HHS Public Access

Author manuscript

Biometrics. Author manuscript; available in PMC 2019 July 01.

Published in final edited form as:

Biometrics. 2015 June ; 71(2): 508–519. doi:10.1111/biom.12294.

Multilevel Quantile Function Modeling with Application to Birth Outcomes

Luke B. Smith[1,*], **Brian J. Reich**[1], **Amy H. Herring**[2], **Peter H. Langlois**[3], and **Montserrat Fuentes**[1]

[1]Department of Statistics, North Carolina State University, Raleigh, North Carolina, 27695-8203, U.S.A.

[2]Department of Biostatistics and Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, 27599-7420, U.S.A.

[3]Texas Department of State Health Services, Austin, Texas 78714-9347, U.S.A.

SUMMARY

Infants born preterm or small for gestational age have elevated rates of morbidity and mortality. Using birth certificate records in Texas from 2002–2004 and Environmental Protection Agency air pollution estimates, we relate the quantile functions of birth weight and gestational age to ozone exposure and multiple predictors, including parental age, race, and education level. We introduce a semi-parametric Bayesian quantile approach that models the full quantile function rather than just a few quantile levels. Our multilevel quantile function model establishes relationships between birth weight and the predictors separately for each week of gestational age and between gestational age and the predictors separately across Texas Public Health Regions. We permit these relationships to vary nonlinearly across gestational age, spatial domain and quantile level and we unite them in a hierarchical model via a basis expansion on the regression coefficients that preserves interpretability. Very low birth weight is a primary concern, so we leverage extreme value theory to supplement our model in the tail of the distribution. Gestational ages are recorded in completed weeks of gestation (integer-valued), so we present methodology for modeling quantile functions of discrete response data. In a simulation study we show that pooling information across gestational age and quantile level substantially reduces MSE of predictor effects. We find that ozone is negatively associated with the lower tail of gestational age in south Texas and across the distribution of birth weight for high gestational ages. Our methods are available in the R package **BSquare**.

Keywords

Birth weight; Discrete; Extremes; Gestational Age; Graphics processing units; Ozone; Quantile

[*] luke_smith@ncsu.edu.

Supplementary Materials

Web Appendices and Figures referenced in Sections 2.5, 3, 4.1, 4.2 and 4.3 are available with this paper at the Biometrics website on Wiley Online Library, as is the code for the simulation study and birth outcomes analysis.

1. Introduction

Infants born preterm (gestational period less than 37 weeks) or small for gestational age (below the 10th percentile of birth weight after controlling for gestational age) have elevated rates of morbidity and mortality (Honein et al., 2009; Pulver et al., 2009; Garite et al., 2004). Reasons for these associations include poorly functioning organs, reduced metabolism, insulin resistance, and increased susceptibility to adverse environmental events later in life (Barker, 2006). Infants who are both preterm and small for gestational age (SGA) are at higher mortality risk than infants with either condition singly (Katz et al., 2013). Narchi et al. (2010) found that adjusting the conditional distribution of birth weight for biological variables better identified at-risk infants.

Our first scientific objective is to better define the conditional distributions of gestational age and birth weight by incorporating personal characteristics from Texan birth certificate records and environmental factors. In a paper with similar aims, Gardosi et al. (1995) used stepwise regression to define the conditional percentiles. We want to understand the relationship between the predictors and the tails of these variables, so we model the conditional quantile functions of the birth outcomes. In a literature review Šrám et al. (2005) ˇ argued that the relationships between air pollution and gestational age and intrauterine growth warrant further analysis. Our second scientific objective is to investigate the effect of maternal exposure to tropospheric ozone, one of the criteria pollutants regulated under the Environmental Protection Agency's Clean Air Act, on SGA and preterm birth (PTB).

Classical frequentist (Koenker and Bassett Jr, 1978; Koenker, 2005) and Bayesian (Yu and Moyeed, 2001) quantile regression models a conditional quantile rather than the conditional mean as a function of predictors. This enables inference of noncentral parts of the distribution, makes fewer assumptions, and is more robust to outliers than mean regression. With these approaches fits at multiple levels can produce "crossing quantiles," where for some values of the predictors the quantile function is decreasing in quantile level. Modeling multiple quantile levels through constraints on the coefficients ensures monotonicity of the quantile function, as in Bondell et al. (2010) and references therein.

The aforementioned approaches model a finite number of quantile levels and do not share information across quantile level. In applications where we expect inference at proximate quantile levels to be similar, it is useful to encourage communication along the distribution. Specifying the full quantile function, which entails separate parameter effects at an uncountable number of quantile levels, fosters this all-encompassing approach. Recent examples of quantile function modeling include Reich et al. (2011), who investigated the effects of temperature on tropospheric ozone using Bernstein polynomials, and Tokdar and Kadane (2011), who analyzed birth weights using stochastic integrals. Reich and Smith (2013) extended quantile function methodology to censored data.

We face three methodological hurdles in our application. PTB and low birth weight are closely related, but distinct, concerns. Researchers prefer to define SGA infants to isolate effects on birth weight from those on gestational age, so it is important to allow the relationship between birth weight and the predictors to vary by gestational age. While

multilevel regression models are well-suited for jointly modeling a collection of distributions, standard hierarchical models assume the predictors affect only the conditional mean of the response. Second, considerable interest lies in the tails (particularly in very premature, SGA, or large-for-gestational age births), so it is important to enable the tails of these distributions to be affected differentially by the predictors relative to the center. Estimation of parameter effects at very low or very high quantiles is generally the purview of extreme value analysis. Multiple conditional extremal methods exist in the literature. Wang and Tsai (2009) modeled the tail index, which determines the thickness of the tails, through a linear log link function of the parameters. Wang et al. (2012) quantile regressed in the shallow tails and extrapolated the results into the deep tails for thickly-tailed data. Our application requires inference along the distribution, so we follow the approaches of Zhou et al. (2012) and Reich et al. (2011), who modeled the middle of the distribution semiparametrically and fit a parametric form above a threshold. In these applications either zero (Zhou et al., 2012) or one (Reich et al., 2011) covariate affected the distribution above the threshold. Our final methodological challenge is modeling a continuous response that has been discretized. The gestational age measurements take values of {25,26, ...,42}. Dichotomizing the response by PTB restricts inference to the cutpoint between 36 and 37 weeks. Previous modeling of quantile functions for discretized data have either jittered the response (Machado and Silva, 2005; Chen and Lazar, 2010) or binned the observations, and then kernel-smoothed the bins (De Gooijer and Yuan, 2011).

The primary contribution of this paper is to introduce a class of multilevel quantile function models that overcomes these methodological challenges. The distribution of birth weight changes smoothly across gestational age, as shown in Figure 1.

We exploit this smoothness by jointly modeling birth weight as a dependent collection of distributions ordered by gestational age. The horizontal lines in Figure 1 represent the thresholds for low birth weight (LBW), defined as 2500 grams, and very low birth weight (VLBW), defined as less than 1500 grams (Rogers and Dunlop, 2006). Most infants born at 25 weeks of gestational age are classified as VLBW, while almost no infants born at 39 weeks and greater are VLBW, so it is imperative to control for gestational age when examining fetal-restricted growth.

Our multilevel approach obviates the choice between the high flexibility and low power associated with separate fits across gestational ages and the high power and inflexibility derived from one fit for all gestational ages. We illustrate another example of our multilevel class by spatially correlating separate distributions of gestational age for each of the eleven Texas Public Health Regions, which are shown in Figure 2.

In both cases we cohere the individual models via Gaussian process priors on the regression coefficients. This class fits separate regression parameters for different gestational ages/ spatial regions, quantile levels and predictors, creating a rich environment for parameter estimation.

Our second methodological contribution is a synthesis of quantile function modeling and conditional extreme value analysis. We adopt a semiparametric approach that models the

middle of the distribution as a linear combination of basis functions and parametrically fits the tails of the distribution via a smooth transition across the semiparametric/parametric threshold. This enhances tail flexibility, ensuring inference on the quantile levels of interest is not perturbed by a few outliers in the tails.

Our final methodological contribution is to extend the quantile function approach to accommodate continuous outcomes that have been discretized or interval-censored into groups. Gestational age from the vital records were recorded by the physician in weeks, not days. Rather than jittering or binning the response, we model the discretized response as a censored realization of a latent continuous process. By modeling the full quantile function we can estimate predictor effects in a computationally stable manner.

