Assessing Effects of Exposures to DDE and PCBs on Premature Delivery via Ordinal Logistic Regression

Raphael Morsomme Rihui Ou Alessandro Zito

Case Study 1 - Stat 723

January 22, 2020

Overview

Introduction

- **Framework**:
  Dichlorodiphenyldichloroethylene (DDE) and *Polychlorinated Biphenyls* (PCBs) are chemicals that persist in the envirnoment and get stored in fatty depositis in the human tissues.
  $\Longrightarrow$ Potential adverse effect on health
- **Question**:
  Is exposure to DDE and PBCs associated with a higher chance of premature delivery in pregnant women?

Pregnancy timeline

- **Dangerous preterm**: delivery at 34 weeks or before (when main organs are underdeveloped)
- **Preterm**: delivery beween 35 and 37 week
- **At term**: delivery after 37 weeks

Data

Data collected by 12 centers contained gestational age (in weeks) of the mother, the DDE and PCBs concentration, socio-economic info and scores (race, occupation, education, income), amount of triglycerides and cholesterol in blood and smoking status.

Preprocessing:

- Drop obs. with gestational age > 45
- Average of standardized PCBs[1]

$$PCB_i = \frac{1}{11} \sum_{j=1}^{11} \frac{PCB_{ij} - mean_i(PCB_{ij})}{sd_i(PCB_{ij})}$$

- Mean imputation of occupation, education and income scores
- Aggregate race into $race = 1$ if white and $race = 0$ if non-white

$\Longrightarrow$ Total obs. $=$ **2336**

[1] This avoids the correlation between the PCBs. See the appendix.

Data

- **Ordinal dependent variable**:

$$gestgroup_i = \begin{cases} \text{Dangerous} & \text{if } \#\text{weeks} \leq 34 \\ \text{Pre term} & \text{if } 34 < \#\text{weeks} \leq 37 \\ \text{At term} & \text{if } 37 < \#\text{weeks} \end{cases}$$

- **Adjusted measure for** *PCB* **and** *DDE* to estimate the level of *exposure*:
  1. Computing total lipids using Phillips et al. (1989) and Bernert et al (2007) formula

$$lipid_i = 2.27 * cholesterol_i + triglycerides_i + 0.623$$

  2. Setting[2]

$$adjDDE_i = \frac{DDE_i}{log(lipid_i)} \qquad adjPCB_i = \frac{PCB_i}{log(lipid_i)}$$

[2]The choice of the log comes from a Box-Cox analysis of the log-likelihood, as in Li et al (2013)

