

Hierarchical Models

Simon Jackman
Department of Political Science
Stanford University

March 25, 2004

Hierarchical models

- What is a *hierarchical* model?

Hierarchical models

- What is a *hierarchical* model? Very broad term, actually refers to lots of set-ups.
- **Multi-level** models are a classic case:

Hierarchical models

- What is a *hierarchical* model? Very broad term, actually refers to lots of set-ups.
- **Multi-level** models are a classic case:

$$y_{ij} \sim N(\mathbf{x}_{ij}\boldsymbol{\beta}_j, \sigma^2)$$

Hierarchical models

- What is a *hierarchical* model? Very broad term, actually refers to lots of set-ups.
- **Multi-level** models are a classic case:

$$y_{ij} \sim N(\mathbf{x}_{ij}\boldsymbol{\beta}_j, \sigma^2)$$

$$\boldsymbol{\beta}_j \sim N(\mathbf{z}_j\boldsymbol{\gamma}, \omega^2)$$

Hierarchical models

- What is a *hierarchical* model? Very broad term, actually refers to lots of set-ups.
- **Multi-level** models are a classic case:

$$y_{ij} \sim N(\mathbf{x}_{ij}\boldsymbol{\beta}_j, \sigma^2)$$

$$\boldsymbol{\beta}_j \sim N(\mathbf{z}_j\boldsymbol{\gamma}, \omega^2)$$

$$\boldsymbol{\gamma} \sim N(\mathbf{g}_0, \mathbf{G}_0)$$

Hierarchical models

- What is a *hierarchical* model? Very broad term, actually refers to lots of set-ups.
- **Multi-level** models are a classic case:

$$y_{ij} \sim N(\mathbf{x}_{ij}\boldsymbol{\beta}_j, \sigma^2)$$

$$\boldsymbol{\beta}_j \sim N(\mathbf{z}_j\boldsymbol{\gamma}, \omega^2)$$

$$\boldsymbol{\gamma} \sim N(\mathbf{g}_0, \mathbf{G}_0)$$

where, say, $i = 1, \dots, n_j$, indexes survey respondents

Hierarchical models

- What is a *hierarchical* model? Very broad term, actually refers to lots of set-ups.
- **Multi-level** models are a classic case:

$$y_{ij} \sim N(\mathbf{x}_{ij}\boldsymbol{\beta}_j, \sigma^2)$$

$$\boldsymbol{\beta}_j \sim N(\mathbf{z}_j\boldsymbol{\gamma}, \omega^2)$$

$$\boldsymbol{\gamma} \sim N(\mathbf{g}_0, G_0)$$

where, say, $i = 1, \dots, n_j$, indexes survey respondents and $j = 1, \dots, m$ indexes geographic units

Hierarchical models

- What is a *hierarchical* model? Very broad term, actually refers to lots of set-ups.
- **Multi-level** models are a classic case:

$$y_{ij} \sim N(\mathbf{x}_{ij}\boldsymbol{\beta}_j, \sigma^2)$$

$$\boldsymbol{\beta}_j \sim N(\mathbf{z}_j\boldsymbol{\gamma}, \omega^2)$$

$$\boldsymbol{\gamma} \sim N(\mathbf{g}_0, G_0)$$

where, say, $i = 1, \dots, n_j$, indexes survey respondents and $j = 1, \dots, m$ indexes geographic units or time.

Hierarchical models

- What is a *hierarchical* model? Very broad term, actually refers to lots of set-ups.
- **Multi-level** models are a classic case:

$$y_{ij} \sim N(\mathbf{x}_{ij}\boldsymbol{\beta}_j, \sigma^2)$$

$$\boldsymbol{\beta}_j \sim N(\mathbf{z}_j\boldsymbol{\gamma}, \omega^2)$$

$$\boldsymbol{\gamma} \sim N(\mathbf{g}_0, G_0)$$

where, say, $i = 1, \dots, n_j$, indexes survey respondents and $j = 1, \dots, m$ indexes geographic units or time.

Hierarchical models

- Estimation and inference for hierarchical models not trivial

Hierarchical models

- Estimation and inference for hierarchical models not trivial
- Nested stochastic structure complicates matters

Hierarchical models

- Estimation and inference for hierarchical models not trivial
- Nested stochastic structure complicates matters
- Issue for likelihood analyses is to ensure that uncertainty in 2nd level structure propagates into inferences about 1st level parameters

Hierarchical models

- Estimation and inference for hierarchical models not trivial
- Nested stochastic structure complicates matters
- Issue for likelihood analyses is to ensure that uncertainty in 2nd level structure propagates into inferences about 1st level parameters
- Switching to a Bayesian setting and using MCMC enables characterization of joint posterior of all model parameters, without any shortcuts/compromises (*a la* MLE).

