Factor Model of Mixtures

Cheng Peng* and Stan Uryasev†

February 1, 2023

Abstract

This paper considers the problem of estimating the distribution of a response variable conditioned on observing some factors. Existing approaches are often deficient in one of the qualities of flexibility, interpretability and tractability. We propose a model that possesses these desirable properties. The proposed model, analogous to classic mixture regression models, models the conditional quantile function as a mixture (weighted sum) of basis quantile functions, with the weight of each basis quantile function being a function of the factors. The model can approximate any bounded conditional quantile model. It has a factor model structure with a closed-form expression. The calibration problem is formulated as convex optimization, which can be viewed as conducting quantile regressions of all confidence levels simultaneously and does not suffer from quantile crossing by design. The calibration is equivalent to minimization of Continuous Probability Ranked Score (CRPS). We prove the asymptotic normality of the estimator. Additionally, based on risk quadrangle framework, we generalize the proposed approach to conditional distributions defined by Conditional Value-at-Risk (CVaR), expectile and other functions of uncertainty measures. Based on CP decomposition of tensors, we propose a dimensionality reduction method by reducing the rank of the parameter tensor and propose an alternating algorithm for estimating the parameter tensor. Our numerical experiments demonstrate the efficiency of the approach.

1 Introduction

This paper proposes a new approach to estimating the distribution of the response variable y conditioned on observing some factors $\boldsymbol{x}$ given data pairs $\{(y_i, \boldsymbol{x}_i)\}_{i=1}^N$. The quantile function of random variable y is defined by $Q_y(p) = \inf_q\{p \leq P(y \leq q)\}$. The conditional quantile function $Q_{y|\boldsymbol{x}}(p)$ is modeled by a mixture (weighted sum) of basis quantile functions, e.g. quantile functions of normal distribution and exponential distribution. The weight of each basis quantile function is a function of the factors. B-Spline (de Boor and Daniel, 1974) is used as a primary example for the weight function in this study. The model formulation is analogous to classic mixture regression model (Quandt, 1972) where the conditional density function is a mixture of basis density functions, while the parameters of each density function are functions of the factors. However, estimation of mixture regression model requires computationally expensive nonconvex optimization.

Our approach has the following key features:

- *Flexibility* – The conditional quantile function is flexible in shape to accommodate fat tails and multimodality. The weight function can capture nonlinear relations between factors and quantiles of the response variable (Section 3). The model can approximate any bounded quantile model in the limit (Section 5).
- *Interpretability* – The model can be viewed as a factor model such that the impact of factors on the shape of conditional quantile function can be traced analytically. Both the conditional quantile function and the quantile curves (hypersurfaces) have closed-form expressions (Section 3).

*Department of Applied Mathematics and Statistics, Stony Brook University, cheng.peng.1@stonybrook.edu
†Department of Applied Mathematics and Statistics, Stony Brook University, stanislav.uryasev@stonybrook.edu

- *Tractability* – The model calibration is formulated as a linear regression problem similar to quantile regression (Koenker and Bassett, 1978), which can be efficiently solved by convex and linear programming. Various constraints and penalties are readily available for variable selection and other purposes (Section 2).

While the quantile functions and weight functions are arbitrary in principle, we primarily focus on quantile functions of common distributions and B-spline (de Boor and Daniel, 1974) with nonnegative coefficients. Based on (Papp, 2011; Papp and Alizadeh, 2014), we prove that the proposed model can approximate any bounded continuous quantile model.

Quantile regression, when conducted simultaneously for multiple confidence levels, suffers from the problem of quantile crossing. The proposed calibration method can be viewed as conducting quantile regressions simultaneously with all confidence levels, but without quantile crossing by definition (Section 2). The calibration problem is equivalent to minimization of Continuous Probability Ranked Score (CRPS, Hersbach (2000)), a popular measure of quality of distributional prediction (Section 2).

In addition, we make the following contributions:

- We propose a dimensionality reduction method by reduced rank tensor and an alternating algorithm for estimation, inspired by Canonical Polyadic (CP) decomposition of tensors (Hitchcock, 1927; Harshman, 1970; Carroll and Chang, 1970) (Section 7).
- We generalize the approach to modeling other functions of uncertainty measures beyond quantile, such as Conditional Value-at-Risk (CVaR, Rockafellar and Uryasev (2000, 2002)) and expectile function Newey and Powell (1987) based on the risk quadrangle framework (Rockafellar and Uryasev, 2013) (Section 8).

Multivariate B-spline is obtained by tensor product of univariate B-splines. The CP decomposition for spline tensors has been applied in Computer Aided Design (CAD) to reduce model complexity (Pan et al., 2016). This study introduces its application to statistical estimation. In high dimensions, the tensor product B-spline has a huge number of basis functions. CP decomposition has the appealing interpretation that it reduces the number of basis by assembling them to new ones. Our approach is different from existing work on dimensionality reduction for quantile regression. Lian et al. (2019) proposes reduced matrix rank regression for homoscedastic multiple quantile modeling. Chen et al. (2021) proposes a different approach to dimensionality reduction for quantile regression, which determines different principle components for different confidence levels. Our approach leads to different factors for different confidence levels, connecting our work with Chen et al. (2021) (Section 7).

The fundamental risk quadrangle (Rockafellar and Uryasev, 2013) provides an axiomatic framework to study uncertainty measures and regressions. Besides quantile regression (Koenker and Bassett, 1978), regressions for other uncertainty measures, such as CVaR (superquantile) regression (Rockafellar and Royset, 2013; Rockafellar et al., 2014; Golodnikov et al., 2019) and expectile regression (Newey and Powell, 1987), have been developed and are studied within the risk quadrangle framework (Rockafellar and Royset, 2018; Kuzmenko, 2020). Similar to quantile regression, these regressions only provide estimation on the uncertainty measure at a given confidence level. By simply replacing the error function in optimization and basis functions in model formulation, our approach can be used to model other uncertainty measures (Section 8).

The remainder of this paper is organized as follows. Section 2 conducts a broad literature review and discuss the relations to our approach. Section 3 formulates the model and gives several illustrative examples. Section 4 formulates the convex optimization problem and discuss its relation to joint quantile regression and CRPS minimization. Section 5 proves the approximation theorem for the model. Section 6 proves the asymptotic normality of the estimator. Section 7 introduces a dimensionality reduction methods by reduced rank tensor and an alternating algorithm for estimation. Section 8 extends the method to estimate other uncertainty measure based on risk quadrangle framework. Section 9 is numerical study with real-world data.

2 Discussion on Related Work

We give a broad review of relevant work and discuss the relation with our approach. For reader's convenience, we list the mathematical formulation of the models reviewed in this section in Appendix A.

Mixture quantile models have been developed where the quantile function is a combination of some basis functions functions. Various types of basis functions have been used, such as orthogonal polynomials (Sillitto, 1969), a mixture of normal and Cauchy distributions with linear and quadratic terms (Karvanen, 2006), a modified logistic distribution (Keelin, 2016), and the quantile functions of the Generalized Beta Distribution of the Second Kind (Peng et al., 2022). However, these methods lack a factor model structure and some do not guarantee that the resulting quantile function is nondecreasing. Our model uses mixture quantiles as the conditional quantile function (Section 3). Note that mixture quantiles and mixture densities are two different families of distributions.

Mixture regression shares a similar idea to our approach, where the conditional distribution is modeled as a linear combination of basis functions. The conditional density function of the widely applied Gaussian mixture regression is modeled by a mixture (weighed sum) of Gaussian densities, where the weight, mean and variance of each component can be modeled as functions of factors. Gaussian mixture regression has been studied under different names, such as switching regression model, mixture of experts, mixture of factor analyzers (Quandt, 1972; DeSarbo and Cron, 1988; Ghahramani et al., 1996; Villani et al., 2009; Yuksel et al., 2012). Similarly, in our model formulation, the weights of quantile functions are functions of the factors (Section 3). Mixture regression is generally computationally expensive to calibrate, while our model can be calibrated by convex optimization. The weight function in mixture regression must sum up to one, which is typically achieved by the softmax transformation, while our model only requires nonnegative weight functions.

Quantile regression estimates the conditional quantile of a given confidence level given some factors. The application of spline functions to conduct nonparametric quantile regression has been widely explored in literature (Koenker et al., 1994; He et al., 1998; Koenker, 2011). However, conducting multiple quantile regressions separately can result in quantile crossing, particularly in nonlinear models, which impedes the interpretation of the results. To mitigate this problem, various methods have been proposed, such as imposing extra constraints (Bondell et al., 2010) or rearrangement (Chernozhukov et al., 2010). The constraints are often imposed at observed data points or at some confidence levels. We present a novel approach to quantile regression that guarantees noncrossing quantiles without any post-processing. Our method estimates the entire conditional distribution of the response variable, not just the quantiles at several confidence levels, as a result of conducting joint quantile regression. The proper choice of basis functions ensures the validity of the noncrossing property at all points (Section 3). The objective function in our approach is a linear combination of pinball losses at different confidence levels (Section 4), which has been adopted in various model calibration methods (Koenker, 1984; Zou and Yuan, 2008; Sottile and Frumento, 2022).

Quantile regression process refers to the regression coefficient as a function of the confidence level of quantile regression. Angrist et al. (2006) shows that the rescaled quantile regression process converges to a zero-mean Gaussian process. A number of independent but closely related studies have explored the use of various basis functions for modeling the quantile regression process, such as Bernstein basis polynomials (Reich et al., 2011), P-spline basis (Lian et al., 2015), common parameterized functions (Frumento and Bottai, 2016), and monotone B-spline (Yuan et al., 2017). By rearranging the model formulation in Section 3, it can be observed that our model coincides with the quantile regression process models if we view the individual polynomial terms as factors.

Nonetheless, there are several notable differences between our approach and existing research. First, the model calibration in Frumento and Bottai (2016) and Reich et al. (2011) involves numerical integration and Bayesian method, respectively, while our approach uses convex optimization, which is more efficient. Second, it it can be challenging to ensure noncrossing quantiles when directly modeling the linear quantile regression process.

Reich et al. (2011) introduces prior latent unconstrained variables; Frumento and Bottai (2016) checks the nonnegativity of derivative; Yuan et al. (2017) uses linear constraints when the feasible set is a bounded convex polytope; Lian et al. (2015) does not guarantee noncrossing quantiles. In contrast, our model guarantees noncrossing conditional quantiles at any two distinct confidence levels and any given factor value. Second, using a spline to model quantile regression process results in bounded conditional quantile, leading to severe underestimation of tail risk. Thus it is important to supplement the spline with quantile functions of common distributions. Furthermore, all aforementioned quantile regression process models study linear models and can be regarded as special cases of our approach. For linear quantile models to be noncrossing everywhere, the coefficients for each factor must be equal across all confidence levels.

