Linear Models

- Variety of linear models
- MLE derivation of parameters and se's [ref]
- Comparison to Bayesian
- **Assumptions of linear models**
- **Relaxing these assumptions**

Linear models

- Statistically, a model is judged based on whether it is linear or not with respect to the PARAMETERS

$$y = \beta_0 + \beta_1 x_1$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

$$y = \beta_0 + \beta_1 x + \beta_2 x^2$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 \cdot x_2$$

$$y = \beta_0 + \beta_1 \ln(x_1) + \beta_2 \exp(x_2)$$

$$y = \beta_0 + \beta_1 I(TRT1) + \beta_2 I(TRT2)$$

Recall for simple linear model

$$y = \beta_0 + \beta_1 x + \epsilon$$

$$\beta_1 = \frac{cov[x, y]}{var[x]}$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

Parameter CI by Fisher Information

$$lnL = -\frac{n}{2}\ln(2\pi\sigma^2) - \sum \frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}$$

$$\frac{\partial\, lnL}{\partial \beta_0} = \frac{1}{\sigma^2}\sum (y_i - \beta_0 - \beta_1 x_i) = \frac{1}{\sigma^2}\left[\sum y_i - n\beta_0 - \beta_1 \sum x_i\right]$$

$$\frac{\partial\, lnL}{\partial \beta_1} = \frac{1}{\sigma^2}\sum x_i (y_i - \beta_0 - \beta_1 x_i)$$

$$= \frac{1}{\sigma^2}\left[\sum x_i y_i - \beta_0 \sum x_i - \beta_1 \sum x_i^2\right]$$

$$\frac{\partial \ln L}{\partial \beta_0} = \frac{1}{\sigma^2}\left[\sum y_i - n\beta_0 - \beta_1 \sum x_i\right]$$

$$\frac{\partial^2 \ln L}{\partial \beta_0^2} = -\frac{1}{\sigma^2}\left[n + \frac{\partial \beta_1}{\partial \beta_0}\sum x_i\right]$$

$$\frac{\partial^2 \ln L}{\partial \beta_0^2} = -\frac{n}{\sigma^2}\left[1 - \frac{\bar{x}^2}{\overline{x^2}}\right]$$

$$\frac{\partial^2 \ln L}{\partial \beta_0^2} = -\frac{n}{\sigma^2}\left[\frac{\overline{x^2} - \bar{x}^2}{\overline{x^2}}\right]$$

$$\frac{\partial^2 \ln L}{\partial \beta_0^2} = \frac{-n\, var[x]}{\sigma^2 \overline{x^2}}$$

From our MLE estimator

$$\beta_1 = \frac{\overline{xy} - \beta_0 \bar{x}}{\overline{x^2}}$$

$$\frac{\partial \beta_1}{\partial \beta_0} = \frac{-\bar{x}}{\overline{x^2}}$$

$$se_{\beta_0} = \frac{1}{\sqrt{I_{\beta_0}}}$$

$$se_{\beta_0} = \sigma \sqrt{\frac{\overline{x^2}}{n\, var[x]}}$$

$$\frac{\partial \ln L}{\partial \beta_1} = \frac{1}{\sigma^2} \left[\sum x_i y_i - \beta_0 \sum x_i - \beta_1 \sum x_i^2 \right]$$

$$\frac{\partial^2 \ln L}{\partial \beta_1^2} = \frac{1}{\sigma^2} \left[\frac{-\partial \beta_0}{\partial \beta_1} \sum x_i - \sum x_i^2 \right]$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

$$\frac{\partial \beta_0}{\partial \beta_1} = -\bar{x}$$

$$\frac{\partial^2 \ln L}{\partial \beta_1^2} = \frac{1}{\sigma^2} \left[\bar{x} \sum x_i - \sum x_i^2 \right]$$

$$\frac{\partial^2 \ln L}{\partial \beta_1^2} = -\frac{n}{\sigma^2} var[x]$$

$$se_{\beta_1} = \frac{\sigma}{\sqrt{n \, var[x]}}$$

Multiple Regression via MLE

- Recall from our Bayesian derivation that we can express the regression likelihood in matrix form

$$\vec{y}|\vec{\beta},\sigma^2 \sim N(\boldsymbol{X}\vec{\beta},\sigma^2)$$

$$L \propto \sigma^{-n} \exp\left[\frac{-(y-\boldsymbol{X}\vec{\beta})^T(y-\boldsymbol{X}\vec{\beta})}{2\sigma^2}\right]$$

$$lnL \propto -n\ln(\sigma) - \frac{(y-\boldsymbol{X}\vec{\beta})^T(y-\boldsymbol{X}\vec{\beta})}{2\sigma^2}$$

$$lnL \propto -n\ln(\sigma) - \frac{1}{2\sigma^2}\left[y^T y - y^T X\beta - \beta^T X^T y + \beta^T X^T X\beta\right]$$

$$lnL \propto -n\ln(\sigma) - \frac{1}{2\sigma^2}\left[y^T y - y^T X\beta - \beta^T X^T y + \beta^T X^T X\beta\right]$$

Vector derivative properties

$$\frac{\partial A\beta}{\partial\beta} = \frac{\partial\beta^T A}{\partial\beta^T} = A \qquad \frac{\partial\beta^T A\beta}{\partial\beta} = \beta^T A^T + \beta^T A$$

$$\frac{\partial lnL}{\partial\beta} \propto \frac{-1}{2\sigma^2}\left[-2y^T X + 2\beta^T X^T X\right] = 0$$

$$y^T X = \beta^T X^T X$$

$$X^T y = X^T X\beta$$

$$\beta = (X^T X)^{-1} X^T y$$

$$\beta_1 = \frac{cov[x, y]}{var[x]}$$

MLE vs Bayes

$$\beta = (X^T X)^{-1} X^T y$$

$$\sigma^2 = (y - X\beta)^T (y - X\beta)/n$$

$$\beta \sim N\left((\sigma^{-2} X^T X + V_b^{-1})^{-1} (\sigma^{-2} X^T \vec{y} + V_b^{-1} \vec{b}_0), \right.$$
$$\left. (\sigma^{-2} X^T X + V_b^{-1})^{-1} \right)$$

$$\sigma^2 \sim IG\left(s_1 + \frac{n}{2},\ s_2 + \frac{1}{2} (\vec{y} - Xb)^T (\vec{y} - Xb) \right)$$

