Comments on "Measurement Error Models in Astronomy" by Brandon C. Kelly

David Ruppert

Operations Research & Information Engineering
and Department of Statistical Science, Cornell University

Jun 15, 2011

- These comments are personal views about measurement error modeling from someone who has
  - worked in statistics for over 35 years
  - worked on measurement error for over 25 years
  - worked in astrostatistics for roughly one year

- There are many special-purpose methods for handling measurement errors for particular models with certain assumptions
- What is proposed here is a general approach that can be used in most, if not all, situations

My Perspective on Measurement Error Modeling

- Bayesian approaches have much to offer
- I prefer structural models
  - a Bayesian approach is inherently structural
- Flexible parametric structural models can avoid problems of low-dimensional parametric structural models
- Careful modeling is essential
  - Example: pitfalls of orthogonal regression

Advantages of Bayesian Models

There are several advantages to taking a Bayesian approach to measurement error modeling

1. Focuses attention on careful modeling
2. Make efficient use of information in data
   - asymptotically efficient
   - optimal according to decision theory
   - all admissible estimators are Bayes for some prior
3. Inference is straightforward using credible intervals
   - these are similar in practice to confidence intervals

Advantages of Bayesian Models, cont.

4. Allows the use of prior information
   - but can use diffuse priors when there is little prior information
5. The true values of mismeasured data can be treated in the same way as unknown parameters
   - To a Bayesian, anything unknown is random
   - and one conditions on everything known
   - MCMC multiply imputes values of unknown true values of mismeasured variables
6. Bayesian analysis works for virtually any problem

Example: measurement error in a nonlinear model

Example: quadratic regression simulation

$Y_i = \alpha + \beta X_i + \gamma X_i^2 + \epsilon_i$ (regression model)

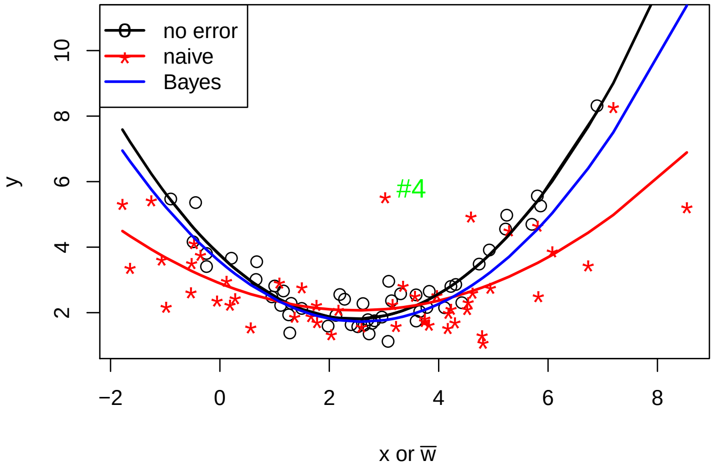
- $\epsilon_i \overset{\text{iid}}{\sim} N(0, \sigma_\epsilon^2)$

$W_{ij} = X_i + U_{ij}, \; j = 1, 2$ (measurement model)

- $U_{ij} \overset{\text{iid}}{\sim} N(0, \sigma_u^2)$
- σ_U^2 unknown
- $\overline{W}_i = (W_{i1} + W_{i2})/2$

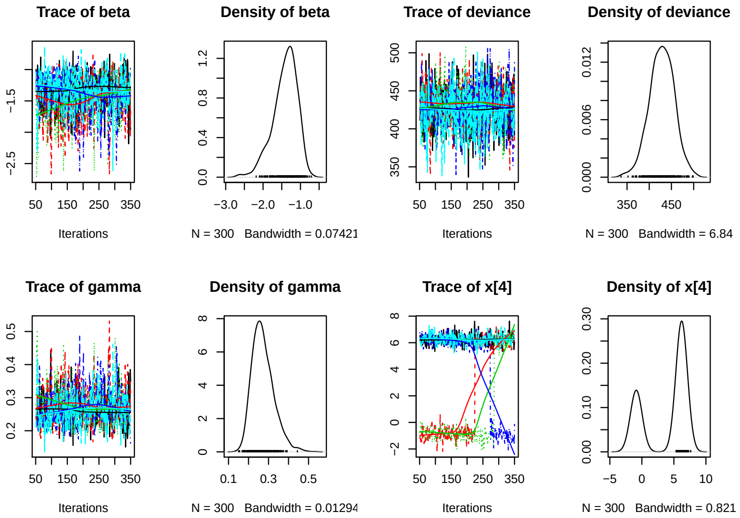
$X_i \overset{\text{iid}}{\sim} N(\mu_x, \sigma_x^2)$ (structural model)

Example: measurement error in a nonlinear model



Example: measurement error in a nonlinear model

R2WinBUGS output: 5 chains each of length 35,000



```
model{
for(i in 1:N){
w1[i] ~ dnorm(x[i],tauw)
w2[i] ~ dnorm(x[i],tauw)
x[i] ~ dnorm(mux,taux)
y[i] ~ dnorm(muy[i],taue)
muy[i] <- alpha + beta*x[i]+ gamma*x[i]*x[i]
}
mux ~ dnorm(0.0,1.0E-6)
alpha ~ dnorm(0.0,1.0E-6)
beta ~ dnorm(0.0,1.0E-6)
gamma ~ dnorm(0.0,1.0E-6)
tauw ~ dgamma(0.1,0.01)
taux ~ dgamma(0.1,0.01)
taue ~ dgamma(0.1,0.01)
}
```

