LA-UR- 01-4406

Hierarchical Bayes Regression with Correlated Measurement Errors

J. Morzinski[1], H. Martz[1], S. Reese[1], L. Denogean[2]
[1]Los Alamos National Laboratory, [2]Cornell University

Introduction

A physical process produces three elements, call them u, v, and w. We can measure these elements, but because of certain relationships between them, it turns out that we are interested in the ratio u/w. We would like to predict that ratio using another ratio, v/w. Thus, our predictor and response variables are correlated. Also, there is measurement error in all three measurements.

The physical process is carried out in groups, and there are anywhere from 2 to 6 samples per group. A traditional practice is to calculate simple linear regressions, considering measurement error, within each group. However, the estimated regression coefficients from these regressions have large variances, and in some cases, a regression is not even possible (e.g., only two samples in a group, and the ratios have the same x-coordinate). The large uncertainty associated with these regressions is cause for concern, so we seek a method for reducing that uncertainty.

Discussion

In order to illustrate hierarchical Bayes regression, we apply the method to the set of data in Table 1 below. We have three groups of data, which are ratios of the measurements resulting from the physical process under consideration, with 3, 4, and 2 samples per group, respectively.

	Group 1			Group 2				Group 3	
u/w	1.29	1.1	1.33	1.4	1.11	0.88	1.83	1.29	1.1
v/w	1.14	0.7	1.33	1.0	0.89	0.75	1.5	1.14	0.7
Var(u/w)	0.0339	0.0246	0.0353	0.0396	0.025	0.0153	0.0683	0.03399	0.0246
Var(v/w)	0.026	0.0099	0.0353	0.02	0.0159	0.0112	0.0456	0.026	0.0099
Cov(u/w, v/w)	0.0135	0.0069	0.016	0.0126	0.0091	0.0058	0.0248	0.0135	0.0069

Table 1.

We consider a full hierarchical Bayesian (HB) model as a means of combining the regression data across groups. HB is a model in which all the unknown parameters, including prior distribution parameters (hyperparameters) are treated as random variables [2]. Let response variable $y = u/w$, observed predictor variable $x = v/w$, and true predictor variable be denoted x^*, then the basic model is the following:

$$y_{ij} = \alpha_i + \beta_i x^*_{ij} + \varepsilon_{ij} + \eta_{ij}, \qquad (1)$$

where, for the j^{th} sample within the i^{th} group,

y_{ij} is the response variable of interest,

x^*_{ij} is the true predictor variable (x_{ij} is the observed predictor measured with error),

α_i is the unknown random intercept,

β_i is the unknown random slope,

ε_{ij} is the error about the regression line, and

η_{ij} is the measurement error in y_{ij}.

In the HB formulation, the unknown parameters α_i and β_i are treated as random variables and given the following distributions:

$$\alpha_i \sim \mathrm{N}(\mu_\alpha, \sigma^2_\alpha), \qquad \beta_i \sim \mathrm{N}(\mu_\beta, \sigma^2_\beta),$$

and error terms have the following distributions:

$$\varepsilon_{ij} \sim \mathrm{N}(0, \sigma^2_{\varepsilon_i}), \qquad \eta_{ij} \sim \mathrm{N}(0, \sigma^2_{\eta_{ij}}).$$

Recalling that the (observed) ratios x_{ij} and y_{ij} are correlated, we use the following formulas to find the variance and the covariance between them. For each j^{th} sample within i^{th} group,

$$\sigma^2_x = Var\left(\frac{v}{w}\right) \approx \left(\frac{\mu_v}{\mu_w}\right)^2 \left(\frac{\sigma^2_v}{\mu^2_v} + \frac{\sigma^2_w}{\mu^2_w}\right),$$

the measurement error in y_{ij} has variance (found similarly)

$$\sigma^2_\eta = Var\left(\frac{u}{w}\right) \approx \left(\frac{\mu_u}{\mu_w}\right)^2 \left(\frac{\sigma^2_u}{\mu^2_u} + \frac{\sigma^2_w}{\mu^2_w}\right),$$

so the response y_{ij} (from the model in equation 1) has total variance

$$\sigma^2_y = \sigma^2_\varepsilon + \sigma^2_\eta,$$

and the covariance of the observed x_{ij} and response y_{ij} is

$$\sigma^2_{xy} = Cov\left(\frac{v}{w}, \frac{u}{w}\right) \approx \frac{\mu_v \mu_u \left(\mu^6_w \left(3\sigma^4_w + 6\mu^2_w \sigma^2_w + \mu^4_w\right) - \left(\mu^2_w + \sigma^2_w\right)^5\right)}{\mu^6_w \left(\mu^2_w + \sigma^2_w\right)^3}.$$