Assumptions of Linear Model

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

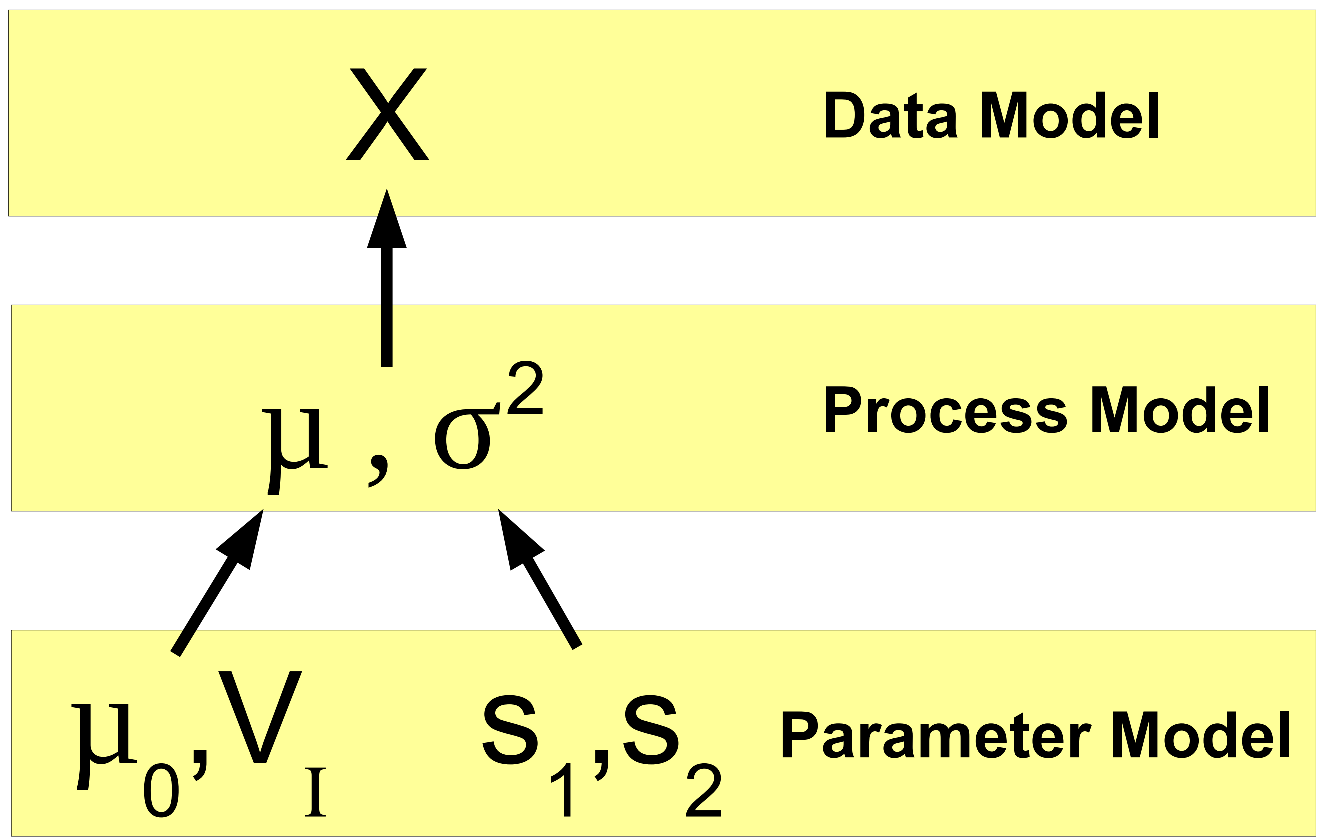
Graph notation

- Focuses on relationships among parameters and data sets rather than distributions
- Can facilitate writing conditional distributions

