Small Area Estimation Using Data from Multiple Sources

Sharon L. Lohr
Arizona State University
June 30, 2009
(joint work with Lynn Ybarra)

Small Area Estimation

- Subpopulation with small (or 0) sample size
- Examples
  - School-age poverty rates for each U.S. county
  - Alcohol use for each school district in Arizona
  - Robbery rates for black males 19-24
  - Characteristics of domestic violence victims

Small Area Estimation

- Want to estimate Y_i in area i, $i = 1, \ldots, m$
- Challenge: direct estimate y_i from survey unreliable in some areas because n too small.
- Approach: borrow strength from other small areas, covariates $\mathbf{x}_i$ known from administrative data
- Rao (2003) *Small Area Estimation*

Fay-Herriot (1979) Model

- $y_i = \mathrm{x}_i'\boldsymbol{\beta} + v_i + e_i,\ i = 1, \ldots, m,$
- $v_i \sim (0, \sigma_v^2)$, $e_i \sim (0, \psi_i)$; independent
- $Y_i = \mathrm{x}_i'\boldsymbol{\beta} + v_i$, assumed to be the population mean of area i
- BLUP of Y_i is

$$\tilde{Y}_{i\mathrm{FH}} = \gamma_{iv} y_i + (1 - \gamma_{iv})\mathrm{x}_i'\boldsymbol{\beta},$$

  where $\gamma_{iv} = \sigma_v^2/(\sigma_v^2 + \psi_i)$

- Inference is conditional on x_i

$$\text{MSE}\,(\tilde{Y}_{i\mathrm{FH}}) = E(\tilde{Y}_{i\mathrm{FH}} - Y_i)^2 = \gamma_{iv}\psi_i$$

What if x_i not known?

- x_i measured in another survey
- Administrative records with errors
- Imputed values for x_i with imputation error

Disease prevalence

- National Health and Nutrition Examination Survey (NHANES): Mobile Exam Centers
- National Health Interview Survey (NHIS): interview

NHANES (2003–2004)	NHIS (2004)
15 psus per year	358 psus
$<$10,000 persons in MEC	95,000 persons

Violent Crime Rates

- U.S. National Crime Victimization Survey
  - General population survey
  - Used to be 50,000 households; n reduced
- FBI Uniform Crime Reports
  - Provided by police agencies
  - Underestimate crime; not all reported to police
  - Missing data; outliers

American Community Survey

- Replaces U.S. Census Long Form
- 3 million households per year
- More up-to-date auxiliary information than U.S. Census, but has sampling, nonsampling error

3 Methods When Auxiliary Information has Error

- Measurement error model
- Multivariate model
- Dual frame model

Why not substitute estimates x_i?

- x_i estimates X_i

$$\tilde{Y}_{i\text{sub}} = \gamma_{iv} y_i + (1 - \gamma_{iv}) \mathrm{x}_i' \boldsymbol{\beta}.$$

- x_i, e_i and v_i independent for each i:

$$E(\tilde{Y}_{i\text{sub}} - Y_i) = (1 - \gamma_{iv}) \{E(\mathrm{x}_i) - \mathrm{X}_i\}' \boldsymbol{\beta},$$

$$\text{MSE}\,(\tilde{Y}_{i\text{sub}}) = \gamma_{iv} \psi_i + (1 - \gamma_{iv})^2 \boldsymbol{\beta}' \mathrm{C}_i \boldsymbol{\beta}.$$