We then define a bivariate normal random variable,

$$z_{ij} = \begin{pmatrix} y_{ij} \\ x_{ij} \end{pmatrix} \sim N(\mu_{ij}, \Sigma_{ij}),$$

where $\mu_{ij} = \begin{pmatrix} \alpha_i + \beta_i x^*_{ij} \\ x^*_{ij} \end{pmatrix}$, the variance/covariance matrix of x_{ij} and y_{ij} is

$\Sigma_{ij} = \begin{bmatrix} \sigma^2_{\varepsilon\, i} + \sigma^2_{\eta\, ij} & \sigma^2_{xy\, ij} \\ \sigma^2_{xy\, ij} & \sigma^2_{x\, ij} \end{bmatrix}$, $\sigma^2_{\eta\, ij}$, $\sigma^2_{xy\, ij}$, and $\sigma^2_{x\, ij}$ are assumed to be known from the data, and $\sigma^2_{\varepsilon\, i}$ is to be estimated.

Figure 1 shows the directed graph that describes the model.

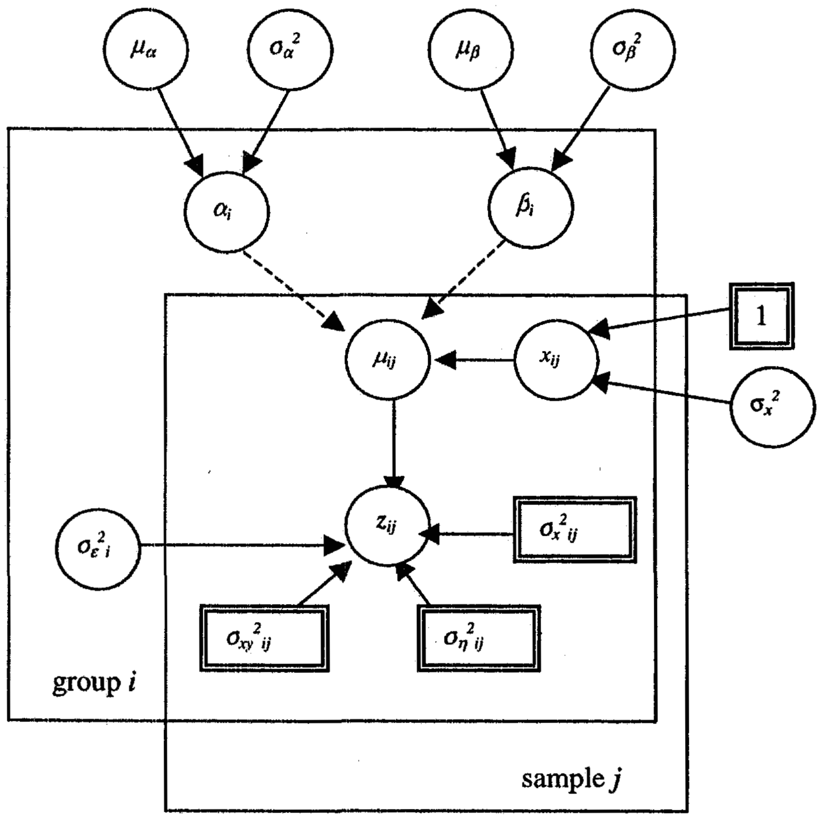


Figure 1. Directed Graph

We implemented this model in WinBUGS, the Windows version of BUGS software (Bayesian inference Using Gibbs Sampling). As more fully explained in [1] and [2], this software assumes a Bayesian or full probability model, in which all quantities are treated as random variables. We define a joint distribution over all unobserved and observed quantities, and then obtain inferences on the main quantities of interest using a Monte Carlo approach to numerical integration (Gibbs sampling).

The WinBUGS code for this model is relatively simple, and is shown in Figure 2 on the following page. We chose to regress on $(x - 1)$ for our example here, rather than just x, in order to reduce the autocorrelations within the parameter chains. Note that WinBUGS uses parameters $tau = (\sigma^2)^{-1}$ and $prec = \Sigma^{-1}$, and we use the following flat hyperpriors:

$x^*_{ij} \sim N(1, \tau_x)$, $\tau_x \sim \text{Gamma}(1.0 \times 10^{-3}, 1.0 \times 10^{-3})$,

$\mu_\alpha \sim N(0, 1.0 \times 10^{-10})$, $\tau_\alpha \sim \text{Gamma}(1.0 \times 10^{-3}, 1.0 \times 10^{-3})$,