$$X \sim N(\mu, \sigma^2)$$

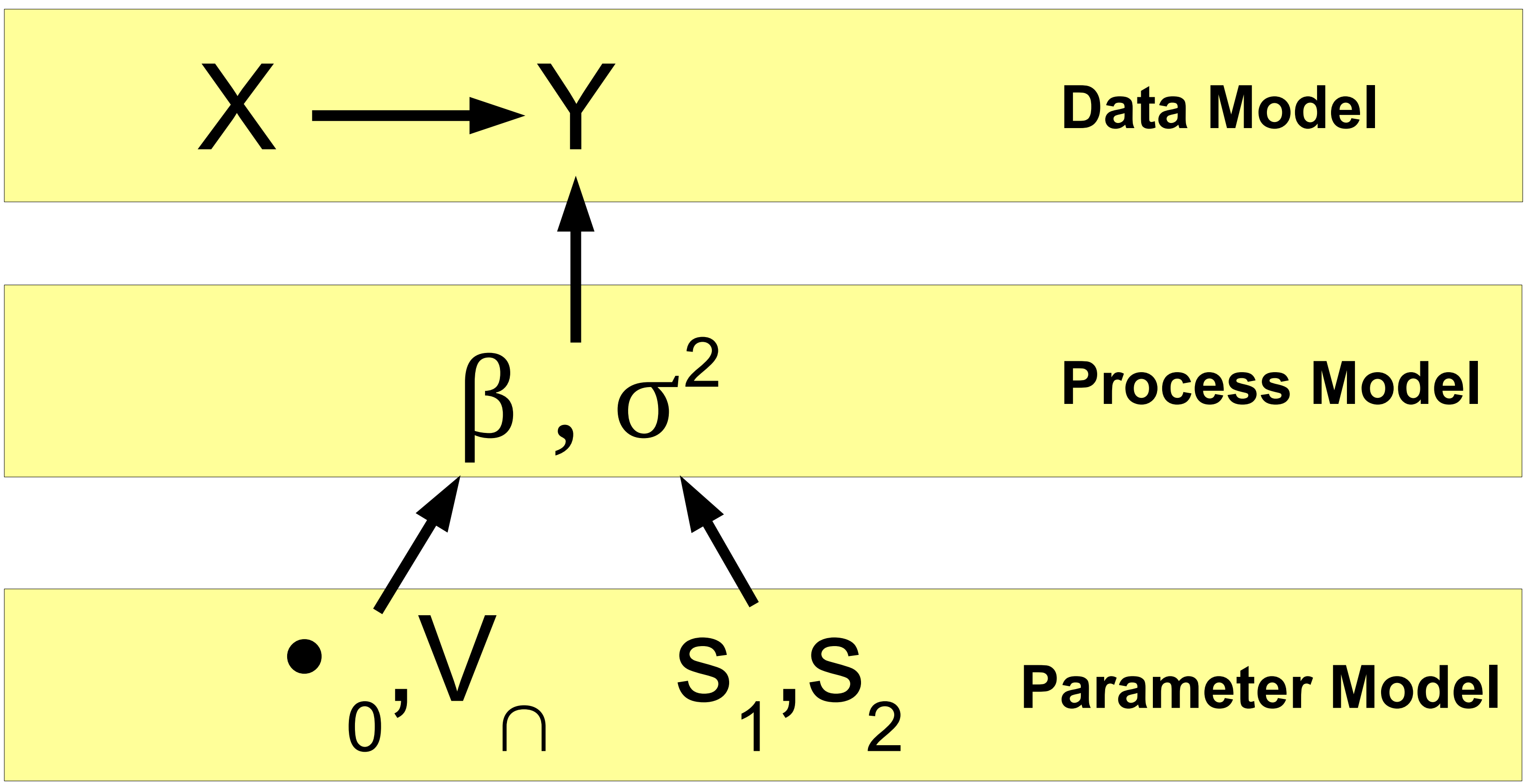
$$\mu \sim N(\mu_0, V_\mu)$$

$$\sigma^2 \sim IG(s_1, s_2)$$

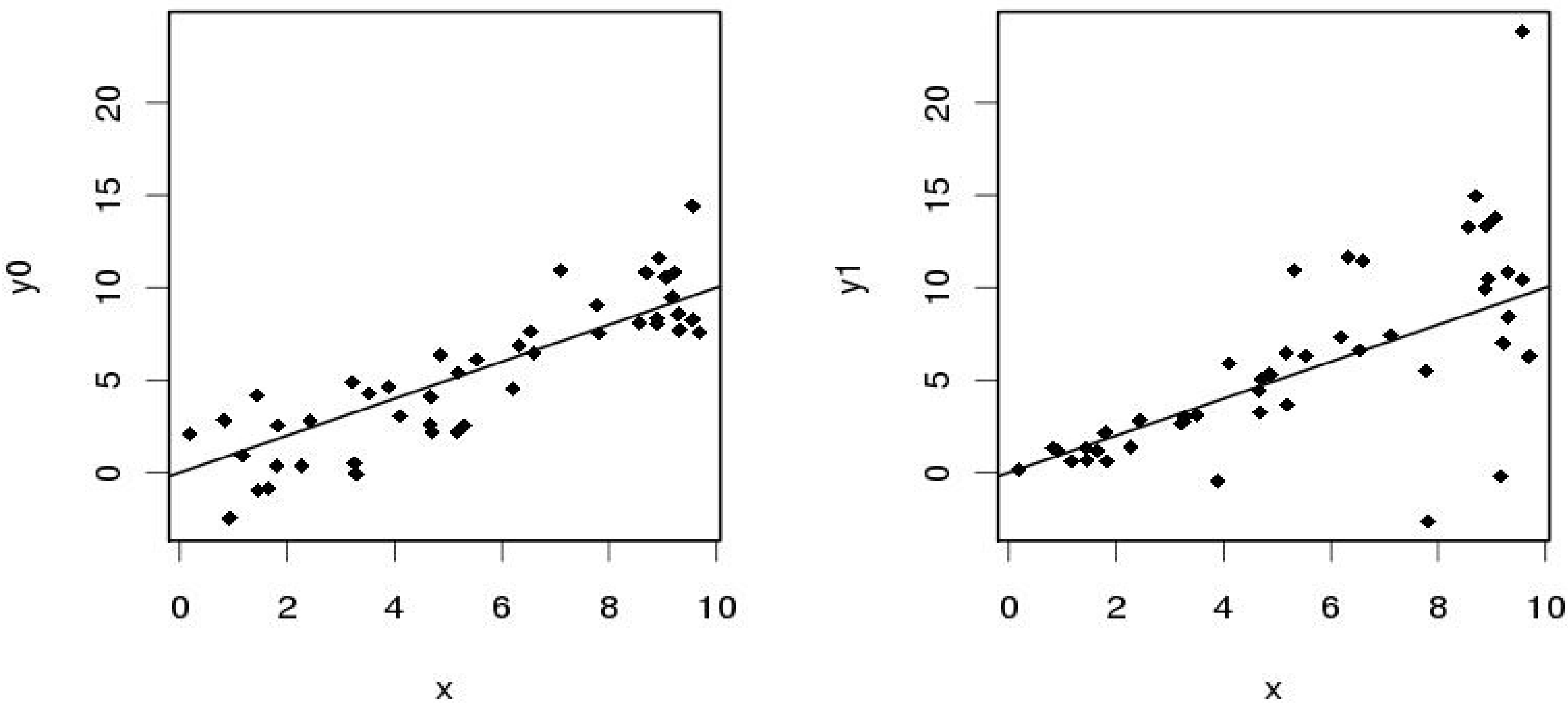


Linear Regression

$$\vec{y} \sim N(\boldsymbol{X}\vec{\beta}, \sigma^2)$$



Heteroskedasticity

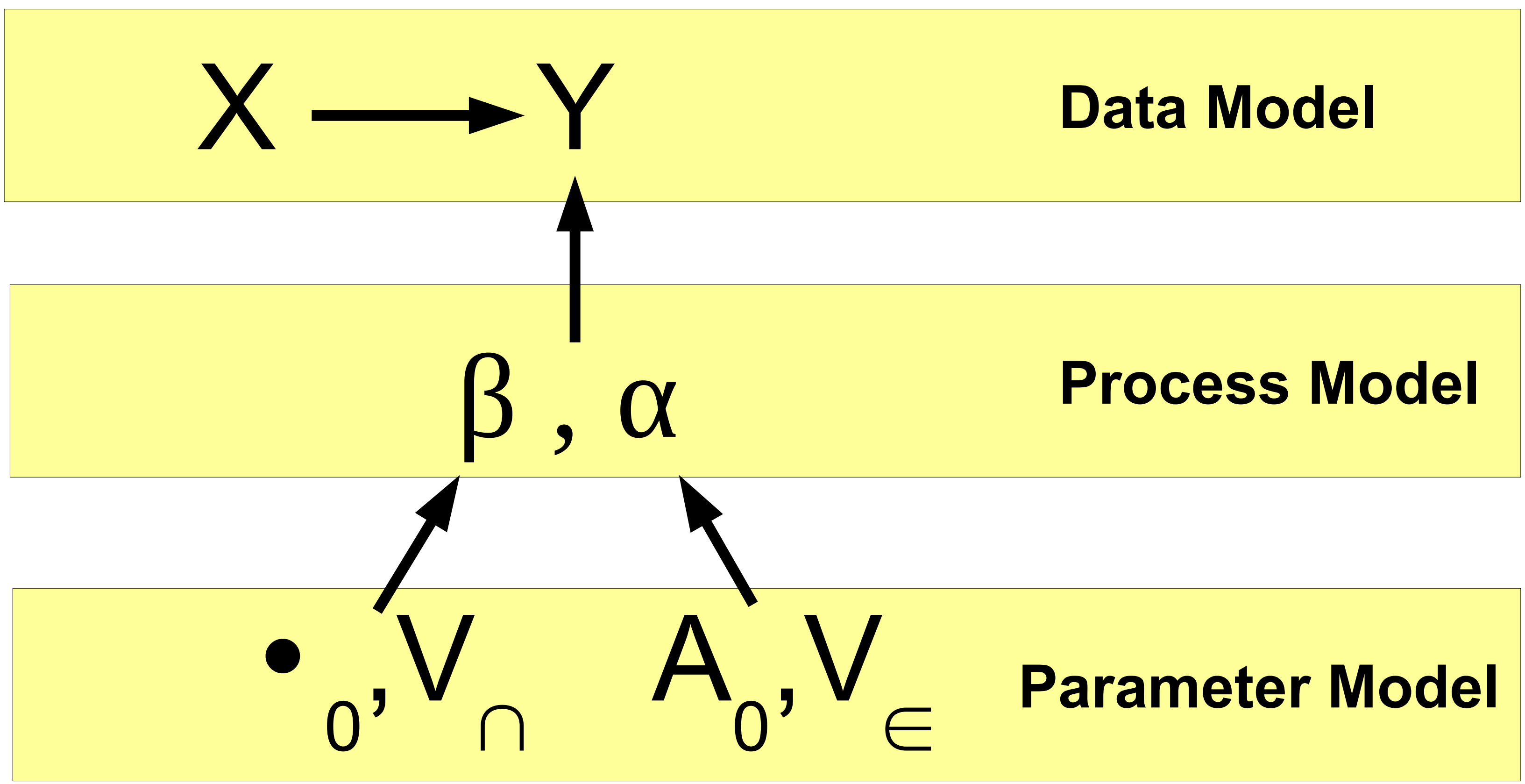


Solutions

1) Transform the data
   1) Pro: No additional parameters
   2) Cons: No longer modeling the original data, likelihood & process model have different meaning, backtransformation non-trivial (Jensen's Inequality)
2) Model the variance
   1) Pro: working with original data and model, no tranf.
   2) Con: additional process model and parameters (and priors)

