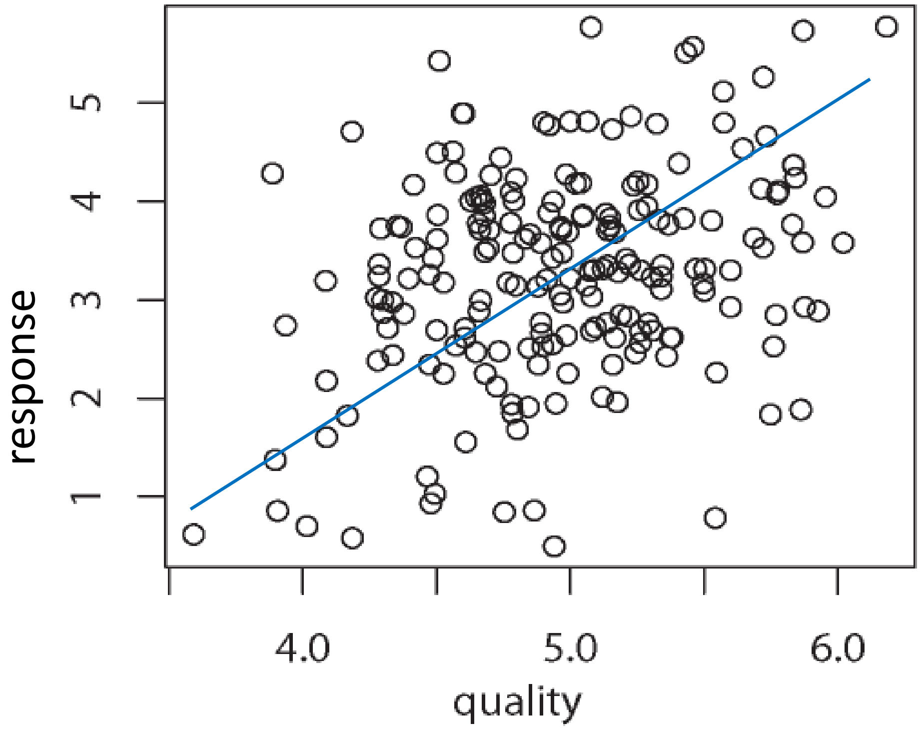
Characterizing Uncertainty:Part 2

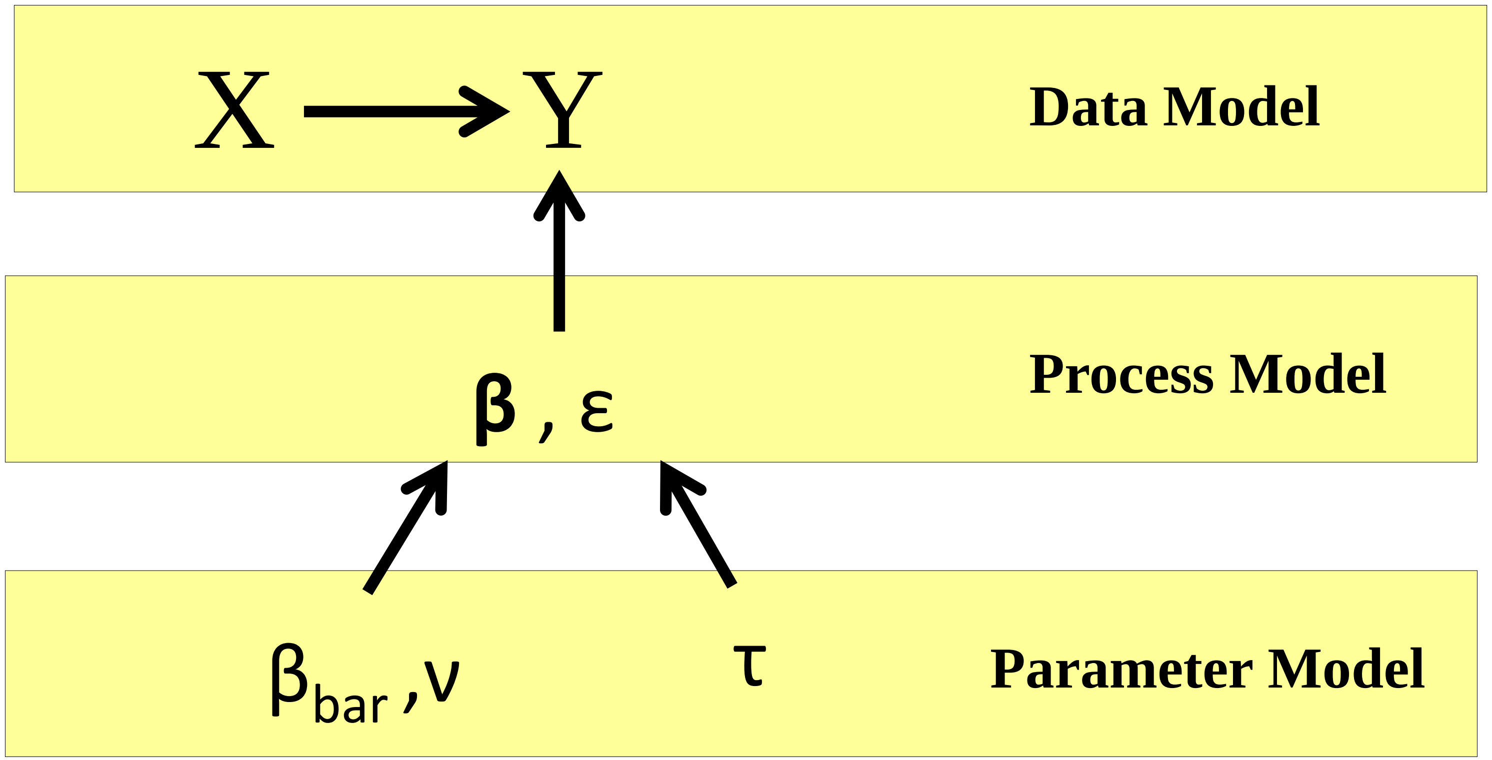
Linear Model

$$y_i \sim \beta_0 + \beta(x_i) + \varepsilon_i$$



Linear Model – Graph Notation

$$y_i \sim \beta_0 + \beta(x_i) + \varepsilon_i$$



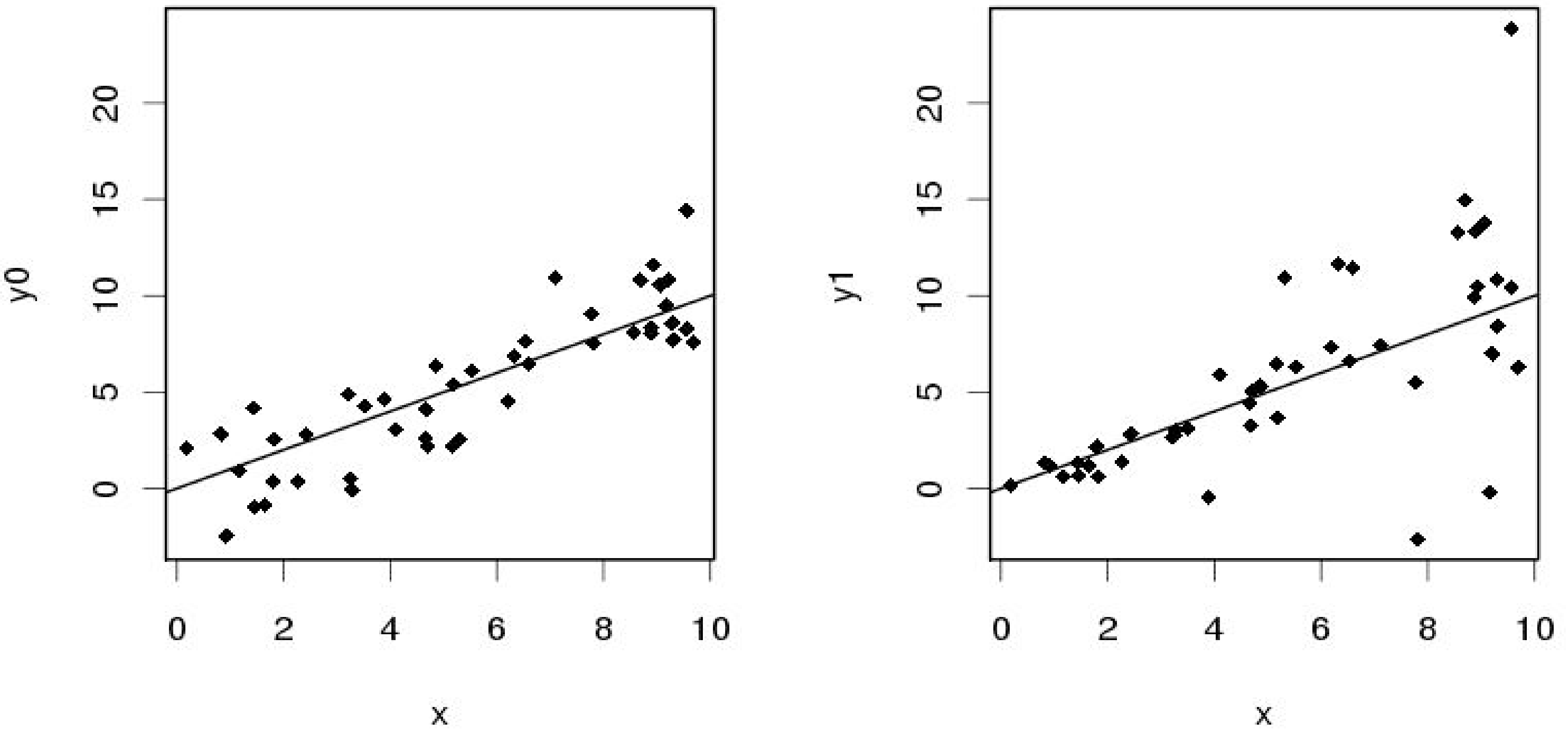
$\varepsilon_i \sim N(0,\tau)$

$\boldsymbol{\beta} \sim N(\beta_{bar}, \nu)$

Classic Assumptions of Linear Model:

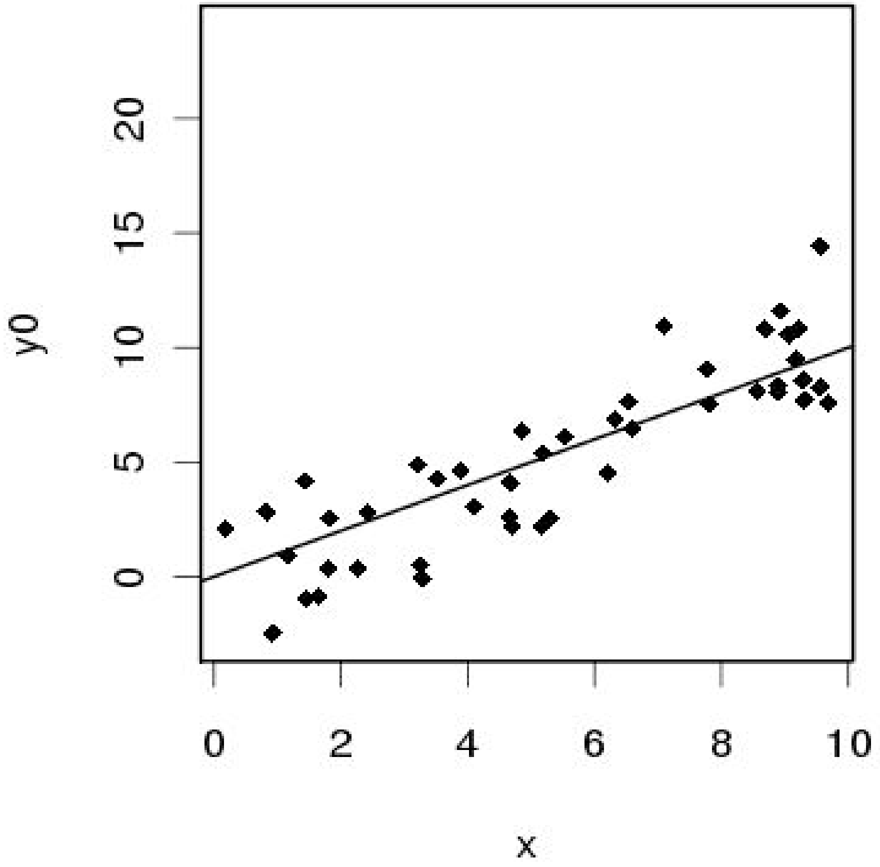
- Normally distributed error
- Homoskedasticity
- No error in X variables
- Error in Y variables is measurement error
- Observations are independent
- No missing data

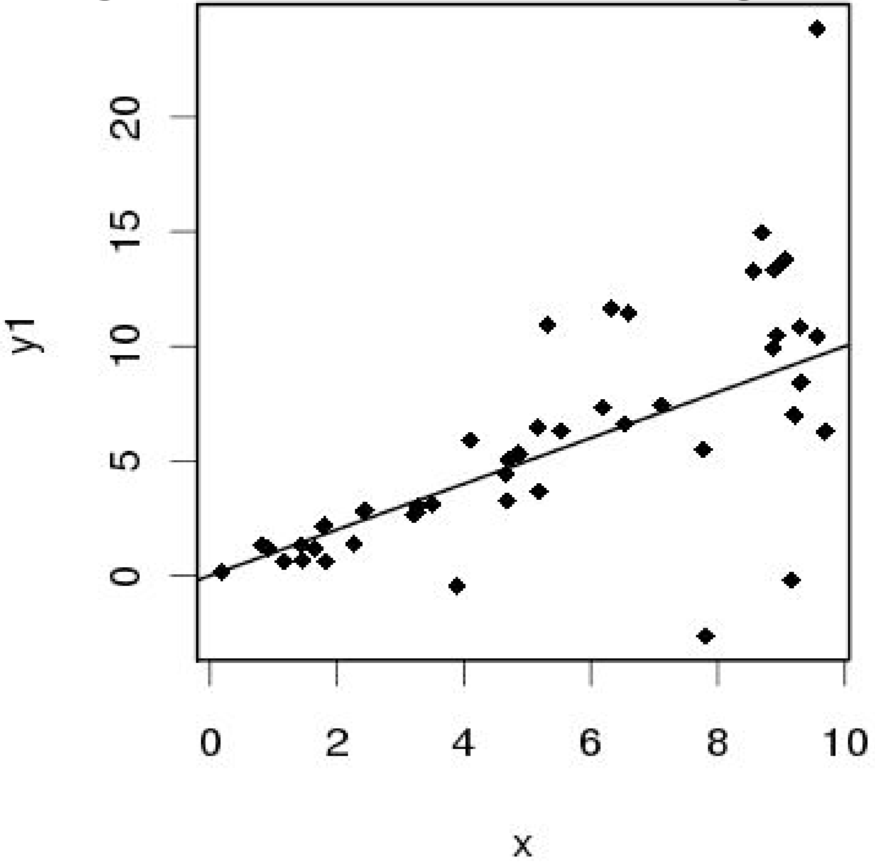
Heteroskedasticity



Heteroskedasticity

Could use different distribution
(lognormal, Poisson, Neg. Binomial)

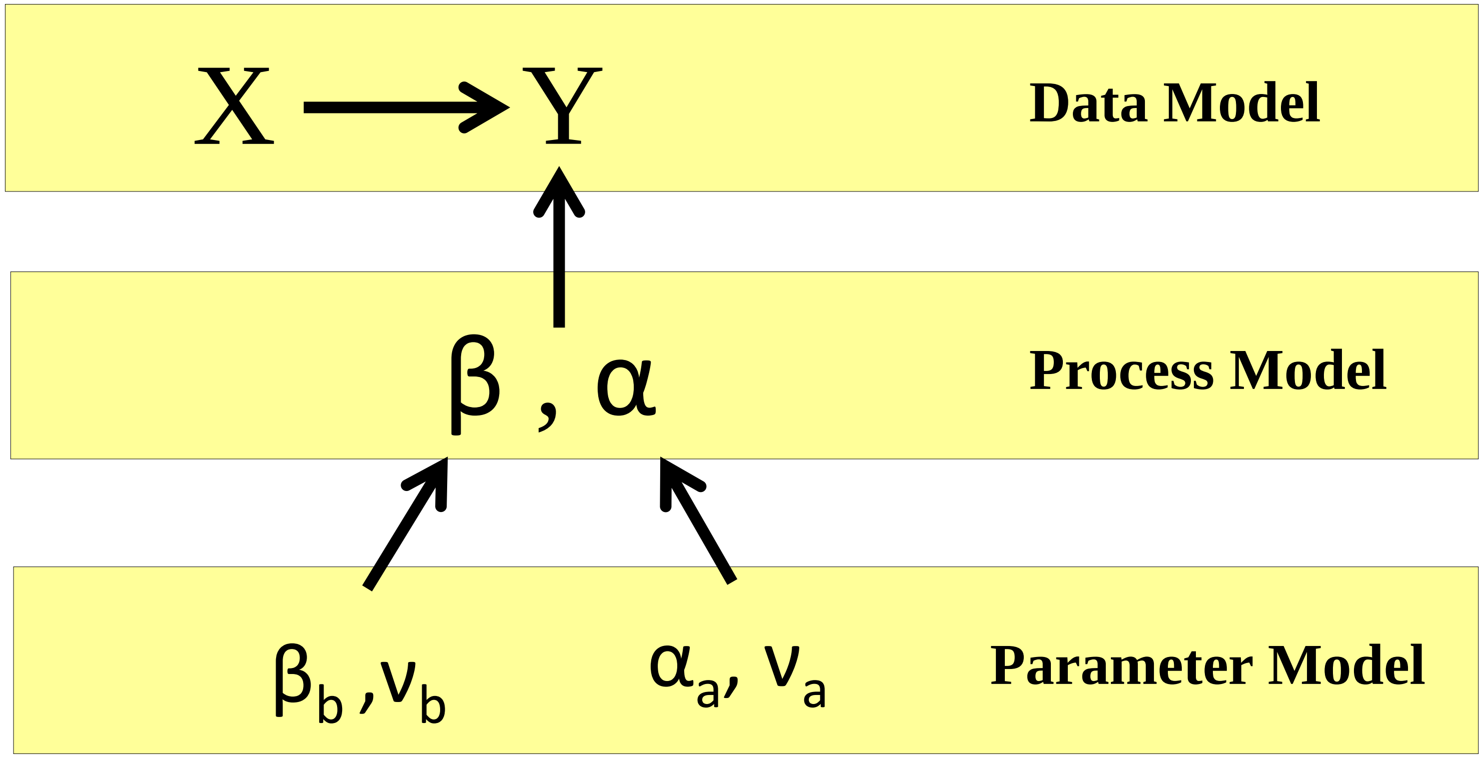




Heteroskedasticity

Model the variance:

$$y \sim N\left(\beta_1 + \beta_2 x, (\alpha_1 + \alpha_2 x)^2\right)$$



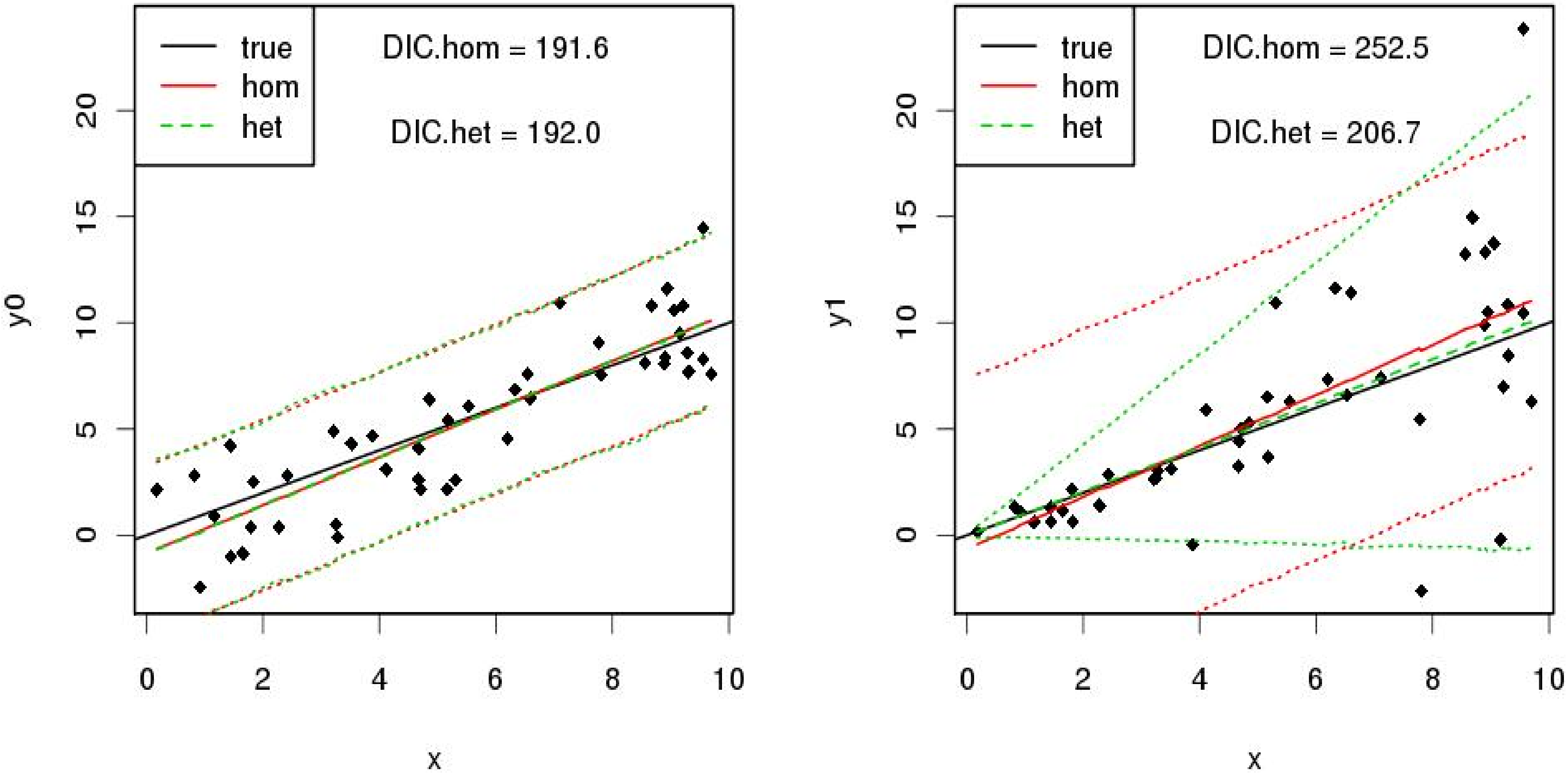
$\boldsymbol{\alpha} \sim N(\alpha_a, \nu_a)$

$\boldsymbol{\beta} \sim N(\beta_b, \nu_b)$

Heteroskedasticity

$$y \sim N(\beta_1 + \beta_2 x, s^2)$$

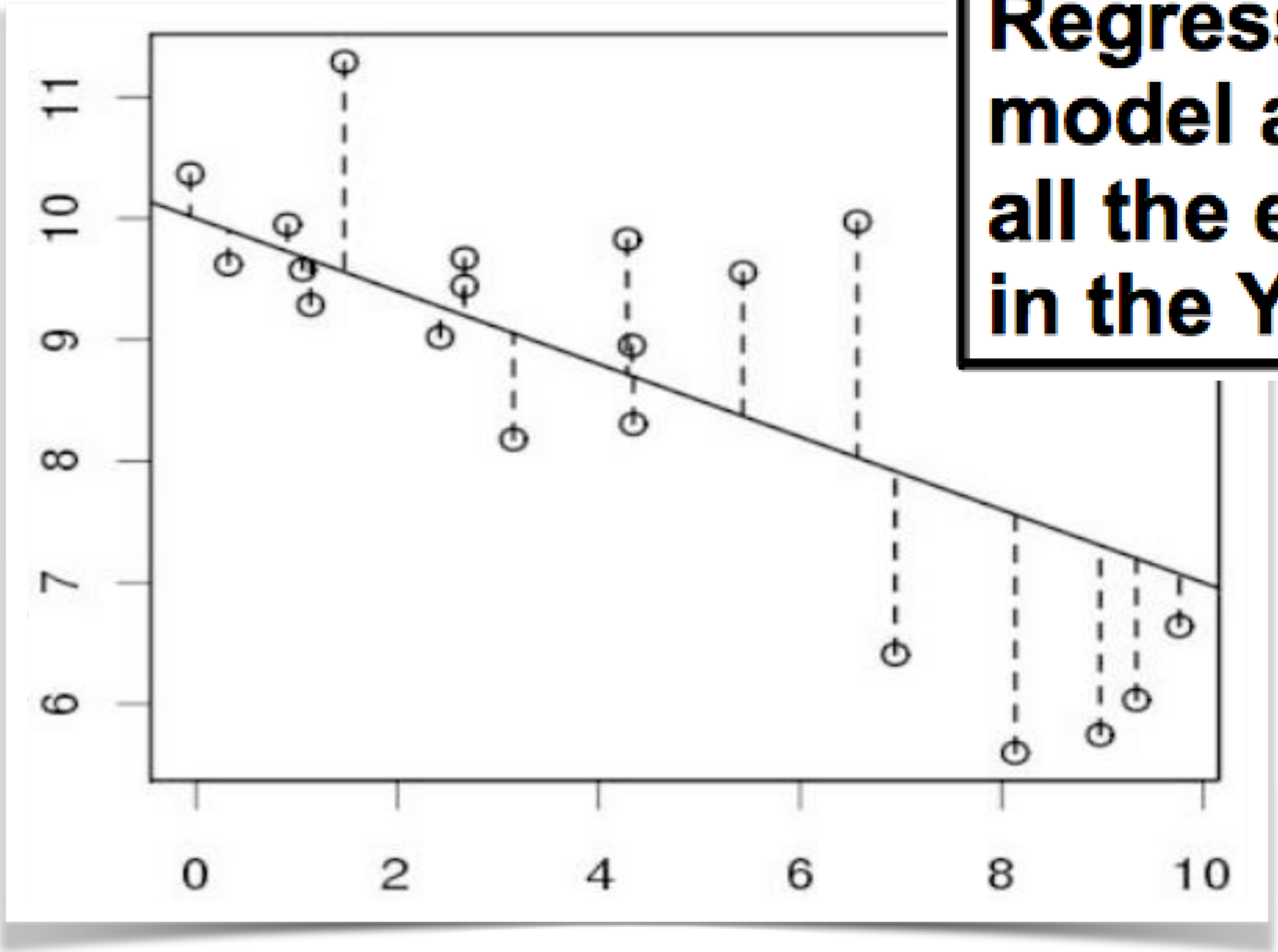
$$y \sim N(\beta_1 + \beta_2 x, (\alpha_1 + \alpha_2 x)^2)$$



Classic Assumptions of Linear Model:

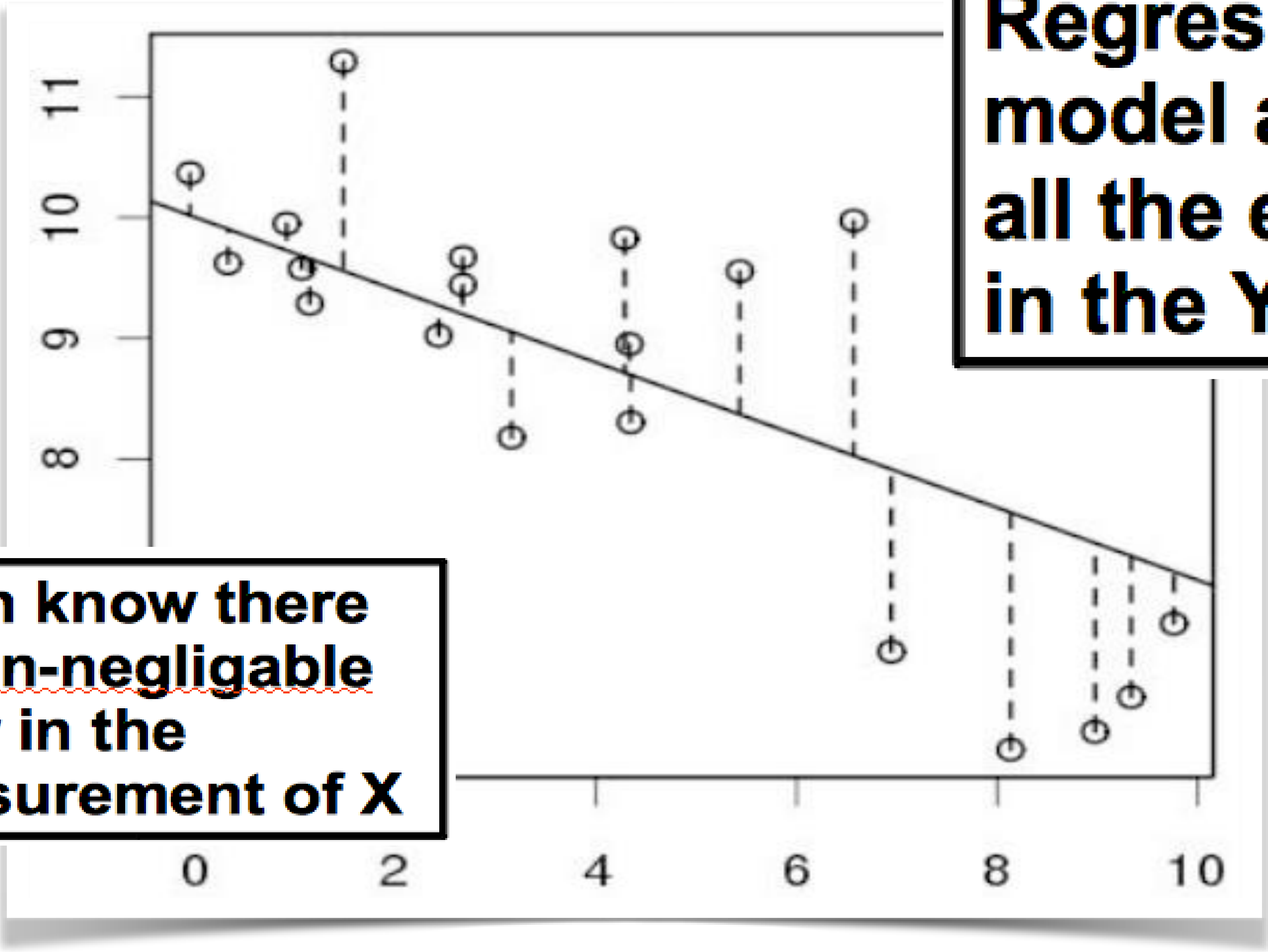
- Normally distributed error
- Homoskedasticity
- **No error in X variables**
- Error in Y variables is measurement error
- Observations are independent
- No missing data

Errors in variables (EIV)



Regression model assumes all the error is in the Y

Errors in variables (EIV)



Regression model assumes all the error is in the Y

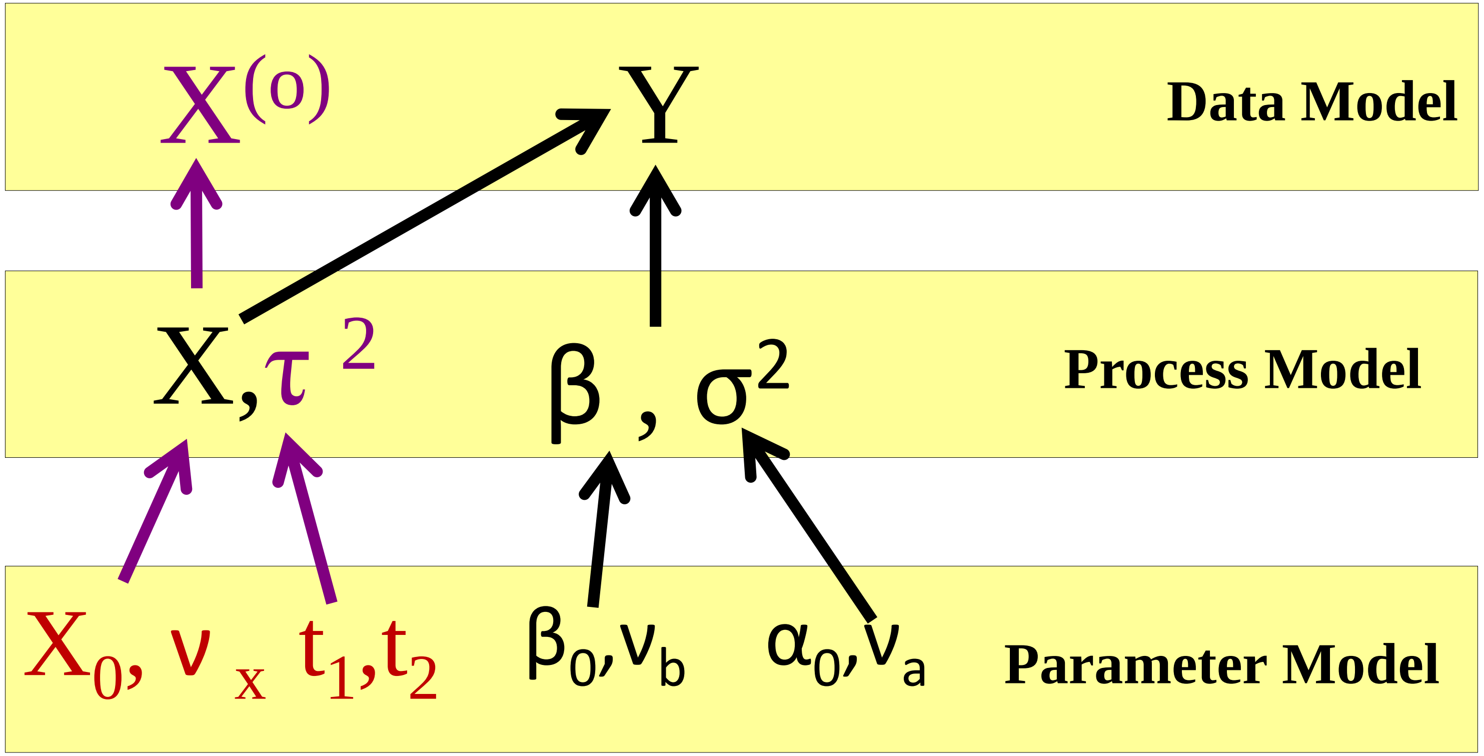
Often know there is non-negligable error in the measurement of X