$\mu_\beta \sim N(0, 1.0 \times 10^{-10})$, $\tau_\beta \sim \text{Gamma}(1.0 \times 10^{-3}, 1.0 \times 10^{-3})$.

```
model
{
for(j in 1:N) {
# jth overall sample
z[j,1:2] ~ dmnorm(mu[j,],prec[j,,])
mu[j,1] <- alpha[test[j]] + beta[test[j]]*(x[j]-1)
mu[j,2] <- x[j]
x[j] ~ dnorm(1, taux)
prec[j,1,1] <- varx[j]/(-covxy[j]*covxy[j] + (vareta[j] + varepsilon[test[j]])*varx[j])
prec[j,1,2] <- -covxy[j]/(-covxy[j]*covxy[j] + (vareta[j] + varepsilon[test[j]])*varx[j])
prec[j,2,1] <- prec[j,1,2]
prec[j,2,2] <- (varepsilon[test[j]]+vareta[j])/(-covxy[j]*covxy[j] + (vareta[j] +
varepsilon[test[j]])*varx[j])
}

for(i in 1:T) {
# ith group
alpha[i] ~ dnorm(mualpha, taualpha)
beta[i] ~ dnorm(mubeta, taubeta)
varepsilon[i] ~ dgamma(1,30)
muavg[i] <- sum(mu[cumsize[i] +1: cumsize[i+1], 1])/(cumsize[i+1] - cumsize[i])
}

taux ~ dgamma(1.0E-3, 1.0E-3)
mualpha ~ dnorm(0,1.0E-10)
taualpha ~ dgamma(1.0E-3, 1.0E-3)
mubeta ~ dnorm(0,1.0E-10)
taubeta ~ dgamma(1.0E-3, 1.0E-3)
}

# Data
list(N = 9, T = 3, cumsize = c(0,3,7,9), test = c(1,1,1,2,2,2,2,3,3), z=structure(.Data =
c(1.29,1.14,1.1,0.7,1.33,1.33,1.4,1,1.11,0.89,
0.88,0.75,1.83,1.5,1.29,1.14,1.1,0.7),.Dim=c(9,2)),varx=c(0.026,0.0099,0.0353,0.02,0.
0159,0.0112,0.0456,0.026,0.0099),vareta=c(0.0339,
0.0246,0.0353,0.0396,0.025,0.0153,0.0683,0.0339,
0.0246),covxy=c(0.0135,0.0069,0.016,0.0126,0.0091,0.0058,0.0248,0.0135,0.0069))

# Inits
list(taux = .01, mualpha = 0, taualpha = .01, mubeta = 0, taubeta = .01, varepsilon =
c(.1,.1,.1))
```

Figure 2. WinBUGS code

Results

WinBUGS generates results in several formats. Table 2 shows results in tabular format, for 100,000 iterations, using the code and data from Figure 2. All values are calculated starting at iteration 501, in order to account for a burn-in of 500 iterations.

node	mean	sd	MC error	2.5%	median	97.5%
alpha[1]	1.241	0.102	7.994E-4	1.04	1.24	1.445
alpha[2]	1.248	0.09836	7.856E-4	1.06	1.246	1.45
alpha[3]	1.253	0.1166	9.257E-4	1.025	1.251	1.491
beta[1]	0.7092	0.7658	0.0106	-0.6834	0.6995	2.115
beta[2]	1.061	0.8381	0.01164	-0.2412	0.9987	2.794
beta[3]	0.7471	0.875	0.01099	-0.8155	0.7447	2.316
mu[1,1]	1.285	0.142	8.938E-4	1.022	1.277	1.587
mu[2,1]	1.093	0.1579	0.001212	0.7702	1.097	1.395
mu[3,1]	1.33	0.1595	0.001157	1.039	1.32	1.671
mu[4,1]	1.267	0.151	8.765E-4	0.9919	1.258	1.592
mu[5,1]	1.168	0.1364	6.253E-4	0.8875	1.171	1.432
mu[6,1]	1.046	0.153	0.001116	0.7334	1.052	1.329
mu[7,1]	1.486	0.2316	0.00185	1.101	1.464	1.995
mu[8,1]	1.295	0.1551	0.00101	1.007	1.287	1.623
mu[9,1]	1.097	0.1615	0.001107	0.7677	1.101	1.406
muavg[1]	1.236	0.1053	6.159E-4	1.029	1.235	1.445
muavg[2]	1.242	0.1014	6.106E-4	1.046	1.24	1.447
muavg[3]	1.196	0.1211	6.352E-4	0.9529	1.198	1.433
varepsilon[1]	0.02347	0.0251	1.469E-4	5.545E-4	0.01548	0.09218
varepsilon[2]	0.02507	0.0253	1.608E-4	6.498E-4	0.0174	0.09275
varepsilon[3]	0.02598	0.02749	1.604E-4	6.278E-4	0.0173	0.1013
x[1]	1.07	0.1214	5.022E-4	0.8426	1.065	1.323
x[2]	0.7833	0.0972	6.617E-4	0.589	0.7842	0.9701
x[3]	1.145	0.1421	8.492E-4	0.8936	1.134	1.449
x[4]	1.01	0.1086	3.789E-4	0.7933	1.01	1.226
x[5]	0.9238	0.1001	3.986E-4	0.7198	0.9269	1.115
x[6]	0.8085	0.09719	6.424-4	0.6128	0.8103	0.9933
x[7]	1.23	0.1637	0.001071	0.9515	1.216	1.585
x[8]	1.066	0.1212	5.394E-4	0.8387	1.06	1.317
x[9]	0.7832	0.09739	6.89E-4	0.5895	0.7842	0.9712