The paper is structured as follows. In Section 2 we describe the hierarchical quantile model. In Section 3 we present the results of a simulation study that explores our three methodological innovations. In Section 4 we analyze the birth outcomes and we conclude in Section 5.

2. Quantile Function Modeling

Denote Y_i as the response (either birth weight or gestational age as described below) and $\mathbf{X}_i = (X_{i1}, ..., X_{iP})$ as the vector of length P containing personal characteristics, environmental variables and intercept of infant i. We can define the model for Y_i by the conditional distribution function $F(y|\mathbf{X}_i) = P(Y_i \leqslant y|\mathbf{X}_i)$ or the density $f(y|\mathbf{X}_i) = \frac{\mathrm{d}}{\mathrm{d}y}F(y|\mathbf{X}_i)$. Alternatively we can specify the conditional quantile function $Q(\tau|\mathbf{X}_i)$ where $Q(\tau|\mathbf{X}_i) = F^{-1}(\tau|\mathbf{X}_i) = \inf\{y : F(y|\mathbf{X}_i) \geqslant \tau\}$. The value $\tau \in (0,1)$ is known as the quantile level and the quantile function is nondecreasing in the quantile level. Birth weight has been previously modeled in the quantile regression (Koenker and Hallock, 2001; Burgette and Reiter, 2012; Tokdar and Kadane, 2011), density estimation (Dunson et al., 2008) and spatial (Kammann and Wand, 2003) settings. In this paper we borrow from all of these domains.

We begin with the class of bounded distributions, where there exist real numbers a and b such that for all $\mathbf{X}_i$, $a < Q(0|\mathbf{X}_i) < Q(1|\mathbf{X}_i) < b$. For now we assume the density of Y_i is absolutely continuous with respect to Lebesgue measure, implying a unique quantile function that is increasing in quantile level. We describe extensions to cases of unbounded distributions and discrete response in Sections 2.3 and 2.4 respectively.

2.1 Individual Quantile Function

In this section we introduce our semiparametric quantile regression model. The most flexible method would allow the predictors to nonlinearly affect the quantile function. This approach is promising for prediction, but the nonlinearity of the predictor effects makes inference challenging, so we model the parameter effects for each predictor to be linear at each quantile level.

We model the projection of the quantile function onto the space of cubic integrated M-splines, known as I-splines of degree 3 (Ramsay, 1988). Let $t = \{t_0, ..., t_K\}$ be an ordered sequence of knots whose minimum value is $t_0 = 0$ and maximum value is $t_K = 1$. In the k^{th}

interval between knots t_k and t_{k+1} I-splines of degree 3 are locally cubic polynomials, and at the knots I-splines are continuous and first-order differentiable. Our quantile function is

$$Q(\tau|\mathbf{X}_i) = \sum_{j=1}^{P} X_{ij}\beta_j(\tau) = \sum_{j=1}^{P} X_{ij} \sum_{m=1}^{M} I_m(\tau)\theta_{mj} \quad (1)$$

where $I_1(\tau) \equiv 1$, $I_m(\tau)$ is the m$^{\text{th}}$ I-spline and θ_{mj} are the regression parameters. As the number of knots increases the space of polynomial splines converges to the space of continuous functions (Schumaker, 1981). The space of cubic monotonic splines converges to the space of continuous monotonic functions almost as quickly as unconstrained cubic splines (DeVore, 1977). A one-unit increase in X_{ij} is associated with a $\beta_j(\tau)$ increase in the τ^{th} quantile of the response. Mean regression is a special case of this model where $\theta_{mj} \equiv 0$ for $j > 1$ and $m > 1$. In this case the effect of the j^{th} predictor is $\beta_j(\tau) \equiv \theta_{j1}$ for all τ and the residual distribution is determined by the intercept function $\beta_1(\tau)$.

The quantile function is increasing in τ if the derivative of the quantile function with respect to τ is positive. This derivative $q(\tau|\mathbf{X}_i) = \frac{\mathrm{d}}{\mathrm{d}\tau}Q(\tau|\mathbf{X}_i)$ is known as the sparsity function (Tukey, 1965; Parzen, 1979) and is the reciprocal of the density for any valid differentiable quantile function (note that $F(Q(\tau|\mathbf{X}_i)|\mathbf{X}_i) = \tau$ and differentiate both sides with respect to τ.) Therefore, we can start with any valid quantile function and find its likelihood, as in Tokdar and Kadane (2011). In our model the sparsity function is

$$q(\tau|\mathbf{X}_i) = \sum_{m=1}^{M} \sum_{j=1}^{P} X_{ij}B_m(\tau)\theta_{mj} \quad (2)$$

where $B_m(\tau) = \frac{\mathrm{d}}{\mathrm{d}\tau}I_m(\tau)$. This produces the likelihood

$$L(\theta, \mathbf{Y}) = \prod_{i=1}^{N} q(U_i|\mathbf{X}_i)^{-1}$$

where $U_i = F(Y_i|\mathbf{X}_i)$.

Similarly to Reich and Smith (2013), we map all predictors into [−1,1] and require $\theta_{m1} > \sum_{j=2}^{P}|\theta_{mj}|$ for m > 1 to ensure a valid quantile process. Let $\tau \in (0,1)$ be arbitrary. Let $X_{ij} = 1$ if the j^{th} predictor has a negative effect at τ and $X_{ij} = -1$ otherwise. This is the "worst case" combination of the predictors that minimizes the sparsity function. Then the sparsity function $q(\tau|\mathbf{X}_i) = \sum_{m=1}^{M} B_m(\tau)\sum_{j=1}^{P} X_{ij}\theta_{mj} \geqslant \sum_{m=1}^{M} B_m(\tau)(\theta_{m1} - \sum_{j=2}^{P}|\theta_{mj}|) > 0$ because $\theta_{m1} > \sum_{j=2}^{P}|\theta_{mj}|$. The derivative of the quantile function is positive for all τ and therefore is increasing in τ. Across basis function and predictor we model latent coefficients

$\theta^{\star}_{mj} \overset{\text{ind}}{\sim} \text{N}(\mu_{mj}, \gamma^2_{mj})$. We let $\theta_{mj} = \theta^{\star}_{mj}$ for all j if the monotonicity constraints are satisfied for the m^{th} basis function and set

$$\theta_{mj} = \begin{cases} 0.001 & j = 1 \\ 0 & \text{otherwise} \end{cases}$$

if $\boldsymbol{\theta}^{\star}$ is outside of the constraint space. We did not have monotonicity issues in our application or simulation study. If the monotonicity constraints are violated, a subset of the predictors can be selected to exert effects solely on the location of the response.

2.2 Multilevel Quantile Model

In the birth weight application we individually relate the distribution of birth weight to the predictors for each of the $G = 18$ gestational ages. Let $\beta_{jg}(\tau) = \sum_{m=1}^{M} \theta_{mjg}$ be the effect of covariate j during week g. To spur communication across gestational ages, we jointly estimate these effects by using a multivariate normal prior distribution.

The quantile function at one gestational age in our global fit is of the form

$$Q(\tau | \mathbf{X}_i, GA = g) = \sum_{j=1}^{P} X_{ij} \beta_{jg}(\tau) = \sum_{j=1}^{P} X_{ij} \sum_{m=1}^{M} I_m(\tau) \theta_{mjg}. \quad (3)$$

We model these parameters collectively to gain power by borrowing information across gestational age. Denote $\boldsymbol{\theta}^{\star}_{mj.}$ as the vector of length G of regression coefficients corresponding to basis function m and predictor j. We assign $\boldsymbol{\theta}^{\star}_{mj.} \sim \text{MVN}(\mu_{mj}, \gamma^{-1}_{mj} \boldsymbol{\Sigma}_{mj})$, with common mean μ_{mj}, precision γ_{mj}, and correlation matrix $\boldsymbol{\Sigma}_{mj}$. Similar to Section 2.1 we ensure the monotonocity constraints are satisfied at each gestational age.

This lower-rank representation of the individual fits reduces the variance of the regression coefficients. The degree of shrinkage varies across basis function and predictor. This enables varying degrees of similarity in regression effects by distribution location (e.g. lower tail vs. middle) and by covariate.