EDA (I) - Exposures and gestational groups by race

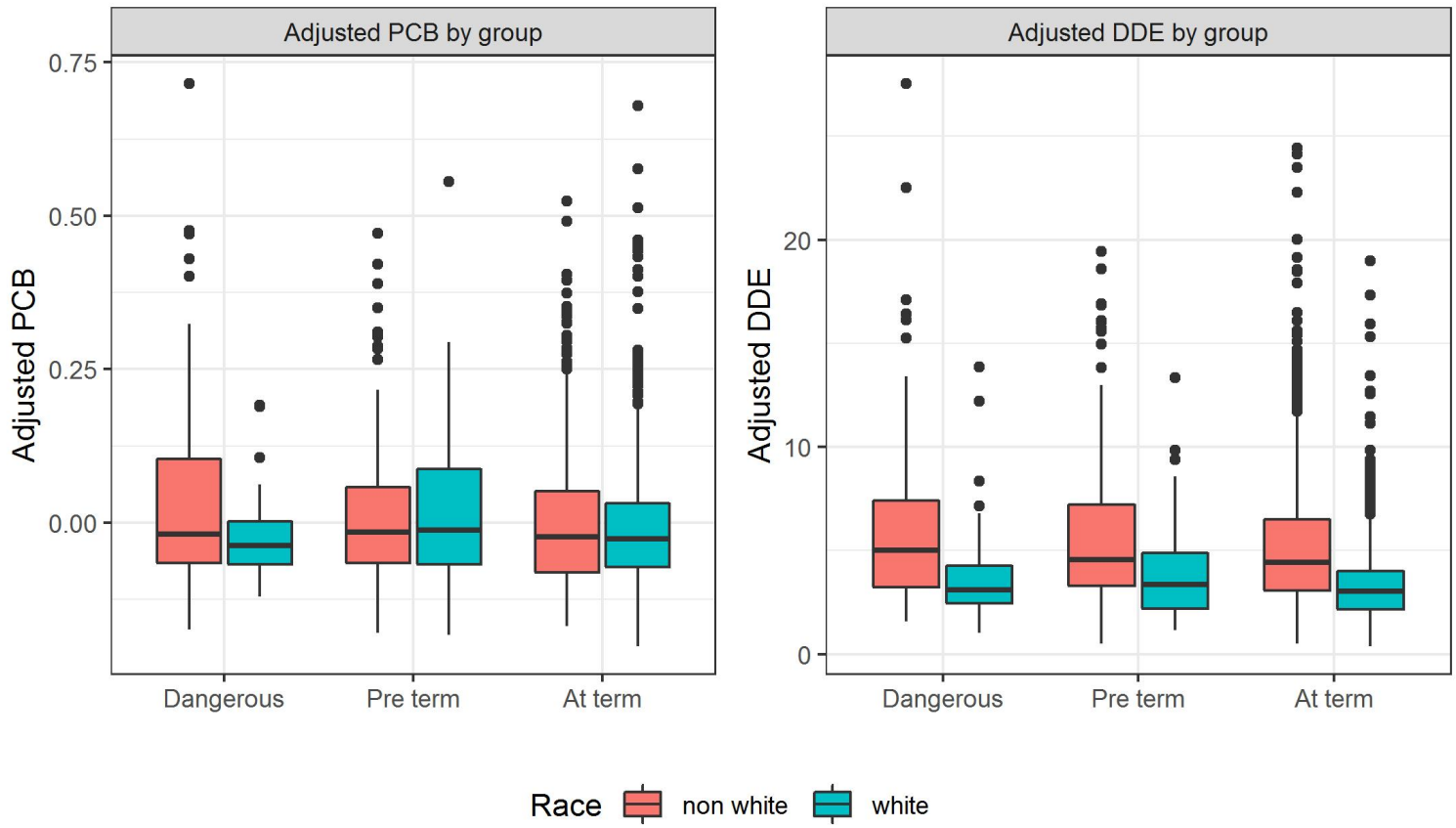

Figure: Relationship between gestation outcome and adjusted exposures, by race.