$$\mathrm{C}_i = \text{MSE}\,(\mathrm{x}_i)$$

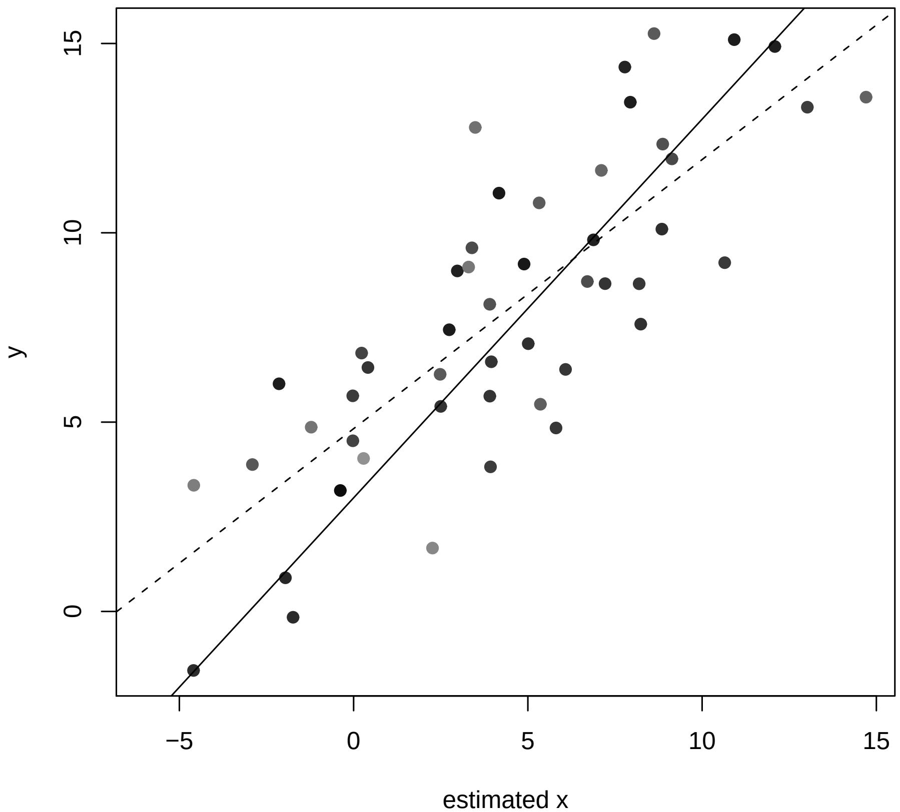
Substituting x_i

- MSE $(\tilde{Y}_{i\mathrm{sub}}) > \psi_i$ if $\boldsymbol{\beta}'\mathrm{C}_i\boldsymbol{\beta} > \sigma_v^2 + \psi_i$
- $\tilde{Y}_{i\mathrm{sub}}$ can be worse than y_i
- If error in x ignored, $\gamma_{iv}\psi_i <$ MSE $(\tilde{Y}_{i\mathrm{sub}})$
- γ_{iv} relies too much on prediction from model
- If x_i unreliable, want to rely more heavily on direct estimator.

Why not condition on x's?

- If x has measurement error, $E[\widehat{\beta}_{1,\mathrm{OLS}}] < \beta_1$
- But—taught in regression class that predictions are ok if there is measurement error (structural)
- OK if c_i same for all i. Buonaccorsi (1995): when c_i's vary, should account for the measurement error when estimating $\boldsymbol{\beta}$ and σ_v^2 and then use the measurement error estimates of $\boldsymbol{\beta}$ and σ_v^2 for prediction.

Estimated slope if x measured with error



solid = true line
dashed = OLS line

darker circles have more precision for x

Measurement Error Model

- FH Model: $y_i = \mathrm{X}_i'\boldsymbol{\beta} + v_i + e_i$
- Estimator x_i: $E(\mathrm{x}_i) = \mathrm{X}_i$
  MSE(x_i) = C_i (design-based survey error)

$$y_i = \mathrm{x}_i'\boldsymbol{\beta} + r_i(\mathrm{x}_i, \mathrm{X}_i) + e_i,$$

$$r_i(\mathrm{x}_i, \mathrm{X}_i) = v_i + (\mathrm{X}_i - \mathrm{x}_i)'\boldsymbol{\beta}.$$

- MSE $(r_i) = \sigma_v^2 + \boldsymbol{\beta}'\mathrm{C}_i\boldsymbol{\beta}$
- Assume X_i and Y_i estimated using independent data sources; Ybarra (2003) looked at x_i, y_i dependent

Which measurement error model?

- Structural and ultrastructural: assume x_i's are independent random vectors
  Ghosh et al. (2006), Torabi et al. (2009): empirical, hierarchical Bayesian methods
- Functional: x_i's unknown constant vectors
  Accords with finite population assumptions: area mean considered fixed but unknown, use sampling design for expected value and variance of an estimator. (Ybarra and Lohr, 2008)

If parameters known

- $\tilde{Y}_{i\mathrm{ME}} = \gamma_i y_i + (1 - \gamma_i)\mathbf{x}_i'\boldsymbol{\beta}$

$$\gamma_i = \frac{\sigma_v^2 + \boldsymbol{\beta}'\mathrm{C}_i\boldsymbol{\beta}}{\sigma_v^2 + \boldsymbol{\beta}'\mathrm{C}_i\boldsymbol{\beta} + \psi_i}$$

- Rely more on y_i if ψ_i small or C_i big
- MSE $(\tilde{Y}_{i\mathrm{ME}}) = \gamma_i \psi_i$
- Best predictor under normality (conditioning on residuals)

$$\tilde{Y}_{i\mathrm{ME}} = y_i - E(e_i \mid y_i - \mathbf{x}_i'\boldsymbol{\beta})$$

Parameters unknown

- Estimate $\boldsymbol{\beta}$, σ_v^2 using only the areas with $\mathbf{x}_i$ precise (or, if possible, known)
- Modified least squares (gives Prasad-Rao type estimators): Use $\widehat{\boldsymbol{\beta}}$ that satisfies

$$\sum_{i=1}^{m} w_i \left(\mathbf{x}_i \mathbf{x}_i' - \mathbf{C}_i \right) \widehat{\boldsymbol{\beta}} = \sum_{i=1}^{m} w_i \mathbf{x}_i y_i$$

(Modify if inverse does not exist)

- $w_i = 1$ or $w_i = 1/(\sigma_v^2 + \psi_i + \boldsymbol{\beta}' \mathbf{C}_i \boldsymbol{\beta})$.