Heteroskedasticity

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



Example: Linear varying SD

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

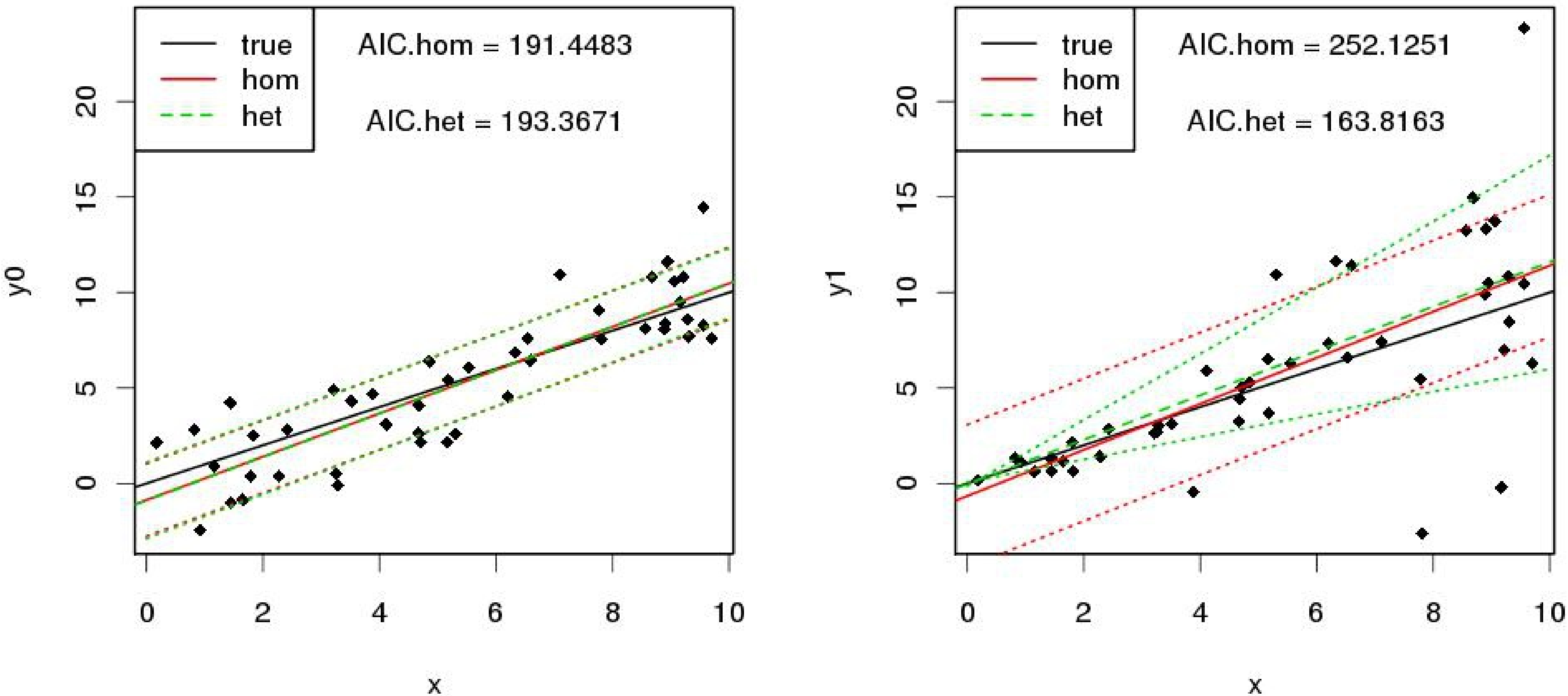
Likelihood (R)

```
LnL = function(theta,x,y){
    beta = theta[1:2]
    alpha = theta[3:4]  ## was sigma = theta[3]
    -sum(dnorm(y,beta[1]+beta[2]*x,alpha[1]+alpha[2]*x,log=TRUE))
}
```

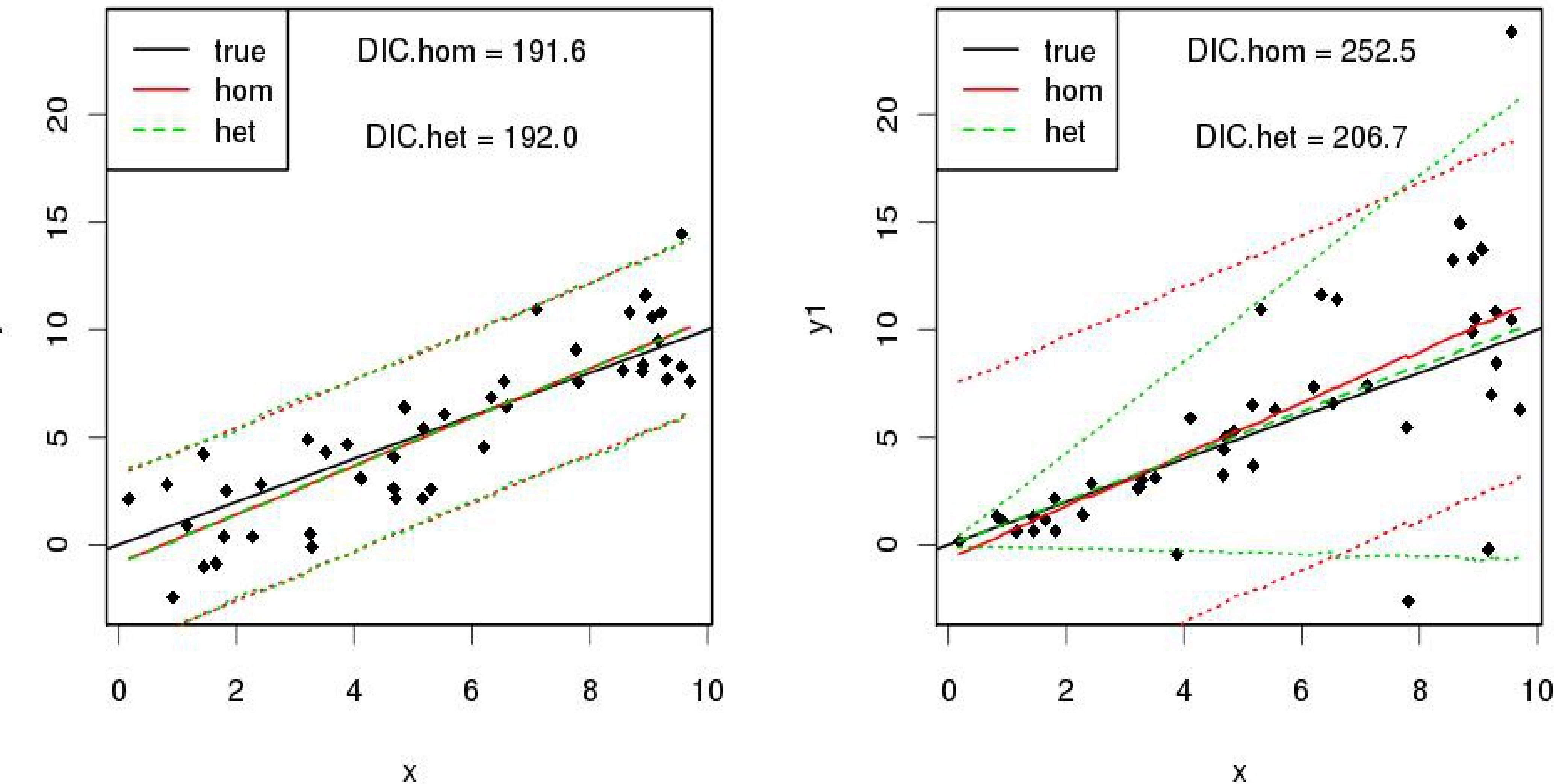
Bayes (JAGS)

```
model{
  for(i in 1:2) { beta[i] ~ dnorm(0,0.001)}  ## priors
  for(i in 1:2) { alpha[i] ~ dlnorm(0,0.001)}  ## was prec ~ gamma(a1,a2)
  for(i in 1:n){
       prec[i] <- 1/pow(alpha[1] + alpha[2]*x[i],2)
       mu[i] <- beta[1]+beta[2]*x[i]
       y[i] ~ dnorm(mu[i],prec[i])
  }
}
```

Likelihood


true
hom
het
AIC.hom = 191.4483
AIC.het = 193.3671
y0
x
0
5
10
15
20
0
2
4
6
8
10
true
hom
het
AIC.hom = 252.1251
AIC.het = 163.8163
y1
x

Bayes



true
hom
het
DIC.hom = 191.6
DIC.het = 192.0
x
true
hom
het
DIC.hom = 252.5
DIC.het = 206.7
y1
x