EDA (II) - Exposure across centers

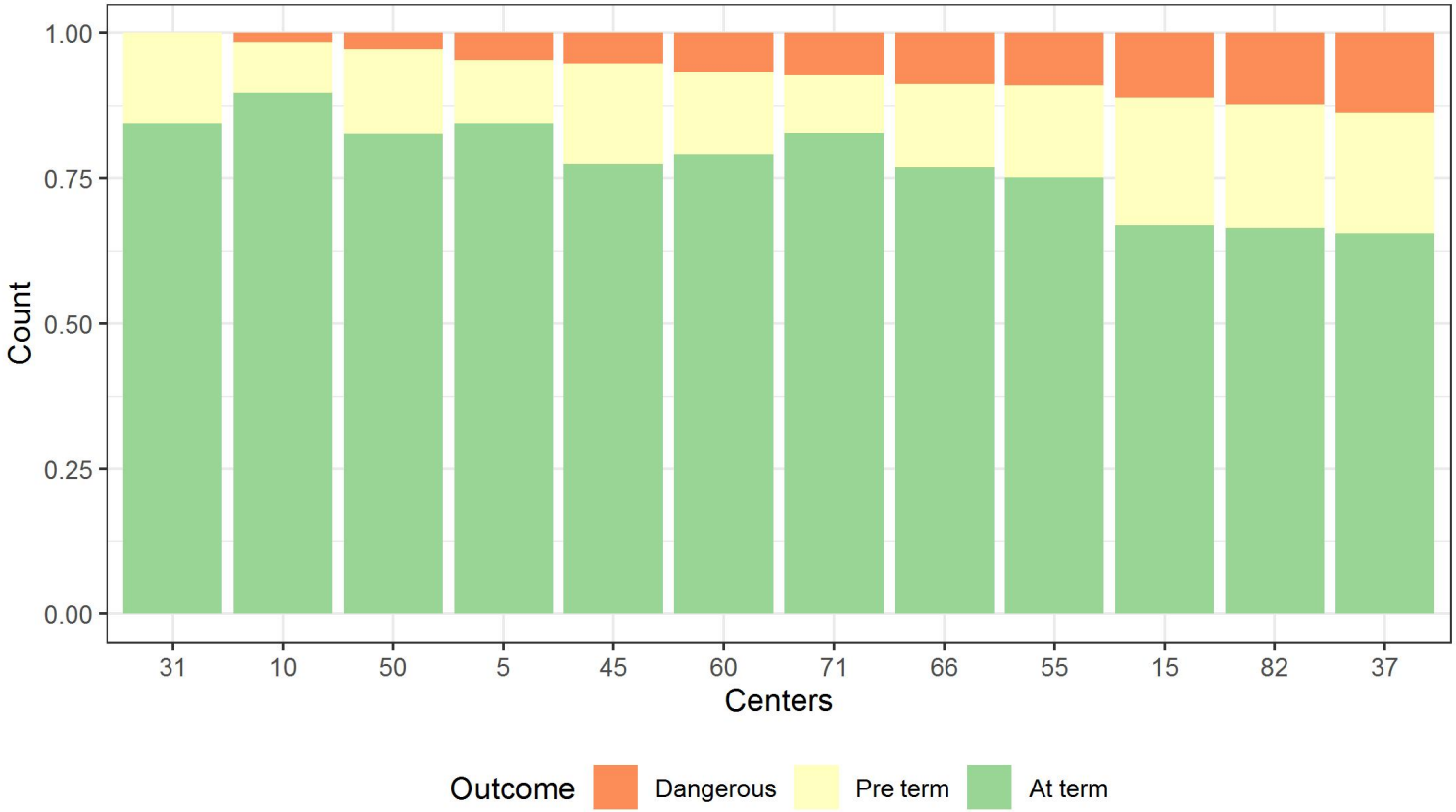

Figure: Distribution of gestation outcome per center.

Model - Ordinal Logistic Regression

We run the following ordinal logistic regression model:

$$\text{logit}(P(\textit{gestgroup} \leq j)) = \beta_{0j} - \mathbf{X}\boldsymbol{\beta}$$

where $j = 0, 1$ corresponds to the outcome level, and $\mathbf{X}$ contains:

- *DDE*, *PCB*, *race*, *center*, *smoke*, the 3 scores, mother age [main effects]
- (*adjDDE* + *adjPCB*) * (*race* + *center*) [interactions].

Model - Inference

Bayesian ordinal logistic regression using the variables selected by AIC-based backward variable selection procedure.

- Maintain *DDE*, *PCB*, *smoke*, *center*, *race*, (*PCB* + *DDE*) * *race*
- Drop 3 scores, (*DDE* + *PCB*) * *center* , mother age
- uniform, and R^2 prior on coefficients
- Model assumptions are checked in the appendix.

Numerical Results

	mean	5%	95%
adjDDE	0.02	-0.01	0.05
adjPCB	1.76	0.72	2.75
adjDDE*white	0.05	-0.02	0.12
adjPCB*white	-1.60	-3.26	0.02

Interpretation:

For a 1 unit increase of adjDDE:

- **Nonwhite**: the odds of having a more dangerous delivery increase by $(e^{(0.02)} - 1) * 100\% = 2.02\%$
- **White**: the same odds increase by $(e^{(0.02+0.05)} - 1) * 100\% = 7.25\%$

For a 0.1 unit increase of adjPCB:

- **Nonwhite**: the odds of having a more dangerous delivery increase by $(e^{(1.76)*0.1} - 1) * 100\% = 19.22\%$
- **White**: the same odds increase by $(e^{0.1*(1.76-1.60)} - 1) * 100\% = 1.595\%$