Continuous Ranked Probability Score (CRPS) is one of the proper scoring rules (Matheson and Winkler, 1976) that is frequently used to measure the quality of a distributional prediction when the response variable is a real number (Hersbach, 2000). Laio and Tamea (2006) proposes an equivalent definition that is useful to formulate convex optimization. Hothorn et al. (2014); Gasthaus et al. (2019); Gouttes et al. (2021); Berrisch and Ziel (2021) use CRPS as the objective function of optimization in learning tasks. Zhang et al. (2022) proposes a related model where the CDF is a linear combination of basis CDFs. Although not mentioned in the paper, their estimation procedure minimizes CRPS.

Quantile model aggregation aims to aggregate multiple models of conditional quantile function to improve overall performance. The aggregation can be performed across different factor values and different quantile levels. Berrisch and Ziel (2021) uses B-splines as weight function across confidence levels, while Fakoor et al. (2021) uses neural networks as weight function across different factor values and different quantile levels. Our model can be adapted for quantile model aggregation. On the other hand, Fakoor et al. (2021) considers a regression model called the deep quantile regression by using constant functions as individual models. The conditional quantile function is a neural network, which requires extra efforts to ensure monotonicity.

The linear and polynomial model in Chernozhukov and Umantsev (2001) can be regarded as a special case of our model. *Conditional transformation model* (Hothorn et al., 2014) finds the optimal conditional transformation of a fixed quantile function such that it best fits the data.

3 Model Description

This section describes the Factor Model of Mixture Quantiles and discusses noncrossing quantile and various perspectives to view the model.

- $I =$ number of basis functions in the mixture
- $p =$ confidence level of a quantile function
- $\boldsymbol{x} =$ vector of factors
- $J =$ number of basis functions
- $\boldsymbol{a} =$ vector of coefficients
- $Q(p) =$ quantile function
- $G(p, \boldsymbol{x}, \boldsymbol{a}) =$ model with parameters $\boldsymbol{a}$ that outputs the p-quantile of the response variable conditioned on observing factor $\boldsymbol{x}$

3.1 Factor Model of Mixture Quantiles

The Factor Model of Mixture Quantiles is defined as

$$G(p, \boldsymbol{x}, \boldsymbol{a}) = \sum_{i=0}^{I} f_i(\boldsymbol{x}, \boldsymbol{a}_i) Q_i(p) \ , \tag{1}$$

where $Q_0(p) = 1$. The basis functions $\{Q_i(p)\}_{i=0}^{I}$ are defined on the unit interval, and are linearly independent, i.e., any $Q_i(p)$ is not equal to a linear combination of other $Q_j(p)$, $j \neq i$. The scale of each basis function $Q_i(p)$ is a function of the factors $\boldsymbol{x}$.

The model formulation (1) is quite general since $\{Q_i(p)\}_{i=1}^{I}$ and $\{f_i(\boldsymbol{x}, \boldsymbol{a}_i)\}_{i=0}^{I}$ can be arbitrary functions, providing great flexibility in modeling heteroskedastic data and nonlinear relations. The weight functions determine how the factors impact the scale of basis quantile functions. It has the appealing interpretation that different factors may impact the different part of the distribution. The conditional quantile function has a closed-form expression if all basis quantile functions do. Monte Carlo simulation can be easily conducted with inverse transform sampling.

We focus on the case where $\{f_i(\boldsymbol{x}, \boldsymbol{a}_i)\}_{i=0}^{I}$ are splines with basis spline functions $\{B_{ij}(\boldsymbol{x})\}_{i,j=0}^{I,J}$. Furthermore, we use the same basis spline functions for all basis quantile functions $\{Q_i(p)\}_{i=1}^{I}$, i.e., $\forall i, j,\ B_{ij}(\boldsymbol{x}) = B_j(\boldsymbol{x})$. Then the model is defined as

$$G(p, \boldsymbol{x}, \boldsymbol{a}) = \sum_{i=0}^{I} \sum_{j=0}^{J} a_{ij} B_j(\boldsymbol{x}) Q_i(p) \ , \tag{2}$$

where $B_0(x) = 1$, $Q_0(p) = 1$. Spline is adaptive to the data and can be optimized with the model in one shot. The linearity with respect to the coefficients not only results in a convenient formulation of convex optimization, but also retains an interpretable factor model structure such that we can analyze the impact of each factor on the shape of the conditional quantile function.

To distinguish $\{Q_i(p)\}_{i=0}^{I}$ and $\{B_j(\boldsymbol{x})\}_{j=0}^{J}$, we hereafter refer to $\{Q_i(p)\}_{i=0}^{I}$ as basis quantile function and $\{B_j(\boldsymbol{x})\}_{j=0}^{J}$ as basis spline functions. The meaning of the terms may be clearer if we expand the summation

$$G(p, \boldsymbol{x}, \boldsymbol{a}) = a_{00} + \sum_{i=1}^{I} a_{i0} Q_i(p) + \sum_{j=1}^{J} a_{0j} B_j(\boldsymbol{x}) + \sum_{i=1}^{I} \sum_{j=1}^{J} a_{ij} B_j(\boldsymbol{x}) Q_i(p) \ . \tag{3}$$

a_{00} is the constant location parameter, $\sum_{j=1}^{J} a_{0j} B_j(\boldsymbol{x})$ determines the conditional location, $\sum_{i=1}^{I} a_{i0} Q_i(p)$ is the base quantile function that does not vary with factors, $\sum_{i=1}^{I} \sum_{j=1}^{J} a_{ij} B_j(\boldsymbol{x}) Q_i(p)$ determines the conditional quantile function.

The positions of knots and maximal degree of polynomials of the basis spline functions $\{Q_i(p)\}_{i=1}^{I}$ are chosen or tuned. Splines are known to have bad estimation on two ends where there is not much data. To mitigate this problem, we can constrain the function to be linear on both ends. The knot selection problem is often nonconvex. Alternatively, it can be addressed by P-spline (Eilers and Marx, 1996, 2003) which uses a large number of knots and penalizes the absolute value of the second derivative. The complexity of the model is determined by the number of basis quantile functions, the number of factors and the form of weight functions. The number of parameters in the model is $(I+1) \times (J+1)$. The degrees of freedom are usually less due to constraints in optimization.

3.2 Noncrossing Quantile Model

We refer to a model as fully noncrossing if it satisfies that the conditional quantile functions of any two different confidence levels do not cross conditioned on any value of factors. That is,

$$\forall \boldsymbol{x}, p_1, p_2,\ 0 < p_1 < p_2 < 1,\ G(p_1, \boldsymbol{x}, \boldsymbol{a}) \leq G(p_2, \boldsymbol{x}, \boldsymbol{a})\ , \tag{4}$$

where $\mathcal{X}$ is the set of all possible values of factors $\boldsymbol{x}$. The condition is often relaxed to that $G(p_1, \boldsymbol{x}, \boldsymbol{a}) \leq G(p_2, \boldsymbol{x}, \boldsymbol{a})$ for a certain $\boldsymbol{x} \in \mathcal{X}$ or for a certain selected confidence levels $\{p_m\}_{m=1}^M$. Even with such simplification, it could require a large number of constraints to guarantee noncrossing for arbitrary $\{Q_i(p)\}_{i=1}^I$ and $\{f_i(\boldsymbol{x}, \boldsymbol{a}_i)\}_{i=0}^I$.

Instead of relaxing the condition (4), we propose using a sufficient condition. We show in Section 5 that such formulation is still highly flexible. Note that the model (2) satisfies noncrossing condition (4) if

- $\{Q_i(p)\}_{i=0}^I$ are nondecreasing;
- $\{B_j(\boldsymbol{x}, \boldsymbol{a})\}_{j=0}^J$ are nonnegative;
- $\boldsymbol{a}$ is nonnegative.

$Q_0(p)$ and $B_0(\boldsymbol{x}, \boldsymbol{a})$ are constants. Nonnegativity constraint is relatively easy to impose in optimization. Although a sufficient condition may seem too restrictive, Theorem 1 shows that the model retains good approximation ability.

For $\{Q_i(p)\}_{i=1}^I$, we propose two types of nondecreasing functions:

- quantile functions of common distributions;
- monotone basis spline functions such as I-spline (Ramsay, 1988).

Quantile functions of common distributions are prefered when the shape of the distribution is known to be close to common distributions. When the distribution is multimodal, splines offer greater flexibility. Since splines are bounded by definition, it works well for bounded variable, but needs to be combined with common quantile functions when the distribution has fat tails.

For $\{B_j(\boldsymbol{x}, \boldsymbol{a})\}_{j=1}^J$, we propose two ways to guarantee nonnegativity:

- nonnegative basis spline functions such as B-spline (de Boor and Daniel, 1974) with nonnegative coefficients de Boor and Daniel (1974); Papp (2011); Papp and Alizadeh (2014);
- arbitrary basis spline functions with constraints that characterize nonnegative polynomials in the optimization (Papp, 2011; Papp and Alizadeh, 2014).

While the latter is more flexible, it is more computationally expensive to optimize. Besides B-spline, other choices include Bernstein polynomial and M-spline (Curry and Schoenberg, 1988). B-spline and M-spline differ by a constant in scale. Papp (2011), Papp and Alizadeh (2014) show that piecewise Bernstein polynomial includes B-spline as a subset. The monotone I-spline used for $\{Q_i(p)\}_{i=1}^I$ is obtained by integrating M-spline. The aforementioned choices guarantee nonnegativity in the domain of the spline. When noncrossing property is desired beyond the domain, we can require the spline to have nonnegative partial derivative on the boundary, and extend the straight lines to infinity. The obtained function is nonnegative everywhere. In fact, one can use neural networks with nonnegative output. But it is beyond the scope of this paper.

3.3 Examples

This subsection provides several examples of the model described in Section 3. We use models from other research to demonstrate the wide applicability of our approach. The connection between the our approach and

related work is discussed in detail in Section 2.

Example 1 *Location-scale model of normal distribution*

$$q(p, \boldsymbol{x}, \boldsymbol{a}) = a_{00} + \sqrt{2}\text{erf}^{-1}(2p-1)a_{10} \ , \tag{5}$$

where q is a parameterized quantile function, $\sqrt{2}\text{erf}^{-1}(2p-1)$ is the quantile function of standard normal distribution.

Example 2 *Logistic-normal mixture of quantiles without factor dependence (Karvanen, 2006; Keelin, 2016; Peng et al., 2022)*

$$q(p, \boldsymbol{x}, \boldsymbol{a}) = a_{00} + \log(\frac{p}{1-p})a_{10} + \sqrt{2}\text{erf}^{-1}(2p-1)a_{20} \ , \tag{6}$$

where q is a parameterized quantile function, $\log(\frac{p}{1-p})$ is the quantile function of standard logistic distribution.