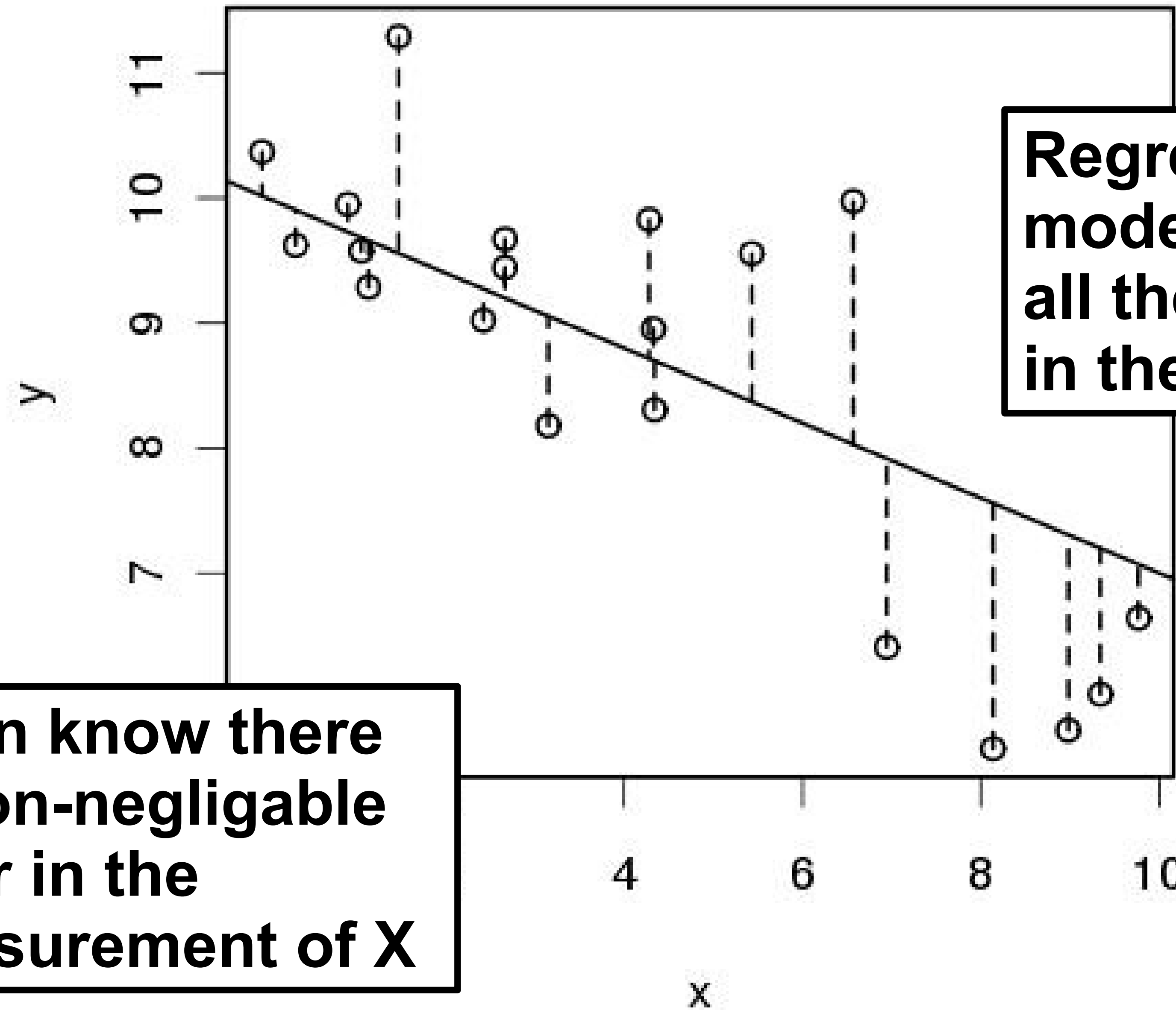
Additional thoughts on modeling variance

- Need not be linear
- Can model in terms of sd, variance, or precision
- Can vary with treatments/factors or categorical variables
  - e.g. can relax the ANOVA assumptions of equal variance among treatments
- Can vary by measurement technique, sensor, etc.

Assumptions of Linear Model

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

Errors in Variables



Regression model assumes all the error is in the Y

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

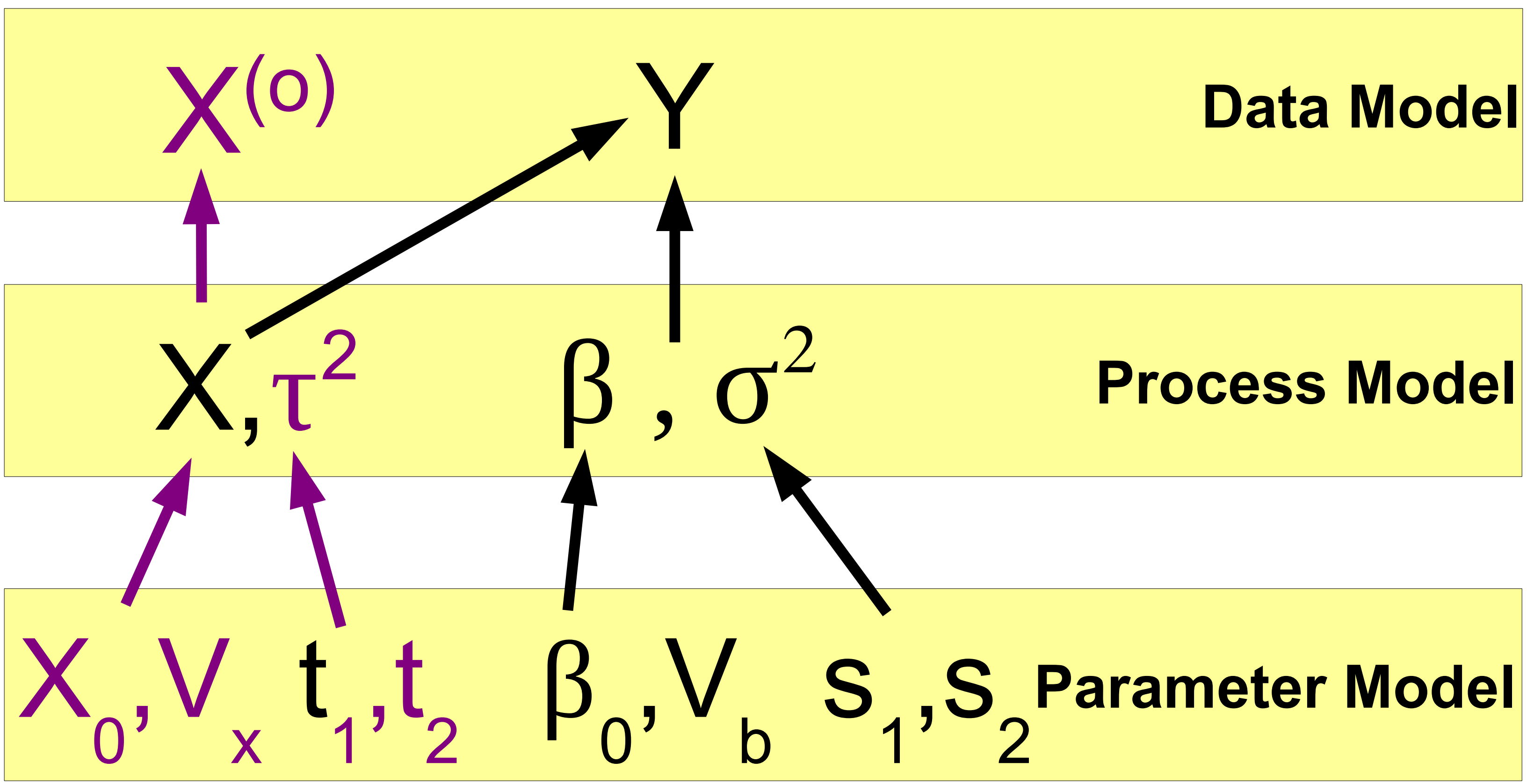
Errors in Variables

$\mu = \beta_1 + \beta_2 x$	**Process model**
$y \sim N(\mu, \sigma^2)$	**Data model for y**
$x^{(o)} \sim N(x, \tau^2)$	**Data model for x**
$\vec{\beta} \sim N(B_{0,} V_B)$	**Prior for beta**
$\sigma^2 \sim IG(s_{1,} s_2)$	**Prior for sigma**
$\tau^2 \sim IG(t_{1,} t_2)$	**Prior for tau**
$x \sim N(X_{0,} V_X)$	**Prior for X**