The correlation $\boldsymbol{\Sigma}_{mj}$ is used to capture patterns in the basis functions after shrinking them to a common mean. In our applications $\boldsymbol{\Sigma}_{mj}$ is used to smooth gestational age quantile functions over spatial health regions and birth weight quantile functions over gestational age. For gestational age we are interested in examining the distribution regionally. To capture spatial dependence in the quantile function we fit an exponential spatial correlation matrix, where $\boldsymbol{\Sigma}_{mj}[u, v] = \exp\{-d(u, v)/\phi_{mj}\}$, $d(u, v)$ is the distance between the centroids of public health regions u and v and ϕ_{mj} is the range parameter. We assign ϕ_{mj} a uniform prior with minimum 0 and maximum equal to half the maximum distance between public health region centroids. This approach enables us to examine large scale spatial patterns in the relationship between gestational age and the predictors. For birth weight we anticipate dependence across gestational age, so we impose an autoregressive order 1 correlation matrix of the form

$\mathbf{\Sigma}_{mj}[u, v] = \rho_{mj}^{|u-v|}$ with correlation parameter ρ_{mj}. We give ρ_{mj} a uniform prior on the unit interval.

2.3 Tail Modeling

For extreme quantile levels, we employ a hybrid approach of quantile regression and extremal analysis. While there are lower bounds for the birth outcomes, modeling the quantile function of birth weight using bounded basis functions is restrictive. If any observations fall below $Q(0|\mathbf{X}_i)$ or above $Q(1|\mathbf{X}_i)$ the likelihood is zero, so bounded basis functions result in a discontinuous likelihood with respect to the parameters. Both expert opinion (Wilcox, 2001) and exploratory analysis suggest the density of birth weight has thick tails, which are best modeled using basis functions that decay slowly in the tails.

To increase the flexibility of our distribution in the tails we set thresholds τ_L and τ_H at extreme quantile levels (e.g. 0.01 and 0.99). We model the middle of the distribution semiparametrically and the values beyond the thresholds with the generalized Pareto distribution (GPD) family. The limiting distribution of exceedances over a threshold is the GPD for most distributions (Coles, 2001), providing motivation for our approach. The GPD has a scale parameter σ, which determines the value of the density at the threshold, and a shape parameter ξ, which determines the tail decay rate. The GPD can model bounded distributions ($\xi < 0$), distributions with light tails ($\xi = 0$) and distributions with heavy tails ($\xi > 0$). These three densities are

$$\zeta_Z(z|\sigma,\xi) = \begin{cases} \sigma^{-1}\left(1+\frac{\xi z}{\sigma}\right)^{-1/\xi-1} \mathbb{1}_{0<z<-\sigma/\xi} & \xi<0 \\ \sigma^{-1}\exp\left(\frac{-z}{\sigma}\right)\mathbb{1}_{z>0} & \xi=0 \\ \sigma^{-1}\left(1+\frac{\xi z}{\sigma}\right)^{-1/\xi-1} \mathbb{1}_{z>0} & \xi>0 \end{cases}$$

where $\mathbb{1}$ is the indicator function. In our application $z = Q(\tau_L|\mathbf{X}_i) - Y_i$ in the lower tail and $z = Y_i - Q(\tau_H|\mathbf{X}_i)$ in the upper tail. The $\xi = 0$ case corresponds to an exponential quantile function and the $\xi > 0$ case corresponds to a Pareto quantile function. These are good fits for light-tailed and heavy-tailed data respectively. This methodology was developed in Zhou et al. (2012) for continuous spatial data without predictors. We extend this approach to incorporate covariates and discrete data.

The quantile distribution $Q^\star$ of birth weight with a Pareto tail is

$$Q^\star(\tau|\mathbf{X}_i) = \begin{cases} Q(\tau_L|\mathbf{X}_i) - \frac{\sigma_L}{\xi_L(\mathbf{X}_i)}\left[(\tau/\tau_L)^{-\xi_L(\mathbf{X}_i)} - 1\right] & \tau < \tau_L \\ Q(\tau|\mathbf{X}_i) & \tau_L \leqslant \tau \leqslant \tau_H \\ Q(\tau_U|\mathbf{X}_i) + \frac{\sigma_U}{\xi_H(\mathbf{X}_i)}\left[\left(\frac{1-\tau}{1-\tau_H}\right)^{-\xi_H(\mathbf{X}_i)} - 1\right] & \tau_H < \tau. \end{cases} \quad (4)$$

The scale parameters (σ_L, σ_H) are the density of the Pareto distribution evaluated at the thresholds ($Q(\tau_L|\mathbf{X}_i), Q(\tau_H|\mathbf{X}_i)$), as can be seen by plugging $z = 0$ into the density. Assuming a continuous density at the breakpoints τ_L and τ_H specifies the scale parameters as

$$\sigma_L|\mathbf{X}_i = \tau_L * q(\tau_L|\mathbf{X}_i),$$

$$\sigma_H|\mathbf{X}_i = (1 - \tau_H) * q(\tau_H|\mathbf{X}_i)$$

where $q(.|\mathbf{X}_i)$ is the sparsity function defined in (2). The scaling factors τ_L and $(1 - \tau_H)$ guarantee the full density of birth weight integrates to 1.

In the most general case we let the shape parameter be a function of the predictors. Shape parameters can be difficult to estimate, so we favor the simplifications $\xi_L(\mathbf{X}_i) = \xi_L$ and $\xi_U(\mathbf{X}_i) = \xi_U$. We assign independent priors for the lower and upper tail shape parameters. Allowing the shape parameter to take on positive and negative values, as in a linear link function, causes instability for shape values near zero. For the individual quantile function model in Section 2.1 we assign $\log(\xi_L) \sim N(\mu_L, \gamma_L)$, $\log(\xi_H) \sim N(\mu_H, \gamma_H)$. For the collective quantile function model in Section 2.2 we index the shape parameters by gestational age and assign $\log(\boldsymbol{\xi}_L) \sim N(\mu_L, \gamma_L^2 \boldsymbol{\Sigma}_{\xi L})$, $\log(\boldsymbol{\xi}_H) \sim N(\mu_H, \gamma_H^2 \boldsymbol{\Sigma}_{\xi H})$.

Threshold selection is a key aspect of extreme value analysis. Recall that one focus of the analysis is inference at the first percentile. We could have selected a lower threshold above 0.01 (e.g. $\tau_L = 0.05$) and extrapolated those effects to the first percentile. We have found that our model performs better when thresholds are set at or beyond the quantile levels of interest. If more flexibility is desired for quantile levels inside the threshold (e.g. $\tau = 0.02$), then more knots can be placed at this part of the distribution. The density is monotonically decreasing beyond the threshold for unbounded tails, another consideration for threshold selection.

This meld of semiparametric and parametric approaches has several attractive properties, the first of which is covariate-dependent threshold selection. Permitting the predictor effects to change with quantile level helps properly identify the extreme values and allows the relationship between the predictors and the response to differ in the tail and the heart of the distribution. Second, a larger percentage of the data are used to inform our tails than in canonical extremal analysis, where observations below a threshold are often discarded. The scale parameter is the reciprocal of the conditional density at the threshold and is influenced by values above and below the threshold. In our model moving from the tail to the middle of the distribution attenuates the effect of the observations on the tail parameters. Antecedents of this idea include (Frigessi et al., 2002) and (Behrens et al., 2004), who utilized mixture models where a light-tailed component dominated in the bulk of the distribution and a Pareto-tailed component determined the density in the tail. Finally, our hierarchical framework enables separate tails for the distribution of birth weight for each gestational age. By shrinking the predictor effects independently across basis function the degree of

shrinkage can differ in the tails and the middle. The opportunity for communication is especially important in the tails, where information is often limited.

2.4 Discrete Data

Here we extend quantile function methodology to permit a discrete response. Treating an infant born at 36 weeks as similar to an infant born at 25 weeks yet qualitatively different from an infant born at 37 weeks is undesirable. Instead of dichotomizing by PTB, we model gestational age as interval-censored values of a continuous latent process. For a reported gestational age g_i we model a continuous value $G_i \in [g_i, g_i + 1)$. We find the values U_{1i} and U_{2i} such that $Q(U_{1i}|\mathbf{X}_i) = g_i$ and $Q(U_{2i}|\mathbf{X}_i) = g_i + 1$. Note that U_{1i} and U_{2i} are the conditional CDF evaluated at the endpoints and $P(g_i \leqslant G_i < g_i + 1|\mathbf{X}_i) = U_{2i} - U_{1i}$. This produces a likelihood of $\prod_{i=1}^{N}(U_{2i} - U_{1i})$.