Example 3 *Linear model with normal noise*

$$q(p, \boldsymbol{x}, \boldsymbol{a}) = a_{00} + xa_{01} + \sqrt{2}\text{erf}^{-1}(2p-1) \ , \tag{7}$$

where q is a parameterized quantile function.

Example 4 *Quantile regression process model (Reich et al., 2011; Lian et al., 2015; Frumento and Bottai, 2016; Yuan et al., 2017)*

$$q_(p, \boldsymbol{x}, \boldsymbol{a}) = \sum_{k_1=1}^{K} B_{k_1}(p)x_1a_{1k_1} + \sum_{k_2=1}^{K} B_{k_2}(p)x_2a_{2k_2} \ , \tag{8}$$

where $B_k(p)$ can be Bernstein basis polynomials or B-spline basis.

Example 5 *Autoregressive conditional heteroskedasticity-1 (ARCH(1)) model with normal noise (Engle, 1982)*

$$q_t(p, \boldsymbol{x}, \boldsymbol{a}) = \sqrt{2}\text{erf}^{-1}(2p-1) \cdot \sqrt{a_{10} + x_{t-1}^2 a_{11}} \ , \tag{9}$$

where q_t is the parameterized quantile function of x_t at time t, x_{t-1} is the response variable at at time $t-1$. At every time t, the quantile function $q_t(p, \boldsymbol{x}, \boldsymbol{a})$ depends on the previous response variable x_{t-1}.

Example 6 *Dynamic quantile model (Gourieroux and Jasiak, 2008)*

$$q_t(p, \boldsymbol{x}, \boldsymbol{a}) = a_{00} + x_{t-1}a_{01} + \log(\frac{p}{1-p}) \cdot (a_{10} + x_{t-1}^2 a_{11}) \ . \tag{10}$$

where q_t is the parameterized quantile function of x_t at time t, x_{t-1} is the response variable at at time $t-1$. At every time t, the quantile function $q_t(p, \boldsymbol{x}, \boldsymbol{a})$ depends on past response variable x_{t-1}. The response variable at time t is generated by plugging a sample p from uniform distribution $U(0,1)$ and the past response variable x_{t-1} to $q_t(p, \boldsymbol{x}, \boldsymbol{a})$.

Example 1, 3 and 5 shows that our model incorporates some most common models as special cases. Maximum Likelihood estimation for these models are well studied. We propose a different estimation method in Section 4. Dynamic models such as Example 5 and 6 provide useful complements to classic time series models. Although we focus on linear function in $\boldsymbol{a}$ in subsequent sections on model calibration, the function $f(\boldsymbol{x}, \boldsymbol{a})$ in Example 5 is nonlinear in both $\boldsymbol{x}$ and $\boldsymbol{a}$. The model can still be calibrated by optimizing our proposed objective function, but the optimization is not convex. Gourieroux and Jasiak (2008) also studies using $|x|$ and $x^+ = \max\{0, x\}$, $x^- = \max\{0, -x\}$ to replace x, so that the factors remain nonnegative. Examples involving spline function can be formulated by replacing factors with splines.

4 Model Calibration by Convex Optimization

This section first reviews the basics of quantile regression (Koenker and Bassett, 1978) and then introduces the convex formulation of model calibration.

4.1 Quantile Regression

Consider a model of p-quantile of response variable $y|\boldsymbol{x}$ conditioned on factor $\boldsymbol{x}$ with parameter $\boldsymbol{a}$

$$q_p(y|\boldsymbol{x}) = g(p, \boldsymbol{x}, \boldsymbol{a}) \ . \tag{11}$$

Quantile regression estimates the parameter $\boldsymbol{a}$ by minimizing the normalized Koenker-Bassett error of the residuals

$$\min_{\boldsymbol{a}} \boldsymbol{\mathcal{E}}_p(y - g(p, \boldsymbol{x}, \boldsymbol{a})) \tag{12}$$

where the error is the expected pinball loss

$$\boldsymbol{\mathcal{E}}_p(Z) = E[\rho_p(Z)] \ , \quad \rho_p(Z) = pZ\mathbb{1}_{\{Z>0\}} - (1-p)Z\mathbb{1}_{\{Z\leq 0\}} \ , \tag{13}$$

Z is a random variable, $\mathbb{1}_{\{\cdot\}}$ is the indicator function that equals 1 when the equation in the bracket is true and 0 if otherwise. Suppose we have N pairs of response variables and factors $\{y_i, \boldsymbol{x}_i\}_{i=1}^N$. The optimization problem is

$$\min_{\boldsymbol{a}} \boldsymbol{\mathcal{E}}_p(y_i - g(p, \boldsymbol{x}_i, \boldsymbol{a})) \ . \tag{14}$$

We write error $\boldsymbol{\mathcal{E}}_p$ instead of average of pinball loss $\frac{1}{N}\sum_{i=1}^N \rho_p(y_i - g(p, \boldsymbol{x}_i, \boldsymbol{a}))$, since some errors cannot be written as an average in the generalization in Section 8.

4.2 Calibration: Optimization Problem Statement

- M = number of grid points of discretized confidence level
- N = sample size
- $\boldsymbol{Y}^N = (y_1, \cdots, y_N)$ = response variables of dependent variable
- $\boldsymbol{x}_i$ = vector of factors corresponding to response variable y_n, $n = 1, \cdots, N$;
- $\boldsymbol{x}^N = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N)$ = set of factor vectors
- $\boldsymbol{a}$ = vector of parameters
- $\mathcal{A}$ = feasible set of $\boldsymbol{a}$ determined by constraints
- $L(\boldsymbol{a}; \boldsymbol{Y}^N, \boldsymbol{x}^N, p)$ = discrete residual random variable that takes with equal probabilities the following values

$$y_n - \sum_{i=0}^{I}\sum_{j=0}^{J} a_{ij} B_j(\boldsymbol{x}_n) Q_i(p) \ , \quad n = 1, \ldots, N$$

The optimization problem statement for finding optimal parameters $\boldsymbol{a}$ is formulated as follows

Problem 1.

$$\min_{\boldsymbol{a}} \int_0^1 w(p)\boldsymbol{\mathcal{E}}_p\left(L(\boldsymbol{a};\boldsymbol{Y}^N,\boldsymbol{x}^N,p)\right)\mathrm{d}p \quad \text{subject to } \boldsymbol{a}\in\mathcal{A}\,, \tag{15}$$

where $w(p)$ is a nonnegative weight function satisfying $\int_0^1 w(p)\mathrm{d}p = 1$.

$w(p)$ can be chosen to focus on the distribution tail or body. Inherent from the pinball loss, the calibration is robust to outliers. While we consider the case where f_{ik} are spline functions, the calibration method works for the general case (1).

Next, we discretize the problem (15) by using a grid on p. The resultant optimization problem is still a convex programming problem.

Problem 2. Discrete variant:

$$\min_{\boldsymbol{a}} \sum_{m=1}^{M} w_m\boldsymbol{\mathcal{E}}_{p_m}\left(L(\boldsymbol{a};\boldsymbol{Y}^N,\boldsymbol{x}^N,p_m)\right) \quad \text{subject to } \boldsymbol{a}\in\mathcal{A}\,, \tag{16}$$

where w_m are nonnegative weights satisfying $\sum_{m=1}^{M} w_m = 1$.

Similar to quantile regression, the problem can be reduced to linear programming. Note that although we select only a finite number of confidence levels in the discrete variant, the procedure still estimates the whole conditional quantile function, and the quantiles are noncrossing. The calibration problem also allows one to focus on the tail of the distribution by assigning higher weights on errors with tail confidence levels. Rockafellar and Uryasev (2000) shows that the optimal objective value of quantile regression equals CVaR of the residuals. The optimal objective value is only approximately $\sum_{m=1}^{M} w_m\mathrm{CVaR}_{p_m}(L(\boldsymbol{a};\boldsymbol{Y}^N,\boldsymbol{x}^N,p_m))$, since the solution $\boldsymbol{a}$ may not give the exact solution to each of the M minimizations.

4.3 Equivalence to Constrained Joint Quantile Regression

Define M discrete random variables $\widehat{L}_m(\boldsymbol{a};\boldsymbol{Y}^N,\boldsymbol{x}^N)$, each taking with equal probabilities the values

$$y_n - \sum_{j=0}^{J}\lambda_{jm}B_j(\boldsymbol{x}_n), \quad n = 1,\ldots,N\,.$$

We can write (16) equivalently as

Problem 3.

$$\min_{\boldsymbol{a},\boldsymbol{\lambda}} \sum_{m=1}^{M} w_m\boldsymbol{\mathcal{E}}_{p_m}\left(\widehat{L}_m(\boldsymbol{a};\boldsymbol{Y}^N,\boldsymbol{x}^N)\right) \tag{17}$$

$$\text{subject to } \boldsymbol{a}\in\mathcal{A} \tag{18}$$

$$\sum_{i=0}^{I} Q_i(p_m)a_i = \lambda_{jm}\,, \quad m = 1,\ldots,M,\ j = 0,\cdots,J\,. \tag{19}$$

This problem formulation makes it clear that the calibration in Problem 2 is equivalent to conducting several quantile regressions in one shot with constraints on the parameters. Since $\sum_{j=0}^{J}\lambda_{jm}B_j(\boldsymbol{x}_n)$ is a spline of scalar

factor x_j, the objective function in (17) is the sum of objective functions of spline quantile regressions with confidence levels p_m, $m = 1, \cdots, M$. $\{\lambda_{jm}\}_{j=0,m=1}^{J,M}$ are associated by systems of equations (19), where the coefficients $\boldsymbol{a}$ are constrained by feasible set $\mathcal{A}$.

As J systems of equations, (19) can be written in matrix format

$$\boldsymbol{Q}\boldsymbol{A} = \boldsymbol{\Lambda} \,. \tag{20}$$

where $\boldsymbol{Q} = [Q_i(p_m)]'_{M\times(I+1)}$, $\boldsymbol{A} = [a_{ij}]_{(I+1)\times(J+1)}$, $\boldsymbol{\Lambda} = [\lambda_{jm}]'_{M\times(J+1)}$.

If the solution to the unconstrained problem (17) is in the feasible set defined by (18)(19), we can solve Problem 3 in the following two steps.

