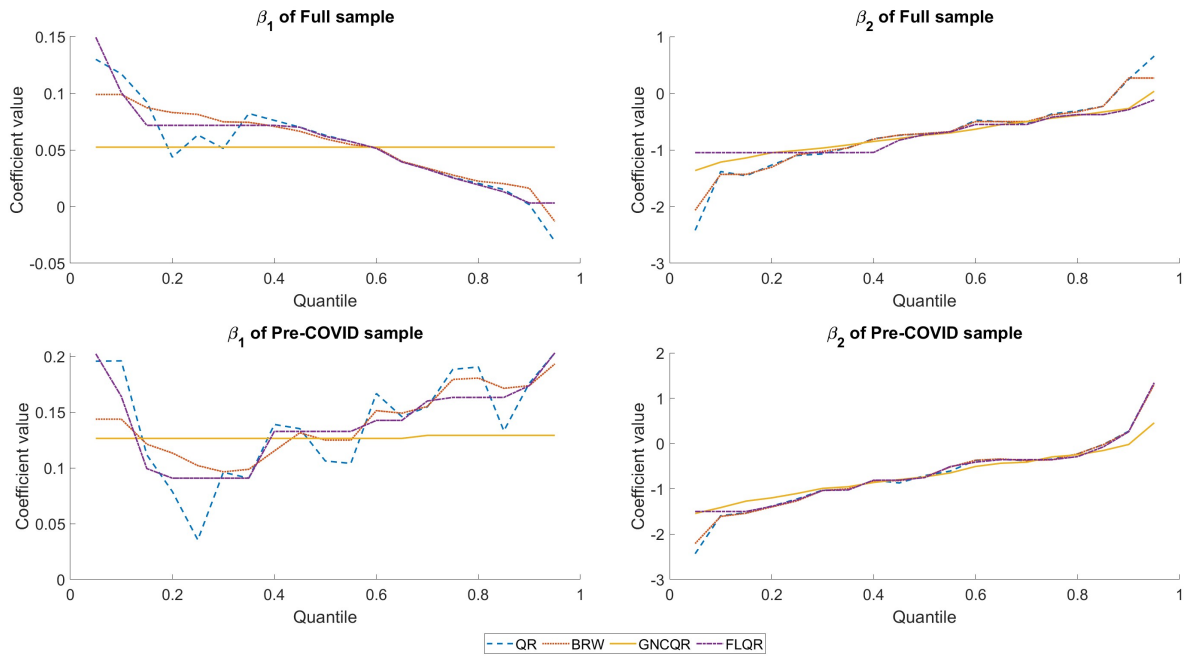


Figure 9: β coefficients for the different estimators at $h=4$

Table 4: Forecast results of the different estimators

	Unsorted			Sorted			
	CRPS	Centre	Left	CRPS	Centre	Left	Logscore
h=1							
<i>QR</i>	0.916	0.173	0.280	0.909	0.172	0.277	-3.854
<i>BRW</i>	0.914	0.172	0.279	0.911	0.173	0.278	-3.904
<i>GNCQR</i>	0.910	0.172	0.274	0.909	0.172	0.274	-3.831
<i>FLQR</i>	0.918	0.173	0.279	0.914	0.173	0.277	-3.919
h=4							
<i>QR</i>	0.559	0.108	0.168	0.554	0.107	0.167	-3.054
<i>BRW</i>	0.554	0.107	0.166	0.553	0.107	0.166	-3.050
<i>GNCQR</i>	0.552	0.107	0.165	0.552	0.107	0.165	-2.709
<i>FLQR</i>	0.560	0.108	0.167	0.557	0.108	0.166	-3.094

to distinguish between quantile varying variables and shrink away any unnecessary quantile variation. In particular, GNCQR does this better than FLQR, which performs the worst on many of the forecast horizons.

The forecast result table also shows how 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, at the longer forecast horizon there is no difference between the sorted and unsorted performance of GNCQR which indicates that the method yields proper densities out of sample as well.

The GNCQR applied to the GaR of US data verified the crucial finding of Adrian et al. (2019), namely that macro-financial linkages are of important for the downside of GDP growth. Notably, the GNCQR completely shrinks the quantile variation in the GDP coefficient, which signifies that, while past values of GDP are important in determining the location of the GDP distribution, most variation in the shape of the distribution is on account of changes in the financial conditions. Furthermore, the method provides gains in a forecasting exercise. This highlights how even in small dimensional settings, one can have better forecast performance through correctly identifying which variables are quantile varying.