Table 2. Results

Figure 3 shows the HB regressions obtained with this method, as well as the classical regressions for each group. The classical regressions give the best fit to the data within each group, of course, but we consider the HB regressions an improvement in predictive ability for population parameters.

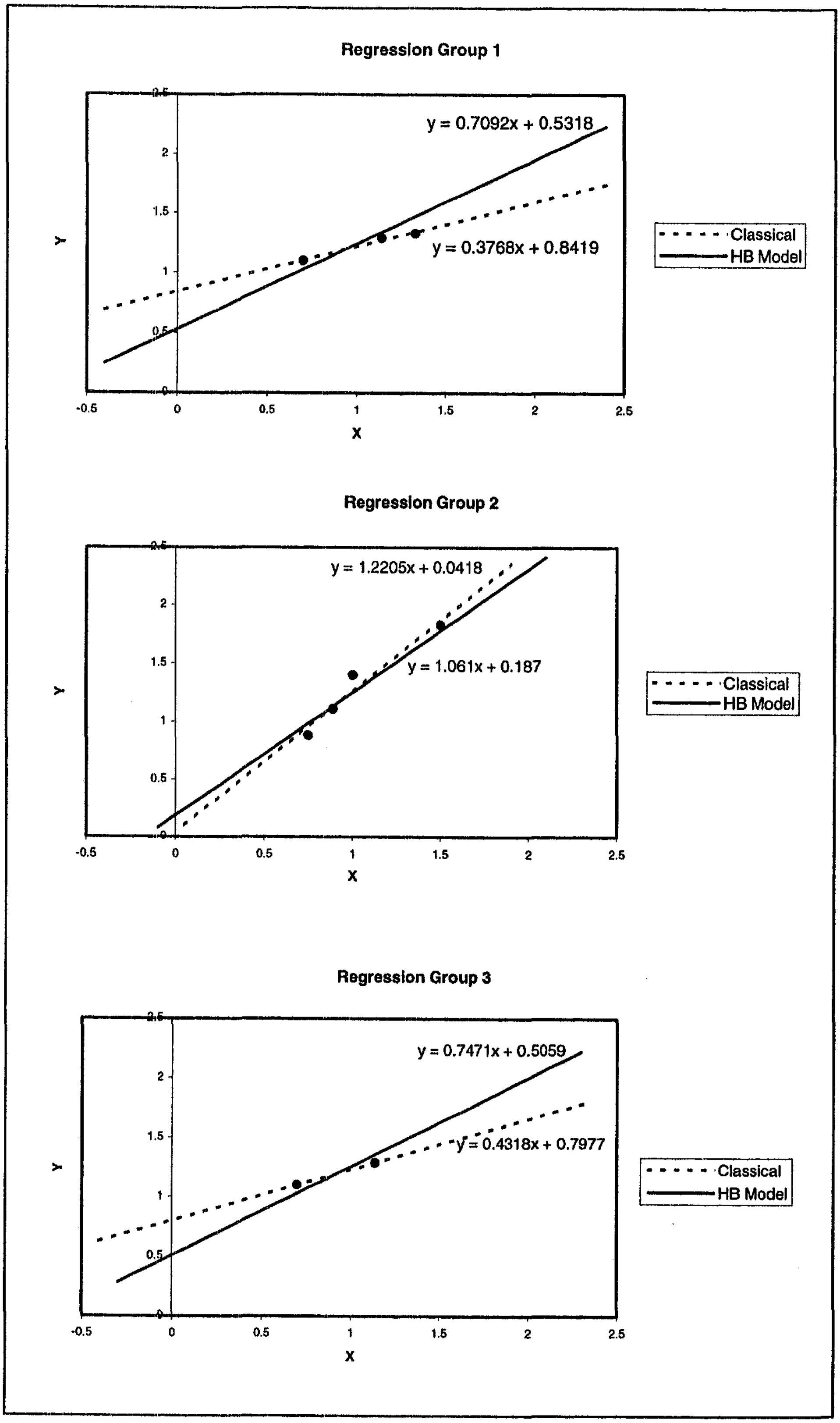


Figure 3. Comparison of regressions ignoring measurement error

Since the individual data points in each group are derived from the same physical process, we assume that they belong to the same super-population, i.e., they have something in common. Given that assumption, we borrow strength by combining data across groups. With that borrowing, the effect of any extreme values is lessened. That is, values that may be affected to a great degree by one or two extremes in the individual groups are shrunk towards the mean in the HB regressions. This is especially useful with sparse data. For example, as seen in Figure 3 for group 3, the classical regression is perfect, because there are only two points. However, because of the effect of borrowing strength, we expect the HB regression to give better estimates of certain population parameters of interest, e.g., the mean ratio y.

Previous analysis techniques for this problem included confidence interval estimates obtained by calculating the mean of y for each group and bounding by 10% on either side. Figure 4 shows our posterior densities of the mean of y for each group. The standard deviations are all less than 10% of the means (see Table 2), indicating that we can obtain a tighter and more accurate level of estimation with the HB approach.