Graphical Results

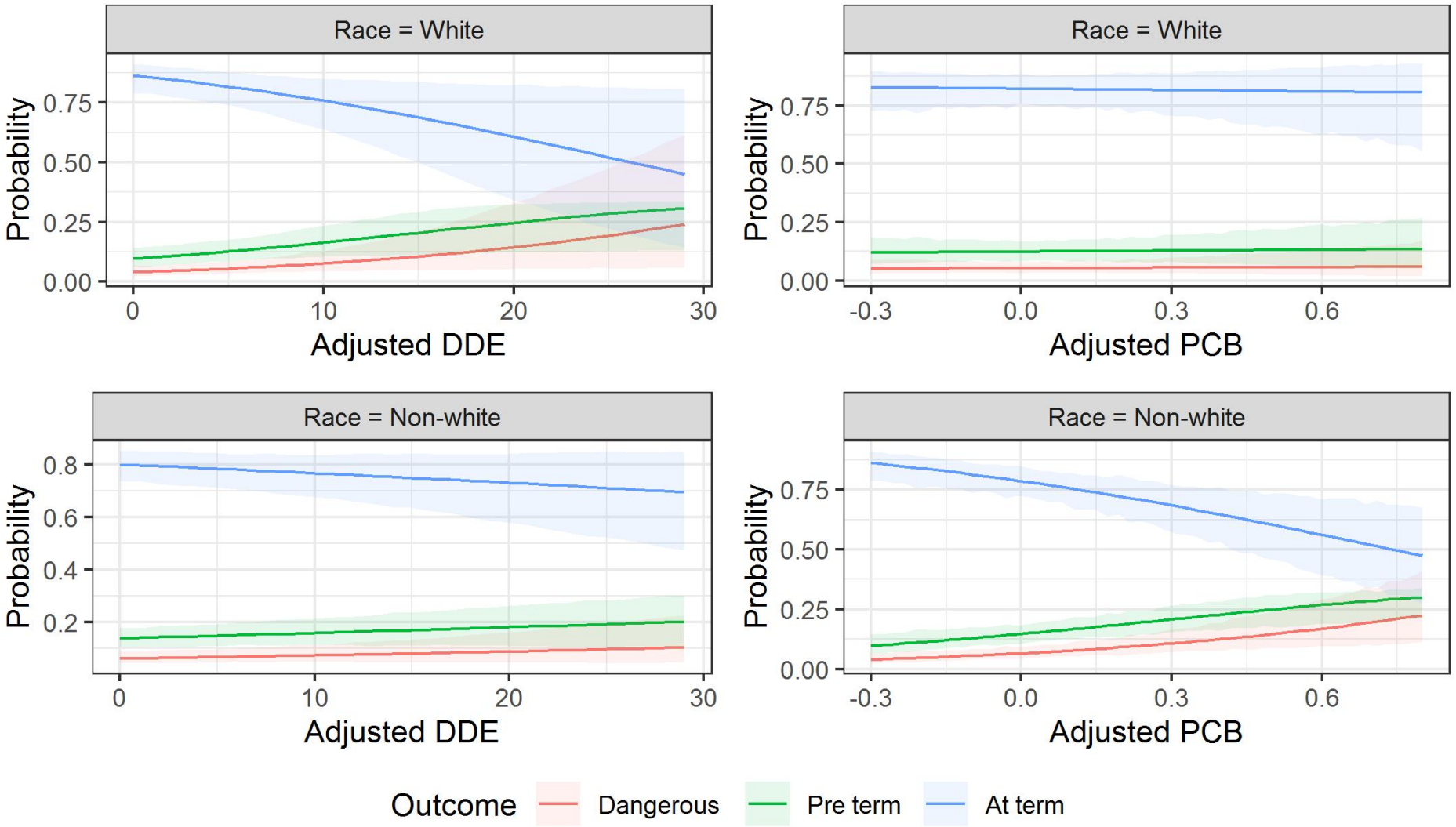

Figure: Estimated probability of gestation outcomes for DDE and PCB, by race.