Unlike models that dichotomize by PTB, this approach enables parameter effects to adapt smoothly across the distribution. Similarly modeling the effects at adjacent weeks can help compensate for measurement error in gestational age, which is generally estimated using a combination of maternal self-report of the last menstrual period and clinician judgment based on interpretation of early ultrasound examinations or other factors. Modeling the latent continuous quantile function is most suitable for a continuous response that has been discretized or censored. This holds for gestational age, but may not be appropriate in other applications.

2.5 Computation

The computationally expensive part of our model is finding the solution U_i such that $U_i = F(Y_i|\mathbf{X}_i)$ for each observation in each likelihood evaluation. The quantile function is locally a cubic polynomial, as it is the sum of cubic polymomials, so a closed-form solution exists for $U_i = F(Y_i|\mathbf{X}_i)$. Cubic roots are numerically unstable and the solution is complicated, so we find U_i through Newtonian recursion, where $U_{t+1} = U_t + [Q(U_t|\mathbf{X}_i) - Y_i]/q(U_t|\mathbf{X}_i)$ is computed until $|Q(U_t|\mathbf{X}_i) - Y_i|$ is less than the error ϵ. For our application we chose $\epsilon = 10^{-5}$.

To our knowledge this paper presents the first Bayesian quantile function model that can accommodate large sample sizes and multiple predictors. For moderate sample sizes (e.g. n = 1000) and a small number of predictors our model runs in a few minutes. To analyze the birth data set, which has 565,703 observations, we used graphics processing units (Zhou et al., 2010). The likelihood is embarrassingly parallel, so we ran our model on a graphics processing unit with 400 arithmetic cores. For a likelihood evaluation of 100,000 observations we attained a better than 20-fold improvement in computation time. With fourteen predictors and eleven separate quantile functions our final analysis ran in less than 27 hours. For the simulation study and the birth outcome analyses the posterior was sampled using the Metropolis within Gibbs algorithm, with details described in Web Appendix A.

3. Simulation Study

Our simulation study is designed to answer three questions about our model. First, we compare a global fit of multiple quantile functions shrunk together by a collective prior

(CP), as in (3), to individual fits of the quantile functions with independent priors (IP) that do not communicate, as in (1). To test this we generate data at 5 gestational ages, each with separate covariate effects.

$$\begin{aligned} Y_i \big| g_i &= 15 + Z(g_i) * Q_t(U_i) + X_i * 0.5 * Z(g_i) * Q_t(U_i) \\ U_i &\overset{\text{iid}}{\sim} \text{Unif}(0,1) \\ X_i &\overset{\text{iid}}{\sim} \text{Unif}(-1,1) \\ Z &\sim \text{N}_5(0, \boldsymbol{\Sigma}) \end{aligned}$$

where $\boldsymbol{\Sigma}$ is the covariance matrix of a 5-dimensional first order autoregressive (AR-1) process with correlation 0.5 and unit variance. We add a scalar so the realization from this process has minimum 1 to ensure monotonicity of the quantile function. The base quantile function Q_t is the Student t quantile function with 10 degrees of freedom. We selected Student's t-distribution because our data are moderately thick-tailed and the Student t-distribution enables comparison between Pareto and exponential tails, our second factor of interest. Finally, we are interested in the loss of information due to the discretization of a continuous process. For each combination of prior and tail type we evaluate one fit with continuous response and one fit with response rounded to the nearest integer. We compare these three factors for sample sizes of n = 200 and n = 400 observations for each of the 5 levels. We ran 100 Monte Carlo replications.

To complete the CP model we assign priors $\mu_{mj} \sim$ Gaussian(0,0.001) and $\gamma_{mj} \sim$ Gamma(1,1). For the IP model we assign Gaussian(0,0.001) priors to the parameters for the constant basis functions and Gaussian $(\mu_j, \gamma_j^{-1}\boldsymbol{\Sigma}_j)$ priors for $\theta_j = \{\theta_{2j}, ..., \theta_{Mj}\}$. The correlation matrices $\boldsymbol{\Sigma}_j$ are AR-1 with correlation parameter ρ_j and are designed to capture residual correlation in the regression coefficients for a predictor after shrinking to a common mean. For the IP model we assign Gaussian(0,0.001) priors to the means μ_j and Gamma(1,1) priors to the precisions γ_j. We chose a lower threshold of $\tau_L = 0.01$ and an upper threshold of $\tau_H = 0.99$. We assign (on the log scale) μ_L and μ_H N(−4, 1.5) priors, where 1.5 is the standard deviation. On the log scale, this assigns roughly 95% of the mass to the interval (−7, −1). Exponentiating this interval gives the range (0.001, 0.4), which encompasses light-tailed and heavy-tailed distributions. For all fits we set $\mu_L = \mu_H = -0.4$ and $\gamma_L = \gamma_H = 0.4$. This prior has 95% of its mass between (0.31, 1.47), a wide range of values. We fit 5, 7, and 9 basis functions. For each Monte Carlo iteration we selected the fit that had the best (highest) log pseudo marginal likelihood (Ibrahim et al., 2005) across number of basis functions.

We compare our model regression estimates to the true regression effects $\beta_1(\tau) = 0.5 * Z(g_i) * Q_t(\tau)$ at the quantile levels $\tau = (0.01, 0.05, 0.10, ..., 0.90, 0.95, 0.99)$. We fit individually fit frequentist quantile regression at each of the 5 gestational ages and each quantile level of interest. We also fit a frequentist model with linear effects for **X**, gestational age and an interaction term to enable the classical model to use all the data across gestational age. We show summaries of mean squared error (MSE) and coverage probabilities at the nominal 95% level averaged across week in Figure 3 for the n = 200 case. In Figure 3 the models using discrete data are not visually distinguished from those using continuous data because

the discrete data results were very similar to the continuous data results in the best fits. This implies that there is little to no degradation in performance due to the weekly censoring for our application. For more extreme censoring, where only a few levels of the response are observed, this may not hold.

The first column of Figure 3 shows MSE results. For the individual fits the classical regression and our methods perform similarly in the middle of the distribution, but our method performs better in the tails. Sharing information across quantile level is most important in the tails, where information is sparse.

For the joint fits the classical methods (which utilize gestational age and an interaction term in the regression model) are slightly better in the middle of the distribution, but our method performs much better in the tails. The joint models all have MSE lower than their independent counterparts, indicating that MSE can be substantially reduced by sharing information across gestational age. The gains from sharing information across gestational age are greatest in the tails. The Pareto MSE is lower than the exponential MSE in the tails, as anticipated with heavy-tailed student t-distributed data. Coverage was at or above 90% for our methods. Coverage was above 95% for frequentist methods in the middle of the distribution, but dropped to 80% in the tails. All of these conclusions held for the n = 400 case, which is presented in Web Figures 1–15.

The simulation study illustrates that fusing quantile regression and extremal inference can enhance inference, especially in the tails. However, the Bayesian models are more computationally demanding. Each simulation run took a few minutes for the Bayesian models and only seconds for the frequentist model. Further, frequentist quantile regression (which does not assume parameter effects are linear at all quantile levels) could perform better in other settings. Fits with 5 basis functions were found to lack flexibility and fits with 9 basis functions suffered from overfitting in the tails, as shown in the Web Appendix B. The 7 basis function fit was most commonly preferred by log pseudo marginal likelihood (LPML) for both the CP and IP models. While MSE and coverage are affected by M, LPML does a good job selecting the best model. In practice, we recommend fits across multiple M.

To test our model for distributions with bounded support we generated responses $Y_i|g_i = 20+20*Q_1(U_i)+10*(Z(g_i)+X_i)*Z(g_i)*Q_2(U_i)$ where $Q_1(U_i)$ is the quantile function of a Beta(1,5) random variable and $Q_2(U_i)$ is the quantile function of a Beta(5,1) random variable. Results are presented in the online appendix. Bayesian methods and frequentist methods again had similar MSE, except in the upper tail. The right skewness of the response makes inference in the upper tail more challenging, and the joint frequentist model had substantial trouble fitting the upper tail. Our methods suffered from undercoverage in the tails for the n = 200 case, but achieved near nominal coverage probability for all quantile levels for n = 400.