Parameters unknown

$$\text{Cov}\begin{pmatrix}\widehat{\sigma}_v^2\\ \widehat{\boldsymbol{\beta}}\end{pmatrix} = \begin{pmatrix}V(\widehat{\sigma}_v^2) & \mathbf{d}_m'\\ \mathbf{d}_m & \mathbf{B}_m\end{pmatrix} + o(m^{-1}) = O(m^{-1})$$

MSE $(\widehat{Y}_{i\text{ME}})$

$$\begin{aligned}
&= \gamma_i\psi_i + (1-\gamma_i)^2 \text{tr}\left\{(\mathbf{C}_i + \mathbf{X}_i\mathbf{X}_i')\mathbf{B}_m\right\}\\
&+ \frac{\psi_i^2}{(\boldsymbol{\beta}'\mathbf{C}_i\boldsymbol{\beta}+\sigma_v^2+\psi_i)^3} E(\widehat{\sigma}_v^2+\widehat{\boldsymbol{\beta}}'\mathbf{C}_i\widehat{\boldsymbol{\beta}}-\sigma_v^2-\boldsymbol{\beta}'\mathbf{C}_i\boldsymbol{\beta})^2\\
&+ 2E\left\{(1-\widehat{\gamma}_i)^2(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\right\}\mathbf{C}_i\boldsymbol{\beta} + o(m^{-1})
\end{aligned}$$

Estimate $\boldsymbol{\beta}$, σ_v^2 without area i, so independent of $(\mathbf{x}_i, y_i)$ (Buonaccorsi 1995)

Multivariate Model

- Consider y_i, x_i as multivariate response
- Fay (1987) and Datta et al. (1991): multivariate response from one survey
- Huang and Bell (2004): consider Current Population Survey and American Community Survey for SAIPE (Bayesian approach)

Multivariate Model

- Allow missing responses
- If $\mathbf{x}$, y both measured:

$$\begin{pmatrix} \mathbf{x}_i \\ y_i \end{pmatrix} = \mathbf{T}_i \boldsymbol{\alpha} + \mathbf{u}_i + \mathbf{e}_i$$

$$\mathbf{u}_i \sim (\mathbf{0}, \boldsymbol{\Sigma}),\ \mathbf{e}_i \sim (\mathbf{0}, \boldsymbol{\Psi}_i)$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}'_{xy} & \boldsymbol{\Sigma}_{yy} \end{bmatrix},\ \ \boldsymbol{\Psi}_i = \begin{bmatrix} \mathbf{C}_i & \boldsymbol{\Psi}_{ixy} \\ \boldsymbol{\Psi}'_{ixy} & \psi_i \end{bmatrix}$$

Predictor (parameters known)

Both x and y observed for area i:

$$
\begin{aligned}
\tilde{Y}_{i\text{MFH}} &= \kappa_i y_i + (1-\kappa_i)(\mathbf{0}_p', 1)\mathbf{T}_i\boldsymbol{\alpha} \\
&\quad + \left\{(1-\kappa_i)\boldsymbol{\Sigma}_{xy}'\mathrm{M}_i - \kappa_i\boldsymbol{\Psi}_{ixy}'\mathrm{M}_i\right\} \\
&\quad \{\mathrm{x}_i - (\mathbf{I}_p|\mathbf{0}_p)\mathbf{T}_i\boldsymbol{\alpha}\}
\end{aligned}
$$

$$
\mathrm{M}_i = (\boldsymbol{\Sigma}_{xx} + \mathrm{C}_i)^{-1}
$$

$$
\kappa_i = \frac{\boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{xy}'\mathrm{M}_i(\boldsymbol{\Sigma}_{xy} + \boldsymbol{\Psi}_{ixy})}{\boldsymbol{\Sigma}_{yy} - (\boldsymbol{\Sigma}_{xy} + \boldsymbol{\Psi}_{ixy})'\mathrm{M}_i(\boldsymbol{\Sigma}_{xy} + \boldsymbol{\Psi}_{ixy}) + \psi_i}
$$

$\mathbf{x}$ is observed in area i but y is not:

$$\tilde{Y}_{i\text{MFH}} = [\mathbf{0}'_p, 1]\mathbf{T}_i\boldsymbol{\alpha} + \boldsymbol{\Sigma}'_{xy}\mathbf{M}_i(\mathbf{x}_i - [\mathbf{I}_p\ \mathbf{0}_p]\mathbf{T}_i\boldsymbol{\alpha})$$

$$\text{MSE}\,(\tilde{Y}_{i\text{MFH}}) = \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}'_{xy}\mathbf{M}_i\boldsymbol{\Sigma}_{xy}$$

y is observed in area i but $\mathbf{x}$ is not:

$$\tilde{Y}_{i\text{MFH}} = \kappa_i y_i + (1 - \kappa_i)[\mathbf{0}'_p, 1]\mathbf{T}_i\boldsymbol{\alpha},$$

$$\text{MSE}\,(\tilde{Y}_{i\text{MFH}}) = \kappa_i\psi_i$$

$$\kappa_i = \boldsymbol{\Sigma}_{yy}/(\boldsymbol{\Sigma}_{yy} + \psi_i)$$

Predictor (parameters known)