Hierarchical models

- Estimation and inference for hierarchical models not trivial
- Nested stochastic structure complicates matters
- Issue for likelihood analyses is to ensure that uncertainty in 2nd level structure propagates into inferences about 1st level parameters
- Switching to a Bayesian setting and using MCMC enables characterization of joint posterior of all model parameters, without any shortcuts/compromises (*a la* MLE).

Why Use a Hierarchical Model?

- Suppose we have data from n clusters of data, with T observations in each cluster.

Why Use a Hierarchical Model?

- Suppose we have data from n clusters of data, with T observations in each cluster.
- An example might be where we've run studies in n different localities.

Why Use a Hierarchical Model?

- Suppose we have data from n clusters of data, with T observations in each cluster.
- An example might be where we've run studies in n different localities.
- For now, consider an extremely simple inferential problem: what is the mean of the dependent variable in each cluster, μ_i ?

Why Use a Hierarchical Model?

- Suppose we have data from n clusters of data, with T observations in each cluster.
- An example might be where we've run studies in n different localities.
- For now, consider an extremely simple inferential problem: what is the mean of the dependent variable in each cluster, μ_i ?

Why Use a Hierarchical Model?

- Seems like a simple problem:

Why Use a Hierarchical Model?

- Seems like a simple problem:

$$\bar{y}_i = T^{-1} \sum_{t=1}^T y_{it}$$

is an unbiased estimate of the mean in each cluster,

Why Use a Hierarchical Model?

- Seems like a simple problem:

$$\bar{y}_i = T^{-1} \sum_{t=1}^T y_{it}$$

is an unbiased estimate of the mean in each cluster, with

$$\text{se}(\bar{y}_i) = \sqrt{\frac{\text{var}(y_i)}{T}}$$

Why Use a Hierarchical Model?

- Seems like a simple problem:

$$\bar{y}_i = T^{-1} \sum_{t=1}^T y_{it}$$

is an unbiased estimate of the mean in each cluster, with

$$\text{se}(\bar{y}_i) = \sqrt{\frac{\text{var}(y_i)}{T}}$$

Why Use a Hierarchical Model?

- But there are other estimators available.

Why Use a Hierarchical Model?

- But there are other estimators available.
- We could *pool* across units, and estimate μ_j with the grand mean

$$\bar{y} = n^{-1}T^{-1} \sum_{i=1}^n \sum_{t=1}^T y_{it}$$

Why Use a Hierarchical Model?

- But there are other estimators available.
- We could *pool* across units, and estimate μ_j with the grand mean

$$\bar{y} = n^{-1}T^{-1} \sum_{i=1}^n \sum_{t=1}^T y_{it}$$

- This estimator uses all the data and so has a small standard error

Why Use a Hierarchical Model?

- But there are other estimators available.
- We could *pool* across units, and estimate μ_j with the grand mean

$$\bar{y} = n^{-1}T^{-1} \sum_{i=1}^n \sum_{t=1}^T y_{it}$$

- This estimator uses all the data and so has a small standard error
- But generally, this estimator will be a **biased** estimator of any the cluster-specific means.

Why Use a Hierarchical Model?

- But there are other estimators available.
- We could *pool* across units, and estimate μ_j with the grand mean

$$\bar{y} = n^{-1}T^{-1} \sum_{i=1}^n \sum_{t=1}^T y_{it}$$

- This estimator uses all the data and so has a small standard error
- But generally, this estimator will be a **biased** estimator of any the cluster-specific means.

Why Use a Hierarchical Model?

$$\bar{y} = n^{-1}T^{-1} \sum_{i=1}^n \sum_{t=1}^T y_{it}$$

- The bias *increases* to the extent that the clusters are *heterogeneous*.

Why Use a Hierarchical Model?

$$\bar{y} = n^{-1}T^{-1} \sum_{i=1}^n \sum_{t=1}^T y_{it}$$

- The bias *increases* to the extent that the clusters are *heterogeneous*. But to the extent that there is little *between-cluster* variation, then estimating each mean with the grand mean isn't so bad.

Why Use a Hierarchical Model?

$$\bar{y} = n^{-1}T^{-1} \sum_{i=1}^n \sum_{t=1}^T y_{it}$$

- The bias *increases* to the extent that the clusters are *heterogeneous*. But to the extent that there is little *between-cluster* variation, then estimating each mean with the grand mean isn't so bad.
- In fact, we might even prefer the pooled estimator on mean square error grounds (trading off some bias for smaller standard errors), depending on the relative size of the *between-cluster* variation to the *within-cluster* variation.