4. Birth Outcome Analyses

4.1 Data Description and Modeling

The birth data consist of live birth certificate records in Texas from 2002 – 2004. Personal characteristics consisted of infant sex, maternal parity (binary for having given birth previously), maternal age and paternal age (less than 40 or 40 and above), maternal education and paternal education (did not finish high school, finished high school or finished some education after high school), and maternal ethnicity (white non-Hispanic, black non-Hispanic, Hispanic, and other). All personal characteristics were treated as categorical variables, as in Hoffman et al. (2008a) and Hoffman et al. (2008b). Data from pregnancies that were missing key covariates, or ended with spontaneous fetal death or induced termination were discarded. We analyze viable births, those occurring in weeks 25–42 (Morgan et al., 2008), leaving 565,703 live births.

Pollution data come from the Environmental Protection Agency's downscaler model (Berrocal et al., 2010). We mapped each maternal address to the nearest 12 kilometer by 12 kilometer grid cell centroid and summed these values for the first and second trimesters.

We fit several different versions of the hierarchical model. For gestational age we compare the collective model of Section 2.2 to individual fits across PHR, as in Section 2.1. We fit Pareto tails and exponential tails with 5 and 7 basis functions for each combination. With most of the births occuring during weeks 31–42, we did not feel comfortable going above 7 basis functions for only 12 levels of gestational age. Some model runs did not converge (GA response, exponential tails, M = 5). Other model runs returned NA for LPML (BW response, exponential tails for M = 7 and 9), indicating the likelihood was computationally zero for some values in the tail. These model runs are omitted. For both analyses we used the same priors and thresholds as in the simulation study. For the range parameters ϕ_{mj} we implemented Uniform(0,r) priors, where r is half of the maximum distance between Texas PHR centroids. We spatially correlated ξ_L and ξ_H with an exponential correlation structure. For birth weight we compare our collective model to individual fits at each gestational age. Allowing the effects of the predictors on birth weight to vary by region and gestational age would have been computationally expensive for 14 predictors, so the predictor effects on birth weight are constant across region. For birth weight we fit 7, 9, and 11 basis functions. Example trace plots are included in Web Appendix C.

4.2 Gestational Age - Personal Characteristic Results

Table 1 shows that for gestational age the best model fit by LPML was 7 basis functions and Pareto tails (9 basis functions and Pareto tails for birth weight).

The effects of black non-Hispanic mothers relative to white non-Hispanic mothers on gestational age are presented in Figure 4.

Black non-Hispanic maternal ethnicity is associated with a more than two week decrease in gestational age at the first percentile, as shown in Web Appendix D. The effect diminishes for large quantiles and there is no significant difference for upper quantiles. This illustrates the ability of quantile function modeling to caputure different covariate effects on different

aspects of the distribution. Other practically significant predictors included maternal parity and maternal education above high school, which were positively associated with gestational age, and maternal age above 40, which was negatively associated with gestational age in the lower tail. All posterior quantile plots for gestational age are available in Web Appendix D.

4.3 Birth Weight Personal Characteristic Results

Infants born to black non-Hispanic mothers weigh between 20–190 grams less than infants born to white non-Hispanic mothers. The effect is larger for higher gestational ages and change little with quantile level. Infants from black non-Hispanic mothers are at high risk for both PTB and SGA, but the relationship between the birth outcomes and risk of morbidity and mortality is subtle. While PTB is associated with negative health outcomes, previous research has shown that certain groups of PTB black non-Hispanic infants have better survival rates than PTB white non-Hispanic infants (Schieve and Handler, 1996). LBW black non-Hispanic infants also have lower mortality rates than their white non-Hispanic counterparts (Alexander et al., 2003). Given health outcome data, the methods presented in this paper could be used to better understand how the predictors affect morbidity and mortality indirectly through the birth outcomes.

While most of the effects of the predictors on gestational age are not practically significant, almost all of the personal characteristics substantially affect birth weight. The uncertainty in the posterior distribution of the regression parameters is smallest for gestational ages with the largest sample sizes, which occur during weeks 37–40. Uncertainty is generally higher in the upper tail than the lower tail. As can be seen in Figure 1, the distribution of birth weight is right-skewed for small gestational ages but becomes more symmetric as gestational age increases. This may explain why posterior variances for birth weight effects are higher in the upper tail than the lower tail for smaller gestational ages. We present the personal characteristic effects in Web Appendix D.

4.4 Ozone Results

For gestational age we permit the ozone effects to vary by region. Ozone levels are higher in cities, so allowing the effects to change regionally allows ozone effects to differ for urban and rural areas. Also, the distribution of unmeasured potential confounders may vary regionally, so allowing the ozone effects to vary spatially may help adjust for this.

As can be seen in Figure 5, second trimester ozone is negatively associated with gestational age in Texas PHRs 5, 6 and 11, located in east and south Texas.

The largest effect in absolute magnitude was 3–4 days for second trimester ozone at the 40^{th} percentile in Region 11. First trimester ozone had negative effects in PHRs 5 and 11. PHRs 2, 3, 5, and 6 showed slight negative associations in the lower tail with first trimester ozone. Previous research using birth records from Harris county, whose county seat is Houston and is located in PHR 6, found evidence that high levels of first trimester ozone increased the probability of PTB (Warren et al., 2012). While almost all of the ozone effects are strongest in the lower tail, in Public Health Region 11 the strongest effects are located around the 40^{th} percentile for both first and second trimester ozone. In our sample 89% of PHR 11 is Hispanic, compared to 49% for Texas. PHR 11 has many migrant workers, high levels of

poverty and high prevalence of birth defects (Hendricks et al., 1999), so it is unsurprising we find that the nature of the relationship between gestational age and ozone differs here from the rest of Texas.

Figure 6 shows the large reduction in variance that is possible by borrowing information across gestational age. By modeling the full quantile function, the individual fits are smoother and generally have smaller variance than frequentist quantile regression. The Bayesian regression parameters change continuously and slowly across quantile level. In contrast, the frequentist effects appear choppy and have larger standard errors because they do not borrow information across quantile level.

We did not find a statistically significant relationship between first trimester ozone and birth weight. Blood exchange between the mother and the placenta increases at the beginning of the second trimester. The negative association between fetal growth and large levels of ozone in the second trimester could be due to impairment of uteroplacental blood flow, which has been shown to be affected by maternal smoking (Mochizuki et al., 1984).

5. Conclusions

In this paper we have presented a novel class of hierarchical quantile function models that retains the interpretability of quantile regression and some of the malleability of density estimation. Our hierarchical framework permits flexible tail inference and can model discretized continuous response. While our application consists only of interval censored data, it is easily extended to data that are left or right censored or a mixture of continuous and discrete responses. Our discrete quantile model could be a viable alternative to other applications, such as Likert scale response.

We conducted a simulation study that demonstrated substantial reductions in MSE for our approach relative to canonical quantile regression. We found that the lower tails of the distributions of gestational age and birth weight for infants born to African American mothers are much lower than the lower tails for infants born to white non-Hispanic mothers. Useful extension of our methods include variable selection for a large number of predictors, incorporating subject-specific random effects, and nonparametric quantile regression.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors are grateful for support from NSF grant DMS-1107046 and NIH grant SR01ES014843. This work is partially supported through cooperative agreement U01DD000494 between the Center for Disease Control and the Texas Department of State Health Services. We thank the editor, associate editor, and two anonymous referees for thoughtful and constructive comments.

References

Alexander GR, Kogan M, Bader D, Carlo W, Allen M, and Mor J (2003). US birth weight/gestational age-specific neonatal mortality: 1995–1997 rates for whites, hispanics, and blacks. Pediatrics 111, e61–e66. [PubMed: 12509596]

Barker DJ (2006). Adult consequences of fetal growth restriction. Clinical obstetrics and gynecology 49, 270–283. [PubMed: 16721106]

Behrens CN, Lopes HF, and Gamerman D (2004). Bayesian analysis of extreme events with threshold estimation. Statistical Modelling 4, 227–244.