- If x, y from independent designs ($\boldsymbol{\Psi}_{ixy} = \mathbf{0}$),

$$\kappa_i = \frac{\boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}'_{xy}\mathrm{M}_i\boldsymbol{\Sigma}_{xy}}{\boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}'_{xy}\mathrm{M}_i\boldsymbol{\Sigma}_{xy} + \psi_i}, \quad \mathrm{M}_i = (\boldsymbol{\Sigma}_{xx} + \mathrm{C}_i)^{-1}$$

- As in ME model, more reliance on y_i if
  - ψ_i small
  - C_i large (in positive definite sense) (implies $\boldsymbol{\Sigma}'_{xy}\mathrm{M}_i\boldsymbol{\Sigma}_{xy}$ small)

Parameters Unknown

Can estimate by ML, REML, PR-type

$$\mathbf{Z}_i = \begin{cases} \mathbf{I}_{p+1} & \text{both } \mathbf{x} \text{ and } y \\ [\mathbf{I}_p \ \ \mathbf{0}_p]' & \mathbf{x} \text{ but no } y \\ \left[\mathbf{0}_p', 1\right]' & y \text{ but no } \mathbf{x} \end{cases}$$

$$\widehat{\boldsymbol{\alpha}}_{\text{OLS}} = \left(\sum_{j=1}^{m} \mathbf{T}_j' \mathbf{Z}_j \mathbf{Z}_j' \mathbf{T}_j \right)^{-1} \sum_{j=1}^{m} \mathbf{T}_j' \mathbf{Z}_j \begin{pmatrix} \mathbf{x}_j \\ y_j \end{pmatrix}$$

$$\widehat{\mathbf{e}}_i = \mathbf{Z}_i' \left[\begin{pmatrix} \mathbf{x}_i \\ y_i \end{pmatrix} - \mathbf{T}_i \widehat{\boldsymbol{\alpha}}_{\text{OLS}} \right]$$

Parameters Unknown

$$\mathbf{R}_{ij} = \mathbf{Z}_i'\mathbf{T}_i \left(\sum_{k=1}^{m} \mathbf{T}_k'\mathbf{Z}_k\mathbf{Z}_k'\mathbf{T}_k \right)^{-1} \mathbf{T}_j'\mathbf{Z}_j$$

$$\begin{aligned}\mathbf{F} \;&=\; \sum_{i=1}^{m} \Big(\mathbf{Z}_i\mathbf{Z}_i' \otimes \mathbf{Z}_i\mathbf{Z}_i' - \mathbf{Z}_i\mathbf{R}_{ii}\mathbf{Z}_i' \otimes \mathbf{Z}_i\mathbf{Z}_i' \\ &\qquad - \mathbf{Z}_i\mathbf{Z}_i' \otimes \mathbf{Z}_i\mathbf{R}_{ii}\mathbf{Z}_i' + \sum_{j=1}^{m} \mathbf{Z}_i\mathbf{R}_{ij}\mathbf{Z}_j' \otimes \mathbf{Z}_i\mathbf{R}_{ij}\mathbf{Z}_j' \Big)\end{aligned}$$

$$\begin{aligned}\text{vec}\,\widehat{\boldsymbol{\Sigma}} \;&=\; \mathbf{F}^{-1} \sum_{i=1}^{m} \text{vec} \Big[\mathbf{Z}_i\widehat{\mathbf{e}}_i\widehat{\mathbf{e}}_i'\mathbf{Z}_i' \\ &\qquad - \mathbf{Z}_i \Big(\boldsymbol{\Psi}_i - \boldsymbol{\Psi}_i\mathbf{R}_{ii} - \mathbf{R}_{ii}\boldsymbol{\Psi}_i + \sum_{j=1}^{m} \mathbf{R}_{ij}\boldsymbol{\Psi}_j\mathbf{R}_{ji} \Big) \mathbf{Z}_i' \Big]\end{aligned}$$

Multivariate: Parameters Unknown

$$\text{MSE}\,(\widehat{Y}_{i\text{MFH}}) = g_{1i} + g_{2i} + g_{3i} + o(m^{-1})$$

Both $\mathbf{x}$ and y measured:

$$\begin{aligned}
g_{1i} &= \kappa_i \psi_i \\
g_{2i} &= [\mathbf{0}_p', 1](\mathbf{I} - \mathbf{\Sigma} \mathbf{Z}_i \mathbf{V}_i^{-1} \mathbf{Z}_i')\mathbf{T}_i \\
& \left(\sum_{j=1}^{m} \mathbf{T}_j' \mathbf{Z}_j \mathbf{V}_j^{-1} \mathbf{Z}_j' \mathbf{T}_j \right)^{-1} \mathbf{T}_i' (\mathbf{I} - \mathbf{Z}_i \mathbf{V}_i^{-1} \mathbf{Z}_i' \mathbf{\Sigma}) [\mathbf{0}_p', 1]' \\
g_{3i} &= \text{tr} \left\{ \frac{\partial \mathbf{b}_i'}{\partial \boldsymbol{\delta}} \mathbf{V}_i \left(\frac{\partial \mathbf{b}_i'}{\partial \boldsymbol{\delta}} \right)' \text{avar}(\widehat{\boldsymbol{\delta}}) \right\}
\end{aligned}$$