1. Solve M spline quantile regressions separately

$$\min_{\lambda} \boldsymbol{\mathcal{E}}_{p_m}\left(\widehat{L}_m(\boldsymbol{a}; \boldsymbol{Y}^N, \boldsymbol{x}^N)\right), \quad m = 1, \cdots, M \,. \tag{21}$$

2. Solve the systems of equations with constraints

$$\begin{aligned} \boldsymbol{a} &\in \mathcal{A} \\ \boldsymbol{Q}\boldsymbol{A} &= \boldsymbol{\Lambda} \,. \end{aligned} \tag{22}$$

In both steps, the solution can be nonunique. The uniqueness of solution to quantile regression is discussed in Koenker (2005). Portnoy (1991) shows that the number of distinct solutions to quantile regression when the confidence level p varies in $(0,1)$ is $\mathcal{O}(N \log N)$ in probability, which grows slower than the upper bound $\binom{N}{M\times(J+1)}$. The uniqueness of solution to the systems of equations depends on $\boldsymbol{Q}$ when $\mathcal{A} = \mathrm{R}^{(I+1)\times(J+1)}$. For proper choice of independent basis quantile functions $\{Q_i(p)\}_{i=1}^{I}$, $\boldsymbol{Q}$ has full column rank when $M \geq I+1$. To check feasibility when $\mathcal{A} = \mathrm{R}_{\geq 0}^{(I+1)\times(J+1)}$, we can solve the linear programming problem $\min_{\boldsymbol{a}\geq 0} ||\boldsymbol{Q}\boldsymbol{A} - \boldsymbol{\Lambda}||_1$.

4.4 Equivalence to CRPS Minimization

Continuous Ranking Probability Score (CRPS) is frequently used to evaluate the quality of a distributional forecast. One may want to directly optimize the measure by which the model is evaluated in model calibration. We show that our calibration method is equivalent to CRPS minimization.

For a cumulative distribution function (CDF) F and an observation y_i of response variable y, CRPS is defined by

$$CRPS(F, y_i) = \int_{\mathcal{R}} \left(F(x) - 1_{\{y_i \leq x\}}\right)^2 \mathrm{d}x \,. \tag{23}$$

This is the squared distance between F and the CDF of a single observation y_i of response variable y. For quantile function Q, i.e., generalized inverse function of F, and response variable y_i, CRPS has the following equivalent definition (Laio and Tamea, 2006)

$$CRPS(Q, y_i) = 2\int_0^1 \rho_p\left(y_i - Q(p)\right) \mathrm{d}p \,. \tag{24}$$

Laio and Tamea (2006) proposes this equivalence without mentioning the condition that $F(x)$ should have a thinner tail than power law $x^{-1/2}$. Fakoor et al. (2021) gives a detailed proof.

Consider calibrating the model $G(p, \boldsymbol{\theta}, \boldsymbol{a})$ by minimization of the sum of CRPS of the response variables

$$\min_{\boldsymbol{a}} \sum_{n=1}^{N} \int_0^1 \rho_p \left(y_n - G(p, \boldsymbol{x}_n, \boldsymbol{a})\right) \mathrm{d}p \,. \tag{25}$$

We see that (25) is equal to the objective function in Problem 1 with uniform weight $w(p)$ by exchanging the integral and sum

$$\int_0^1 \boldsymbol{\mathcal{E}}_p \left(y_n - G(p, \boldsymbol{x}_n, \boldsymbol{a})\right) \mathrm{d}p = \int_0^1 \sum_{n=1}^{N} \rho_p \left(y_n - G(p, \boldsymbol{x}_n, \boldsymbol{a})\right) \mathrm{d}p \,. \tag{26}$$

5 Approximation Theorem

Our model can approximate any bounded conditional quantile model to arbitrary precision as the number of knots tends to infinity. Since our model has nonnegative parameters, classic approximation theorem of polynomials cannot be directly applied.

- $\mathcal{X}$ = hyperrectangle of domain of factors
- $T = (t_{ij})_{i=1,\cdots,I+1,j=1,\cdots,J}$ = matrix representation of subdivision of $[0,1] \times \mathcal{X}$
- $||T|| = \max_{i,j} |t_{i+1,j} - t_{i,j}|$ mesh size of subdivision Z
- $\mathrm{cone}(U)$ = cone of nonnegative linear combination of functions in a set of basis functions U
- int = interior of a set
- $\mathcal{P}(U, T)$ = set of piecewise functions where each piece (in the scaled representation) defined on a subdivision defined by Z is in $\mathrm{cone}(T)$
- $G_{[0,1]\times\mathcal{X}}$ = cone of all continuous and nonnegative conditional quantile models defined on $[0,1] \times \mathcal{X}$ that is nondecreasing in the first variable

Theorem 1 *Consider I-spline basis* $\{Q_i(p)\}_{i=0}^{\mathrm{I}}$ *defined on* $[0,1]$ *and B-spline basis* $\{B_j(\boldsymbol{x})\}_{j=0}^{\mathrm{J}}$ *defined on* $\mathcal{X}$. *Furthermore, let* T_i *be an asymptotically nested sequence of subdivisions with mesh sizes approaching zero. Then the set* $\cup_i \mathcal{P}(Q_i(p) \otimes B_j(\boldsymbol{x}), T_i)$ *is a dense subcone of* $G_{[0,1]\times\mathcal{X}}$.

Theorem 1 is an application of Papp (2011); Papp and Alizadeh (2014).

6 Asymptotic Property

This section contains the asymptotic properties of the estimator.

- $\widehat{\mathcal{E}}_N(\boldsymbol{a}) = \sum_{m=1}^{M} w_m \boldsymbol{\mathcal{E}}_{p_m} \left(L(\boldsymbol{a}; \boldsymbol{Y}^N, \boldsymbol{x}^N, p_m)\right)$
- $\widehat{\boldsymbol{a}}_N = \mathrm{argmin}_{\boldsymbol{a}\in\mathcal{A}} \widehat{\mathcal{E}}_N(\boldsymbol{a})$
- $\boldsymbol{a}_0$ = true parameter

Theorem 2 *Assume that* $\mathcal{A}$ *is a compact set. If there is a function* $\widehat{\mathcal{E}}_0(\boldsymbol{a})$ *such that (i)* $\widehat{\mathcal{E}}_N(\boldsymbol{a})$ *converges uniformly in probability to* $\widehat{\mathcal{E}}_0(\boldsymbol{a})$*; (ii)* $\widehat{\mathcal{E}}_0(\boldsymbol{a})$ *is uniquely minimized at* $\widehat{\boldsymbol{a}}$*; (iii)* $\widehat{\mathcal{E}}_0(\boldsymbol{a})$ *is continuous at* $\widehat{\boldsymbol{a}}$*, then* $\widehat{\boldsymbol{a}}_{\mathrm{N}} \xrightarrow{d} \boldsymbol{a}_0$.

Theorem 3 *Suppose that the conditions Theorem 2 are satisfied, and (i) $\boldsymbol{a}_0$ is an interior point of $\mathcal{A}$; (ii) $\widehat{\mathcal{E}}_{\mathrm{N}}(\boldsymbol{a})$ is twice continuously differentiable in a neighborhood of $\boldsymbol{a}_0$; (iii) $\sqrt{n}\nabla_{\boldsymbol{a}}\widehat{\mathcal{E}}_{\mathrm{N}}(\boldsymbol{a}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Omega)$; (iv) there is $H(\boldsymbol{a})$ that is continuous at $\boldsymbol{a}_0$ and $\sup_{\boldsymbol{\theta}_0 \in \mathcal{A}} ||\nabla_{\boldsymbol{aa}}\widehat{\mathcal{E}}_{\mathrm{N}}(\boldsymbol{a}_0) - H(\boldsymbol{a})|| \xrightarrow{p} \mathbf{0}$; (v) $H = H(\boldsymbol{a}_0)$ is nonsingular. Then*

$$\sqrt{N}(\widehat{\boldsymbol{a}}_{\mathrm{N}} - \boldsymbol{a}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, H^{-1}\Omega H^{-1}) \,. \tag{27}$$

Theorem 2, 3, appeared in Frumento and Bottai (2016), are applications of Newey and McFadden (1994).

7 Dimensionality Reduction by Reduced Rank Tensor

This section introduces the basics of CP decomposition of tensors, and introduces an alternating algorithm for dimensionality reduction. For a comprehensive review of tensor decomposition, readers are referred to Kolda and Bader (2009); Rabanser et al. (2017). Spline-based methods are often deemed inadequate for high-dimensional problems due to the exponential growth of parameters in multivariate splines, which poses challenges in both statistical inference and optimization. Our proposed dimensionality reduction method, however, reduces the growth rate to linear. Additionally, our proposed alternating algorithm addresses a smaller convex optimization problem in each iteration, and differs from the classic Alternating Least-Squares Algorithm through the minimization of a distinct objective function and the addition of a nonnegative constraint in each step, given that our model incorporates nonnegative parameters.

7.1 Tensor and CP Decomposition

- $\boldsymbol{A}$ = parameter tensor
- K = number of scalar factors
- R = rank of parameter tensor
- $\otimes$ = tensor product of two vectors
- $\bigotimes_{k=1}^{K}$ = tensor product of K vectors
- $\cdot$ = sum of elements of the elementwise product of two tensors of the same size

CP decomposition for a rank-R tensor $\boldsymbol{A}$ is defined as

$$\boldsymbol{A} = \sum_{r=1}^{R} \boldsymbol{u}_r^{(0)} \otimes \cdots \otimes \boldsymbol{u}_r^{(K)} \,, \tag{28}$$

where $\{\boldsymbol{u}_r^k\}_{r=1,k=0}^{R,K}$ are vectors.

Let $\boldsymbol{U}_k = (\boldsymbol{u}_1^{(k)}, \cdots, \boldsymbol{u}_R^{(k)})$, $k = 1, \cdots, K$. CP decomposition is often estimated by Alternating Least-squares Algorithm, which estimates $\boldsymbol{U}_k$ one by one with fixed $\{\boldsymbol{U}_{k'}\}_{k' \neq k}$. Each step is a convex optimization problem.

The following equation is useful for subsequent interpretation of reduced rank method

$$\bigotimes_{k=0}^{K} \boldsymbol{u}_r^{(k)} \cdot \bigotimes_{k=0}^{K} \boldsymbol{v}_r^{(k)} = \prod_{k=0}^{K} \boldsymbol{u}_r^{(k)} \cdot \boldsymbol{v}_r^{(k)} \,, \tag{29}$$

i.e., $\boldsymbol{u}_r^{(0)} \otimes \cdots \otimes \boldsymbol{u}_r^{(K)} \cdot \boldsymbol{v}_r^{(0)} \otimes \cdots \otimes \boldsymbol{v}_r^{(K)} = \left(\boldsymbol{u}_r^{(0)} \cdot \boldsymbol{v}_r^{(0)}\right) \cdots \left(\boldsymbol{u}_r^{(K)} \cdot \boldsymbol{v}_r^{(K)}\right)$.