Berrocal VJ, Gelfand AE, and Holland DM (2010). A bivariate space-time downscaler under space and time misalignment. The annals of applied statistics 4, 1942. [PubMed: 21853015]

Bondell HD, Reich BJ, and Wang H (2010). Noncrossing quantile regression curve estimation. Biometrika 97, 825–838. [PubMed: 22822254]

Burgette LF and Reiter JP (2012). Modeling adverse birth outcomes via confirmatory factor quantile regression. Biometrics 68, 92–100. [PubMed: 21689080]

Chen J and Lazar NA (2010). Quantile estimation for discrete data via empirical likelihood. Journal of Nonparametric Statistics 22, 237–255.

Coles S (2001). An Introduction to Statistical Modeling of Extreme Values. Springer, London.

De Gooijer JG and Yuan A (2011). Kernel-smoothed conditional quantiles of correlated bivariate discrete data. Statistica Sinica 21,.

DeVore RA (1977). Monotone approximation by splines. SIAM Journal on Mathematical Analysis 8, 891–905.

Dunson DB, Herring AH, and Siega-Riz AM (2008). Bayesian inference on changes in response densities over predictor clusters. Journal of the American Statistical Association 103, 1508–1517.

Frigessi A, Haug O, and Rue H (2002). A dynamic mixture model for unsupervised tail estimation without threshold selection. Extremes 5, 219–235.

Gardosi J, Mongelli M, Wilcox M, and Chang A (1995). An adjustable fetal weight standard. Ultrasound in Obstetrics & Gynecology 6, 168–174. [PubMed: 8521065]

Garite TJ, Clark R, and Thorp JA (2004). Intrauterine growth restriction increases morbidity and mortality among premature neonates. American journal of obstetrics and gynecology 191, 481–487. [PubMed: 15343225]

Hendricks KA, Simpson JS, and Larsen RD (1999). Neural tube defects along the texas-mexico border, 19931995. American Journal of Epidemiology 149, 1119–1127. [PubMed: 10369506]

Hoffman CS, Mendola P, Savitz DA, Herring AH, Loomis D, Hartmann KE, Singer PC, Weinberg HS, and Olshan AF (2008b). Drinking water disinfection by-product exposure and duration of gestation. Epidemiology 19, 738–746. [PubMed: 18633329]

Hoffman CS, Mendola P, Savitz DA, Herring AH, Loomis D, Hartmann KE, Singer PC, Weinberg HS, and Olshan AF (2008a). Drinking water disinfection by-product exposure and fetal growth. Epidemiology 19, 729–737. [PubMed: 18633330]

Honein MA, Kirby RS, Meyer RE, Xing J, Skerrette NI, Yuskiv N, Marengo L, Petrini JR, Davidoff MJ, Mai CT, et al. (2009). The association between major birth defects and preterm birth. Maternal and child health journal 13, 164–175. [PubMed: 18484173]

Ibrahim JG, Chen M-H, and Sinha D (2005). Bayesian survival analysis. Wiley Online Library.

Kammann EE and Wand MP (2003). Geoadditive models. Journal of the Royal Statistical Society: Series C (Applied Statistics) 52, 1–18.

Katz J, Lee AC, Kozuki N, Lawn JE, Cousens S, Blencowe H, Ezzati M, Bhutta ZA, Marchant T, Willey BA, et al. (2013). Mortality risk in preterm and smallfor-gestational-age infants in low-income and middle-income countries: a pooled country analysis. The Lancet.

Koenker R (2005). Quantile regression. Number 38. Cambridge university press.

Koenker R and Hallock KF (2001). Quantile regression. Journal of Economic Perspectives 15, 143–156.

Koenker RW and Bassett G Jr (1978). Regression quantiles. Econometrica 46, 33–50.

Machado JAF and Silva JS (2005). Quantiles for counts. Journal of the American Statistical Association 100, 1226–1237.

Mochizuki M, Maruo T, Masuko K, Ohtsu T, et al. (1984). Effects of smoking on fetoplacental-maternal system during pregnancy. American journal of obstetrics and gynecology 149, 413–420. [PubMed: 6203408]

Morgan MA, Goldenberg RL, and Schulkin J (2008). Obstetrician-gynecologists' practices regarding preterm birth at the limit of viability. Journal of Maternal-Fetal and Neonatal Medicine 21, 115–121. [PubMed: 18240080]

Narchi H, Skinner A, and Williams B (2010). Small for gestational age neonates-are we missing some by only using standard population growth standards and does it matter? Journal of Maternal-Fetal and Neonatal Medicine 23, 48–54. [PubMed: 19565425]

Parzen E (1979). Nonparametric statistical data modeling. Journal of the American Statistical Association 74, 105–121.

Pulver LS, Guest-Warnick G, Stoddard GJ, Byington CL, and Young PC (2009). Weight for gestational age affects the mortality of late preterm infants. Pediatrics 123, e1072–e1077. [PubMed: 19482740]

Ramsay J (1988). Monotone regression splines in action. Statistical Science pages 425–441.

Reich BJ, Cooley D, Foley KM, Napelenok S, and Shaby BA (2011). Extreme value analysis for evaluating ozone control strategies.

Reich BJ, Fuentes M, and Dunson DB (2011). Bayesian spatial quantile regression. Journal of the American Statistical Association 106,.

Reich BJ and Smith LB (2013). Bayesian quantile regression for censored data. Biometrics.

Rogers JF and Dunlop AL (2006). Air pollution and very low birth weight infants: A target population? Pediatrics 118, 156–164. [PubMed: 16818561]

Schieve LA and Handler A (1996). Preterm delivery and perinatal death among black and white infants in a chicago-area perinatal registry. Obstetrics & Gynecology 88, 356–363. [PubMed: 8752239]

Schumaker S (1981). Spline Functions: Basic Theory. John Wiley & Sons, Inc., United States of America.

Šrám RJ, Binková B, Dejmek J, and Bobak M (2005). Ambient air pollution and pregnancy outcomes: a review of the literature. Environmental health perspectives 113, 375. [PubMed: 15811825]

Tokdar S and Kadane JB (2011). Simultaneous linear quantile regression: A semiparametric bayesian approach. Bayesian Analysis 6, 1–22. [PubMed: 22247752]

Tukey JW (1965). Which part of the sample contains the information? Proceedings of the National Academy of Sciences of the United States of America 53, 127. [PubMed: 16578583]

Wang H and Tsai C-L (2009). Tail index regression. Journal of the American Statistical Association 104, 1233–1240.

Wang HJ, Li D, and He X (2012). Estimation of high conditional quantiles for heavy-tailed distributions. Journal of the American Statistical Association 107, 1453–1464.

Warren J, Fuentes M, Herring A, and Langlois P (2012). Spatial-temporal modeling of the association between air pollution exposure and preterm birth: Identifying critical windows of exposure. Biometrics 68, 1157–1167. [PubMed: 22568640]

Wilcox AJ (2001). On the importance-and the unimportance-of birthweight. International journal of epidemiology 30, 1233–1241. [PubMed: 11821313]

Yu K and Moyeed RA (2001). Bayesian quantile regression. Statistics and Probability Letters 54, 437–447.

Zhou H, Lange K, and Suchard MA (2010). Graphics processing units and high-dimensional optimization. Statistical Science 25,.

Zhou J, Chang HH, and Fuentes M (2012). Estimating the health impact of climate change with calibrated climate model output. Journal of agricultural, biological, and environmental statistics 17, 377–394.