It is often reasonable to assume that the true covariates $\xi_1, \ldots, \xi_n$ come from some probability distribution

- This assumption justifies using a structural model
- Any Bayesian model must be structural because
  - to a Bayesian, any unknown is random

Parsimonious structural models

- Structural models often assume that the distribution of the true covariates is Gaussian or in some other low-dimensional parametric family
  - This is done for simplicity and parsimony
- Often conclusions are robust (= insensitive) to this assumption
- But not always

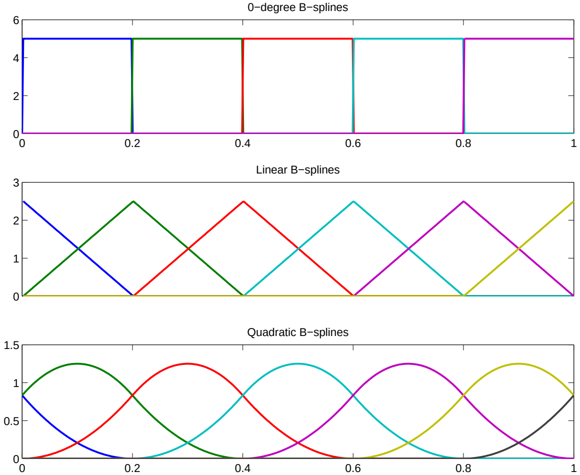
Flexible structural models

Sometimes we are worried about the nonrobustness of a structural model

- An alternative is to use a high-dimensional parametric family to model the true covariate distribution
- Flexible parametric families are, in effect, nonparametric
- Two examples
  - splines
  - mixture distributions
- One needs to guard against overfitting = undersmoothing
  - Bayesian methods can do this automatically

- B-splines are nonnegative and have minimal support
- They can be normalized to be densities
- A convex combination of normalized B-splines is a density

B-splines


0−degree B−splines
6
4
2
0
0
0.2
0.4
0.6
0.8
1
Linear B−splines
3
2
1
0
Quadratic B−splines
1.5
1
0.5
0
0
0.2
0.4
0.6
0.8
1

Density Estimation by Splines

Staudenmayer, J., Ruppert, D., and Buonaccorsi, J. (2008)

Density estimation in the presence of heteroskedastic measurement error, *JASA*, 103, 726–736.

- Estimates the variance function
  - this is the conditional variance of the measurement error given the true covariate value
- Uses splines
- Could be part of a structural model
- Bayesian
  - so can easily be modified to handle similar problems

Focus on Modeling, not Algorithms

- Carefully statistical modeling is always important
- Focusing on algorithms/estimators is potentially a distraction
- The distinction between measurement error and equation error was an important conceptual advance

- In simple cases, one replaces a constant error variance by an average variance
- The Akritas and Bershady (1996) is a nice example
- For nonparametric modeling (local estimation) this approach fails
- Staudenmayer and Ruppert found that for nonparametric density estimation using the average variance
  - overcorrects where the actual variance is smaller than average
  - undercorrects where the actual variance is larger than average
- Their Bayesian estimator does not have these flaws

Orthogonal Regression Setup

The orthogonal regression (OR) model is

$$\begin{aligned} y_{\text{true}} &= \beta_0 + \beta_1 X \qquad \text{(no equation error)} \\ Y &= y_{\text{true}} + \epsilon \\ W &= X + U \end{aligned}$$

It is assumed that we "know"

$$\eta = \frac{\text{var}(Y|X)}{\text{var}(W|X)} = \frac{\sigma_\epsilon^2}{\sigma_U^2}$$

Orthogonal Regression Estimator

The OR estimator can be viewed as a functional estimator that treats $X_1, \ldots, X_n$ as unknown parameters

$\beta_0, \beta_1, X_1, \ldots, X_n$ are estimated by minimizing

$$\sum_{i=1}^{n} \left\{ \eta^{-1}(Y_i - \beta_0 - \beta_1 X_i)^2 + (W_i - X_i)^2 \right\}$$

$$\propto \sum_{i=1}^{n} \left\{ \sigma_\epsilon^{-2}(Y_i - \beta_0 - \beta_1 X_i)^2 + \sigma_U^{-2}(W_i - X_i)^2 \right\}$$

over $(\beta_0, \beta_1, X_1, \ldots, X_n)$

- The danger is that it is easy to misapply OR in the presence of equation error
- This leads to overcorrection if one uses

$$\eta = \frac{\sigma_\epsilon^2}{\sigma_U^2}$$

- Instead one should use

$$\eta_{\mathsf{EE}} := \frac{\mathsf{var}(Y|X)}{\mathsf{var}(W|X)} = \frac{\sigma_Q^2 + \sigma_\epsilon^2}{\sigma_U^2}$$

  - σ_Q^2 is the equation error variance

Bayesian Inference for Very Large Data Sets

- Simple problems (measurement error in linear models)
  - look for low dimensional sufficient statistics
- Bliznyuk, Ruppert, Shoemaker (et al) (2008, 2011, 2012), JCGS
  - Bayesian inference with computationally expensive posterior densities
  - radial basis function emulator of log-posterior
  - adaptive design
  - focuses on high posterior density region
  - might be 0.1% of volume of parameter space
- Emulators for high-dimensional parameter spaces will be challenging
  - and for measurement error models true covariate values are parameters

A Bayesian framework focuses on the three components of the model

- structural model for the true covariates
- measurement model for the measurement errors
- regression model for the conditional distribution of the response given the true covariate values

After modeling, inference is relatively automatic (and efficient)