7.2 Reduced Rank Method for Parameter Tensor

- K = number of scalar factors
- L = number of basis spline functions for each scalar factor
- $J = K \times L$
- $\boldsymbol{A}$ = parameter tensor with the same elements as parameter vector $\boldsymbol{a}$
- $\boldsymbol{B}_k(x) = (B_{k1}(x), \cdots, B_{kL}(x))$ = vector of univariate spline basis of scalar factor x_k
- $\bigotimes_{k=1}^{K} \boldsymbol{B}_k(x_k)$ = tensor product of K vectors of univariate spline basis of scalar factors
- $\boldsymbol{Q}(p) = (Q_0(p), \cdots, Q_I(p))$ = vector of basis quantile functions

For B-spline, L = degree of univariate polynomial+number of knots+1. Multivariate B-spline basis is obtained by tensor product of univariate B-spline basis.

With the above notations, we can write the model (2) in tensor format

$$G(p, \boldsymbol{x}, \boldsymbol{a}) = \boldsymbol{A} \cdot \left(\boldsymbol{Q}(p) \otimes \bigotimes_{k=1}^{K} \boldsymbol{B}_k(x_k) \right) . \tag{30}$$

Note that in our model formulation, $\boldsymbol{A} \geq 0$. Consider a nonnegative version of CP decomposition (28) on the parameter tensor $\boldsymbol{A}$ where $\boldsymbol{u}_r^{(k)} \geq 0$. With (28)(29), we have

$$G(p, \boldsymbol{x}, \boldsymbol{a}) = \sum_{r=1}^{R} \left(\boldsymbol{u}_r^{(0)} \cdot \boldsymbol{Q} \right) \left(\boldsymbol{u}_r^{(1)} \cdot \widehat{\boldsymbol{B}}_1 \right) \cdots \left(\boldsymbol{u}_r^{(K)} \cdot \boldsymbol{B}_K \right) . \tag{31}$$

Thus reduced rank method has a straightforward interpretation. The model is reduced to a linear combination of R functions from $(I+1)(J+1)$. In the extreme case where $R = 1$, the model reduced to a heteroscedastic model where the conditional quantile function is $\boldsymbol{u}_r^{(0)} \cdot \boldsymbol{Q}$, whose scale is controlled by a scalar function $\prod_{k=1}^{K} \boldsymbol{u}_r^{(k)} \cdot \boldsymbol{B}_k$. It can be regarded as obtaining R new basis quantile functions $\{\boldsymbol{u}_r^{(0)} \cdot \boldsymbol{Q}\}_{r=1}^{R}$, likewise for basis spline functions. The new basis, though having a smaller number, can still have undesirable smoothness condition. We expect better performance when it is used along with penalties in P-spline. The reduced basis functions are a linear combination of all original basis functions, while sparse optimization leads to a linear combination of a subset by forcing zeros among the parameters. We find that the solution to Problem 2 is often sparse with many small nonzero values. Thus the low rank decomposition is expected to produce good approximation, although the decomposition (28) is not always valid for any $\boldsymbol{A}$ for a low rank R.

7.3 Alternating Algorithm

We propose an alternating algorithm to find $\boldsymbol{A}$. Define $\widetilde{L}\left(\{\boldsymbol{U}^{(k)}\}_{k=0}^{K}; \boldsymbol{Y}^N, \boldsymbol{x}^N, p \right)$ = discrete residual random variable that takes with equal probabilities the following values

$$y_n - \sum_{r=1}^{R} \left(\boldsymbol{u}_r^{(0)} \cdot \boldsymbol{Q}(p) \right) \left(\boldsymbol{u}_r^{(1)} \cdot \boldsymbol{B}_1(x_n) \right) \cdots \left(\boldsymbol{u}_r^{(K)} \cdot \boldsymbol{B}_K(x_n) \right) , \quad n = 1, \ldots, N .$$

Algorithm 1 Alternating Algorithm for Estimating Reduced Rank Parameter Tensor

Input: Data $\{(y_i, \boldsymbol{x}_i)\}_{i=1}^N$; confidence levels $\{p_m\}_{m=1}^M$; model $G(p, \boldsymbol{\theta}, \boldsymbol{a})$; error function $\boldsymbol{\mathcal{E}}_p$; threshold ϵ

1: Initialize $\{\boldsymbol{U}^{(k)}\}_{k=0}^K$

2: **repeat**

3: **for** $k = 0$ to K **do**

4: Update $\boldsymbol{U}_k^{(s)}$ to $\boldsymbol{U}_k^{(s+1)}$ with fixed $\{\boldsymbol{U}_{k_t}^{(s)}\}_{k_t \neq k}$

$$\boldsymbol{U}_k^{(s+1)} = \underset{\boldsymbol{U}_k \geq 0}{\operatorname{argmin}} \sum_{m=1}^{M} w_m \boldsymbol{\mathcal{E}}_{p_m} \left(\widetilde{L}(\{\boldsymbol{U}_k^{(s)}\}_{k=0}^K; \boldsymbol{Y}^N, \boldsymbol{x}^N, p_m) \right) \tag{32}$$

5: **end for**

6: $s += 1$

7: **until** $||\boldsymbol{A}^{(s+1)} - \boldsymbol{A}^{(s)}||_F \leq \epsilon$

Output: Parameter tensor $\boldsymbol{A}$

Other stopping rules can be used as well, such as the number of iterations, the difference between consecutively updated objective values. Each step in the algorithm is a convex optimization problem. The objective function is nonincreasing in each step. The algorithm will find a local minimum at the end. Different initialization is needed for finding a smaller local minimum.

8 Generalization Based on Risk Quadrangle Framework

This section introduces CVaR and the risk quadrangle framework (Rockafellar and Uryasev, 2013) and uses the framework to extend our approach to estimation of conditional functions of other uncertainty measures such as CVaR.

8.1 Risk Quadrangle

Conditional Value-at-Risk (CVaR) For a continuous random variable Z, the p-Conditional Value-at-Risk (Rockafellar and Uryasev, 2000, 2002) is defined as the average value that exceeds the p-quantile of Z

$$\text{CVaR}_p(Z) = \bar{q}_p(Z) = E[Z|Z > q_p(X)] \,. \tag{33}$$

As an uncertainty measure, it takes into account not only the probability, but also the scale of extreme losses.

Error and statistic The risk quadrangle is a general framework describing the relation between risk, deviation, regret, error and statistics. This section presents the basics of the regression theory in risk quadrangle framework necessary for understanding the subsequent generalization of the model.

A functional of random variable Z is called a regular measure of error if it satisfies the following conditions: it has values in $[0, \infty)$; it is closed convex in Z with $\boldsymbol{\mathcal{E}}(0) = 0$, $\boldsymbol{\mathcal{E}}(Z) > 0$ when $Z \neq 0$; for sequences of random variables $\{Z_k\}_{k=1}^{\infty}$, if $\lim_{k\to\infty} \boldsymbol{\mathcal{E}}(Z_k) = 0$, then $\lim_{k\to 0} E[Z_k] = 0$.

The statistic associated with Z by $\boldsymbol{\mathcal{E}}$ is defined by

$$S(Z) = \arg\min_{C \in \mathcal{R}} \boldsymbol{\mathcal{E}}(Z - C) \,. \tag{34}$$

Two prominent examples are the quantile-based quadrangle and the superquantile-based Quadrangle.

Quantile-based quadrangle (at any confidence level $p \in (0,1)$):

$$\mathcal{S}(Z) = \mathrm{VaR}_p(Z) = q_p(Z) = \text{quantile}, \quad \boldsymbol{\mathcal{E}}_p(Z) = E[\frac{p}{1-p}Z_+ + Z_-] = \text{normalized Koenker-Bassett error}. \tag{35}$$

Superquantile-based quadrangle (at any confidence level $p \in (0,1)$):

$$\mathcal{S}(Z) = \mathrm{CVaR}_p(Z) = \bar{q}_p(Z) = \text{superquantile}, \quad \boldsymbol{\mathcal{E}}_p(Z) = \frac{1}{1-p}\int_0^1 \max\{0, \bar{q}_\beta(Z)\}\mathrm{d}\beta - EZ. \tag{36}$$

Regression in risk quadrangle framework We can conduct the following regression to estimate the statistic S associated with Y by $\boldsymbol{\mathcal{E}}_p$

$$\min_{g \in \mathcal{C}} \boldsymbol{\mathcal{E}}(Y - g(\boldsymbol{x}))\,, \tag{37}$$

for given random variables $\boldsymbol{x}, Y$, and some given class $\mathcal{C}$ of functions g. With g obtained in (37), we can estimate the statistic of Y conditioned on observing $\boldsymbol{x}$ by plugging in values of $\boldsymbol{x}$ to g.

8.2 Generalization

By replacing the error $\boldsymbol{\mathcal{E}}$ in Problem 1 with other errors pamameterized by a confidence level, our approach can be generalized to model other functions of uncertainty measures. Examples of risk quadrangle parameterized by a parameter in $(0,1)$ include quantile-based (Rockafellar and Uryasev (2013)), superquantile (CVaR)-based (Rockafellar et al. (2014); Golodnikov et al. (2019)), and expectile-based quadrangle (Kuzmenko (2020)).

An example of Factor Model of Mixture CVaRs is given as follows.

Example 7 *Two-factor model of normal-logistic CVaR mixtures*

$$\bar{q}(p, \boldsymbol{x}, \boldsymbol{a}) = a_{00} + x_1 a_{01} + \bar{Q}_{\mathtt{N}}(p) x_1 a_{11} + \frac{H(p)}{1-p} x_2 a_{21}\,, \tag{38}$$

where $\bar{q}$ is a parameterized CVaR function, $H(p) = -plog(p) - (1-p)\log(1-p))$, $\frac{H(p)}{1-p}$ is the CVaR function of standard logistic distribution, $\bar{Q}_{\mathtt{N}}(p) = f_{\mathtt{N}}(\sqrt{2}\mathrm{erf}^{-1}(2p-1))/(1-p)$ is the CVaR function of standard normal distribution, $f_{\mathtt{N}}$ is the density function of standard normal distribution (Norton et al. (2021)).

9 Numerical Experiments

This section presents numerical experiments with real-world data. The response variable is the daily return of the S&P 500 index, while the factors are the daily returns of the Euronext 100 index and the SSE Composite Index. The data, covering the period from 01/01/2019 to 01/01/2021, is obtained from Yahoo Finance and excludes days with missing values for any of the indices. For convenience in cross-validation, a few earliest days are excluded so that the remaining data can be divided into ten equal parts. The data is randomly permuted before the experiment is conducted.

The model calibration is conducted with R package Portfolio Safeguard (Zabarankin et al., 2016). The benchmark models, Gaussian mixture regression and Generalized additive model, are implemented in R packages Mixtools (Benaglia et al., 2009) and GAMLSS (Rigby and Stasinopoulos, 2005).