Conclusions

To summarize:

- Effect of the chemicals is race-dependent
- DDE has more impact on white people
- PCB has more impact on non-white people

References

Li, D; Longnecker, M.P.; and Dunson, D.B.
Lipid Adjustment for Chemical Exposures: Accounting for Concomitant Variables.
Epidemiology, Nov 2013

Phillips, D; Pirke, J., Burse, V.; Bernert, J.; Henderson, L.; Needham, L.
Chlorinated hydrocarbon levels in human serum: Effects of fasting and feeding.
Archives of Environmental Contamination and Toxicology, 1989

Bernert, JT.; Turner, WE.;, Patterson, DG. Jr;, Needham, LL.
Calculation of serum total lipid concentrations for the adjustment of persistent organohalogen toxicant measurements in human samples.
Chemosphere, 2007

Liu, D.; and Zhang, H.;
Residuals and Diagnostics for Ordinal Regression Models: A Surrogate Approach
Journal of the Americal Statistical Association, 2018

Appendix

More EDA

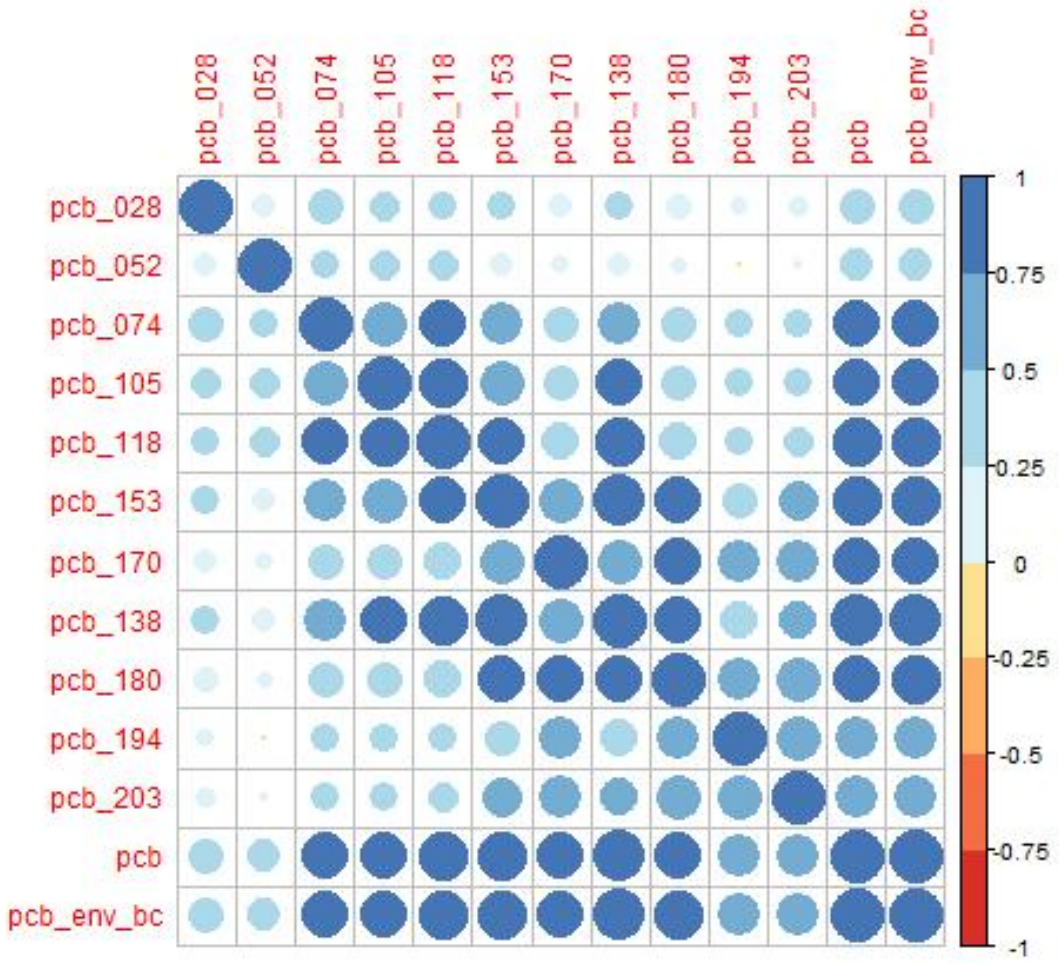

Figure: Correlation plot across PCBs

Appendix

Frequentist Model Checking

We can check the assumption of the (frequentist) ordinal logistic model by looking at the Surrogate residuals (Liu and Zhang, 2018).