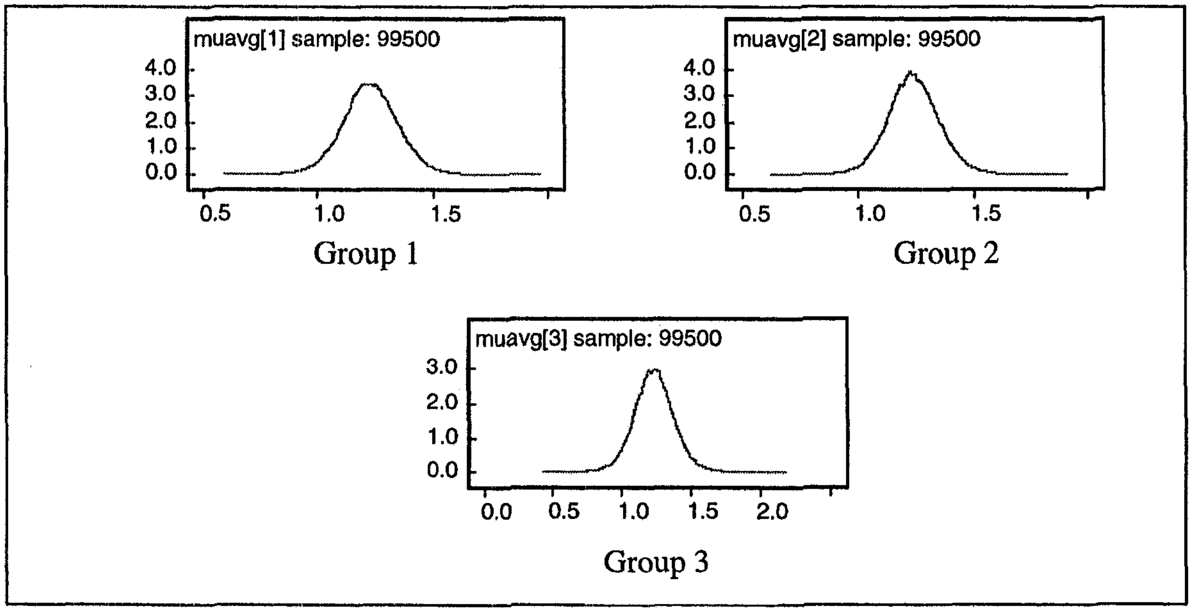


Figure 4. Distribution of the mean of y

Issues

There are several issues that we intend to continue to explore.

- Model Validation: Simulate data under our HB model and compare to real data.
- Subpopulations of groups: It may make sense, based on physical characteristics, to combine some groups.
- Multiple regressors: There are, in fact, other physical measurements of this process; can anything be gained by using other regressors?

- Nonlinear regression models: What criteria do we use for model selection?
- Identification of prior/hyperprior distributions: What happens if we choose different prior distributions (how sensitive are our results to different priors)?

Conclusions

A large data set is traditionally partitioned into smaller groups, and individual regressions are done within each group. The uncertainty on these regressions is high, so we looked for a way to combine the data. We demonstrated the use of HB, in the context of WinBUGS, as a nice statistical framework for combining and improving the regressions. This method is particularly useful when regressions on the individual groups are weak. The strength of HB is its potential for reducing the uncertainty in predictions.