Fold	1	2	3	4	5	Mean
Factor model of mixture quantiles	0.830	0.730	0.870	0.760	0.780	0.794
Gaussian mixture regression	0.840	0.680	0.890	0.750	0.910	0.814
Generalized additive model	0.920	0.730	0.900	0.830	0.850	0.846

(a) This table presents the out-of-sample coverage rate of our model and the benchmarks in 5-fold cross-validation. The coverage rate is calculated by the percentage of data in validation set that falls within the predicted conditional interval of 0.1-quantile and 0.9-quantile.

Fold	1	2	3	4	5	Mean
Factor model of mixture quantiles	0.0052	0.0076	0.0054	0.0071	0.0068	0.0064
Gaussian mixture regression	0.4790	0.7568	0.5373	0.6871	0.5765	0.6073
Generalized additive model	0.0053	0.0077	0.0054	0.0073	0.0063	0.0064

(b) This table presents the out-of-sample CRPS of our model and the benchmarks in 5-fold cross-validation. The CRPS is calculated by sum of CRPS of each distributional prediction on data in validation set.

Table 1

9.1 Factor Model of Mixture Quantiles

The first experiment focuses on the Factor Model of Mixture Quantiles (Section 3) and its calibration by convex optimization (Section 4).

Model Consider the Two-Factor Model of Mixture Quantiles

$$G(p, \boldsymbol{x}, \boldsymbol{a}) = f_0(x_1, x_2) + Q_{\mathrm{N}}(p) f_1(x_1, x_2) + Q_{\mathrm{E}_1}(p) f_2(x_1, x_2) + Q_{\mathrm{E}_2}(p) f_3(x_1, x_2) \,, \tag{39}$$

where $Q_{\mathrm{N}}(p) = \sqrt{2}\mathrm{erf}^{-1}(2p-1)$, $Q_{\mathrm{E}_1}(p) = -\ln(2-2p)$, $Q_{\mathrm{E}_1}(p) = \ln(2p)$. $Q_{\mathrm{N}}(p)$ is the quantile function of standard normal distribution; $Q_{\mathrm{E}_1}(p)$ is the quantile function of right-side transformed exponential distribution supported on $(0.5, 1)$; $Q_{\mathrm{E}_2}(p)$ is the quantile function of left-side transformed exponential distribution supported on $(0, 0.5)$. For each the bivariate B-splines $\{f_i\}_{i=0,1,2,3}$, we use basis of degree three with six knots. The interpretation is that $f_1(x_1, x_2)$ determines the body of the distribution, while $f_2(x_1, x_2)$ and $f_3(x_1, x_2)$ determine the left and right tails, respectively.

Calibration The confidence levels used in the optimization are $0.05, 0.15, 0.25, \cdots, 0.95$ with equal weights. We choose 5 equidistant knots for each factor. We add a penalty term on the squared difference between adjacent coefficients (Eilers and Marx, 1996, 2003) for each spline function $\{f_i\}_{i=0,1,2,3}$ to alleviate overfit. The penalty coefficients are fixed at 1000 and not optimized.

Benchmarks We choose two benchmark models, namely Gaussian mixture regression and generalized additive model. For Gaussian mixture regression, we use BIC as the model selection criteria, which is one of the built-in methods in the package. The weights of the components are not functions of the factors. For generalized additive model, the mean and standard deviation of the normal distribution are functions of the factors. The additive formula of the function is univariate cubic spline for each factor and an additional linear interaction term as the additive terms.

Performance Measure We conduct 5-fold cross-validation to test the out-of-sample CRPS and coverage rate. The CRPS is calculated by sum of CRPS of each distributional prediction on data in validation set. The coverage rate is calculated by the percentage of data in validation set that falls within the predicted conditional interval of 0.1-quantile and 0.9-quantile. A small CRPS and coverage rate close to 0.8 indicate a high-quality distributional prediction.

Results Results are presented in Tables 1a and 1b. Our approach exhibits improved out-of-sample coverage rate compared to both benchmark models and comparable out-of-sample CRPS to the generalized additive model. Gaussian mixture regression, on the other hand, exhibits poor performance in terms of CRPS.

We can obtain the p-quantile surface of the response variable by varying the values of (x_1, x_2) with a fixed p. We use the model in the 5th fold of the cross-validation as an example. Figure 1 shows three selected quantile surfaces for visualization, demonstrating the highly nonlinear relations incorporated in the model. The surfaces does not to cross each other.

9.2 Dimensionality Reduction by Reduced Rank Tensor for Factor Model of Mixture CVaRs

The second experiment examines the Reduced Rank Method (Section 7) for the Factor Model of Mixture CVaRs (Section 8).

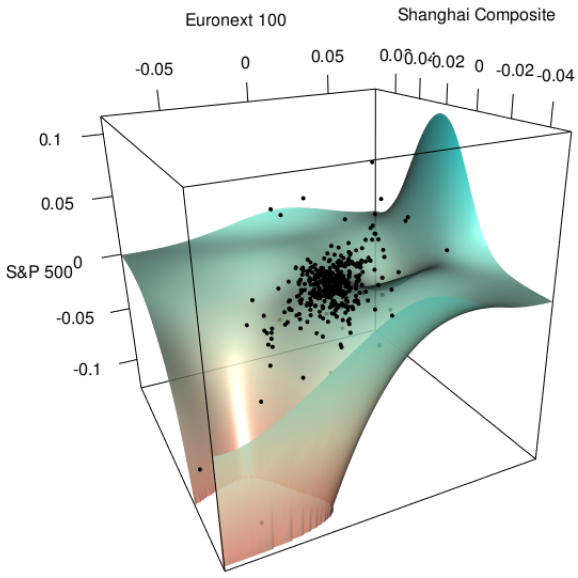
Model Consider the Factor Model of Mixture CVaRs in tensor format

$$G(p, \boldsymbol{x}, \boldsymbol{a}) = \boldsymbol{A} \cdot \left(\bar{\boldsymbol{Q}}(p) \otimes \bigotimes_{k=1}^{K} \boldsymbol{B}_k(x_k) \right) , \tag{40}$$

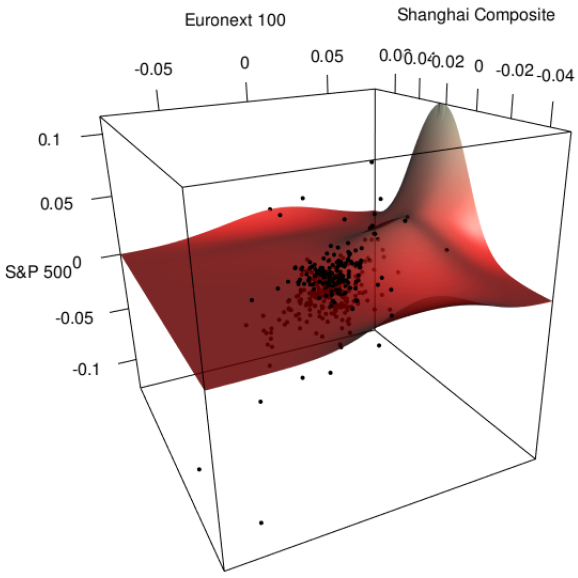
where $\bar{\boldsymbol{Q}}(p)$ is the vector of basis CVaR functions. We use I-spline basis of degree three with five knots and CVaR function of exponential distribution as basis CVaR functions. for splines $\boldsymbol{B}_k(x_k)$, we still use B-spline of degree three with six knots.

Calibration The experiment considers cases where the rank, R in (31), equals 1, 3, and 10. For each case, 21 optimization steps are conducted using the alternating algorithm. The error function for the regression is defined by replacing $\boldsymbol{\mathcal{E}}_p$ in Problem 1 with the error defined in (36). Since CVaR is often used to measure tail risk, we choose more tail confidence levels in the optimization. The confidence levels are $0.25, 0.55, 0.60, 0.65, \cdots, 0.95$. The full data set from Section 9.1 is used. No penalty is applied to the smoothness of splines as in Section 9.1, since the aim is to compare the models with varying ranks.

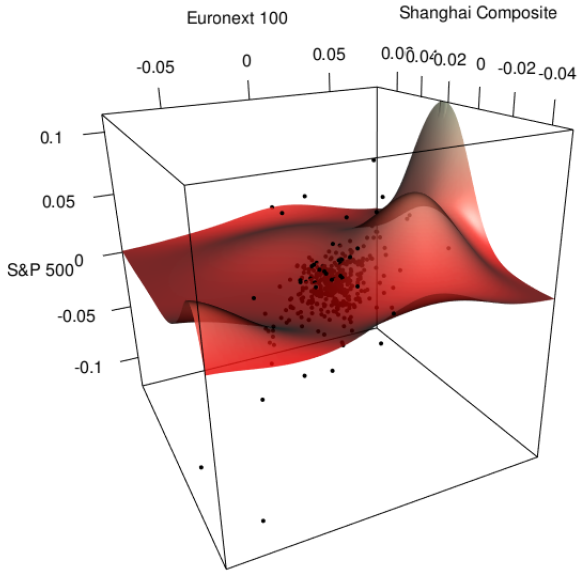
Results The objective values at each step are displayed in Figure 2. A higher rank corresponds to a smaller final objective value and a greater number of steps to converge. Nonetheless, the optimization converges to the optimal objective value in a few steps for all three cases, demonstrating the efficiency of the algorithm for estimation.



(a) Quantile surface of confidence level 0.05



(b) Quantile surface of confidence level 0.5



(c) Quantile surface of confidence level 0.95

Figure 1: This figure shows the quantile surfaces of confidence levels 0.05, 0.5, 0.95 in the numerical experiment in Section 9.1. The black dots are the data points.

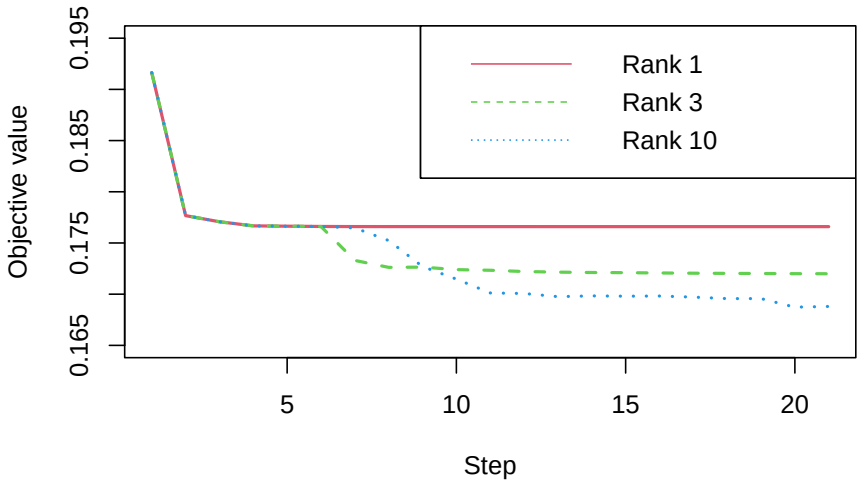


Figure 2: Stepwise objective value of alternating algorithm of rank 1, 3 and 10

References

C. de Boor and James W. Daniel. Splines with nonnegative b-spline coefficients. *Mathematics of Computation*, 28(126):565–568, 1974. ISSN 00255718, 10886842. URL http://www.jstor.org/stable/2005928.