Why Use a Hierarchical Model?

$$\bar{y} = n^{-1}T^{-1} \sum_{i=1}^n \sum_{t=1}^T y_{it}$$

- The bias *increases* to the extent that the clusters are *heterogeneous*. But to the extent that there is little *between-cluster* variation, then estimating each mean with the grand mean isn't so bad.
- In fact, we might even prefer the pooled estimator on mean square error grounds (trading off some bias for smaller standard errors), depending on the relative size of the *between-cluster* variation to the *within-cluster* variation.

A Bias-Variance Tradeoff

- At one extreme, we have no pooling, and zero bias in the estimates of μ_j .

A Bias-Variance Tradeoff

- At one extreme, we have no pooling, and zero bias in the estimates of μ_j . But these estimates could have large standard errors, say, if the *within-cluster* variation is large (e.g., T is small).

A Bias-Variance Tradeoff

- At one extreme, we have no pooling, and zero bias in the estimates of μ_j . But these estimates could have large standard errors, say, if the *within-cluster* variation is large (e.g., T is small).
- At the other extreme, we have complete pooling, and we have potentially lots of bias if the *between-variation* is large, but potentially big gains in efficiency (say, if n is large relative to T).

A Bias-Variance Tradeoff

- At one extreme, we have no pooling, and zero bias in the estimates of μ_j . But these estimates could have large standard errors, say, if the *within-cluster* variation is large (e.g., T is small).
- At the other extreme, we have complete pooling, and we have potentially lots of bias if the *between-variation* is large, but potentially big gains in efficiency (say, if n is large relative to T).
- Famous result by Charles Stein (1955): there exists a biased estimator of μ_j that we would prefer to the simple “no pooling estimator”

A Bias-Variance Tradeoff

- At one extreme, we have no pooling, and zero bias in the estimates of μ_j . But these estimates could have large standard errors, say, if the *within-cluster* variation is large (e.g., T is small).
- At the other extreme, we have complete pooling, and we have potentially lots of bias if the *between-variation* is large, but potentially big gains in efficiency (say, if n is large relative to T).
- Famous result by Charles Stein (1955): there exists a biased estimator of μ_j that we would prefer to the simple “no pooling estimator” on mean-squared error criteria (the gains in efficiency outweigh the bias).

A Bias-Variance Tradeoff

- At one extreme, we have no pooling, and zero bias in the estimates of μ_j . But these estimates could have large standard errors, say, if the *within-cluster* variation is large (e.g., T is small).
- At the other extreme, we have complete pooling, and we have potentially lots of bias if the *between-variation* is large, but potentially big gains in efficiency (say, if n is large relative to T).
- Famous result by Charles Stein (1955): there exists a biased estimator of μ_j that we would prefer to the simple “no pooling estimator” on mean-squared error criteria (the gains in efficiency outweigh the bias). See also wonderful “baseball paper” by Efron and Morris (1975).

A Bias-Variance Tradeoff

- At one extreme, we have no pooling, and zero bias in the estimates of μ_j . But these estimates could have large standard errors, say, if the *within-cluster* variation is large (e.g., T is small).
- At the other extreme, we have complete pooling, and we have potentially lots of bias if the *between-variation* is large, but potentially big gains in efficiency (say, if n is large relative to T).
- Famous result by Charles Stein (1955): there exists a biased estimator of μ_j that we would prefer to the simple “no pooling estimator” on mean-squared error criteria (the gains in efficiency outweigh the bias). See also wonderful “baseball paper” by Efron and Morris (1975).

An Optimal Estimator?

- Is there a compromise estimator?

An Optimal Estimator?

- Is there a compromise estimator?
- i.e., make $\hat{\mu}_j$ close to the “no pooling” estimator if the between-cluster variation is large, but close to the grand mean if the between-cluster variation is small?

An Optimal Estimator?

- Is there a compromise estimator?
- i.e., make $\hat{\mu}_j$ close to the “no pooling” estimator if the between-cluster variation is large, but close to the grand mean if the between-cluster variation is small?
- Yes:

An Optimal Estimator?

- Is there a compromise estimator?
- i.e., make $\hat{\mu}_j$ close to the “no pooling” estimator if the between-cluster variation is large, but close to the grand mean if the between-cluster variation is small?
- Yes: the so-called *random effects* estimator.

An Optimal Estimator?