$\mathbf{b}_i' = \mathbf{\Sigma} \mathbf{Z}_i' \mathbf{V}_i^{-1}$, $\mathbf{\Sigma} = \mathbf{\Sigma}(\boldsymbol{\delta})$ (Datta and Lahiri 2000)

Dual Frame Approach

- Assume response variable is same from both surveys ($Y_i = X_i$), samples independent
- Elliott and Davis (2005)
- Hartley (1962) dual frame estimator when frame populations coincide:

$$\tilde{Y}_{i\mathrm{DF}} = \frac{C_i}{C_i + \psi_i} y_i + \frac{\psi_i}{C_i + \psi_i} x_i$$

Dual Frame / Multivariate / Measurement Error

- $\tilde{Y}_{i\text{DF}} = \frac{C_i}{C_i + \psi_i} y_i + \frac{\psi_i}{C_i + \psi_i} x_i$
- Model for pop means $Y_i = X_i$; $\sigma_v^2 = 0$
- Under model,

$$\tilde{Y}_{i\text{DF}} = \tilde{Y}_{i\text{MFH}} = \tilde{Y}_{i\text{ME}}$$

- If model does not hold, $\tilde{Y}_{i\text{DF}}$ biased

Comparison

Both $\mathbf{x}$ and y are measured in area i, model

$$\begin{pmatrix} \mathbf{x}_i \\ y_i \end{pmatrix} = \begin{pmatrix} \mathbf{X}_i \\ Y_i \end{pmatrix} + \mathbf{e}_i = \mathbf{u}_i + \mathbf{e}_i,$$

$\mathbf{u}_i \sim N(\mathbf{0}, \mathbf{\Sigma})$, $\mathbf{e}_i \sim N(\mathbf{0}, \mathbf{\Psi}_i)$, all $\mathbf{u}_i$ and $\mathbf{e}_j$ independent

$$y_i \mid \mathbf{X}_i \sim N(\mathbf{\Sigma}'_{xy}\mathbf{\Sigma}^{-1}_{xx}\mathbf{X}_i, \mathbf{\Sigma}_{yy} - \mathbf{\Sigma}'_{xy}\mathbf{\Sigma}^{-1}_{xx}\mathbf{\Sigma}_{xy} + \psi_i)$$

Conditional Bias

Method	$E[\tilde{Y}_i - Y_i \mid \mathbf{X}_i]$
FH (no $\mathbf{X}$s)	0
FH ($\mathbf{X}$ known)	0
MFH	$-(1-\kappa_i)\mathbf{X}_i'(\mathbf{C}_i + \mathbf{\Sigma}_{xx})^{-1}\mathbf{C}_i\boldsymbol{\beta}$
ME	0

Method	Multiplier of y_i (γ_i, κ_i)
FH (no xs)	$\dfrac{\Sigma_{yy}}{\Sigma_{yy} + \psi_i}$
FH (x known)	$\dfrac{\Sigma_{yy} - \Sigma'_{xy}\Sigma_{xx}^{-1}\Sigma_{xy}}{\Sigma_{yy} - \Sigma'_{xy}\Sigma_{xx}^{-1}\Sigma_{xy} + \psi_i}$
MFH	$\dfrac{\Sigma_{yy} - \Sigma'_{xy}(\mathrm{C}_i + \Sigma_{xx})^{-1}\Sigma_{xy}}{\Sigma_{yy} - \Sigma'_{xy}(\mathrm{C}_i + \Sigma_{xx})^{-1}\Sigma_{xy} + \psi_i}$
ME	$\dfrac{\Sigma_{yy} + \Sigma'_{xy}\Sigma_{xx}^{-1}(\mathrm{C}_i - \Sigma_{xx})\Sigma_{xx}^{-1}\Sigma_{xy}}{\Sigma_{yy} + \Sigma'_{xy}\Sigma_{xx}^{-1}(\mathrm{C}_i - \Sigma_{xx})\Sigma_{xx}^{-1}\Sigma_{xy} + \psi_i}$

g_1 term of MSE $=$ (multiplier) ψ_i

- FH (x known) $\leq$ MFH $\leq \left\{ \begin{array}{l} \text{ME} \\ \text{FH (no } \mathrm{x}\text{)} \end{array} \right\}$
- MFH best predictor in model used for comparison
- Gain from ME depends on $\mathrm{C}_i - \boldsymbol{\Sigma}_{xx}$ (want small)
- Penalty for estimating x_i rather than knowing it: MSE $(\tilde{Y}_{i\mathrm{MFH}})$ $-$ MSE $(\tilde{Y}_{i\mathrm{FH}})$

$$= \frac{\psi_i^2 \boldsymbol{\beta}' \mathrm{C}_i \mathrm{M}_i \boldsymbol{\Sigma}_{xx} \boldsymbol{\beta}}{(\sigma_v^2 + \psi_i)(\boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}'_{xy} \mathrm{M}_i \boldsymbol{\Sigma}_{xy} + \psi_i)}$$