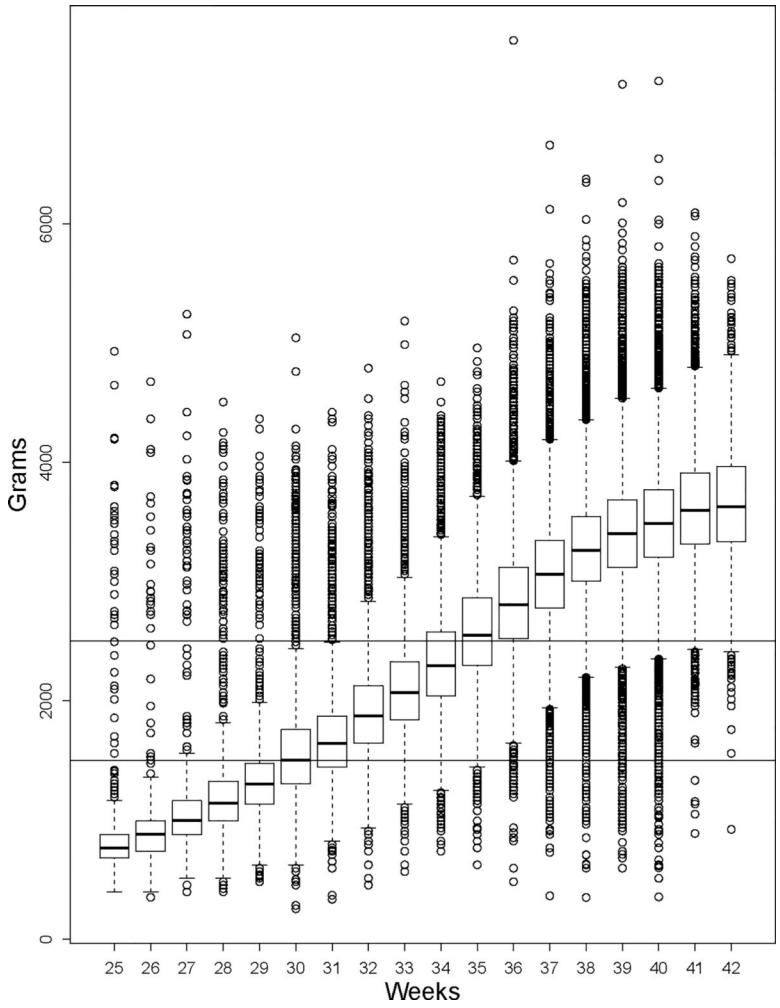

Figure 1.
Boxplots of birth weight by week of gestational age.

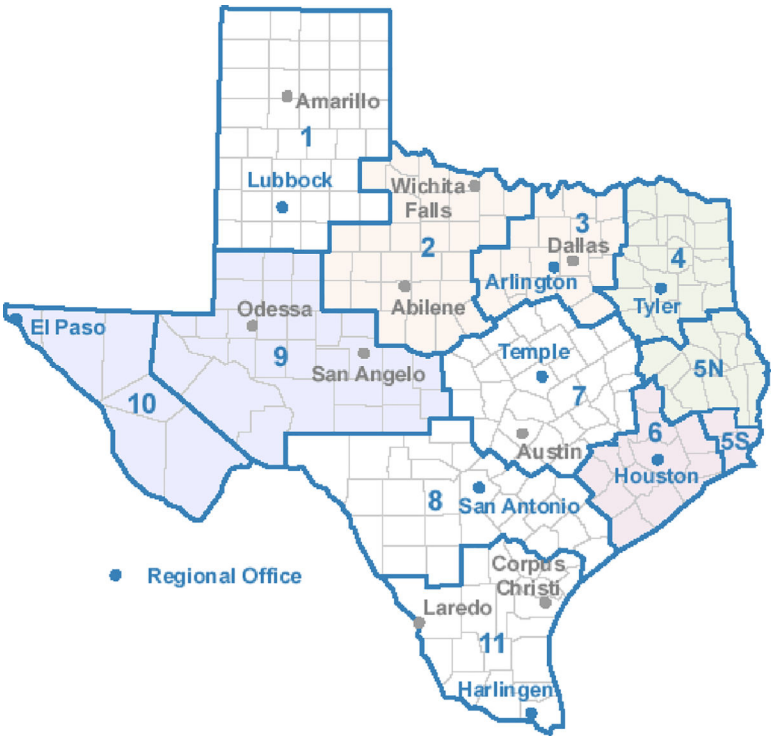

Figure 2.
Texas Public Health Regions.

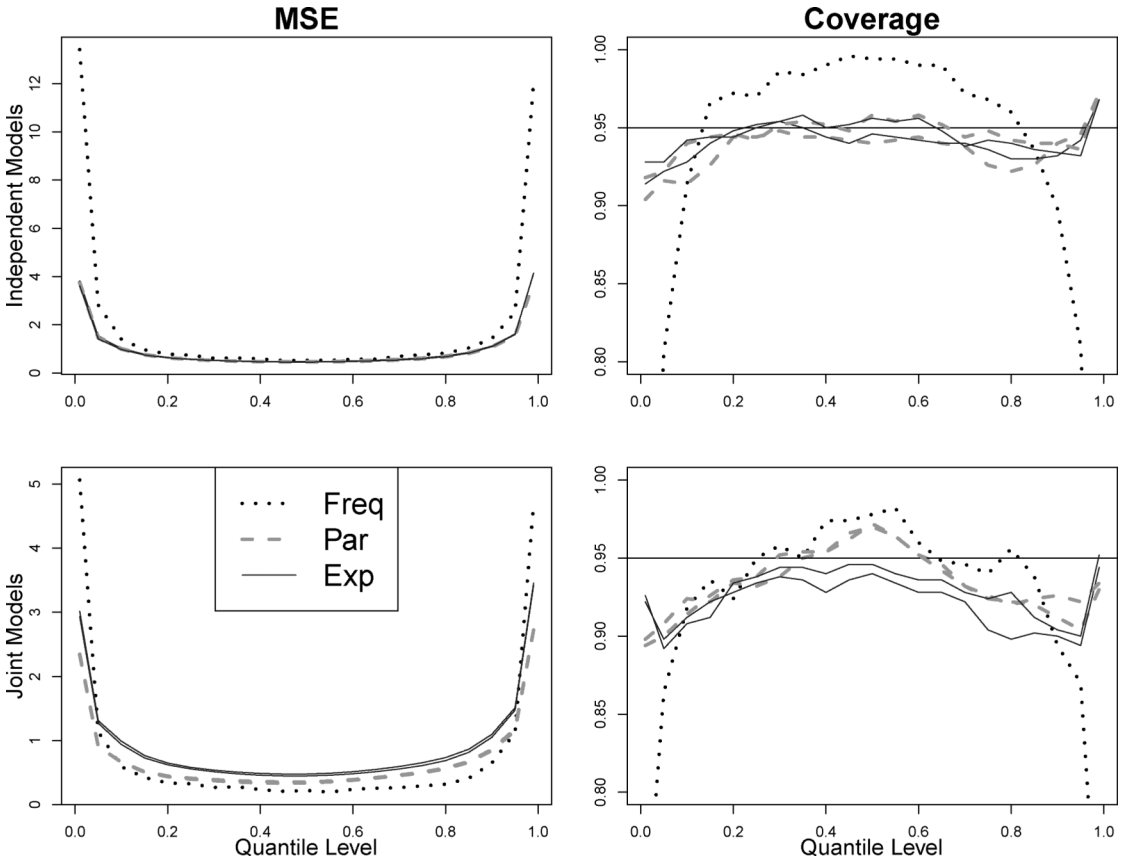

Figure 3.

MSE and coverage probabilities for t-distributed response where n = 200 at each gestational age and M was selected by log pseudo marginal likelihood. Results above are for the classical frequentist estimator, spline estimator with Pareto tails, and spline estimator with exponential tails, titled “Freq”, “Par”, and “Exp” respectively. The maximum Monte Carlo standard error of MSE was 0.03 for Bayesian estimators and 0.61 for the frequentist estimator.