- Is there a compromise estimator?
- i.e., make $\hat{\mu}_j$ close to the “no pooling” estimator if the between-cluster variation is large, but close to the grand mean if the between-cluster variation is small?
- Yes: the so-called *random effects* estimator.
- This estimator dominates the *fixed effects* (or no pooling estimator) in terms of mean-square error, by buying just the right amount of bias relative to the efficiency gain, cluster-by-cluster.

An Optimal Estimator?

- Is there a compromise estimator?
- i.e., make $\hat{\mu}_j$ close to the “no pooling” estimator if the between-cluster variation is large, but close to the grand mean if the between-cluster variation is small?
- Yes: the so-called *random effects* estimator.
- This estimator dominates the *fixed effects* (or no pooling estimator) in terms of mean-square error, by buying just the right amount of bias relative to the efficiency gain, cluster-by-cluster.

Simple Hierarchical Model

$$y_{it} \sim N(\theta_i, V_i) \text{ (level 1)}$$

Simple Hierarchical Model

$$y_{it} \sim N(\theta_i, V_i) \text{ (level 1)}$$

$$\theta_i \stackrel{\text{iid}}{\sim} N(\theta, \sigma^2) \text{ (level 2)}$$

Simple Hierarchical Model

$$y_{it} \sim N(\theta_i, V_i) \text{ (level 1)}$$

$$\theta_i \stackrel{\text{iid}}{\sim} N(\theta, \sigma^2) \text{ (level 2)}$$

- Given the stochastic structure, the θ_i are sometimes called *random effects*

Simple Hierarchical Model

$$y_{it} \sim N(\theta_i, V_i) \text{ (level 1)}$$
$$\theta_i \stackrel{\text{iid}}{\sim} N(\theta, \sigma^2) \text{ (level 2)}$$

- Given the stochastic structure, the θ_i are sometimes called *random effects*
- When $\sigma^2 = 0$ we get the fully pooled model:

Simple Hierarchical Model

$$y_{it} \sim N(\theta_i, V_i) \text{ (level 1)}$$

$$\theta_i \stackrel{\text{iid}}{\sim} N(\theta, \sigma^2) \text{ (level 2)}$$

- Given the stochastic structure, the θ_i are sometimes called *random effects*
- When $\sigma^2 = 0$ we get the fully pooled model:

$$\theta_i = \theta \quad \forall i = 1, \dots, n$$

Simple Hierarchical Model

$$y_{it} \sim N(\theta_i, V_i) \text{ (level 1)}$$
$$\theta_i \stackrel{\text{iid}}{\sim} N(\theta, \sigma^2) \text{ (level 2)}$$

- Given the stochastic structure, the θ_i are sometimes called *random effects*
- When $\sigma^2 = 0$ we get the fully pooled model:

$$\theta_i = \theta \quad \forall i = 1, \dots, n$$

i.e., no between-cluster heterogeneity

Simple Hierarchical Model

$$y_{it} \sim N(\theta_i, V_i) \text{ (level 1)}$$

$$\theta_i \stackrel{\text{iid}}{\sim} N(\theta, \sigma^2) \text{ (level 2)}$$

- Given the stochastic structure, the θ_i are sometimes called *random effects*
- When $\sigma^2 = 0$ we get the fully pooled model:

$$\theta_i = \theta \quad \forall i = 1, \dots, n$$

i.e., no between-cluster heterogeneity

Simple Hierarchical Model

$$y_{it} \sim N(\theta_i, V_i) \text{ (level 1)}$$

Simple Hierarchical Model

$$y_{it} \sim N(\theta_i, V_i) \text{ (level 1)}$$

$$\theta_i \stackrel{\text{iid}}{\sim} N(\theta, \sigma^2) \text{ (level 2)}$$

Simple Hierarchical Model

$$y_{it} \sim N(\theta_i, V_i) \text{ (level 1)}$$

$$\theta_i \stackrel{\text{iid}}{\sim} N(\theta, \sigma^2) \text{ (level 2)}$$

Simple Hierarchical Model

$$y_{it} \sim N(\theta_i, V_i) \text{ (level 1)}$$

$$\theta_i \stackrel{\text{iid}}{\sim} N(\theta, \sigma^2) \text{ (level 2)}$$

- Unknown parameters are θ (the mean of the distribution for the random effects, the θ_i) and σ^2 (the variance of this distribution).