% areas with $c = 3$		Empirical mean squared error					
		y_i	FH	FH, no $\mathbf{x}$	SUB	ME	MFH
0	$c_i = 0$	9.6	3.2	8.2	3.2	3.1	3.1
20	$c_i = 0$	9.5	3.1	8.0	3.4	3.4	3.4
	$c_i = 3$	10.8	3.3	9.2	10.2	7.7	7.9
50	$c_i = 0$	9.8	3.2	8.5	4.5	3.6	3.9
	$c_i = 3$	9.7	3.2	8.1	7.3	7.1	6.6
80	$c_i = 0$	10.0	3.2	8.8	5.2	3.8	4.4
	$c_i = 3$	9.7	3.2	8.1	6.7	7.2	6.6
100	$c_i = 3$	9.9	3.2	8.3	6.7	7.4	6.8

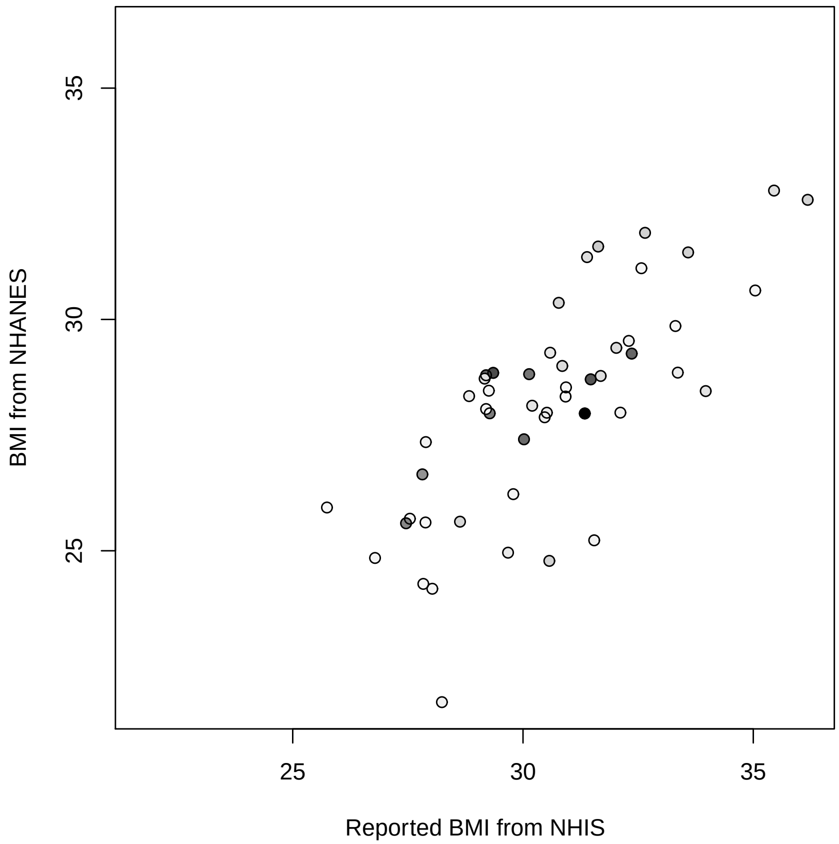
Body mass index for demographic groups

- Small areas: 50 demographic subgroups
  - Race/ethnicity:
    Mexican American, Other Hispanic, White non-Hispanic, Black non-Hispanic, Other
  - Age: 20–29, 30–39, 40–49, 50–59, 60–84
  - Sex
- Most detailed available on public use data set