References:

1. Spiegelhalter DJ, Thomas A, Best N, April 2000, WinBUGS ver 1.3 User Manual.
2. Gilks WR, Richardson S and Spiegelhalter DJ (Eds.) (1996) *Markov chain Monte Carlo in Practice*. Chapman & Hall, London.

Hierarchical Bayes Regression with Correlated Measurement Errors

J. Morzinski*

H. Martz*

S. Reese*

L. Denogean‡

* Los Alamos National Laboratory

‡ Cornell University

Introduction

A physical process produces three elements, call them u, v, and w. We can measure these elements, but we are interested in the ratio u/w. We would like to predict that ratio using another ratio, v/w.

Complications:

- The variables are correlated
- Measurements are made with error
- Multiple Groups

STATISTICAL SCIENCES

Standard Approach

A traditional practice is to calculate simple linear regressions within each group. However, there are several problems with this method applied to the data obtained from the physical process under consideration:

- Sparse data
- Estimated regression coefficients from these regressions have large variances

Los Alamos
National Laboratory

DECISION APPLICATIONS

STATISTICAL SCIENCES

New Approach

Hierarchical Bayes (HB) is a model in which all of the unknown parameters, including prior distribution parameters (hyperparameters), are treated as random variables.

We will use a full HB model as a means of combining the regression data across groups.

Los Alamos
National Laboratory

DECISION APPLICATIONS

Basic Model

$$y_{ij} = \alpha_i + \beta_i x^*_{ij} + \varepsilon_{ij} + \eta_{ij},\ where$$

y_{ij} is the response variable of interest,

x^*_{ij} is the true predictor variable (x_{ij} is the observed predictor measured with error),

α_i is the unknown random intercept,

β_i is the unknown random slope,

ε_{ij} is the error about the regression line, and

η_{ij} is the measurement error in y_{ij}.

Los Alamos
National Laboratory

DECISION APPLICATIONS

HB Formulation

- In the HB formulation, the unknown parameters α_i and β_i (intercept and slope) are treated as random variables and given the following distributions:

$\alpha_i \sim N(\mu_\alpha, \sigma_\alpha^2)$ $\qquad$ $\beta_i \sim N(\mu_\beta, \sigma_\beta^2)$

and error terms have the following distributions:

$\varepsilon_{ij} \sim N(0, \sigma_{\varepsilon_i}^2)$ $\qquad$ $\eta_{ij} \sim N(0, \sigma_{\eta_{ij}}^2)$

Los Alamos
National Laboratory

DECISION APPLICATIONS

Variance Equations

Since the ratios x_{ij} and y_{ij} are correlated, we use the following formulas to find the variance and the covariance between them. For each j^{th} sample within i^{th} group,

$$\sigma_x^2 = Var\left(\frac{v}{w}\right) \approx \left(\frac{\mu_v}{\mu_w}\right)^2 \left(\frac{\sigma_v^2}{\mu_v^2} + \frac{\sigma_w^2}{\mu_w^2}\right)$$

the measurement error in y_{ij} has variance

$$\sigma_\eta^2 = Var\left(\frac{u}{w}\right) \approx \left(\frac{\mu_u}{\mu_w}\right)^2 \left(\frac{\sigma_u^2}{\mu_u^2} + \frac{\sigma_w^2}{\mu_w^2}\right)$$

Los Alamos National Laboratory

DECISION APPLICATIONS

Covariance

The response y_{ij} (from the basic model equation) has total variance

$$\sigma_y^2 = \sigma_\varepsilon^2 + \sigma_\eta^2$$

and the covariance of the observed x_{ij} and response y_{ij} is

$$\sigma_{xy}^2 = Cov\left(\frac{v}{w}, \frac{u}{w}\right) \approx \frac{\mu_v \mu_u \left(\mu_w^6 \left(3\sigma_w^4 + 6\mu_w^2\sigma_w^2 + \mu_w^4\right) - \left(\mu_w^2 + \sigma_w^2\right)^5\right)}{\mu_w^6 \left(\mu_w^2 + \sigma_w^2\right)^3}$$

Los Alamos National Laboratory

DECISION APPLICATIONS

Data

	Group 1			Group 2				Group 3	
y	1.29	1.1	1.33	1.4	1.11	0.88	1.83	1.29	1.1
x	1.14	0.7	1.33	1.0	0.89	0.75	1.5	1.14	0.7
Var(η)	0.034	0.025	0.035	0.040	0.025	0.015	0.068	0.034	0.025
Var(x)	0.026	0.010	0.035	0.020	0.016	0.011	0.046	0.026	0.010
Cov(x,y)	0.014	0.007	0.016	0.013	0.009	0.006	0.025	0.014	0.007

WinBUGS Implementation

We implemented this model in WinBUGS (**B**ayesian inference **U**sing **G**ibbs **S**ampling).