Errors in Variables

$$\vec{y} \sim N(\boldsymbol{X}\vec{\beta}, \sigma^2)$$

$$x^{(o)} \sim N(x, \tau^2)$$



Full Posterior

$$p(\vec{\beta}, \sigma^2, \tau^2, \boldsymbol{X} | \vec{y}, \boldsymbol{X}^{(o)}) \propto N(y | \beta_0 + \beta_1 x, \sigma^2)$$

$$N(x^{(o)} | x, \tau^2)\, N(\vec{\beta} | B_{0,} V_B)$$

$$IG(\sigma^2 | s_{1,} s_2)\, IG(\tau^2 | t_{1,} t_2)$$

$$N(x | X_{0,} V_X)$$

Conditionals

$$p(\vec{\beta} | \ldots) \propto N(y | \beta_0 + \beta_1 x, \sigma^2)\, N(\vec{\beta} | B_{0,} V_B)$$

$$p(\sigma^2 | \ldots) \propto N(y | \beta_0 + \beta_1 x, \sigma^2)\, IG(\sigma^2 | s_{1,} s_2)$$

$$p(\tau^2 | \ldots) \propto N(x^{(o)} | x, \tau^2)\, IG(\tau^2 | t_{1,} t_2)$$

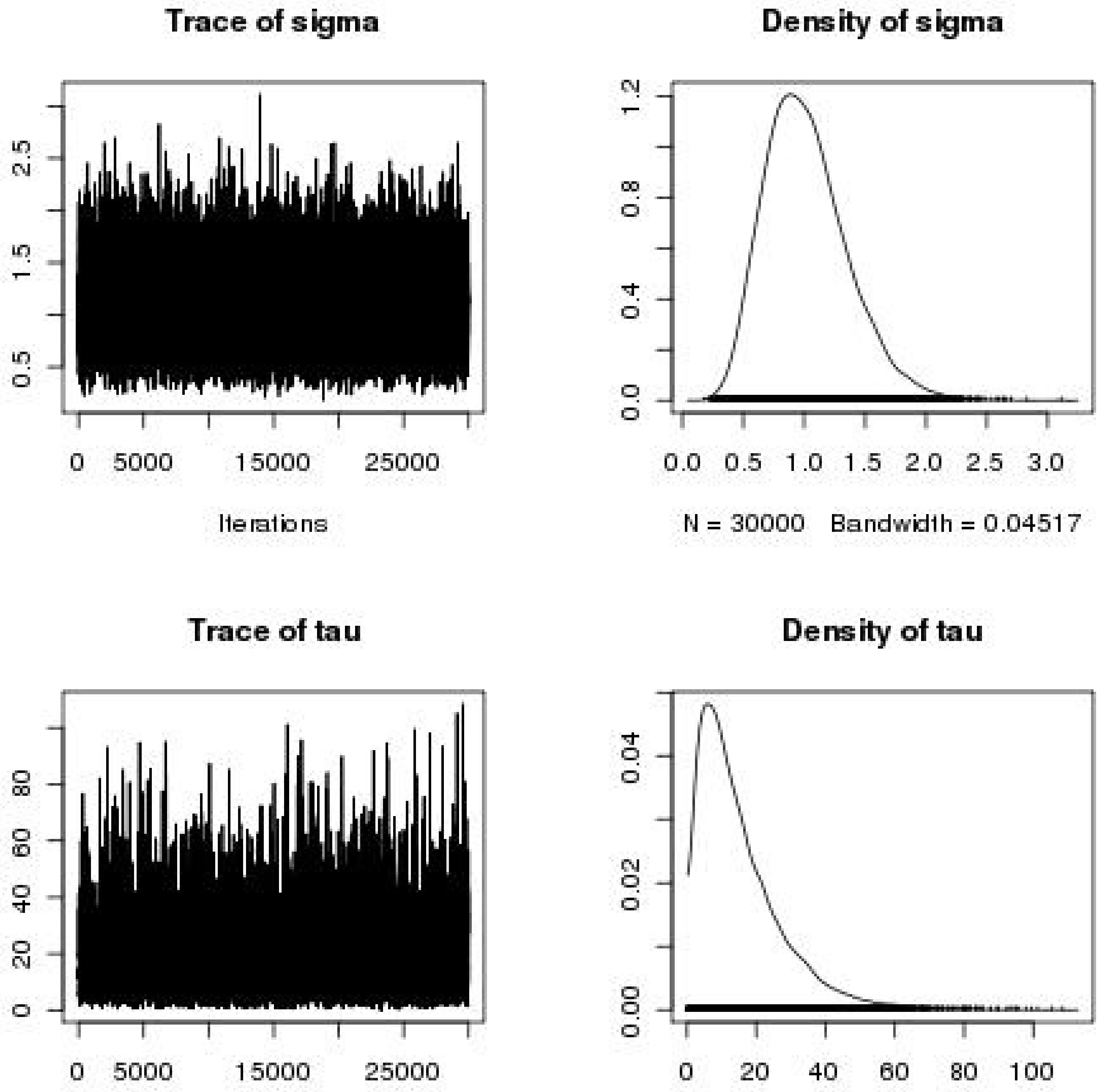
$$p(\boldsymbol{X} | \ldots) \propto N(x^{(o)} | x, \tau^2)\, N(y | \beta_0 + \beta_1 x, \sigma^2)\, N(x | X_{0,} V_X)$$

Conceptually within the MCMC

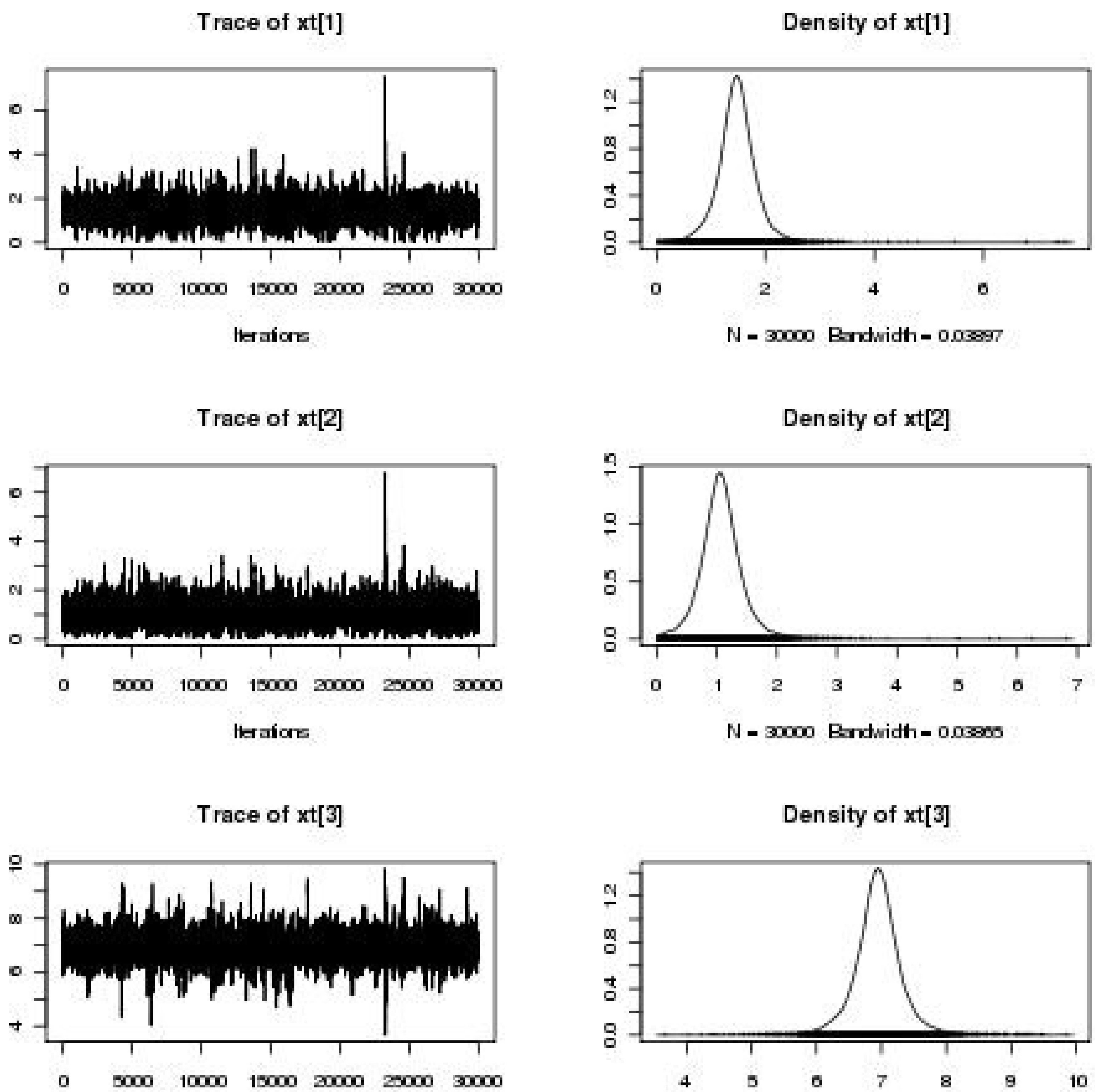
- Update the regression model given the current values of X
- Update the observation error in X based on the difference between the current and observed values of X
- Update the values of X based on the observed values of X <u>and</u> the regression model