Errors in variables

Model covariate (X) as random variable

$$Y \sim N(\beta X, \sigma^2)$$

$$X0 \sim N(X, \tau 2)$$

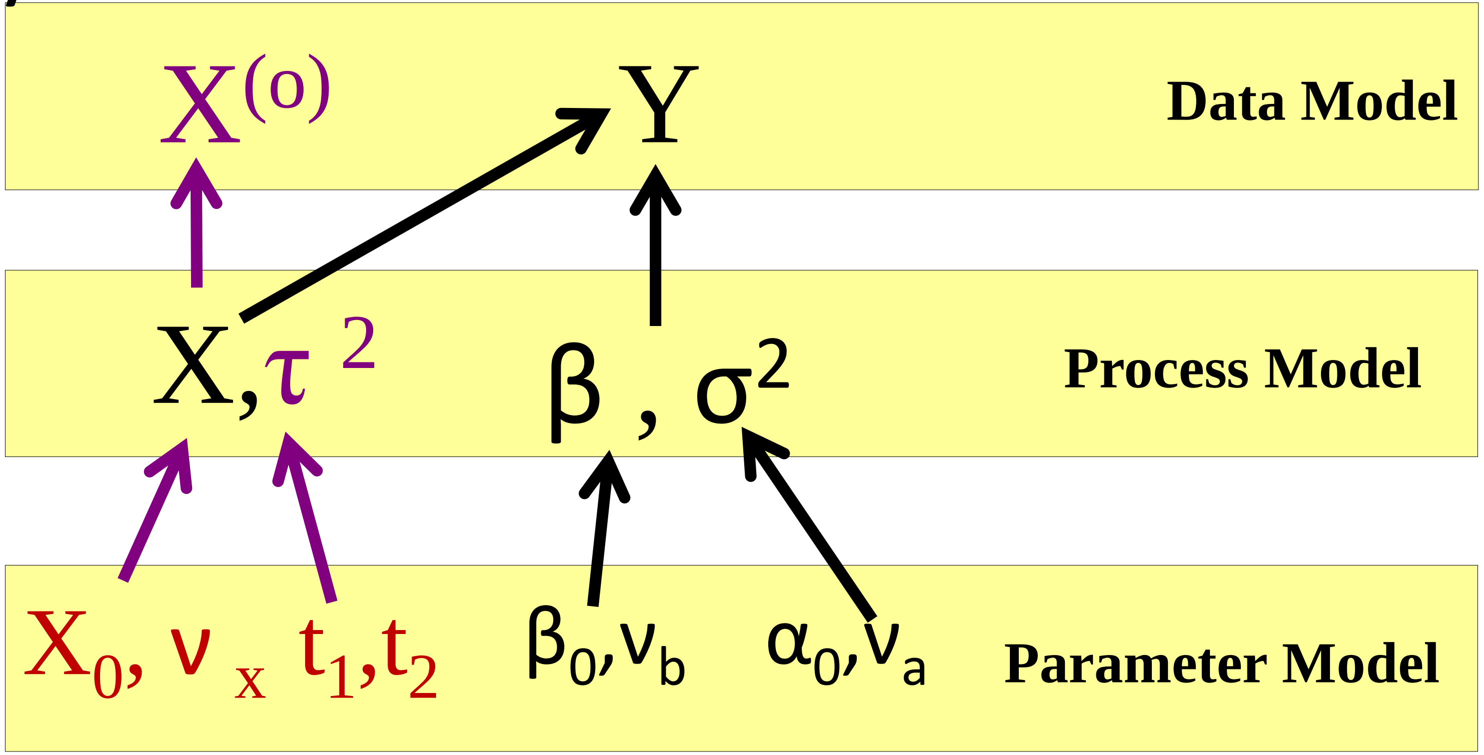


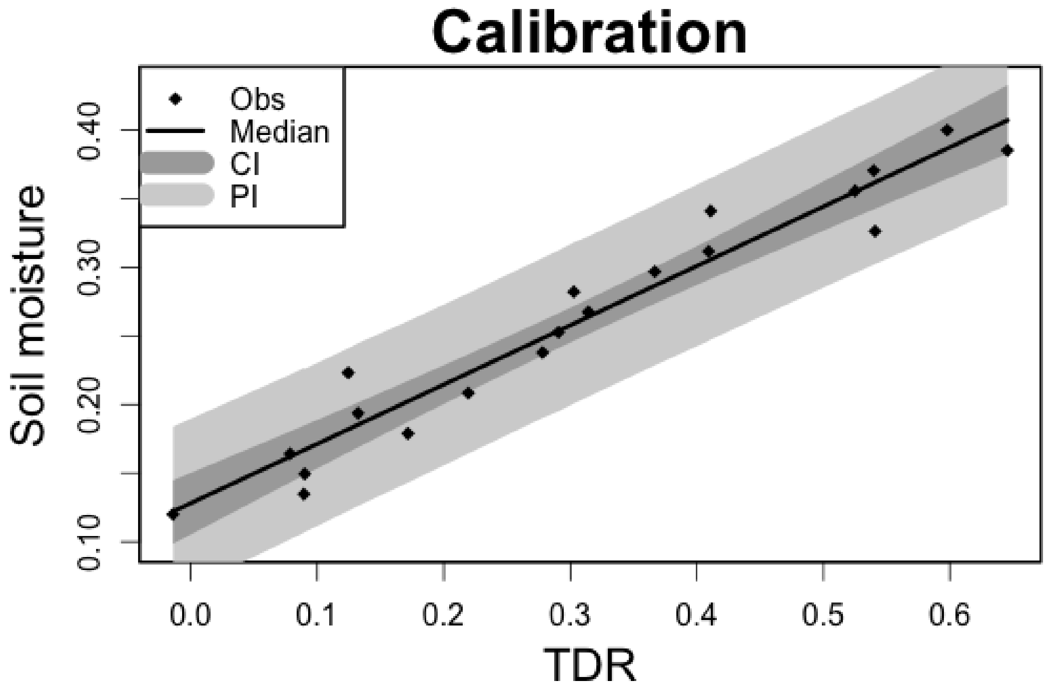
Errors in variables

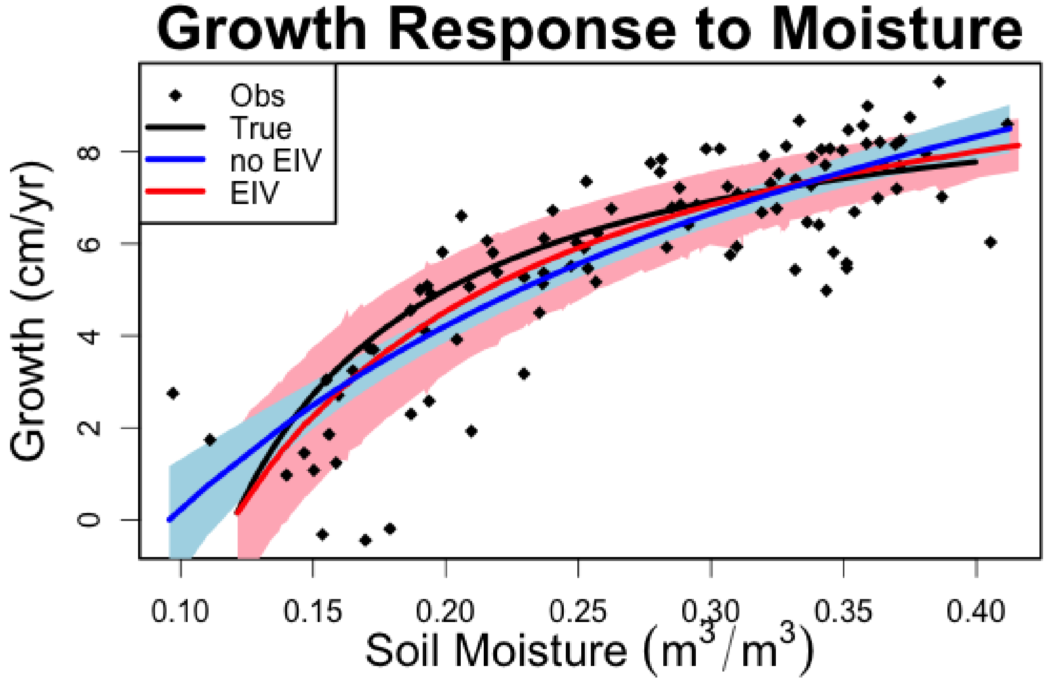
This model could also be a calibration model - if covariate is proxy measure.

$$Y \sim N(\beta X, \sigma^2)$$

$$X0 \sim N(X, \tau 2)$$




Calibration
Obs
Median
CI
PI
Soil moisture
0.10
0.20
0.30
0.40
0.0
0.1
0.2
0.3
0.4
0.5
0.6
TDR


Growth Response to Moisture
Obs
True
no EIV
EIV
Growth (cm/yr)
0
2
4
6
8
0.10
0.15
0.20
0.25
0.30
0.35
0.40
Soil Moisture (m³/m³)

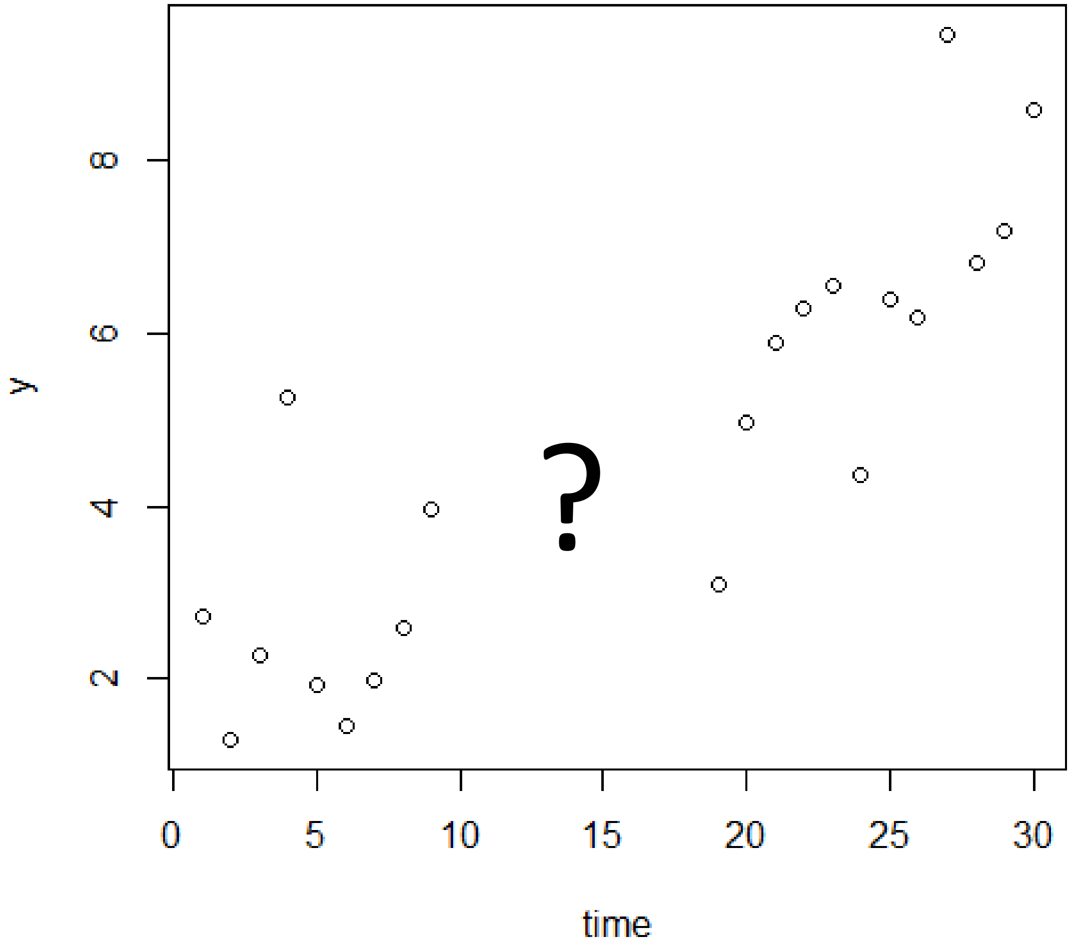
Classic Assumptions of Linear Model:

- Normally distributed error
- Homoskedasticity
- No error in X variables
- **Error in Y variables is measurement error**
- **Observations are independent**
- **No missing data**