- This software assumes a full probability model, where all quantities are treated as random variables.
- A joint distribution is defined over all observed and unobserved quantities.

STATISTICAL SCIENCES

Adjustment to Model Notation

We define a bivariate normal random variable,

$$z_{ij} = \begin{pmatrix} y_{ij} \\ x_{ij} \end{pmatrix} \sim N(\mu_{ij}, \Sigma_{ij})$$

where $\mu_{ij} = \begin{pmatrix} \alpha_i + \beta_i x^*_{ij} \\ x^*_{ij} \end{pmatrix}$, the variance/covariance matrix of x_{ij} and y_{ij} is

$\Sigma_{ij} = \begin{bmatrix} \sigma^2_{\varepsilon i} + \sigma^2_{\eta ij} & \sigma^2_{xy ij} \\ \sigma^2_{xy ij} & \sigma^2_{x ij} \end{bmatrix}$, $\sigma^2_{\eta ij}$, $\sigma^2_{xy ij}$ and $\sigma^2_{x ij}$ are assumed to be known from the data

Los Alamos National Laboratory

DECISION APPLICATIONS

STATISTICAL SCIENCES

Priors

$x^*_{ij} \sim N(1, \tau_x)$, $\tau_x \sim \text{Gamma}(1.0 \times 10^{-3}, 1.0 \times 10^{-3})$

$\mu_\alpha \sim N(0, 1.0 \times 10^{-10})$, $\tau_\alpha \sim \text{Gamma}(1 \times 10^{-3}, 1 \times 10^{-3})$

$\mu_\beta \sim N(0, 1.0 \times 10^{-10})$, $\tau_\beta \sim \text{Gamma}(1 \times 10^{-3}, 1 \times 10^{-3})$

Los Alamos National Laboratory

DECISION APPLICATIONS

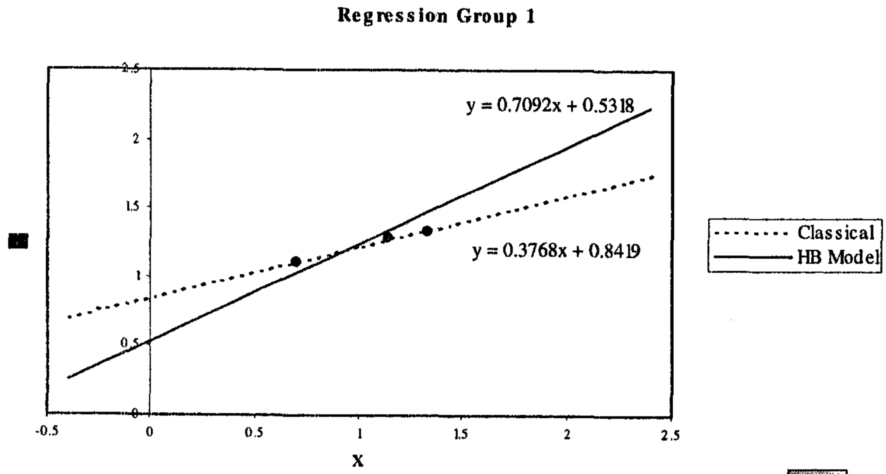
Results—Regression Parameters

Node	Mean	SD	MC Error	2.5%	Median	97.5%
alpha[1]	1.241	0.102	7.99E-4	1.04	1.24	1.445
alpha[2]	1.248	0.0984	7.86E-4	1.06	1.246	1.45
alpha[3]	1.253	0.1166	9.26E-4	1.025	1.51	1.491
beta[1]	0.7092	0.7658	0.0106	-0.6834	0.6995	2.115
beta[2]	1.061	0.8381	0.01164	-0.2412	0.9987	2.794
beta[3]	0.7471	0.875	0.01099	-0.8155	0.7447	2.316

Results—Mean of Predicted *y*

Node	Mean	SD	MC Error	2.5%	Median	97.5%
muavg1	1.236	0.1053	6.159E-4	1.029	1.235	1.445
muavg2	1.242	0.1014	6.106E-4	1.046	1.24	1.447
muavg3	1.196	0.1211	6.352E-4	0.9529	1.198	1.433