5 Conclusion

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

Through a comprehensive 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 fits that are either better or nearly as good as those of the BRW estimator (Bondell et al., 2010). When looking at the variable selection properties of the different estimators we found that traditional Fused LASSO (FLQR) estimator overshrinks while BRW estimator undershrinks compared to GNCQR.

The trade-off between in-sample and out-of sample fits of GNCQR and FLQR is clearly shown for both the cross-section and the time series applications. But there is a difference between the estimators in how the deterioration of in-sample fit occurs as the value of the hyperparameter increases: the in-sample fit profile of GNCQR deteriorates more gradually as the hyperparameter value is increased. We provide an intuitive explanation for why this is: as we increase α , we introduce bias in the quantile property and gradually start to introduce the quantile subset property.

Looking at coefficients of a ‘heat or eat’ dilemma on English cross-section data, we demonstrate an advantage of GNCQR over other shrinkage estimators: the estimator only imposes additional shrinkage if it yields improvements above BRW. Except for 2 regions of England, the optimal hyperparameter of the GNCQR is around $\alpha = 1$, i.e., roughly equivalent to BRW. This highlights the power of the BRW estimator, and why it is a good estimator to use if one is not interested in optimising the degree of shrinkage in the quantile profile.

In the US GaR application using time series data we can see a clear advantage of GNCQR: the coefficients’ quantile profile is more likely to yield no quantile variation. FLQR, on the other hand, is more likely to shrink quantile variation at the lower tail of the NFCI coefficient in the canonical GaR.

In sum, the non-crossing constraints of Equation (5) bestow upon the quantile estimator

additional attractive properties: (1) it can distinguish variables with quantile variation from those without, and shrinks the correct variable; and (2) it renders the estimated quantile profiles less ‘jagged’, i.e., it removes sudden reversions in the difference in the β coefficient.

The findings of this paper reach beyond the estimator proposed here. In particular, because of the equivalence between non-crossing and fused-shrinkage, one can extend the methodology to obtain non-crossing Bayesian quantile regression. To our knowledge, currently the most popular way to obtain non-crossing Bayesian quantiles involve post-processing methods such as Rodrigues and Fan (2017), and our proposal provides clear advantages in comparison.

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 making the hyperparameter have more than just one dimension: making the hyperparameter parameter specific has been shown to offer better performance in other contexts (Zou, 2006). Naturally, doing so will require a better hyperparameter tuning procedure, since grid-search is a computation-intensive choice even with just 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 like in Kohns and Szendrei (2021) and Szendrei and Varga (2023), while being principled about the degree of non-crossing one needs to impose. We leave this work for future research.

A promising direction for research is exploring how the non-crossing equivalence to fused shrinkage extends beyond quantile regression. Of particular interest is investigating non-crossing expectile regressions. Understanding whether non-crossing expectiles can be achieved through fused shrinkage could lead to new insights and methodologies.

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 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.
- 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.
- Beatty, T. K., L. Blow, and T. F. Crossley (2014). Is there a ‘heat-or-eat’ trade-off in the uk? *Journal of the Royal Statistical Society Series A: Statistics in Society* 177(1), 281–294.
- Bergstra, J. and Y. Bengio (2012). Random search for hyper-parameter optimization. *Journal of machine learning research* 13(2).
- Bhattacharya, J., T. DeLeire, S. Haider, and J. Currie (2003). Heat or eat? cold-weather shocks and nutrition in poor american families. *American Journal of Public Health* 93(7), 1149–1154.
- Boeing, H., A. Bechthold, A. Bub, S. Ellinger, D. Haller, A. Kroke, E. Leschik-Bonnet, M. J. Müller, H. Oberritter, M. Schulze, et al. (2012). Critical review: vegetables and fruit in the prevention of chronic diseases. *European journal of nutrition* 51, 637–663.
- Bondell, H. D., B. J. Reich, and H. Wang (2010). Noncrossing quantile regression curve estimation. *Biometrika* 97(4), 825–838.
- Burlinson, A., A. Davillas, and C. Law (2022). Pay (for it) as you go: Prepaid energy meters and the heat-or-eat dilemma. *Social Science & Medicine* 315, 115498.
- Cerqueira, V., L. Torgo, and I. Mozetič (2020). Evaluating time series forecasting models: An empirical study on performance estimation methods. *Machine Learning* 109, 1997–2028.
- Chernozhukov, V., I. Fernandez-Val, and A. Galichon (2009). Improving point and interval estimators of monotone functions by rearrangement. *Biometrika* 96(3), 559–575.
- Cullen, J. B., L. Friedberg, and C. Wolfram (2004). Consumption and changes in home energy costs: How prevalent is the ‘heat or eat’ decision? *University of California, Berkeley manuscript*.
- Figueres, J. M. and M. Jarociński (2020). Vulnerable growth in the euro area: Measuring the financial conditions. *Economics Letters* 191, 109126.