- Overall, integrate over the possible values of X

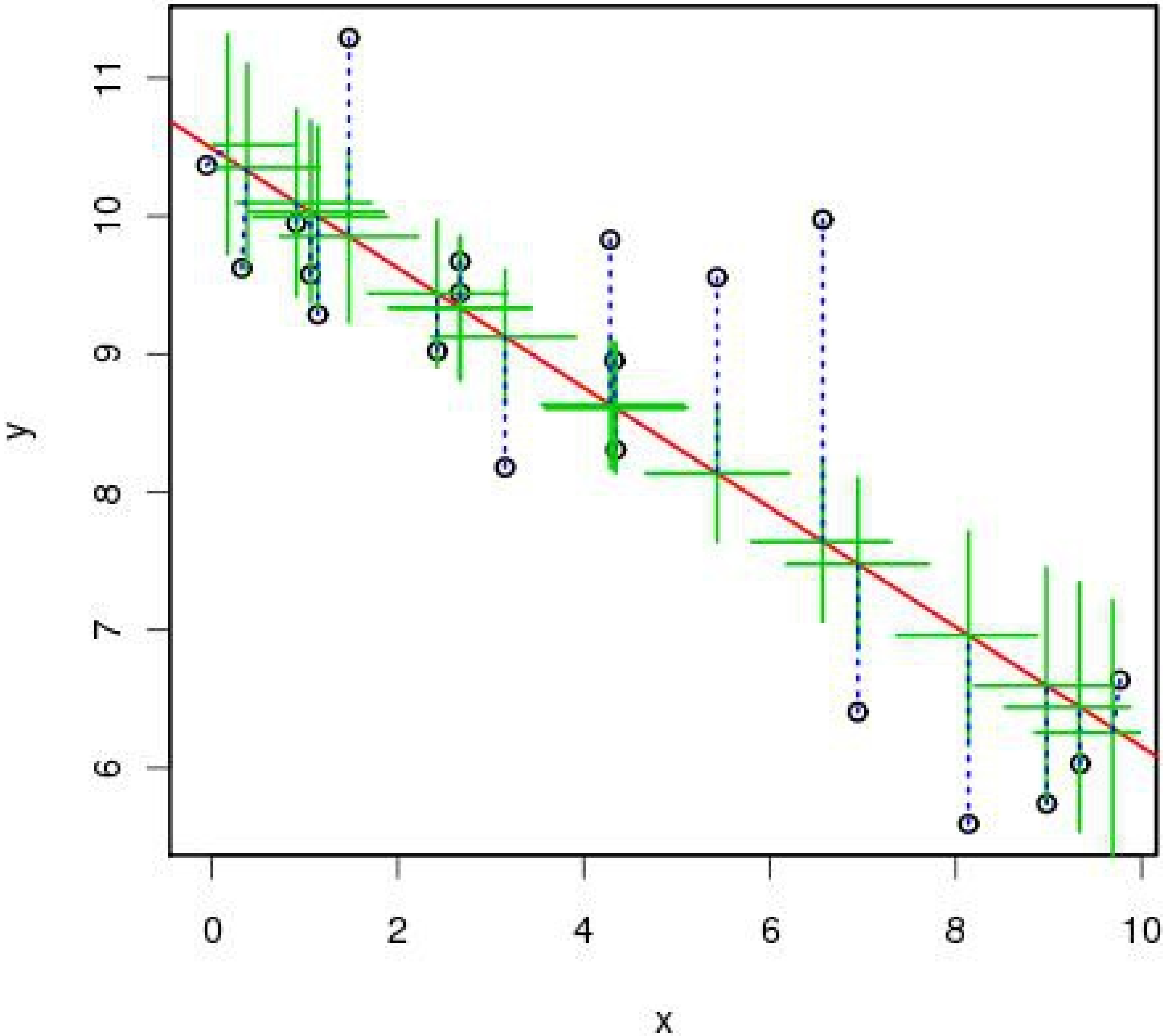
```
model {
  ## priors
  for(i in 1:2) { beta[i] ~ dnorm(0,0.001)}
  sigma ~ dgamma(0.1,0.1)
  tau ~ dgamma(0.1,0.1)
  for(i in 1:n) { x[i] ~ dunif(0,10)}

  for(i in 1:n){
    xo[i] ~ dnorm(x[i],tau)
    mu[i] <- beta[1]+beta[2]*x[i]
    y[i] ~ dnorm(mu[i],sigma)
  }
}
```



Trace of sigma
Iterations
Density of sigma
N = 30000 Bandwidth = 0.04517
Trace of tau
Density of tau



Trace of xt[1]
Iterations
Density of xt[1]
N = 30000 Bandwidth = 0.03897
Trace of xt[2]
Iterations
Density of xt[2]
N = 30000 Bandwidth = 0.03865
Trace of xt[3]
Density of xt[3]



11
10
9
8
7
6
y
0
2
4
6
8
10
x

Additional Thoughts on EIV

$$x^{(o)} \sim g(x|\theta)$$

- Errors in X's need not be Normal
- Errors need not be additive
- Can account for known biases