Simple Hierarchical Model

$$y_{it} \sim N(\theta_i, V_i) \text{ (level 1)}$$

$$\theta_i \stackrel{\text{iid}}{\sim} N(\theta, \sigma^2) \text{ (level 2)}$$

- Unknown parameters are θ (the mean of the distribution for the random effects, the θ_i) and σ^2 (the variance of this distribution).
- We don't estimate the θ_i directly; estimate the parameters of the distribution from which they come;

Simple Hierarchical Model

$$y_{it} \sim N(\theta_i, V_i) \text{ (level 1)}$$

$$\theta_i \stackrel{\text{iid}}{\sim} N(\theta, \sigma^2) \text{ (level 2)}$$

- Unknown parameters are θ (the mean of the distribution for the random effects, the θ_i) and σ^2 (the variance of this distribution).
- We don't estimate the θ_i directly; estimate the parameters of the distribution from which they come; then given these parameters and the observed data, we form expectations for the θ_i

Simple Hierarchical Model

$$y_{it} \sim N(\theta_i, V_i) \text{ (level 1)}$$

$$\theta_i \stackrel{\text{iid}}{\sim} N(\theta, \sigma^2) \text{ (level 2)}$$

- Unknown parameters are θ (the mean of the distribution for the random effects, the θ_i) and σ^2 (the variance of this distribution).
- We don't estimate the θ_i directly; estimate the parameters of the distribution from which they come; then given these parameters and the observed data, we form expectations for the θ_i (BLUPs -- best linear unbiased predictions).

Simple Hierarchical model

$$y_{it} \sim N(\theta_i, V_i) \text{ (level 1)}$$

Simple Hierarchical model

$$y_{it} \sim N(\theta_i, V_i) \text{ (level 1)}$$

$$\theta_i \stackrel{\text{iid}}{\sim} N(\theta, \sigma^2) \text{ (level 2)}$$

Simple Hierarchical model

$$y_{it} \sim N(\theta_i, V_i) \text{ (level 1)}$$

$$\theta_i \stackrel{\text{iid}}{\sim} N(\theta, \sigma^2) \text{ (level 2)}$$

can be re-arranged to yield the estimates

$$\theta_i | \mathbf{y}, \theta, \sigma^2 \sim N(\theta_i^*, V_i(1 - B_i))$$

$$\theta_i^* = (1 - B_i)\bar{y}_i + B_i\theta$$

$$B_i = V_i / (V_i + \sigma^2)$$

Simple Hierarchical model

$$y_{it} \sim N(\theta_i, V_i) \text{ (level 1)}$$

$$\theta_i \stackrel{\text{iid}}{\sim} N(\theta, \sigma^2) \text{ (level 2)}$$

can be re-arranged to yield the estimates

$$\theta_i | \mathbf{y}, \theta, \sigma^2 \sim N(\theta_i^*, V_i(1 - B_i))$$

$$\theta_i^* = (1 - B_i)\bar{y}_i + B_i\theta$$

$$B_i = V_i / (V_i + \sigma^2)$$

Simple Hierarchical model

- Each study's conditional mean is a **weighted average** of the observed study mean and the overall mean.

Simple Hierarchical model

- Each study's conditional mean is a **weighted average** of the observed study mean and the overall mean.
- B_i are *shrinkage factors*.

Simple Hierarchical model

- Each study's conditional mean is a **weighted average** of the observed study mean and the overall mean.
- B_i are *shrinkage factors*. The larger B_i , the more θ_i^* is shrunk back to the grand mean θ .

Simple Hierarchical model

- Each study's conditional mean is a **weighted average** of the observed study mean and the overall mean.
- B_i are *shrinkage factors*. The larger B_i , the more θ_i^* is shrunk back to the grand mean θ .
- Less study-specific precision, more shrinkage

Simple Hierarchical model

- Each study's conditional mean is a **weighted average** of the observed study mean and the overall mean.
- B_i are *shrinkage factors*. The larger B_i , the more θ_i^* is shrunk back to the grand mean θ .
- Less study-specific precision, more shrinkage
- More study-specific precision, less shrinkage

Simple Hierarchical model

- Each study's conditional mean is a **weighted average** of the observed study mean and the overall mean.
- B_i are *shrinkage factors*. The larger B_i , the more θ_i^* is shrunk back to the grand mean θ .
- Less study-specific precision, more shrinkage
- More study-specific precision, less shrinkage