Body mass index for demographic groups

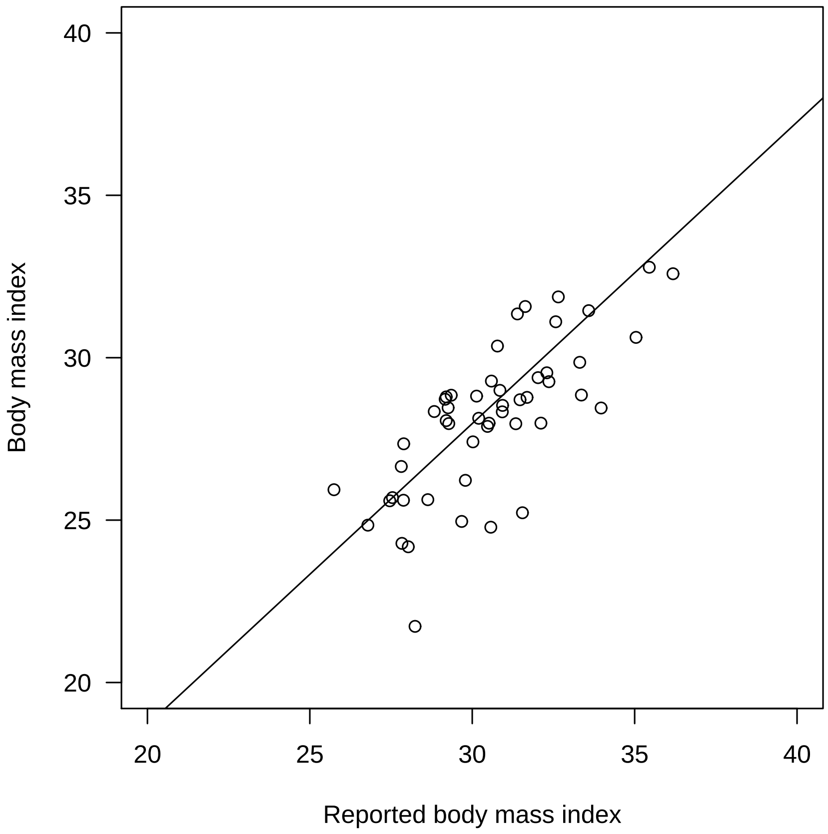
- NHANES measures height, weight, calculates BMI = height/weight^2, in units kg/m^2.
- NHIS asks height, weight, calculates Reported BMI (RBMI)
- NHANES group sample sizes: 8–433
- NHIS group sample sizes: 95–3087
- Expect high correlation for BMI, RBMI not same quantity (corr = .75)

Plot of y (NHANES) vs. x (NHIS)