Richard E. Quandt. A new approach to estimating switching regressions. *Journal of the American Statistical Association*, 67(338):306–310, 1972. ISSN 01621459. URL http://www.jstor.org/stable/2284373.

Roger Koenker and Gilbert Bassett. Regression quantiles. *Econometrica*, 46(1):33–50, 1978. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1913643.

Dávid Papp. Optimization models for shape-constrained function estimation problems involving nonnegative polynomials and their restrictions. 2011.

Dávid Papp and Farid Alizadeh. Shape-constrained estimation using nonnegative splines. *Journal of Computational and Graphical Statistics*, 23(1):211–231, 2014. doi: 10.1080/10618600.2012.707343. URL https://doi.org/10.1080/10618600.2012.707343.

Hans Hersbach. Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, 15(5):559 – 570, 2000. doi: 10.1175/1520-0434(2000)015⟨0559:DOTCRP⟩2.0.CO;2. URL https://journals.ametsoc.org/view/journals/wefo/15/5/1520-043420000150559dotcrp20co2.xml.

Frank L. Hitchcock. The expression of a tensor or a polyadic as a sum of products. *Journal of Mathematics and Physics*, 6(1-4):164–189, 1927. doi: https://doi.org/10.1002/sapm192761164. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sapm192761164.

Richard A. Harshman. Foundations of the parafac procedure: Models and conditions for an "explanatory" multi-model factor analysis. 1970.

J. Douglas Carroll and Jih Jie Chang. Analysis of individual differences in multidimensional scaling via an n-way generalization of "eckart-young" decomposition. *Psychometrika*, 35:283–319, 1970.

R. Tyrrell Rockafellar and Stan Uryasev. The fundamental risk quadrangle in Risk Management, Optimization and Statistical Estimation. *Surveys in Operations Research and Management Science 18*, 2013.

R. Tyrrell Rockafellar and Stanislav Uryasev. Optimization of Conditional Value-at-Risk. *Journal of Risk*, 2: 21–41, 2000.

R. Tyrrell Rockafellar and Stanislav Uryasev. Conditional Value-at-Risk for general loss distributions. *Journal of Banking and Finance*, pages 1443–1471, 2002.

Whitney K. Newey and James L. Powell. Asymmetric least squares estimation and testing. *Econometrica*, 55(4):819–847, 1987. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1911031.

Maodong Pan, Weihua Tong, and Falai Chen. Compact implicit surface reconstruction via low-rank tensor approximation. *Computer-Aided Design*, 78:158–167, 2016. ISSN 0010-4485. doi: https://doi.org/10.1016/j.cad.2016.05.007. URL https://www.sciencedirect.com/science/article/pii/S0010448516300276. SPM 2016.

Heng Lian, Weihua Zhao, and Yanyuan Ma. Multiple quantile modeling via reduced-rank regression. *Statistica Sinica*, 29(3):1439–1464, 2019. ISSN 10170405, 19968507. URL https://www.jstor.org/stable/26706009.

Liang Chen, Juan J. Dolado, and Jesús Gonzalo. Quantile factor models. *Econometrica*, 89(2):875–910, 2021. doi: https://doi.org/10.3982/ECTA15746. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA15746.

R. Tyrrell Rockafellar and Johannes O. Royset. *Superquantiles and Their Applications to Risk, Random Variables, and Regression*, chapter Chapter 8, pages 151–167. 2013. doi: 10.1287/educ.2013.0111. URL https://pubsonline.informs.org/doi/abs/10.1287/educ.2013.0111.

R.T. Rockafellar, J.O. Royset, and S.I. Miranda. Superquantile regression with applications to buffered reliability, uncertainty quantification, and conditional value-at-risk. *European Journal of Operational Research*, 234(1):140–154, 2014. ISSN 0377-2217. doi: https://doi.org/10.1016/j.ejor.2013.10.046. URL https://www.sciencedirect.com/science/article/pii/S0377221713008692.

Alex Golodnikov, Viktor Kuzmenko, and Stan Uryasev. Cvar regression based on the relation between cvar and mixed-quantile quadrangles. *Journal of Risk and Financial Management*, 12(3), 2019. ISSN 1911-8074. doi: 10.3390/jrfm12030107. URL https://www.mdpi.com/1911-8074/12/3/107.

R Tyrrell Rockafellar and Johannes O Royset. Superquantile/cvar risk measures: Second-order theory. *Annals of Operations Research*, 262(1):3–28, 2018.

VM Kuzmenko. A new family of expectiles and its properties. *Cybernetics and Computer Technologies*, 2020.

G. P. Sillitto. Derivation of approximants to the inverse distribution function of a continuous univariate population from the order statistics of a sample. *Biometrika*, 56(3):641–650, 1969. ISSN 00063444. URL http://www.jstor.org/stable/2334672.

Juha Karvanen. Estimation of quantile mixtures via l-moments and trimmed l-moments. *Computational Statistics & Data Analysis*, 51(2):947–959, 2006. ISSN 0167-9473. doi: https://doi.org/10.1016/j.csda.2005.09.014. URL https://www.sciencedirect.com/science/article/pii/S0167947305002513.

Thomas W. Keelin. The metalog distributions. *Decision Analysis*, 13(4):243–277, 2016. doi: 10.1287/deca.2016.0338. URL https://doi.org/10.1287/deca.2016.0338.

Cheng Peng, Yizhou Li, and Stan Uryasev. Mixture quantiles calibrated with constrained linear regression. 2022.

Wayne S DeSarbo and William L Cron. A maximum likelihood methodology for clusterwise linear regression. *Journal of classification*, 5(2):249–282, 1988.

Zoubin Ghahramani, Geoffrey E Hinton, et al. The em algorithm for mixtures of factor analyzers. Technical report, Technical Report CRG-TR-96-1, University of Toronto, 1996.

Mattias Villani, Robert Kohn, and Paolo Giordani. Regression density estimation using smooth adaptive gaussian mixtures. *Journal of Econometrics*, 153(2):155–173, 2009. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2009.05.004. URL https://www.sciencedirect.com/science/article/pii/S0304407609001419.

Seniha Esen Yuksel, Joseph N. Wilson, and Paul D. Gader. Twenty years of mixture of experts. *IEEE Transactions on Neural Networks and Learning Systems*, 23(8):1177–1193, 2012. doi: 10.1109/TNNLS.2012.2200299.

X. He, P. Ng, and S. Portnoy. Bivariate quantile smoothing splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(3):537–550, 1998.

Roger Koenker. Additive models for quantile regression: Model selection and confidence bandaids. *Brazilian Journal of Probability and Statistics*, 25(3):239 – 262, 2011. doi: 10.1214/10-BJPS131. URL https://doi.org/10.1214/10-BJPS131.

Roger Koenker, Pin Ng, and Stephen Portnoy. Quantile smoothing splines. *Biometrika*, 81(4):673–680, 1994.

Howard D Bondell, Brian J Reich, and Huixia Wang. Noncrossing quantile regression curve estimation. *Biometrika*, 97(4):825–838, 2010.

Victor Chernozhukov, Iván Fernández-Val, and Alfred Galichon. Quantile and probability curves without crossing. *Econometrica*, 78(3):1093–1125, 2010. doi: https://doi.org/10.3982/ECTA7880. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA7880.

Roger Koenker. A note on l-estimates for linear models. *Statistics & Probability Letters*, 2(6):323–325, 1984. ISSN 0167-7152. doi: https://doi.org/10.1016/0167-7152(84)90040-3. URL https://www.sciencedirect.com/science/article/pii/0167715284900403.

Hui Zou and Ming Yuan. Composite quantile regression and the oracle model selection theory. *The Annals of Statistics*, 36(3):1108 – 1126, 2008. doi: 10.1214/07-AOS507. URL https://doi.org/10.1214/07-AOS507.

Gianluca Sottile and Paolo Frumento. Robust estimation and regression with parametric quantile functions. *Computational Statistics & Data Analysis*, 171:107471, 2022. ISSN 0167-9473. doi: https://doi.org/10.1016/j.csda.2022.107471. URL https://www.sciencedirect.com/science/article/pii/S0167947322000512.

Joshua Angrist, Victor Chernozhukov, and Iván Fernández-Val. Quantile regression under misspecification, with an application to the u.s. wage structure. *Econometrica*, 74(2):539–563, 2006.

Brian J. Reich, Montserrat Fuentes, and David B. Dunson. Bayesian spatial quantile regression. *Journal of the American Statistical Association*, 106(493):6–20, 2011. doi: 10.1198/jasa.2010.ap09237. URL https://doi.org/10.1198/jasa.2010.ap09237. PMID: 23459794.

Heng Lian, Jie Meng, and Zengyan Fan. Simultaneous estimation of linear conditional quantiles with penalized splines. *Journal of Multivariate Analysis*, 141:1–21, 2015. ISSN 0047-259X. doi: https://doi.org/10.1016/j.jmva.2015.06.010. URL https://www.sciencedirect.com/science/article/pii/S0047259X15001554.

Paolo Frumento and Matteo Bottai. Parametric modeling of quantile regression coefficient functions. *Biometrics*, 72(1):74–84, 2016. doi: https://doi.org/10.1111/biom.12410. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/biom.12410.

Yuan Yuan, Nan Chen, and Shiyu Zhou. Modeling regression quantile process using monotone b-splines. *Technometrics*, 59(3):338–350, 2017. doi: 10.1080/00401706.2016.1211553. URL https://doi.org/10.1080/00401706.2016.1211553.

James E. Matheson and Robert L. Winkler. Scoring rules for continuous probability distributions. *Management Science*, 22(10):1087–1096, 1976. doi: 10.1287/mnsc.22.10.1087. URL https://doi.org/10.1287/mnsc.22.10.1087.

Francesco Laio and Stefania Tamea. Verification tools for probabilistic forecasts of continuous hydrological variables. *Hydrology and Earth System Sciences*, 11:1267–1277, 2006.