Latent Variables

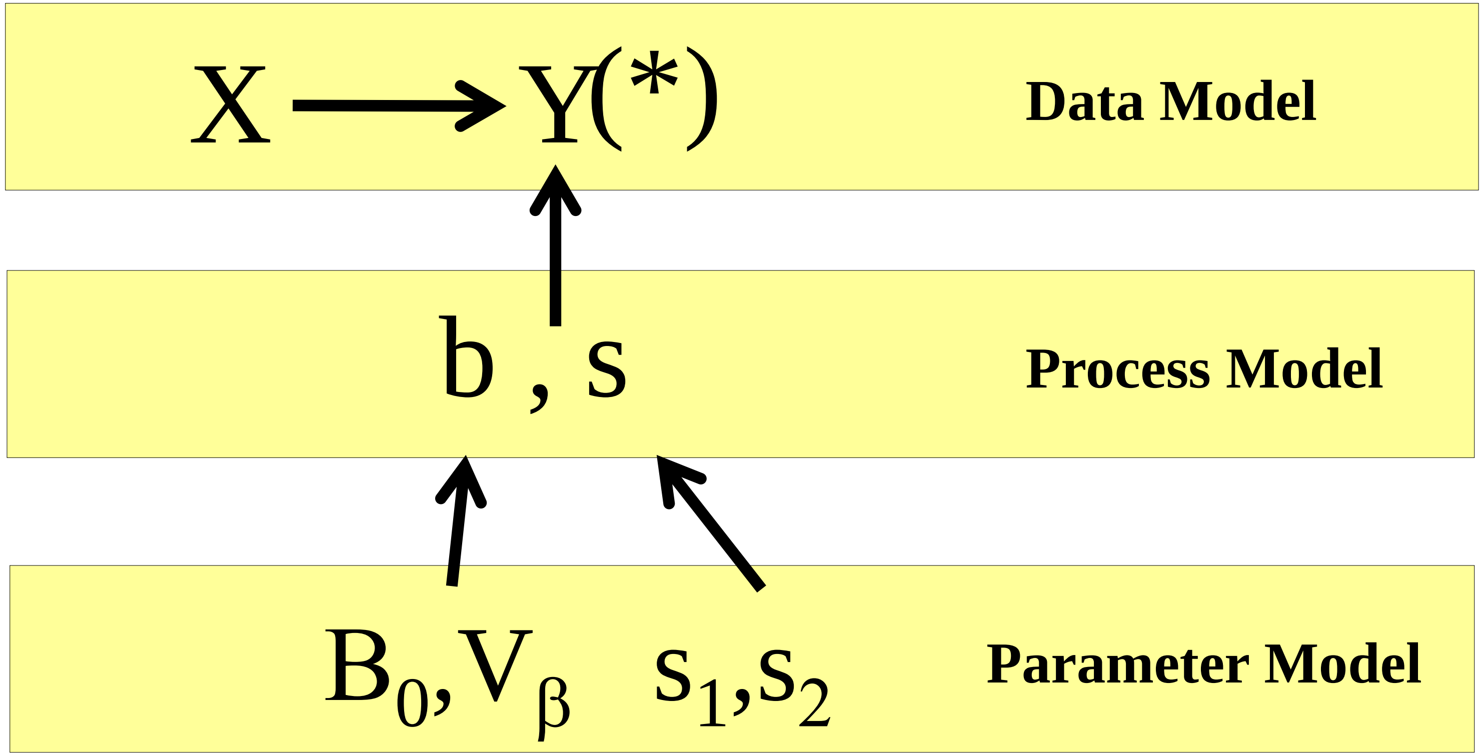
- Any variable not directly observed
  - **Missing data**
  - Variable measured with error
  - Proxy measures