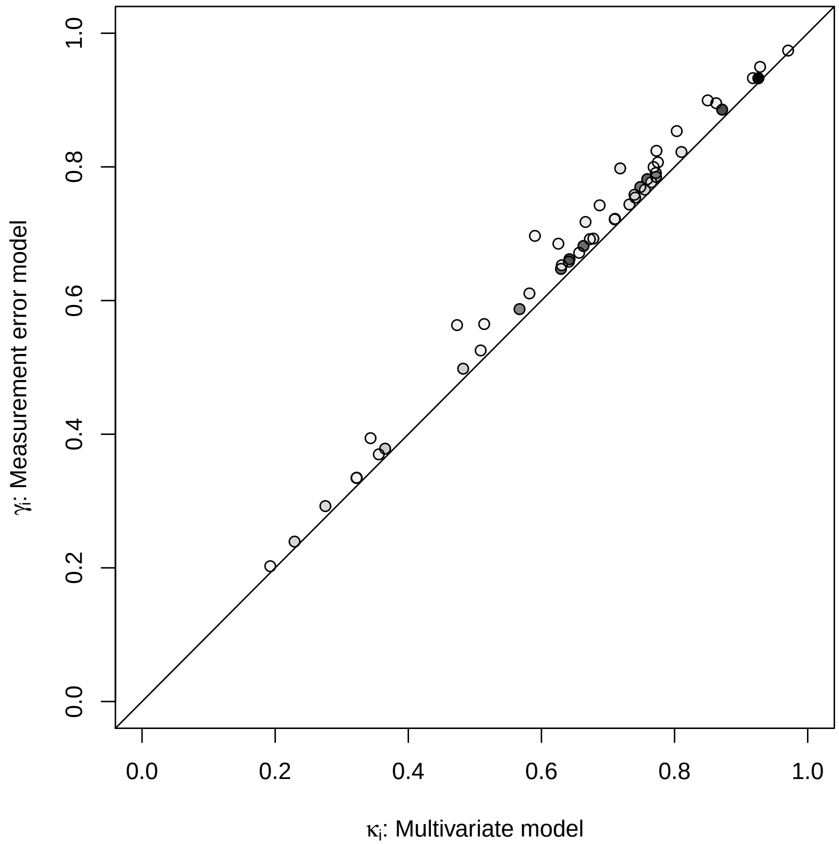


Shading $\propto$ NHIS sample size

WLS regression line, accounts for ME

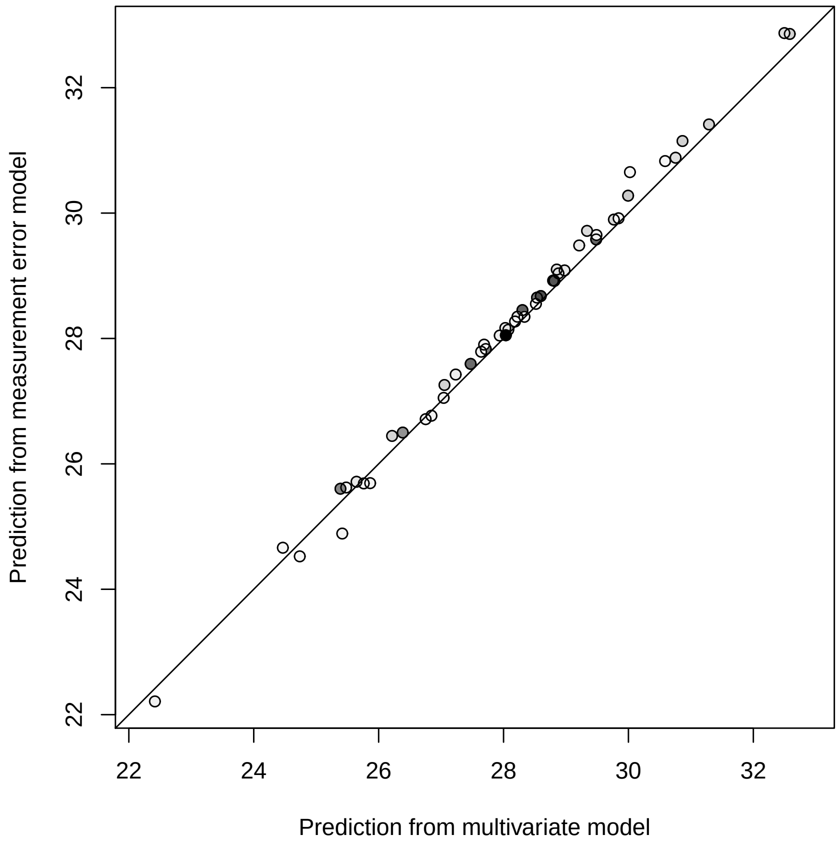


Multiplier of y_i



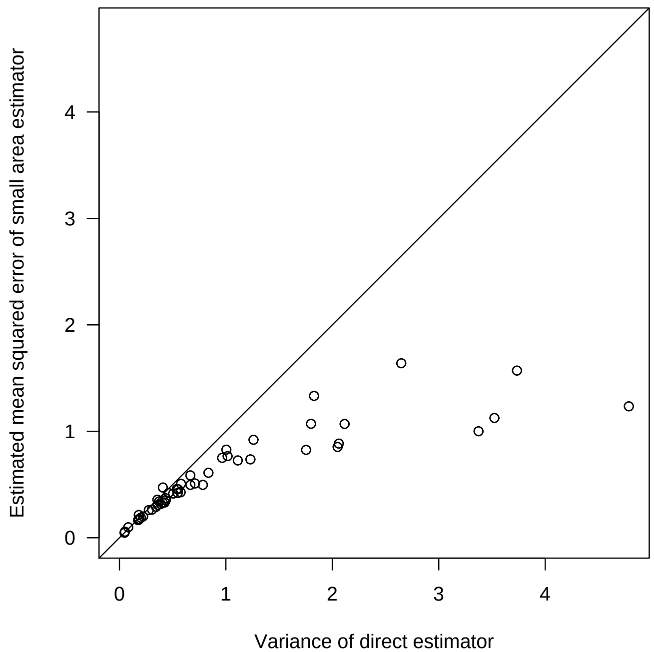
$\gamma_i \geq \kappa_i$:
ME model relies more on direct estimator

Predictions are similar

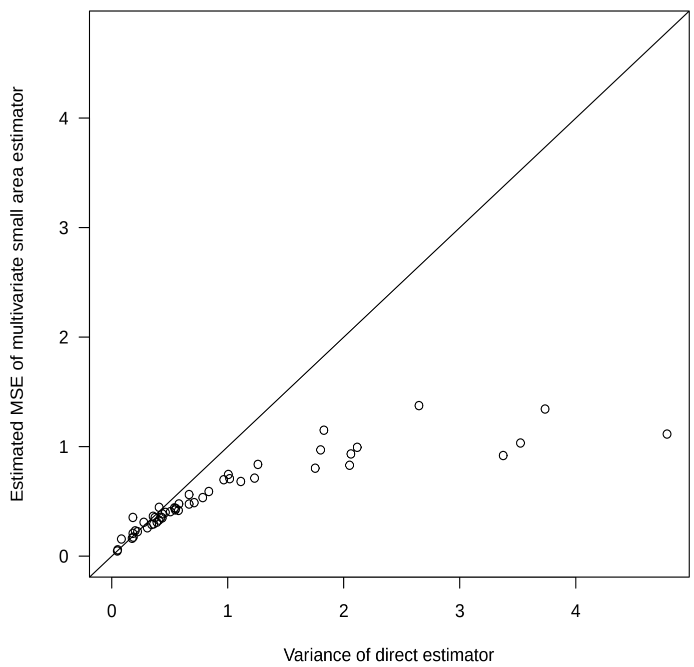


Estimated MSE

ME



MFH



Conclusions

- Measurement error, Multivariate Fay-Herriot models can reduce MSE
- Both better than ignoring error in x
- MFH shrinks y *and* x, relies less on direct est.
- Stability of estimators (Σ, σ_v^2, β)
- MSE estimation (Jiang, Lahiri, Wan 2002 jackknife)
- Model diagnostics
- Robustness (Carroll, Eltinge, Rupert 1993)