DDE, PCB and Gestational Age

Evan Knox, Carol Wang, and Justin Weltz

January 21, 2020

Overview

Background

- Question: is exposure to pesticides (measured by DDE and PCB's) related to risk of premature birth?
- 2380 women in dataset, includes primary variables along with center of entry, basic demographics, smoking status, SES scores, triglycerides, cholesterol, and albumin
- Albumin was mostly missing, excluded from analysis; SES scores were variably missing by center, multiply imputed and included

Categories of gestational age from NIH

≤ 28	$(28, 32]$	$(32, 37]$	> 37
extremely preterm	very preterm	moderate/late preterm	not preterm

- In data, 37 weeks is roughly the 20th percentile and 35 weeks is 10th; 32 weeks is 2nd percentile

Problems With Traditional Linear Model

- Gestation length isn't a typical distribution; won't observe lengths past 45 weeks or so, long left tail
- Modeling full distribution of gestation lengths isn't main goal; less reason to worry about pesticide effects on longer gestation due to ability to induce
- Need modeling paradigm(s) that can prioritize prediction of rare, left-tail events

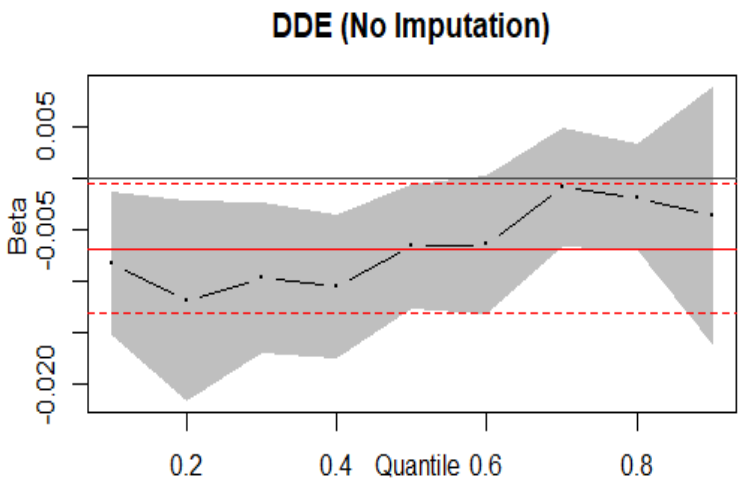
Quantile Regression

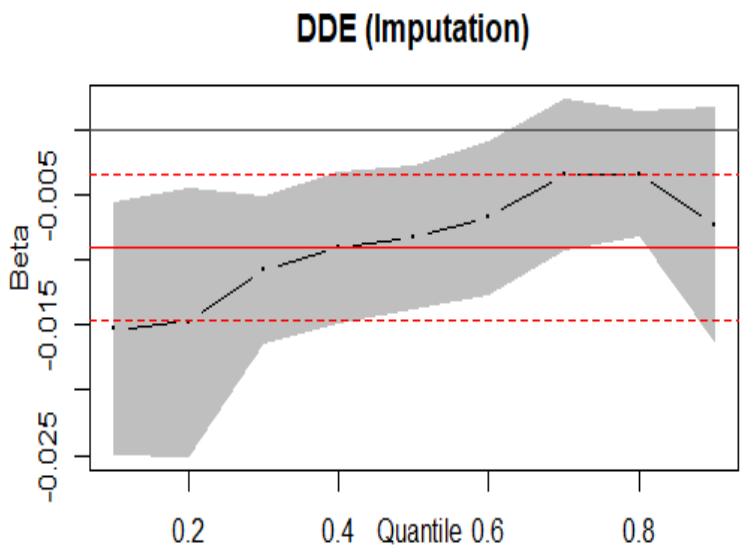
- No assumptions on the error distribution
- Focus on the effect of dde and pcb on specific quantiles
- Minimize