$$x^{(o)} \sim N(\alpha_0 + \alpha_1 x, \tau^2)$$

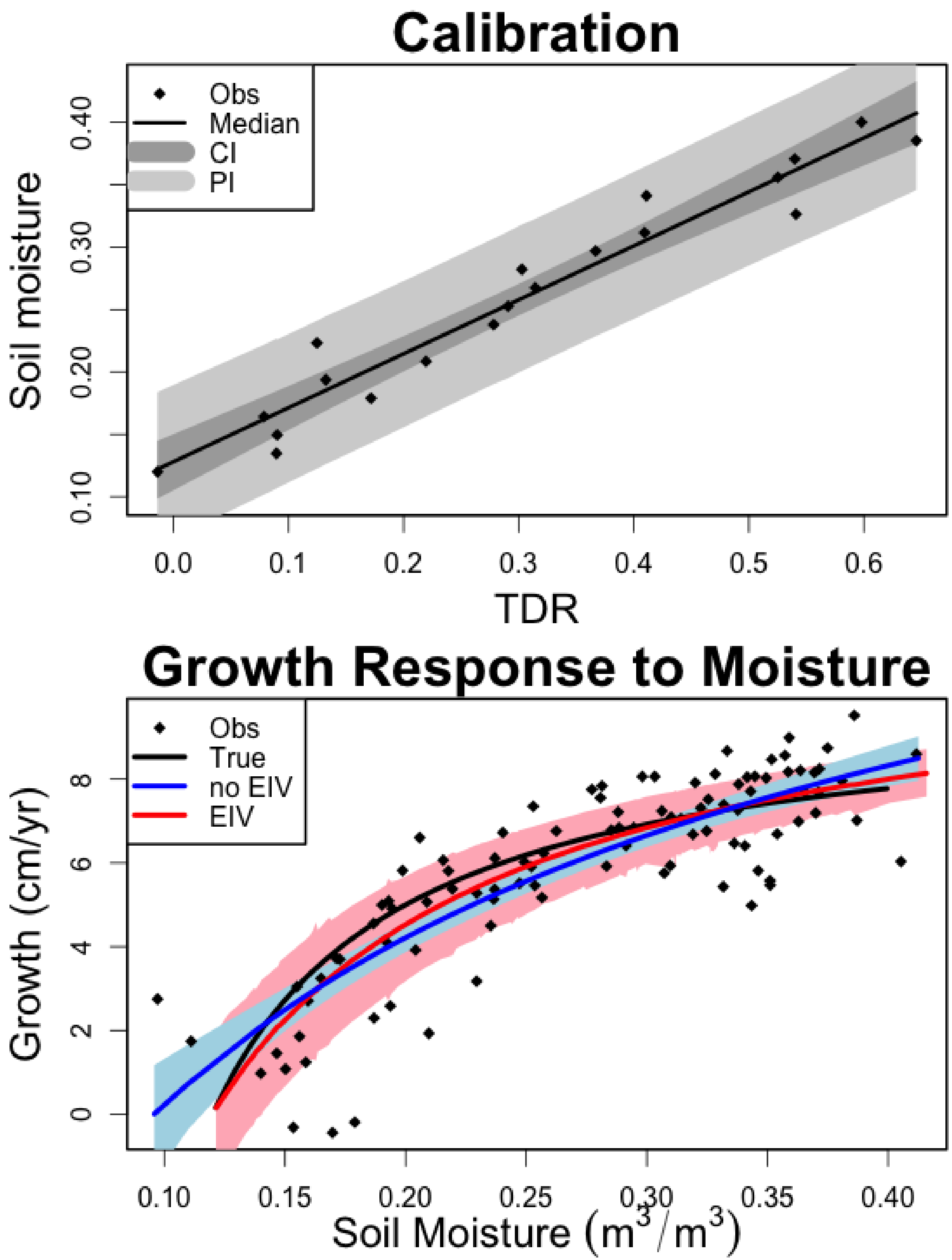
Additional Thoughts on EIV

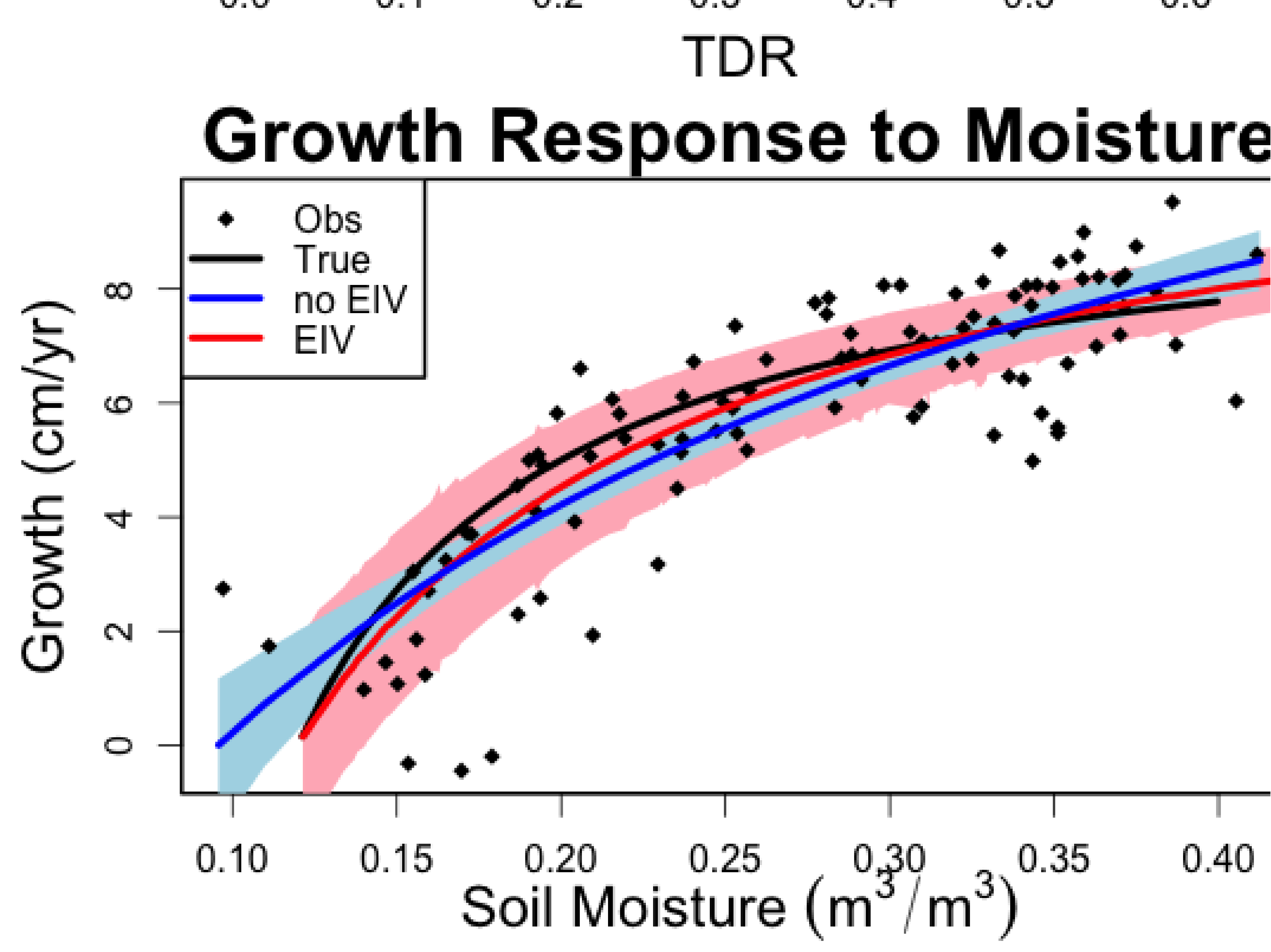
$$x^{(o)} \sim g(x|\theta)$$

- Errors in X's need not be Normal
- Errors need not be additive
- Can account for known biases

$$x^{(o)} \sim N(\alpha_0 + \alpha_1 x, \tau^2)$$

- Observed data can be a different type (proxy)
- Very useful to have informative priors


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 (m^3/m^3)


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)

Latent Variables

- Variables that are not directly observed
- Values are inferred from model
  - Parameter model: prior on value
  - Data and Process models provide constraint

$$p(\boldsymbol{X}|\ldots) \propto N(y|\beta_0+\beta_1 x, \sigma^2)\, N(x^{(o)}|x, \tau^2)\, N(x|X_{0,} V_X)$$

- MCMC integrates over (by sampling) the values the unobserved variable could take on
- Contribute to uncertainty in parameters, model
- Ignoring this variability can lead to falsely <u>overconfident</u> conclusions

Assumptions of Linear Model

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