- Frank, D. A., N. B. Neault, A. Skalicky, J. T. Cook, J. D. Wilson, S. Levenson, A. F. Meyers, T. Heeren, D. B. Cutts, P. H. Casey, et al. (2006). Heat or eat: the low income home energy assistance program and nutritional and health risks among children less than 3 years of age. *Pediatrics* 118(5), e1293–e1302.
- 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 (2023). Aggregate skewness and the business cycle. *Review of Economics and Statistics*, 1–37.
- 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.
- 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: journal of the Econometric Society*, 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.

- Liu, Y. and Y. Wu (2009). Stepwise multiple quantile regression estimation using non-crossing constraints. *Statistics and its Interface* 2(3), 299–310.
- 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.
- Murray, C. J., M. A. Richards, J. N. Newton, K. A. Fenton, H. R. Anderson, C. Atkinson, D. Bennett, E. Bernabé, H. Blencowe, R. Bourne, et al. (2013). Uk health performance: findings of the global burden of disease study 2010. *The lancet* 381(9871), 997–1020.
- Newey, W. K. and J. L. Powell (1987). Asymmetric least squares estimation and testing. *Econometrica: Journal of the Econometric Society*, 819–847.
- Powell, D. (2020). Quantile treatment effects in the presence of covariates. *Review of Economics and Statistics* 102(5), 994–1005.
- Racine, J. (2000). Consistent cross-validated model-selection for dependent data: hv-block cross-validation. *Journal of econometrics* 99(1), 39–61.
- Rodrigues, T. and Y. Fan (2017). Regression adjustment for noncrossing bayesian quantile regression. *Journal of Computational and Graphical Statistics* 26(2), 275–284.
- Shao, J. (1997). An asymptotic theory for linear model selection. *Statistica sinica*, 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*, 595–620.
- Szendrei, T. and K. Varga (2023). Revisiting vulnerable growth in the euro area: Identifying the role of financial conditions in the distribution. *Economics Letters*, 110990.
- University of Essex (2023). Understanding Society: Waves 1-13, 2009-2022 and Harmonised BHPS: Waves 1-18, 1991-2009. 18th Edition.
- Wager, S. (2020). Cross-validation, risk estimation, and model selection: Comment on a paper by rosset and tibshirani. *Journal of the American Statistical Association* 115(529), 157–160.

- Yang, Y. (2005). Can the strengths of aic and bic be shared? a conflict between model identification 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)

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

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

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

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

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

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

$$\begin{aligned} \hat{\beta}_{JWB} &= \underset{\beta}{\operatorname{argmin}} \sum_{q=1}^Q \sum_{t=1}^T \rho_q(y_t - x_t^T \beta_{\tau_q}) \\ \text{s.t. } k^* &\geq \sum_{k=1}^K (\gamma_{k,\tau_q}^+ + \gamma_{k,\tau_q}^-) \end{aligned}$$

- Generalised Non-Crossing Quantile Regression

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

B 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 $n \rightarrow \infty$. 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 $h\nu$ -block CV of Racine (2000), we are implicitly expressing a preference in favour of model selection. If the interest is purely data fit, it might be beneficial to opt for a leave-one-out CV. We retain consideration of the GNCQR using other CV methods to future research.

Given the results of Stone (1977) and Shao (1997), one can also rely on information criteria for hyperparameter selection. The equation for AIC and BIC for quantile regression can be found 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 for 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 entails that the optimal hyperparameter choice remains valid.^{20, 21}

²⁰For 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).