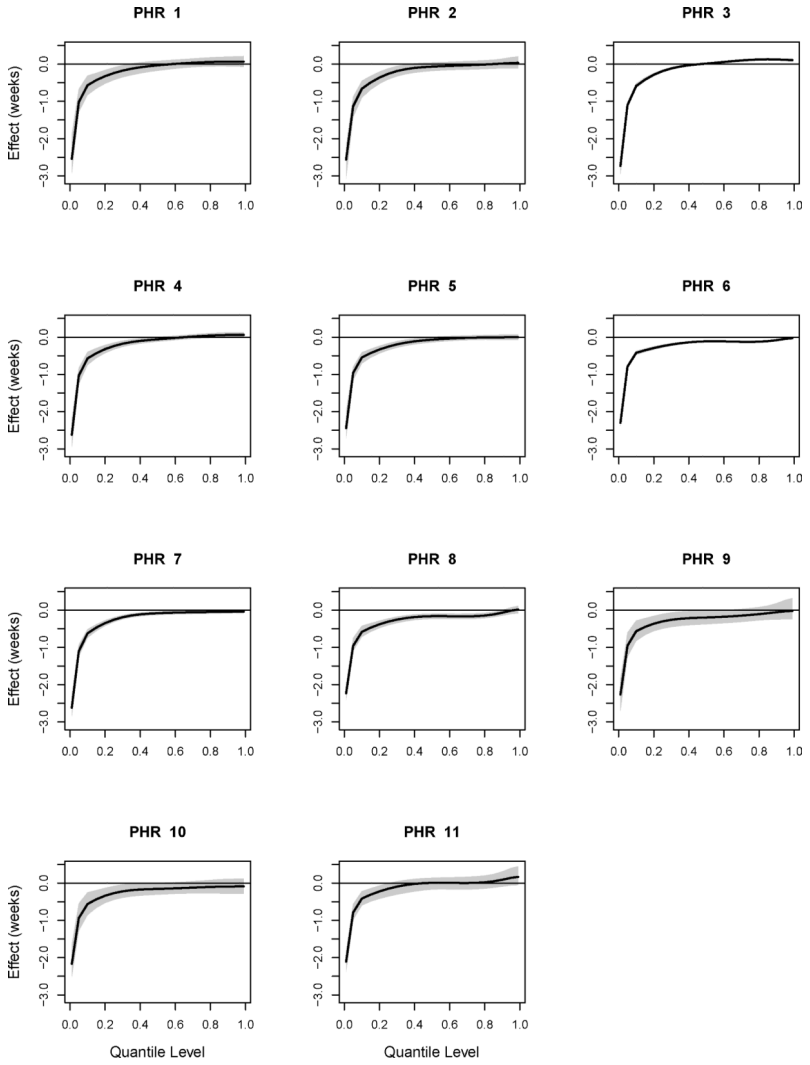

Figure 4.

95% credible limits for the posterior distribution of the difference in gestational age between black non-Hispanic and white non-Hispanic mothers by Public Health Region (PHR).

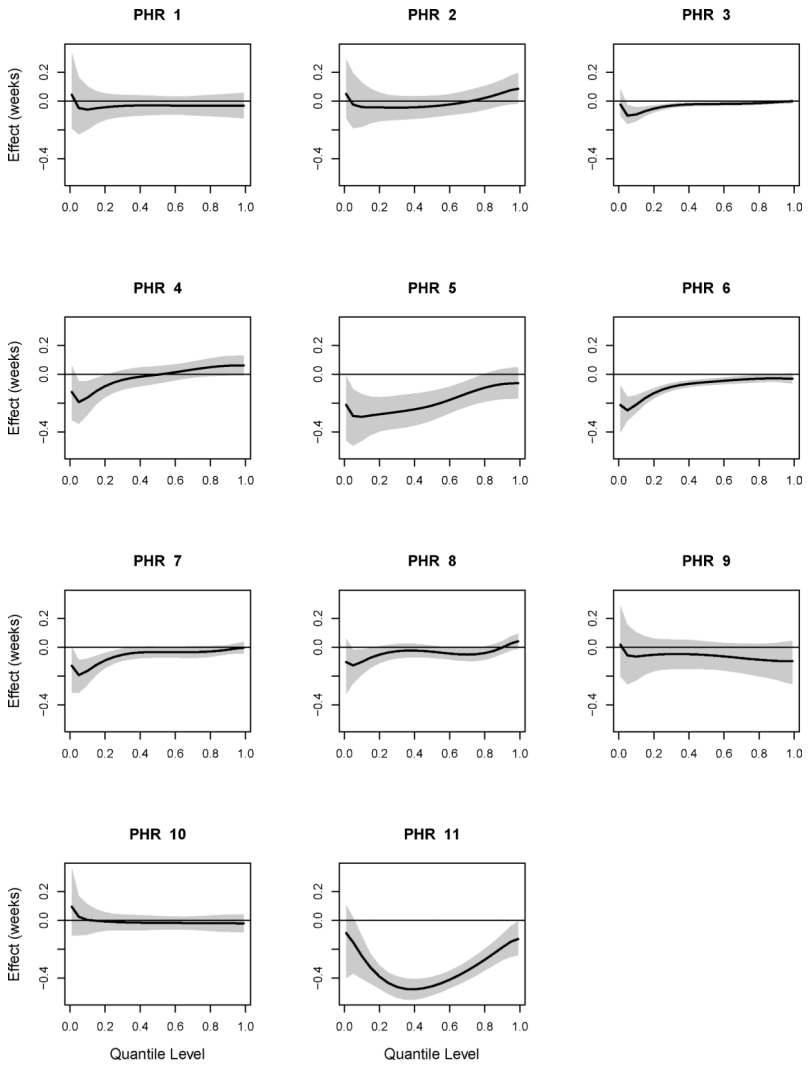

Figure 5.

95% credible limits for the posterior distribution of the effect of a one-unit increase in second trimester ozone exposure on gestational age by Public Health Region. All ozone values were linearly transformed into [−1,1], so a one-unit increase can be roughly thought of as an increase from low levels to middle levels of exposure, or middle to high.

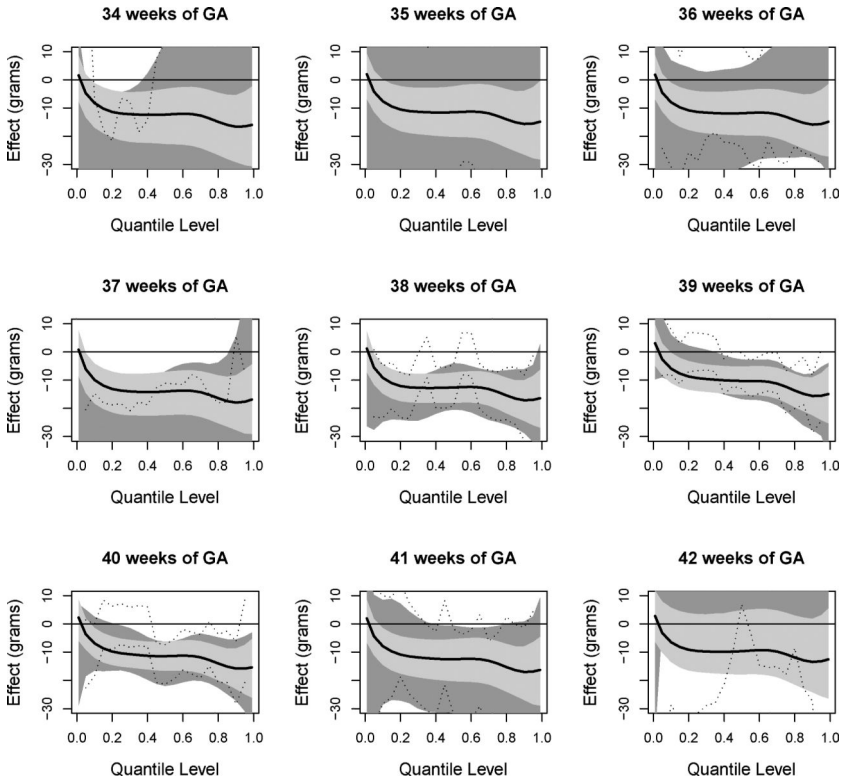

Figure 6.

95% credible limits for the posterior distribution of the effect of a one-unit increase in second trimester ozone exposure on birth weight for gestational age of 34–42 weeks. All ozone values were linearly transformed into [−1,1], so a one-unit increase can be roughly thought of as an increase from low levels to middle levels of exposure, or middle to high. Light gray regions correspond to posterior credible sets for individual fits at each gestational age while dark gray regions correspond to the collective fit across gestational age. Dashed lines indicate limits of 95% frequentist confidence intervals from individual fits.

Table 1

Log pseudo marginal likelihood (LPML) of model fits for the birth outcomes, with higher values corresponding to better fits. Minimum values for gestational age (−5,175,982) and birth weight (−4,094,739) were subtracted from the values shown below for clarity. Model types include exponential tail (Exp), Pareto tail (Par), and independent models for each Public Health Region and gestational age (Ind). Bolded values signify the best fit.

	Number of Basis Functions	Model Type	LPML
Gestational age	5	Exp	1,751,313
	5	Exp	0
	7	Exp	1,869,945
	7	Par	**4,190,201**
	7	Ind	4,190,080
Birth weight	7	Par	1661
	9	Par	**2410**
	9	Ind	0
	11	Exp	838
	11	Par	2397