Estimation for Hierarchical Model

$$y_{it} \sim N(\theta_i, V_i) \text{ (level 1)}$$

Estimation for Hierarchical Model

$$y_{it} \sim N(\theta_i, V_i) \text{ (level 1)}$$

$$\theta_i \stackrel{\text{iid}}{\sim} N(\theta, \sigma^2) \text{ (level 2)}$$

Estimation for Hierarchical Model

$$y_{it} \sim N(\theta_i, V_i) \text{ (level 1)}$$

$$\theta_i \stackrel{\text{iid}}{\sim} N(\theta, \sigma^2) \text{ (level 2)}$$

Likelihood function

$$L(\theta, \sigma^2 | \mathbf{y}, \mathbf{V}) \propto \prod_{i=1}^n \frac{1}{\sqrt{V_i + \sigma^2}} \exp \left[-\frac{1}{2} \sum_{i=1}^n \frac{(y_i - \theta)^2}{V_i + \sigma^2} \right]$$

Likelihood function

$$L(\theta, \sigma^2 | \mathbf{y}, \mathbf{V}) \propto \prod_{i=1}^n \frac{1}{\sqrt{V_i + \sigma^2}} \exp \left[-\frac{1}{2} \sum_{i=1}^n \frac{(y_i - \theta)^2}{V_i + \sigma^2} \right]$$

After optimizing wrt θ and σ^2 :

$$\hat{\theta}_i = (1 - \hat{B}_i)y_i + \hat{B}_i\hat{\theta},$$

Likelihood function

$$L(\theta, \sigma^2 | \mathbf{y}, \mathbf{V}) \propto \prod_{i=1}^n \frac{1}{\sqrt{V_i + \sigma^2}} \exp \left[-\frac{1}{2} \sum_{i=1}^n \frac{(y_i - \theta)^2}{V_i + \sigma^2} \right]$$

After optimizing wrt θ and σ^2 :

$$\hat{\theta}_i = (1 - \hat{B}_i)y_i + \hat{B}_i\hat{\theta}, \hat{B}_i = \frac{V_i}{V_i + \hat{\sigma}^2}.$$

Likelihood function

$$L(\theta, \sigma^2 | \mathbf{y}, \mathbf{V}) \propto \prod_{i=1}^n \frac{1}{\sqrt{V_i + \sigma^2}} \exp \left[-\frac{1}{2} \sum_{i=1}^n \frac{(y_i - \theta)^2}{V_i + \sigma^2} \right]$$

After optimizing wrt θ and σ^2 :

$$\hat{\theta}_i = (1 - \hat{B}_i)y_i + \hat{B}_i\hat{\theta}, \hat{B}_i = \frac{V_i}{V_i + \hat{\sigma}^2}.$$

Classical Inference for Hierarchical Model

Asymptotic justification: as $n \rightarrow \infty$ (large number of units)

$$\hat{\theta} \sim N \left(\theta, \left[\sum_{i=1}^n \frac{1}{V_i + \hat{\sigma}^2} \right]^{-1} \right)$$

$$\hat{\theta}_i \sim N \left(\theta_i, V_i(1 - \hat{B}_i) \right),$$

Classical Inference for Hierarchical Model

Asymptotic justification: as $n \rightarrow \infty$ (large number of units)

$$\hat{\theta} \sim N \left(\theta, \left[\sum_{i=1}^n \frac{1}{V_i + \hat{\sigma}^2} \right]^{-1} \right)$$

$$\hat{\theta}_i \sim N \left(\theta_i, V_i(1 - \hat{B}_i) \right),$$

- Draper notes that these expressions do not account fully for the uncertainty in σ^2 , and therefore underestimate the true sampling variances.

Classical Inference for Hierarchical Model

Asymptotic justification: as $n \rightarrow \infty$ (large number of units)

$$\hat{\theta} \sim N \left(\theta, \left[\sum_{i=1}^n \frac{1}{V_i + \hat{\sigma}^2} \right]^{-1} \right)$$

$$\hat{\theta}_i \sim N \left(\theta_i, V_i(1 - \hat{B}_i) \right),$$

- Draper notes that these expressions do not account fully for the uncertainty in σ^2 , and therefore underestimate the true sampling variances.
- See Browne and Draper (2000), who conclude that Bayesian simulation is probably the best way to estimate these models.

Example: meta-analysis of aspirin and heart attacks (Draper et al., 1992)

Data: 6 studies of effects of aspirin on survivorship after acute myocardial infarction

Study	Aspirin		Placebo	
	Patients	Mortality (%)	Patients	Mortality (%)
UK-1	615	7.97	624	10.74
CDPA	758	5.80	771	8.30
GAMS	317	8.52	309	10.36
UK-2	832	12.26	850	14.82
PARIS	810	10.49	406	12.81
AMIS	2267	10.85	2257	9.70
Total	5599	9.88	5217	10.73

n.b., the large AMIS study finds a negative effect, and tends to dominate any

Example: meta-analysis of aspirin and heart attacks (Draper et al., 1992)

Data: 6 studies of effects of aspirin on survivorship after acute myocardial infarction (heart attack).