²¹While our analysis assumes that the relative ranking of hyperparameters remains valid, further investigation might reveal contexts in which Bates' nested cross-validation method improves hyperparameter tuning.

C In-sample and Out-of-sample fits for the different regions

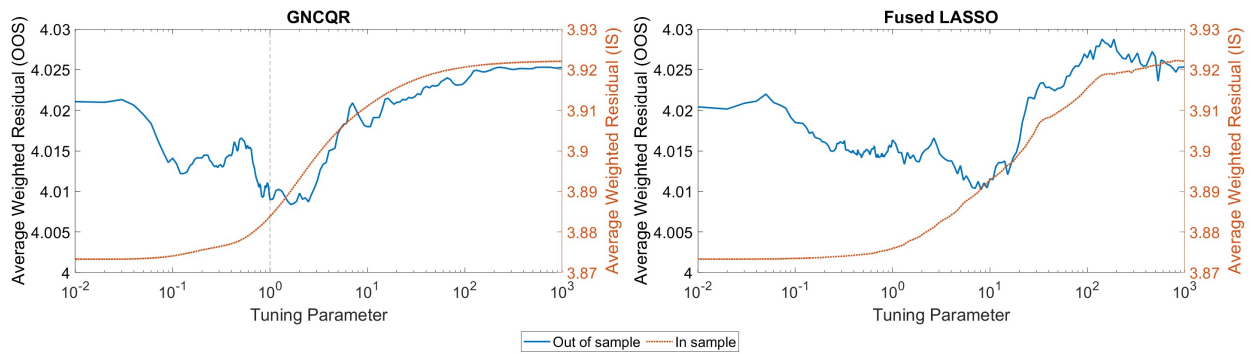


Figure 10: In-sample and Out-of-sample fits for GNCQR and Fused LASSO for North East

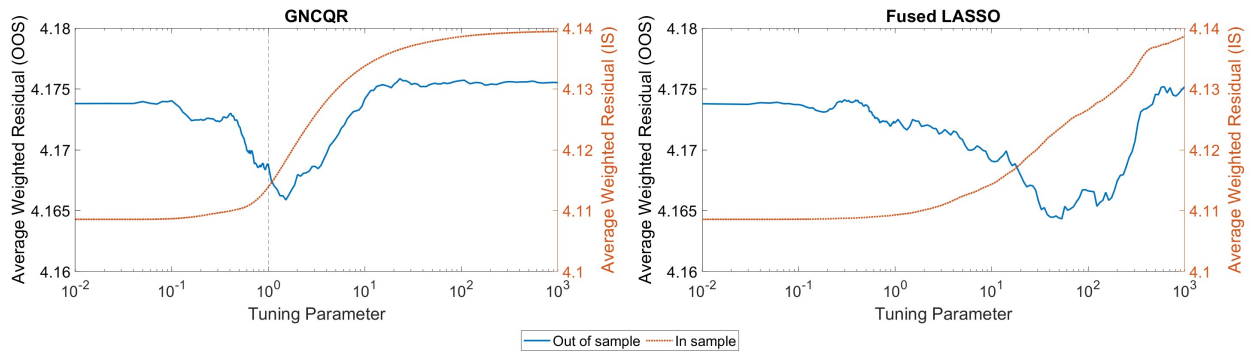


Figure 11: In-sample and Out-of-sample fits for GNCQR and Fused LASSO for North West

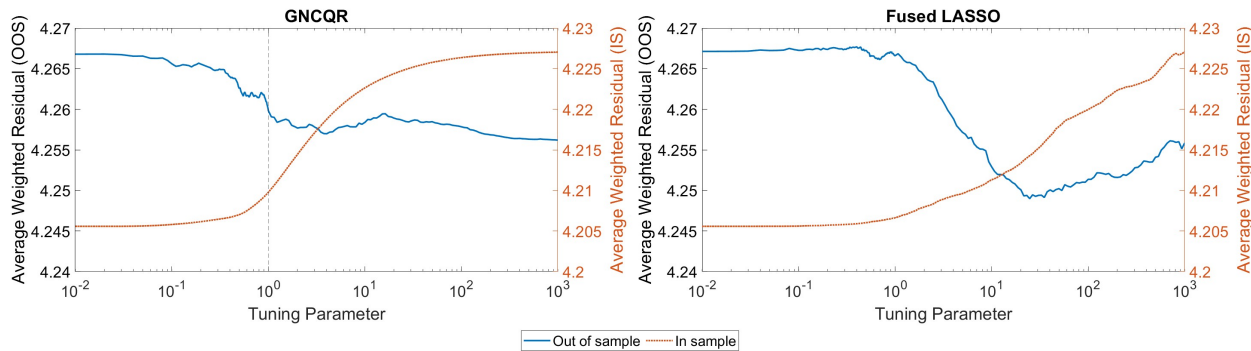


Figure 12: In-sample and Out-of-sample fits for GNCQR and Fused LASSO for Yorkshire and the Humber