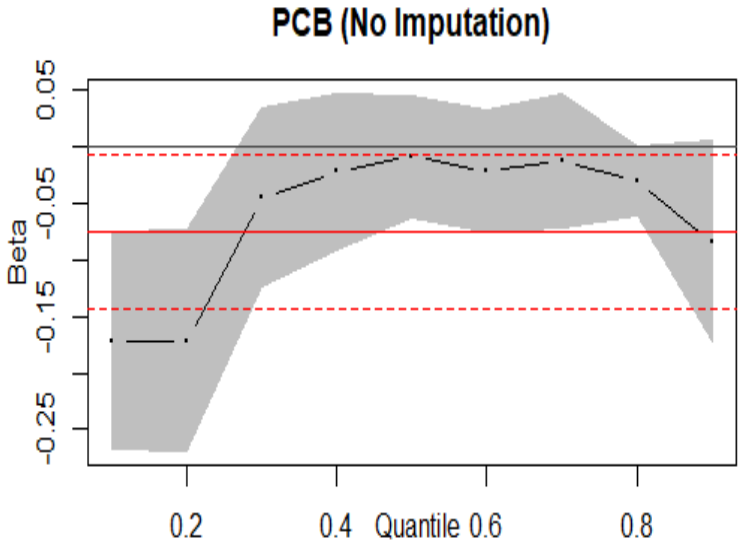
$$\tau \sum_{y_i > \hat{\beta_\tau^T} X_i} |y_i - \hat{\beta_\tau^T} X_i| + (1-\tau) \sum_{y_i < \hat{\beta_\tau^T} X_i} |y_i - \hat{\beta_\tau^T} X_i|$$

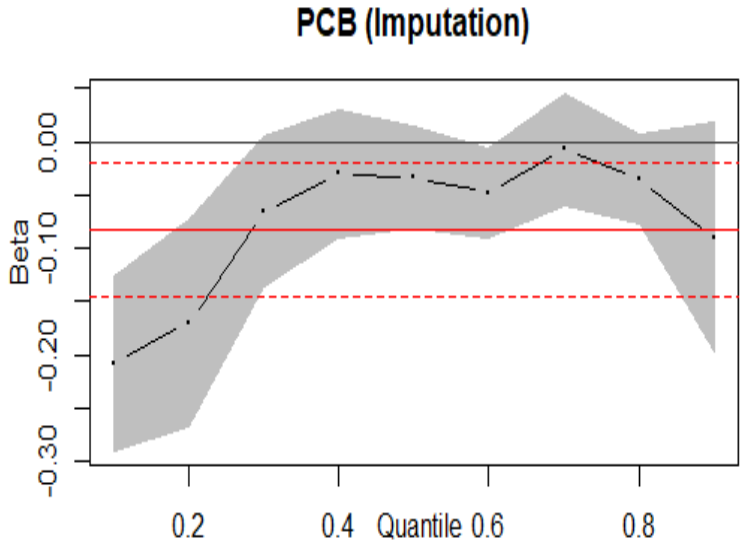
- If $\tau = 0.5$, this problem equates to minimizing absolute error, $\sum |y_i - \hat{\beta_\tau^T} X_i|$, which means the estimator will be the median or the 50% quantile!

Quantile Regression Results









Ordinal Logistic Regression

- Assumptions
  - Ordinal dependent variable
  - Continuous, ordinal or categorical independent variables
  - No collinearity - sample correlation
  - Proportional odds - Brant test
- Model the effects of DDE and PCB on log odds

$$logit(P(Y \leq m_j)) = \beta_{j0} - \eta_1 x_1 - \cdots - \eta_p x_p$$

- Categories defined by NIH guidelines as before:

≤ 28	$(28, 32]$	$(32, 37]$	> 37
extremely preterm	very preterm	moderate/late preterm	not preterm

Ordinal Logistic Regression Results

Holding constant all other variables, for a one unit increase in the predictor, the odds of earlier delivery (e.g., extremely preterm vs very preterm) is multiplied by:

predictor	estimate	25%	97.5%
DDE	1.004	0.998	1.010
sum of PCB	1.084	1.009	1.161
PC1 of PCB	1.064	1.007	1.122

Table: No Imputation

predictor	estimate	25%	97.5%
DDE	1.005	1.000	1.010
sum of PCB	1.088	1.022	1.156
PC1 of PCB	1.068	1.017	1.119

Table: Imputation

The End