If the model assumptions are correct, then the surrogate residuals R_S will have three properties:

1. $E(R_S|X) = 0$
2. $Var(R_S|X) = c$, the conditional variance of R_S is constant
3. The emiprical distribution of R_S resembles an explicit distribution that is related to the link function $G^{-1}(\cdot)$. Specifically, $R_S \sim G(c + \int u dG(u))$.

Appendix

Frequentist Model Checking

Assumptions (i) and (ii) are checked with the Surrogate residuals plot. Both are satisfied in this case.

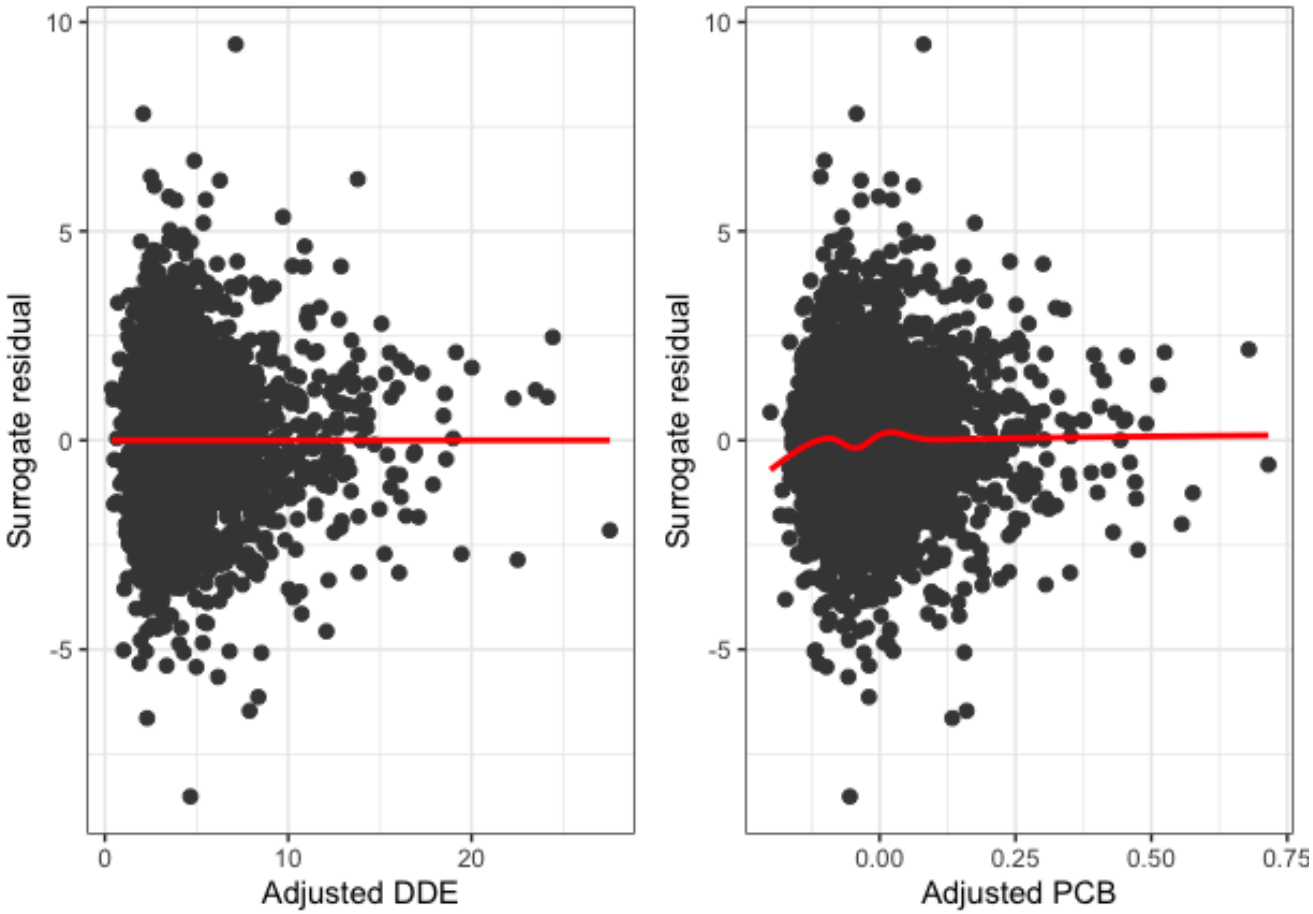

Figure: Surrogate residuals of DDE and PCB

Appendix

Frequentist Model Checking

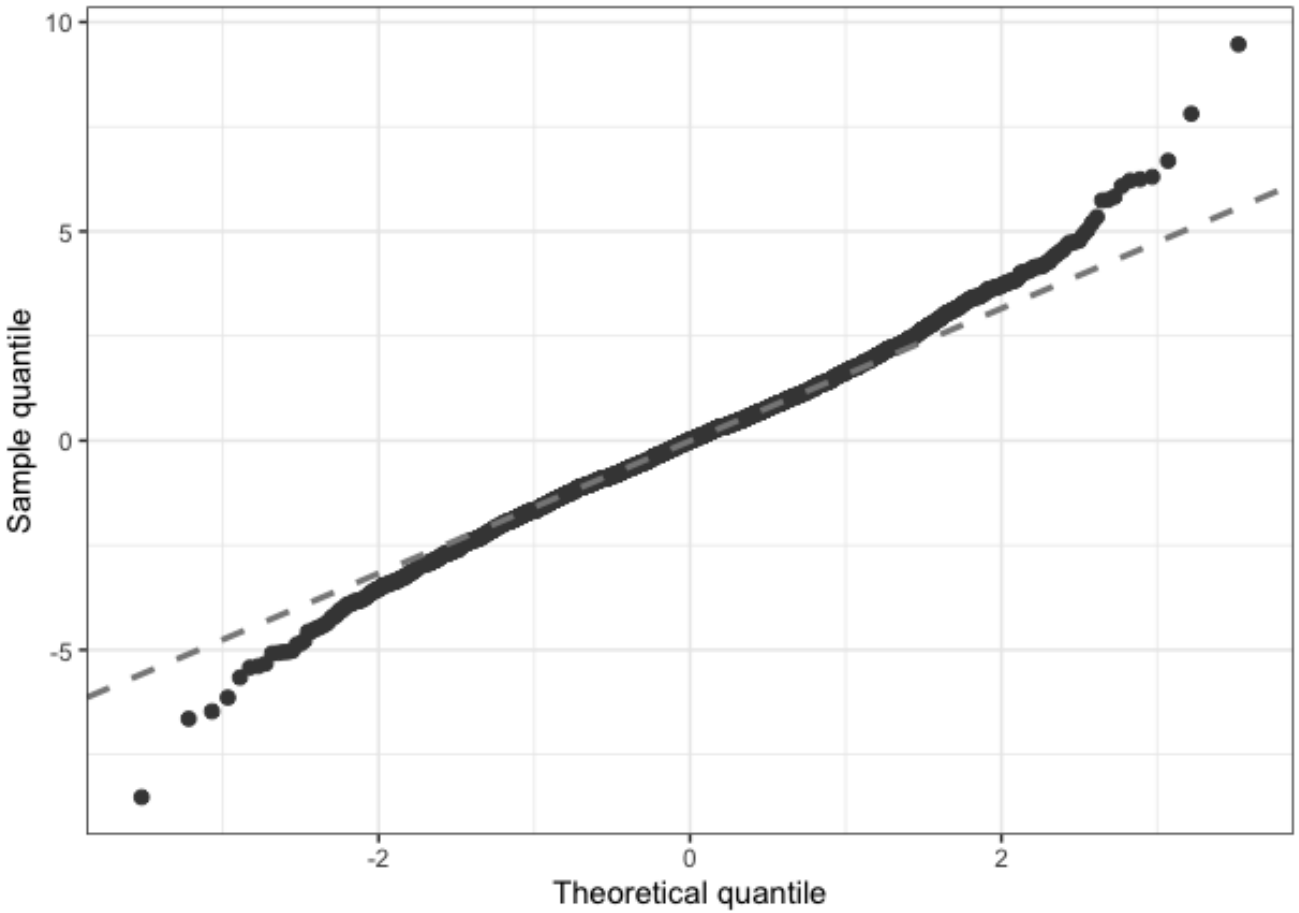

Figure: QQ plot of the Surrogate residuals