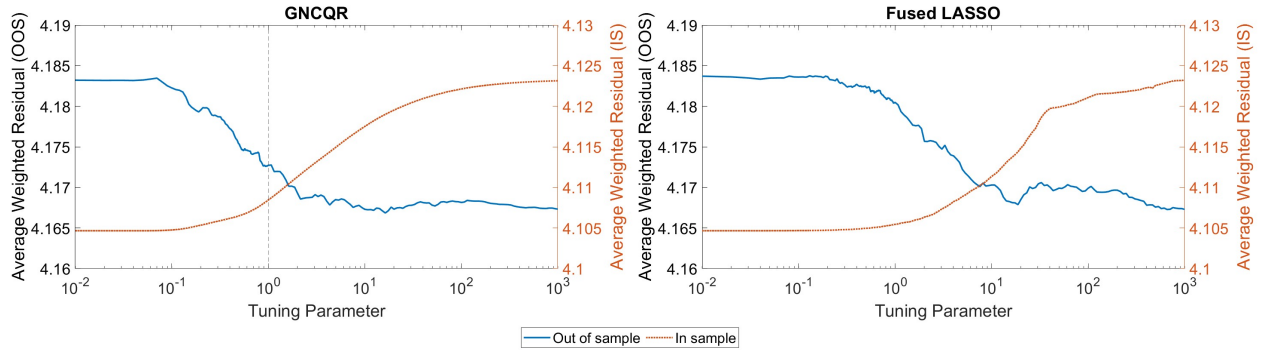


Figure 13: In-sample and Out-of-sample fits for GNCQR and Fused LASSO for East Midlands

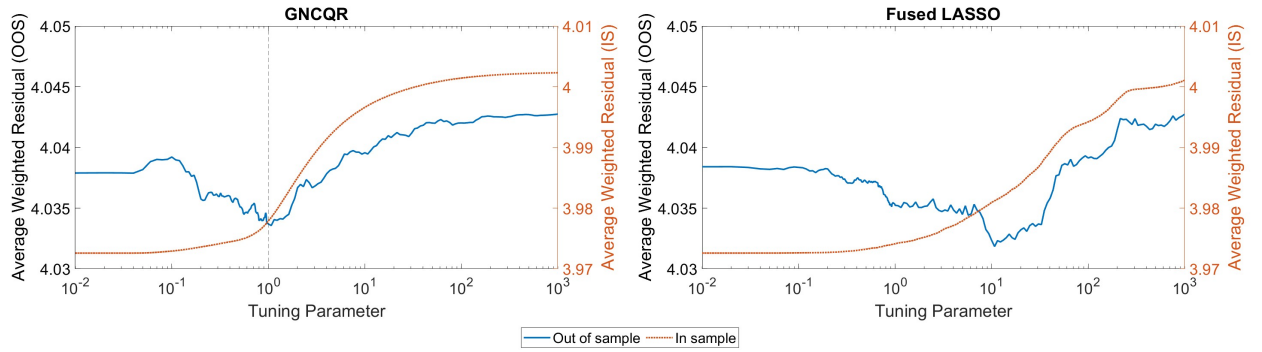


Figure 14: In-sample and Out-of-sample fits for GNCQR and Fused LASSO for West Midlands