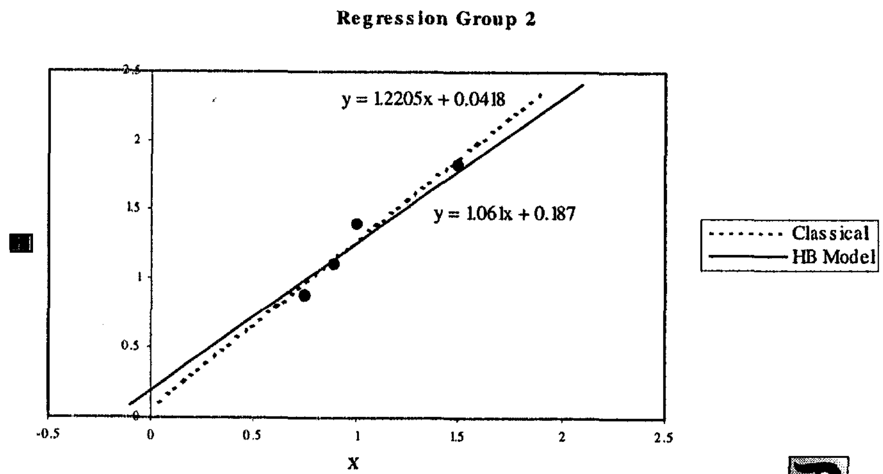
Group 1 Regression Comparisons



Los Alamos
National Laboratory

DECISION APPLICATIONS

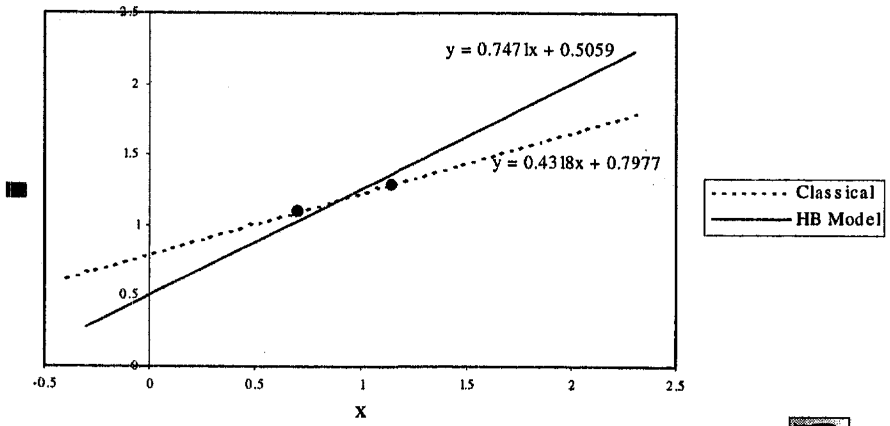
Group 2 Regression Comparisons



Los Alamos
National Laboratory

DECISION APPLICATIONS

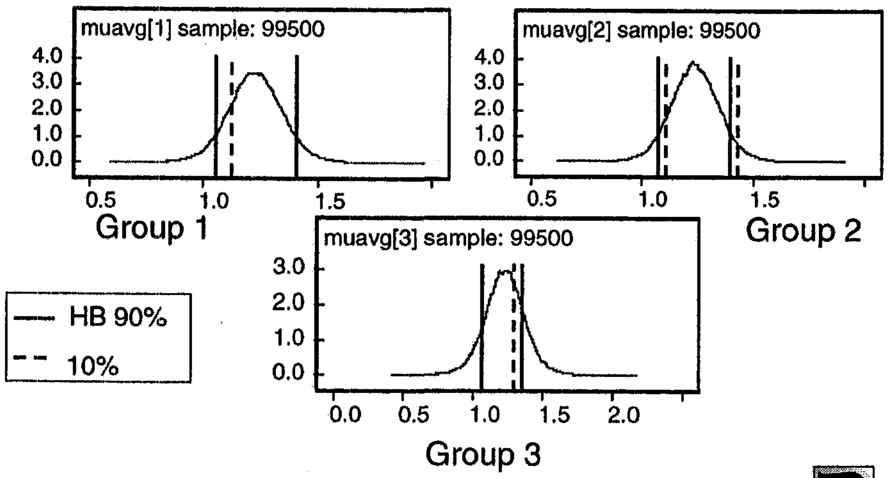
Group 3 Regression Comparisons



Los Alamos National Laboratory

Decision Applications

Prediction Interval Comparisons



Los Alamos National Laboratory

Decision Applications

Issues

There are still several issues to be explored:

- Subpopulations: based on physical characteristics, we may combine some groups
- Multiple regressors: can anything be gained by using other regressors?
- Nonlinear regression models: what criteria do we use for model selection?
- Identification of prior/hyperprior distributions: sensitivity analysis
- Model Validation.

Conclusion

- We demonstrated the use of HB in the context of WinBUGS. This method is particularly useful when regressions on the individual groups are weak or the data is sparse. The strength of the HB is in its ability to reduce and accurately represent the uncertainty in predictions.

- We consider the HB approach an improvement over the traditional method of performing individual regressions within each group.