Study	Aspirin		Placebo	
	Patients	Mortality (%)	Patients	Mortality (%)
UK-1	615	7.97	624	10.74
CDPA	758	5.80	771	8.30
GAMS	317	8.52	309	10.36
UK-2	832	12.26	850	14.82
PARIS	810	10.49	406	12.81
AMIS	2267	10.85	2257	9.70
Total	5599	9.88	5217	10.73

n.b., the large AMIS study finds a negative effect, and tends to dominate any

Example: meta-analysis of aspirin and heart attacks (Draper et al., 1992)

Data: 6 studies of effects of aspirin on survivorship after acute myocardial infarction (heart attack). Some patients given aspirin; some given a placebo.

Study	Aspirin		Placebo	
	Patients	Mortality (%)	Patients	Mortality (%)
UK-1	615	7.97	624	10.74
CDPA	758	5.80	771	8.30
GAMS	317	8.52	309	10.36
UK-2	832	12.26	850	14.82
PARIS	810	10.49	406	12.81
AMIS	2267	10.85	2257	9.70
Total	5599	9.88	5217	10.73

n.b., the large AMIS study finds a negative effect, and tends to dominate any

pooled analysis.

Example: meta-analysis of aspirin and heart attacks

Data: each study yields a difference of means y_i and an accompanying standard error

Example: meta-analysis of aspirin and heart attacks

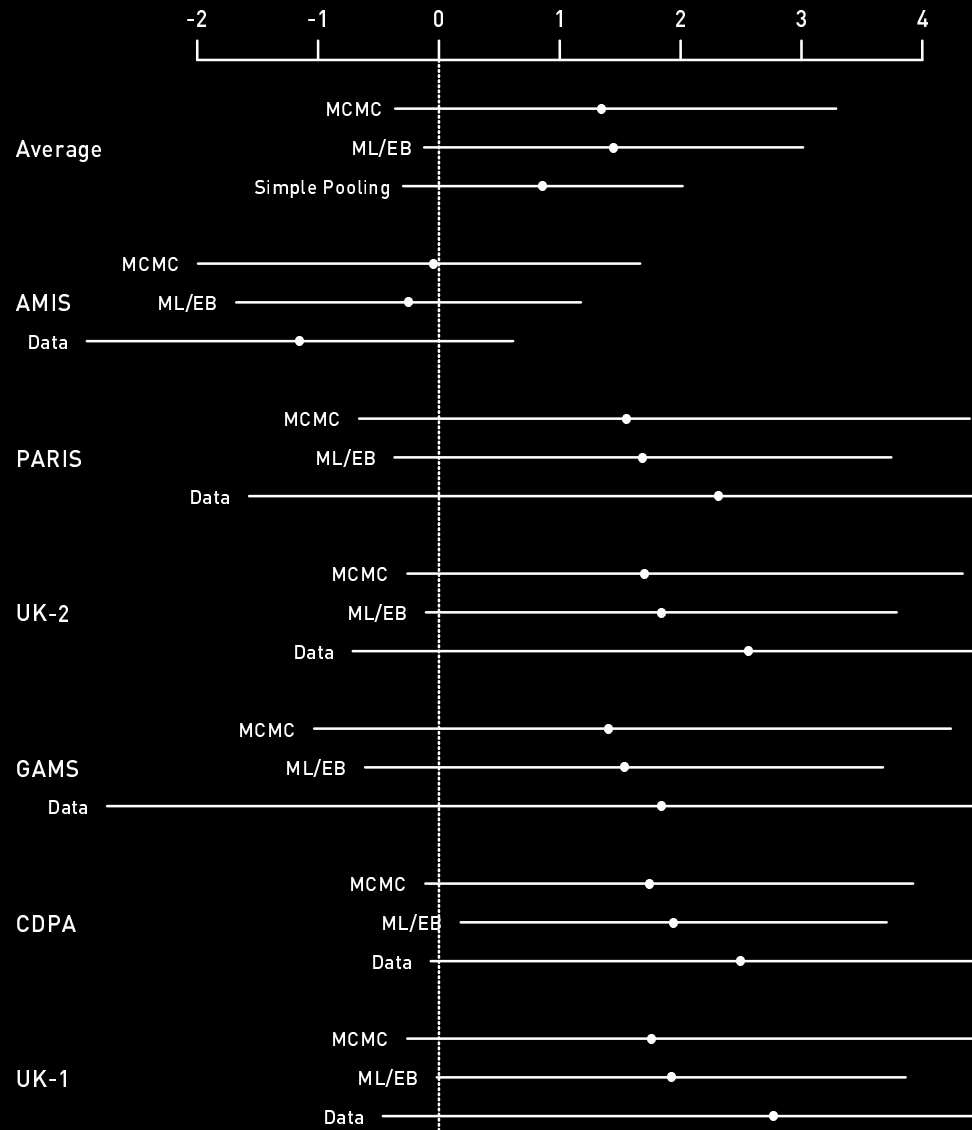
Data: each study yields a difference of means y_i and an accompanying standard error