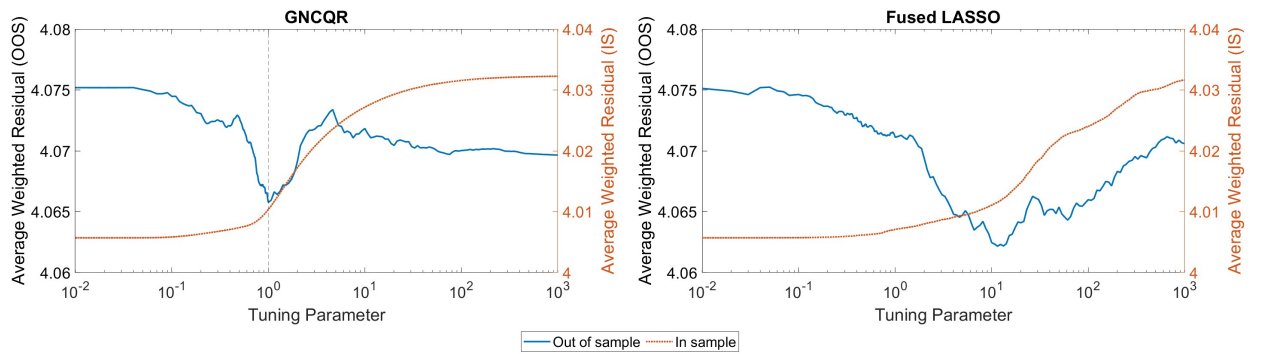


Figure 15: In-sample and Out-of-sample fits for GNCQR and Fused LASSO for East of England

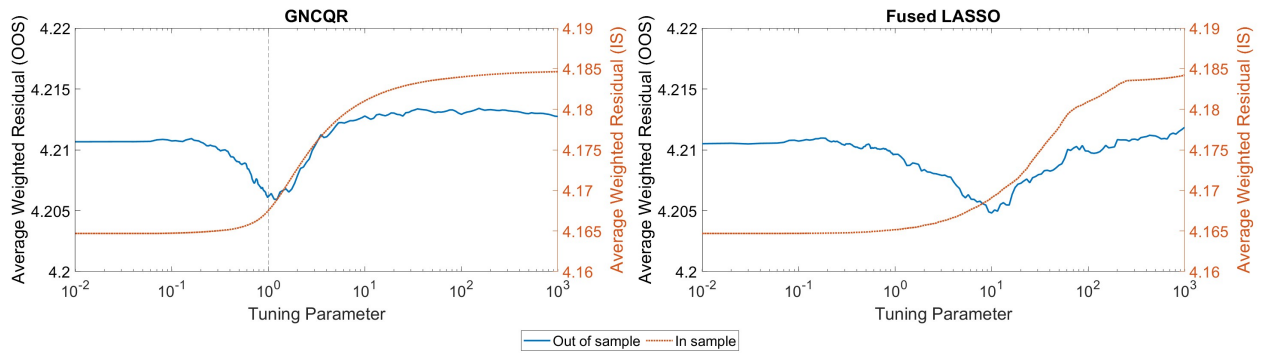


Figure 16: In-sample and Out-of-sample fits for GNCQR and Fused LASSO for South East

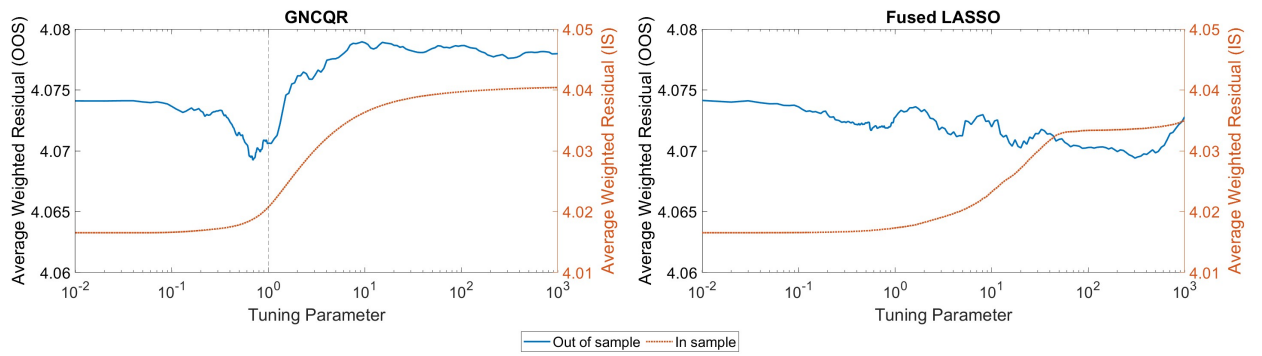


Figure 17: In-sample and Out-of-sample fits for GNCQR and Fused LASSO for South West