Missing Data

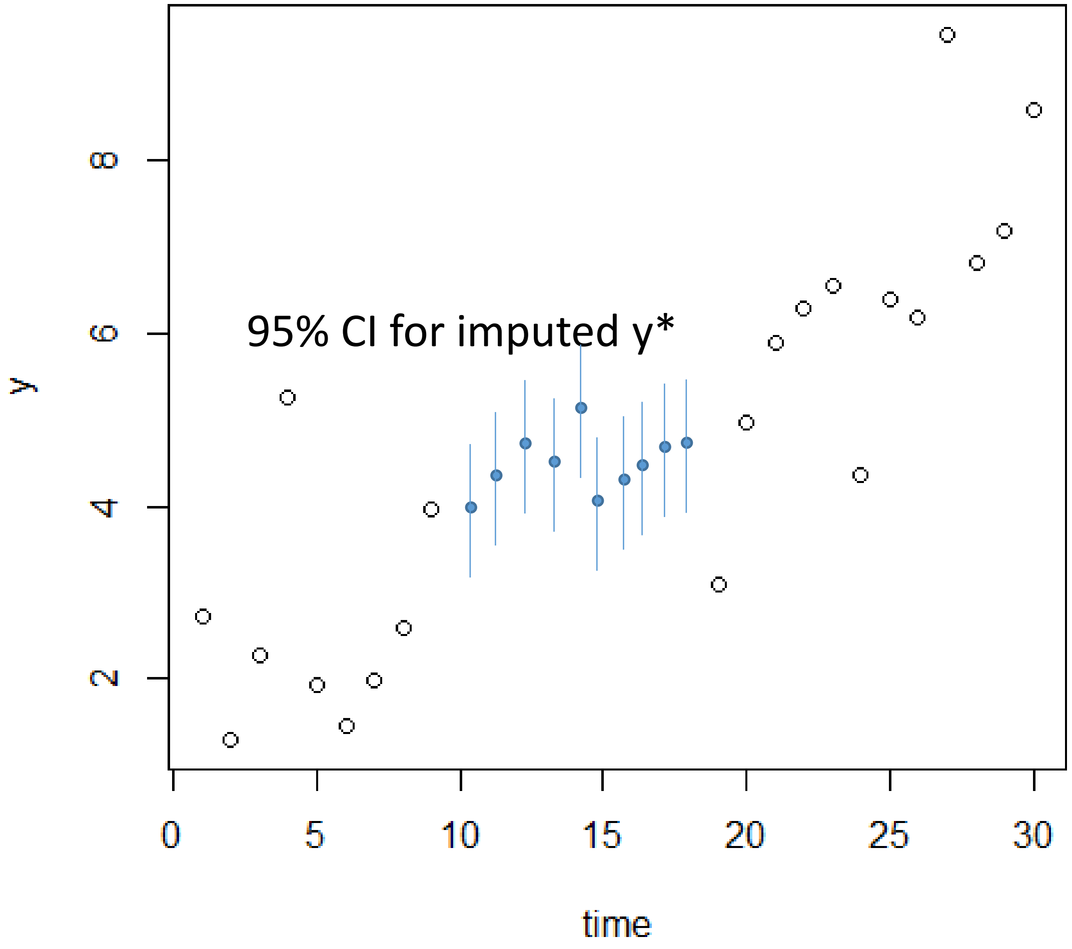


Missing Data Model

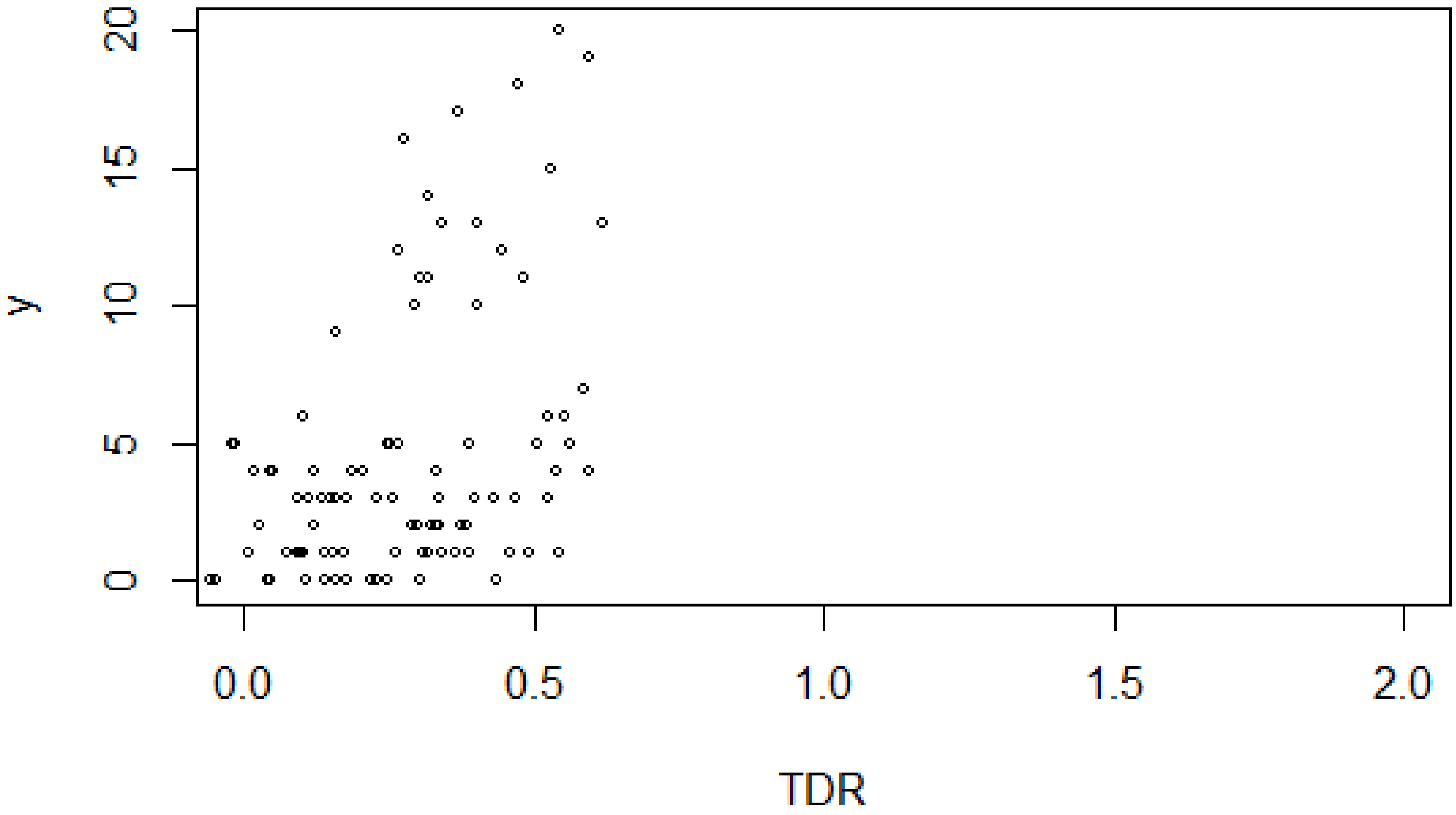
$$y* \sim N(\beta X, s^2)$$



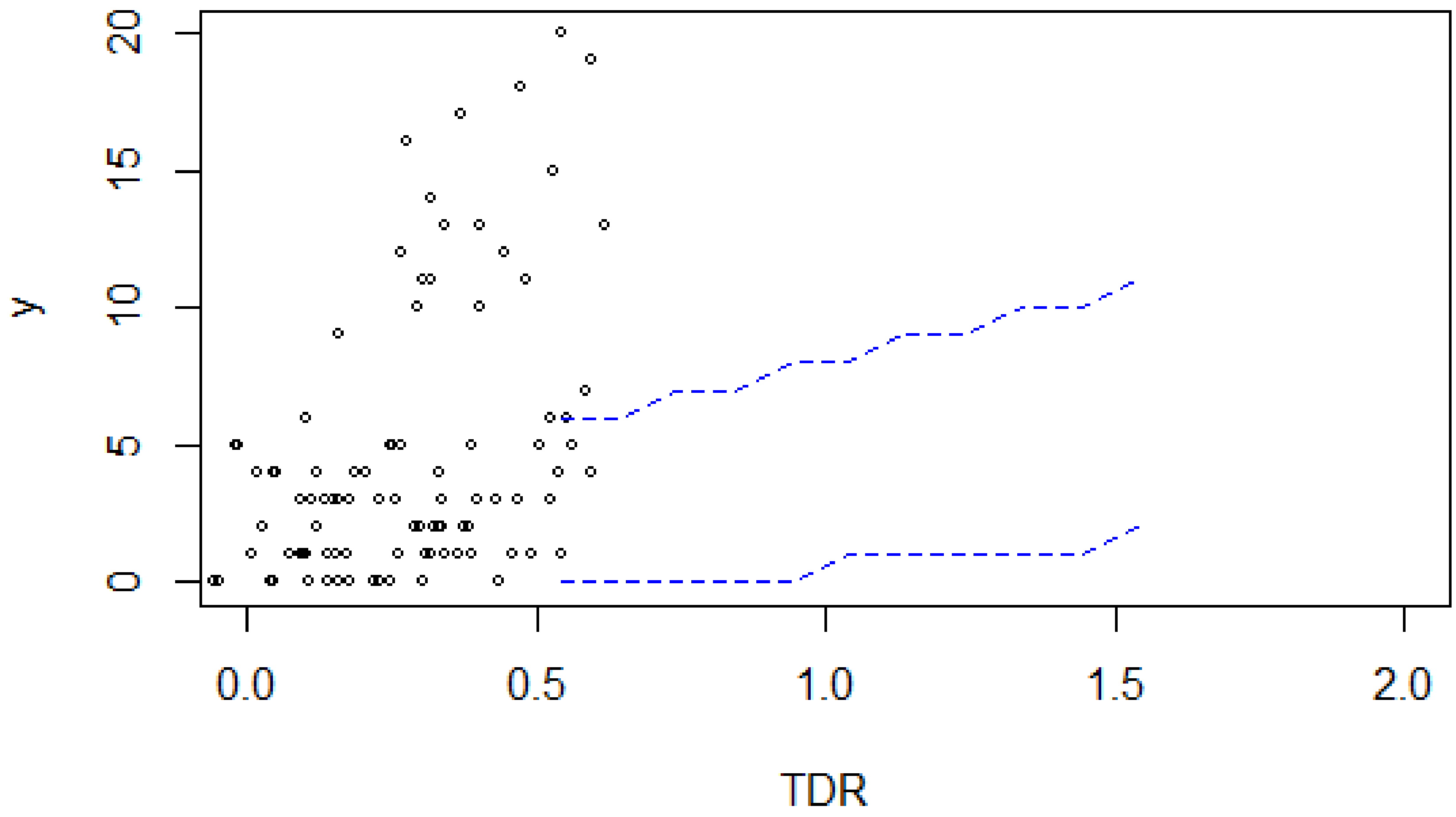
Missing Data



Missing Data



Missing Data



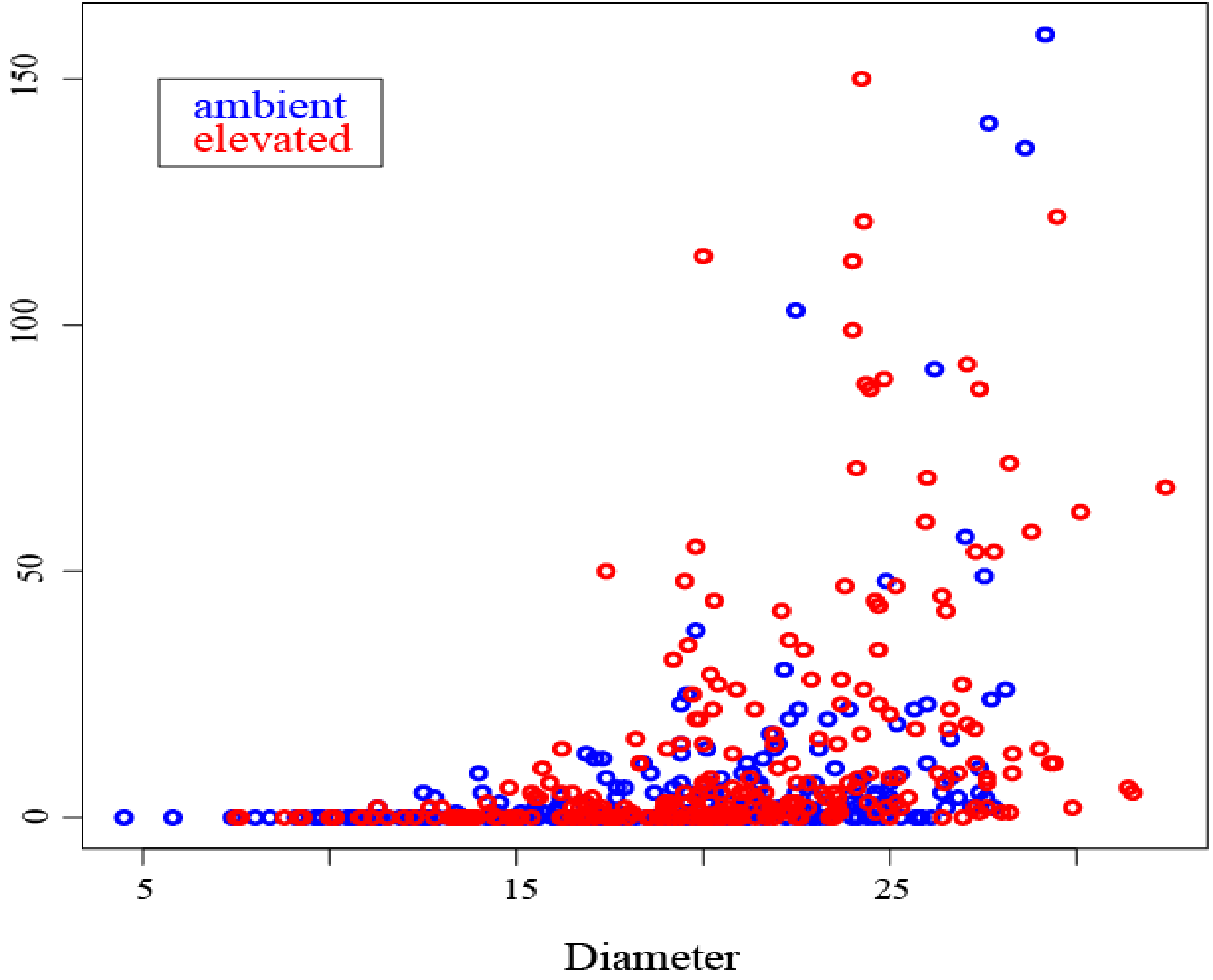
Latent Variables

- Any variable not directly observed
  - Missing data
  - Variable measured with error (e.g., detection models)
  - **Proxy measures –estimating variables never observed**

Fecundity of trees is often related to tree size:

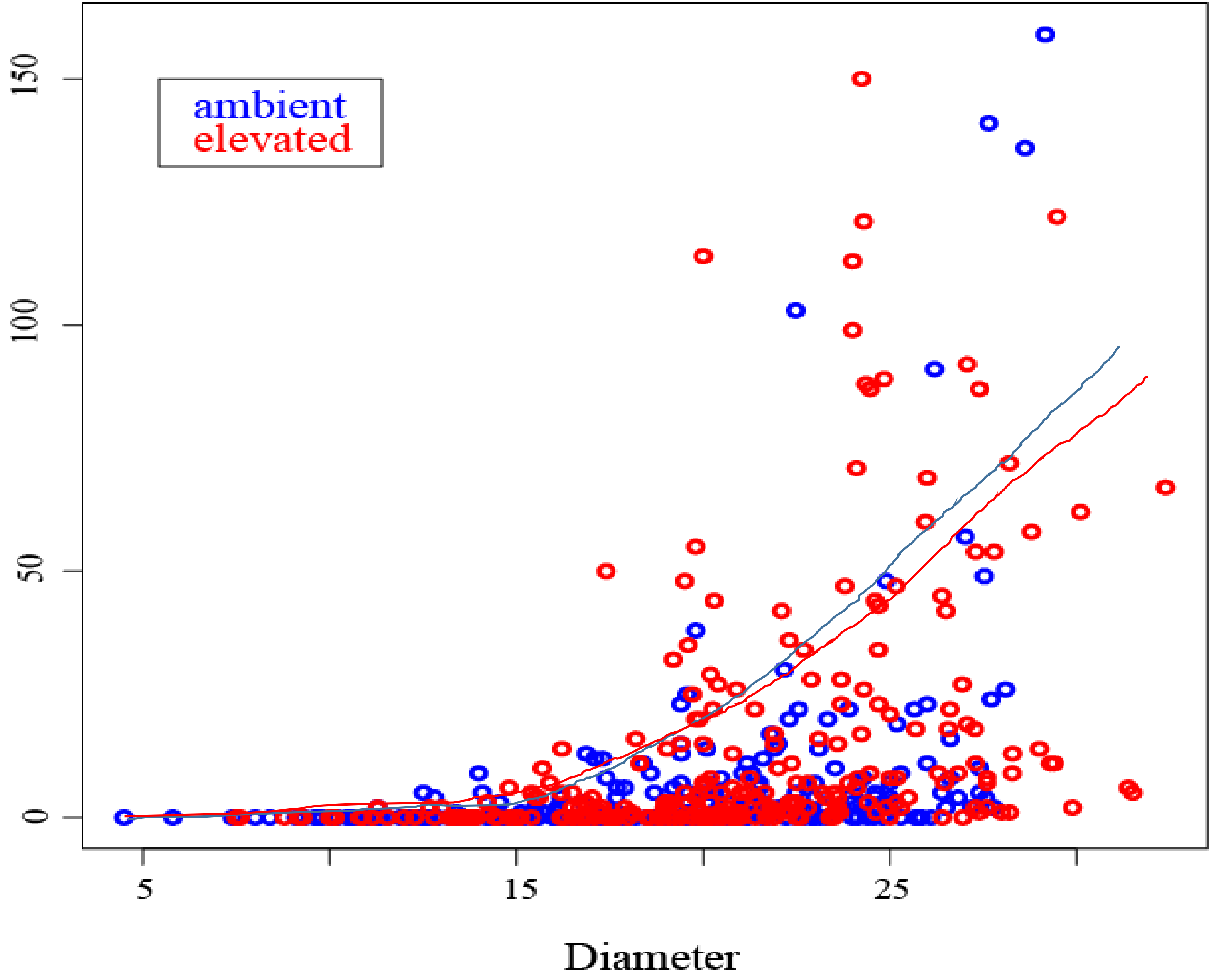
seeds= fxn(diam)

Total Seed Cones

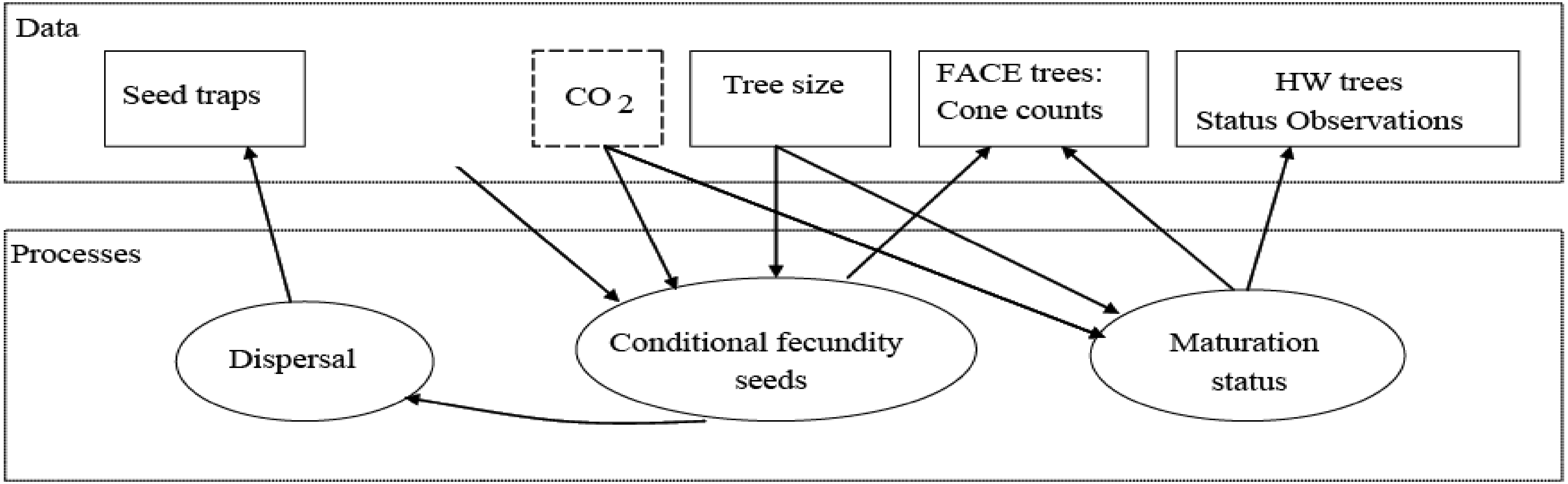


Regression would suggest more cones per diameter for ambient trees (blue line).

Total Seed Cones

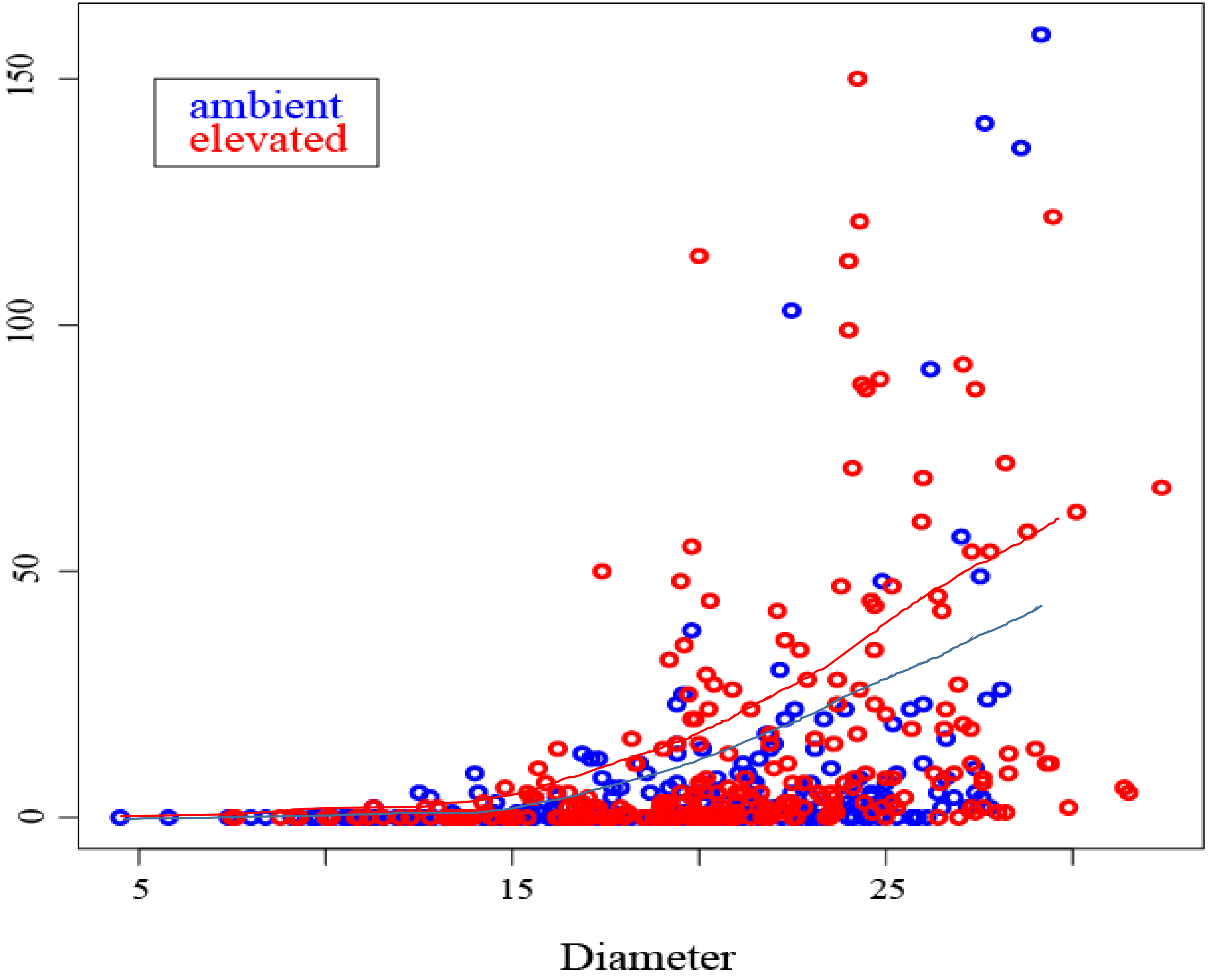


Modeling Fecundity



Accounting for tree maturation and fecundity highlight that many large ambient trees produce 0 seeds.

True Fecundity



Latent Variables=

when variable of interest is not exactly what you measure

Ignoring variable latency (e.g., modeling a derived response or flawed observation) can lead to <u>incorrect</u> or <u>overconfident</u> conclusions