Torsten Hothorn, Thomas Kneib, and Peter Bühlmann. Conditional transformation models. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 76(1):3–27, 2014. ISSN 13697412, 14679868. URL http://www.jstor.org/stable/24772743.

Jan Gasthaus, Konstantinos Benidis, Yuyang Wang, Syama Sundar Rangapuram, David Salinas, Valentin Flunkert, and Tim Januschowski. Probabilistic forecasting with spline quantile function rnns. In Kamalika Chaudhuri and Masashi Sugiyama, editors, *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pages 1901–1910. PMLR, 16–18 Apr 2019. URL https://proceedings.mlr.press/v89/gasthaus19a.html.

Adèle Gouttes, Kashif Rasul, Mateusz Koren, Johannes Stephan, and Tofigh Naghibi. Probabilistic time series forecasting with implicit quantile networks, 2021. URL https://arxiv.org/abs/2107.03743.

Jonathan Berrisch and Florian Ziel. Crps learning. *Journal of Econometrics*, 2021. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2021.11.008. URL https://www.sciencedirect.com/science/article/pii/S0304407621002724.

Qian Zhang, Anuran Makur, and Kamyar Azizzadenesheli. Functional linear regression of cdfs, 2022. URL https://arxiv.org/abs/2205.14545.

Rasool Fakoor, Taesup Kim, Jonas Mueller, Alexander J. Smola, and Ryan J. Tibshirani. Flexible model aggregation for quantile regression, 2021. URL https://arxiv.org/abs/2103.00083.

Victor Chernozhukov and Len Umantsev. Conditional value-at-risk: Aspects of modeling and estimation. *Empirical Economics*, 26(1):271–292, 2001.

Paul H. C. Eilers and Brian D. Marx. Flexible smoothing with B-splines and penalties. *Statistical Science*, 11(2):89 – 121, 1996. doi: 10.1214/ss/1038425655. URL https://doi.org/10.1214/ss/1038425655.

Paul H.C. Eilers and Brian D. Marx. Multivariate calibration with temperature interaction using two-dimensional penalized signal regression. *Chemometrics and Intelligent Laboratory Systems*, 66(2):159–174, 2003. ISSN 0169-7439. doi: https://doi.org/10.1016/S0169-7439(03)00029-7. URL https://www.sciencedirect.com/science/article/pii/S0169743903000297.

J. O. Ramsay. Monotone regression splines in action. *Statistical Science*, 3(4):425–441, 1988. ISSN 08834237. URL http://www.jstor.org/stable/2245395.

H. B. Curry and I. J. Schoenberg. *On Pólya Frequency Functions IV: The Fundamental Spline Functions and their Limits*, pages 347–383. Birkhäuser Boston, Boston, MA, 1988. ISBN 978-1-4899-0433-1. doi: 10.1007/978-1-4899-0433-1_17. URL https://doi.org/10.1007/978-1-4899-0433-117.

Robert F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007, 1982. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1912773.

C. Gourieroux and J. Jasiak. Dynamic quantile models. *Journal of Econometrics*, 147(1):198–205, 2008. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2008.09.028. URL https://www.sciencedirect.com/science/article/pii/S0304407608001358. Econometric modelling in finance and risk management: An overview.

Roger Koenker. *Quantile Regression*. Econometric Society Monographs. Cambridge University Press, 2005. doi: 10.1017/CBO9780511754098.

Stephen Portnoy. Asymptotic behavior of the number of regression quantile breakpoints. *SIAM Journal on Scientific and Statistical Computing*, 12(4):867–883, 1991. doi: 10.1137/0912047. URL https://doi.org/10.1137/0912047.

Whitney K. Newey and Daniel McFadden. Chapter 36 large sample estimation and hypothesis testing. volume 4 of *Handbook of Econometrics*, pages 2111–2245. Elsevier, 1994. doi: https://doi.org/10.1016/S1573-4412(05)80005-4. URL https://www.sciencedirect.com/science/article/pii/S1573441205800054.

Tamara G. Kolda and Brett W. Bader. Tensor decompositions and applications. *SIAM Review*, 51(3):455–500, 2009. doi: 10.1137/07070111X. URL https://doi.org/10.1137/07070111X.

Stephan Rabanser, Oleksandr Shchur, and Stephan Günnemann. Introduction to tensor decompositions and their applications in machine learning, 2017. URL https://arxiv.org/abs/1711.10781.

Matthew Norton, Valentyn Khokhlov, and Stan Uryasev. Calculating cvar and bpoe for common probability distributions with application to portfolio optimization and density estimation. *Annals of Operations Research*, 299(1):1281–1315, 2021.

Michael Zabarankin, Stan Uryasev, et al. *Statistical decision problems*. Springer, 2016.

Tatiana Benaglia, Didier Chauveau, David R. Hunter, and Derek S. Young. mixtools: An r package for analyzing mixture models. *Journal of Statistical Software*, 32(6):1–29, 2009. doi: 10.18637/jss.v032.i06. URL https://www.jstatsoft.org/index.php/jss/article/view/v032i06.

R. A. Rigby and D. M. Stasinopoulos. Generalized additive models for location, scale and shape,(with discussion). *Applied Statistics*, 54:507–554, 2005.

Appendix

A Related Work

A.1 Mixture Quantile Models

Mixture quantile models have been developed where the quantile function $Q_Y(p)$ is a combination of some basis functions $Q_i(p)$ functions

$$Q_Y(p) = \sum_{i=1}^{I} a_i Q_i(p) \ . \tag{41}$$

A.2 Mixture Regression

Mixture regression shares a similar idea to our approach in that the conditional distribution is modeled as a linear combination of basis functions. The conditional density function of the widely applied Gaussian mixture regression is modeled by a mixture (weighed sum) of Gaussian densities $f_{\mathcal{N}(\mu,\sigma)}(y)$, where the weight $w(\boldsymbol{x})$, mean $\mu(\boldsymbol{x})$ and variance $\sigma(\boldsymbol{x})$ of each component can be modeled as functions of factors $\boldsymbol{x}$

$$f(y|\boldsymbol{x}) = w_i(\boldsymbol{x}) f_{\mathcal{N}(\mu(\boldsymbol{x}),\sigma(\boldsymbol{x}))}(y) \ . \tag{42}$$

A.3 Quantile Regression Process Model

Quantile regression process refers to the regression coefficient as a function of the confidence level of quantile regression. Angrist et al. (2006) shows that the rescaled quantile regression process converges to a zero-mean Gaussian process. There are several pieces of independent but closely related research that model the quantile regression process with basis functions. The model formulation is as follows

$$G(p, \boldsymbol{x}, \boldsymbol{a}) = \sum_{i=1}^{I} \sum_{j=1}^{J} a_{ij} Q_{ij}(p) x_i \ . \tag{43}$$

For $Q_{ij}(p)$, Reich et al. (2011) uses Bernstein basis polynomials; Lian et al. (2015) uses P-spline basis and proves its convergence to Gaussian process; Frumento and Bottai (2016) uses common parameterized functions as basis; Yuan et al. (2017) uses monotone B-spline. By expanding the polynomial terms $B_j(\boldsymbol{x})$ and rearranging the model formulation (2), we see that our model coincides with the quantile regression process models if we view the individual terms of polynomials as factors.

A.4 Conditional Transformation Model

Conditional transformation model (Hothorn et al., 2014) finds the optimal conditional transformation of a fixed quantile function such that it best fits the data. Conditional transformation model is defined by

$$F_Y(y, \boldsymbol{x}) = Q^{-1}(h(y, \boldsymbol{x}, \boldsymbol{a})) \ , \tag{44}$$

where $F_Y(y, \boldsymbol{x})$ is the CDF of response variable y conditioned on $\boldsymbol{x}$, $h(y, \boldsymbol{x}, \boldsymbol{a})$ is the transformation conditioned on $\boldsymbol{x}$, $Q(p)$ is a fixed quantile function. In conditional transformation model, $h(y, \boldsymbol{x}, \boldsymbol{a})$ is modeled by splines.

Take the inverse on both sides of (44) ,

$$G(p, \boldsymbol{x}, \boldsymbol{a}) = h^{-1}(Q(p), \boldsymbol{x}, \boldsymbol{a}) \ . \tag{45}$$

$h^{-1}(y, \boldsymbol{x}, \boldsymbol{a})$ transforms a fixed quantile function to a new one conditioned on factors. $G(p, \boldsymbol{x}, \boldsymbol{a})$ is not linear with respect to the unknown parameters $\boldsymbol{a}$. The optimization problem in their paper is not a convex program. Alternatively, we can model h^{-1} by spline function instead of h. This immediately makes the calibration method in Section 4 applicable.

A.5 Quantile Model Aggregation

Quantile model aggregation concerns aggregating multiple models of conditional quantile function to enhance the overall performance. Consider i quantile models $g_i(p, \boldsymbol{x}, \boldsymbol{a}_i)$. The aggregated model is defined by

$$G(p, \boldsymbol{x}, \boldsymbol{a}) = \sum_{i=1}^{I} w_i(p, \boldsymbol{x}, \boldsymbol{a}_i) g_i(p, \boldsymbol{x}) \ , \tag{46}$$

where $\sum_{i=1}^{I} w_i(p, \boldsymbol{x}, \boldsymbol{a}_i) = 1$. The weight is a function of both $\boldsymbol{x}$ and p, meaning that we expect the individual models $g_i(p, \boldsymbol{x}, \boldsymbol{a}_i)$ to have different performances not only across different factor values but also different quantile levels. The aggregation strategy is supposed to give us the optimal ensembled model.

For $w_i(p, \boldsymbol{x}, \boldsymbol{a}_i)$, CRPS learning in Berrisch and Ziel (2021) discards the variable $\boldsymbol{x}$ and uses B-splines as $w_i(p)$, while the medium locally weighted ensembling strategy in Fakoor et al. (2021) uses neural networks. The softmax transformation are used to ensure that the weights sum up to one. However, such constraint seems unnecessary, since a nonnegatively weighted sum of quantile function is a quantile function.

Recall our model (1)

$$G(p, \boldsymbol{x}, \boldsymbol{a}) = \sum_{i=0}^{I} f_i(\boldsymbol{x}, \boldsymbol{a}_i) Q_i(p) \ , \tag{47}$$

where $Q_0(p) = 1$. We can immediately adapt it for quantile model aggregation by modeling each weight function $w_i(p, \boldsymbol{x}, \boldsymbol{a}_i)$ in (46) as

$$w_i(p, \boldsymbol{x}, \boldsymbol{a}_i) = \sum_{j=0}^{J} f_{ij}(\boldsymbol{x}, \boldsymbol{a}_{ij}) Q_j(p) \ . \tag{48}$$

The calibration method for model aggregation remains exactly the same as our model.