Study	y_i	se	Z_i	p_i
UK-1	2.77	1.65	1.68	.047
CDPA	2.50	1.31	1.91	.028
GAMS	1.84	2.34	0.79	.216
UK-2	2.56	1.67	1.54	.062
PARIS	2.31	1.98	1.17	.129
AMIS	-1.15	0.90	-1.27	.898
Total	0.86	0.59	1.47	.072

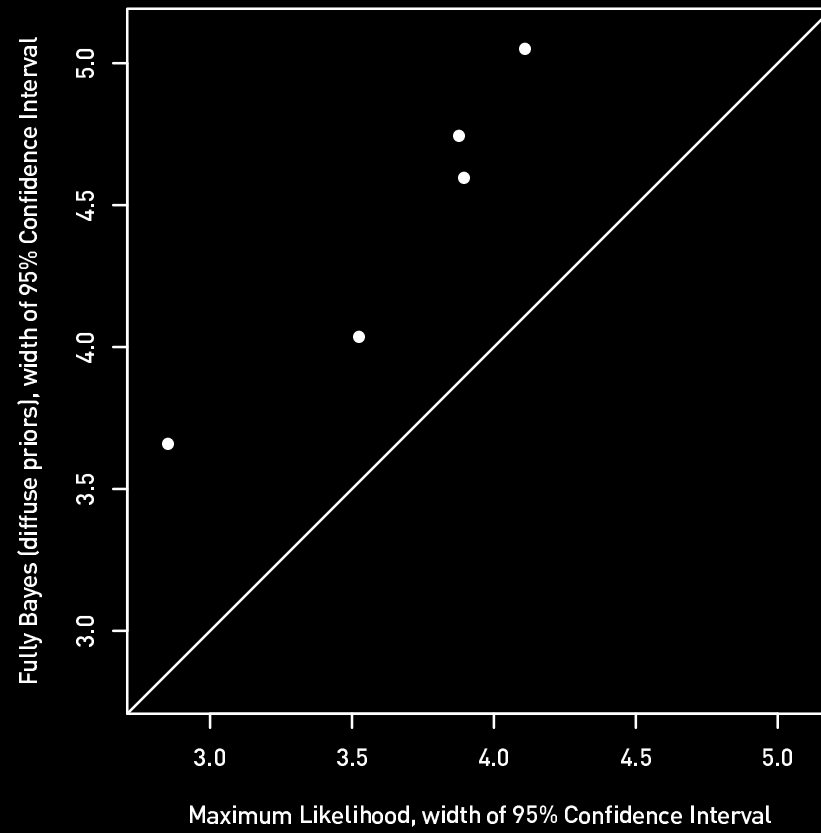
i.e., simple pooling finds a small effect, statistically indistinguishable from zero at conventional levels of significance.

Estimation and Inference for Hierarchical Model via Bayesian Simulation

- See WinBUGS job, aspirin.



Estimated Treatment Effects (95% bounds)



Example: Hierarchical Model for Economic Growth

Corporatism example.

References

Browne, William J. and David Draper. 2000. “Implementation and Performance Issues in the Bayesian and Likelihood Fitting of Multilevel Models.” *Computational Statistics*. To appear.

Draper, D., Donald P. Gaver Jr, Prem K. Goel, Joel B. Greenhouse, Larry V. Hedges, Carl N. Morris, John R. Tucker and Christine M. Waternaux. 1992. *Combining Information: Statistical Issues and Opportunities for Research*. Number 1 in Contemporary Statistics. Alexandria, Virginia: American Statistical Association.

Efron, Bradley and Carl Morris. 1975. “Data Analysis Using Stein’s Estimator and Its Generalizations.” *Journal of the American Statistical Association* 70:311--319.

Stein, C. 1955. Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution. In *Proceedings of the Third Berkeley Symposium*

of Mathematical Statistics and Probability. Vol. 1 Berkeley: University of California Press pp. 197--206.