# Should we use the survey weights to weight?

Roderick Little

UNIVERSITY OF MICHIGAN

# Outline of talk

1. Big picture: design vs. model-based inference, weighting vs. prediction
2. Comparisons of weighting and prediction
3. Weighting and prediction for nonresponse
4. Robust modeling strategies
5. Variance estimation and inference

# Outline of talk

1. Big picture: design vs. model-based inference, weighting vs. prediction
2. Comparisons of weighting and prediction
3. Weighting and prediction for nonresponse
4. Robust modeling strategies
5. Variance estimation and inference

# Design vs. model-based survey inference

- Design-based (Randomization) inference
  - Survey variables $Y$ fixed, inference based on distribution of sample inclusion indicators, $I$
- Model-based inference: Survey variables $Y$ also random, assigned statistical model, often with fixed parameters. Two variants:
  - Superpopulation: Frequentist inference based on repeated samples from sample and superpopulation (hybrid approach)
  - Bayes: add prior for parameters; inference based on posterior distribution of finite population quantities
- key distinction in practice is randomization or model

# My overarching philosophy: calibrated Bayes

- Survey inference is not fundamentally different from other problems of statistical inference
  - But it has particular features that need attention
- Statistics is basically prediction: in survey setting, predicting survey variables for non-sampled units
- Inference should be model-based, Bayesian
- Seek models that are "frequency calibrated":
  - Incorporate survey design features
  - Properties like design consistency are useful
  - "objective" priors generally appropriate
    - Little (2004, 2006); Little & Zhang (2007)

# Weighting

- A pure form of design-based estimation is to weight sampled units by inverse of inclusion probabilities $w_i = 1/\pi_i$
  - Sampled unit *i* "represents" $w_i$ units in the population
- More generally, a common approach is:

$$w_i = w_{is} \times w_{in}(w_{is}) \times w_{ip}(w_{is}, w_{in})$$

$$w_{is} = \text{ sampling weight}$$

$$w_{in}(w_{is}) = \text{ nonresponse weight}$$

$$w_{ip}(w_{is}, w_{in}) = \text{ post-stratification weight}$$

# Prediction

- The goal of model-based inference is to predict the non-sampled values

$$\hat{T} = \sum_{i \in s} y_i + \sum_{i \in \bar{s}} \hat{y}_i$$

$$\hat{y}_i = \text{ prediction based on model } M$$

- Prediction approach captures design information with covariates, fixed and random effects, in the prediction model
- (objective) Bayes is superior conceptual framework, but superpopulation models are also useful
- Compare weighting and prediction approaches, and argue for model-based prediction

# The common ground

- Weighters can’t ignore models
- Modelers can’t ignore weights

# Weighters can’t ignore models

- Weighting units yields design-unbiased or design-consistent estimates
  - In case of nonresponse, under “quasirandomization” assumptions
- Simple, prescriptive
  - Appearance of avoiding an explicit model
- But poor precision, confidence coverage when “implicit model” is not reasonable
  - Extreme weights a problem, solutions often ad-hoc
  - Basu’s (1971) elephants

# Ex 1. Basu's inefficient elephants

$\left(y_1,\ldots,y_{50}\right) =$ weights of $N = 50$ elephants

Objective: $T = y_1 + y_2 + \ldots + y_{50}$. Only one elephant can be weighed!

- Circus trainer wants to choose "average" elephant (Sambo)
- Circus statistician requires "scientific" prob. sampling:

  Select Sambo with probability 99/100

  One of other elephants with probability 1/4900

  Sambo gets selected! Trainer: $\hat{t} = y_{(\text{Sambo})} \times 50$

  Statistician requires unbiased Horvitz-Thompson (1952)

estimator:
$$\hat{T}_{HT} = \begin{cases} y_{(\text{Sambo})} / 0.99 \text{ (!!);} \\ 4900 y_{(i)}, \text{if Sambo not chosen (!!!)} \end{cases}$$

HT estimator is unbiased on average but always crazy!

Circus statistician loses job and becomes an academic

# What went wrong?

- HT estimator optimal under an implicit model that $y_i / \pi_i$ have the same distribution
- That is clearly a silly model given this design …
- Which is why the estimator is silly

# Modelers can’t ignore weights

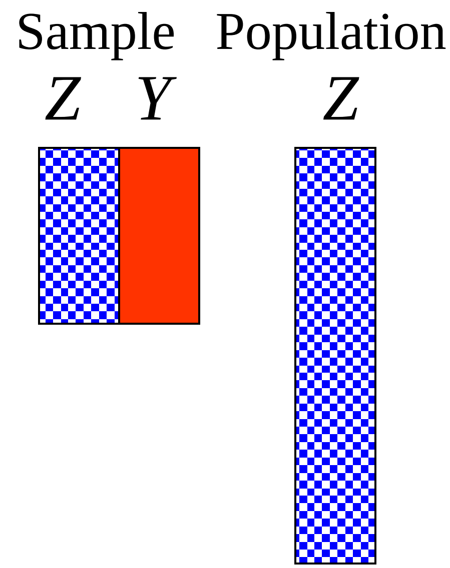
- All models are wrong, some models are useful
- Models that ignore features like survey weights are vulnerable to misspecification
  - Inferences have poor properties
  - See e.g. Kish & Frankel (1974), Hansen, Madow & Tepping (1983)
- But models <u>can</u> be successfully applied in survey setting, with attention to design features
  - Weighting, stratification, clustering

# Outline of talk

1. Big picture: design vs. model-based inference, weighting vs. prediction
2. Comparisons of weighting and prediction
3. Weighting and prediction for nonresponse
4. Robust modeling strategies
5. Variance estimation and inference

# Ex 2. One categorical post-stratifier $Z$

$$\bar{y}_{\text{mod}} = \bar{y}_{\text{wt}} = \sum_{j=1}^{J} P_j \bar{y}_j = \sum_{j=1}^{J} w_j n_j \bar{y}_j \Big/ \sum_{j=1}^{J} w_j n_j$$




In post-stratum $j$:

$P_j =$ population proportion

$n_j =$ sample count, $\bar{y}_j =$ sample mean of $Y$

$w_j = nP_j / n_j$

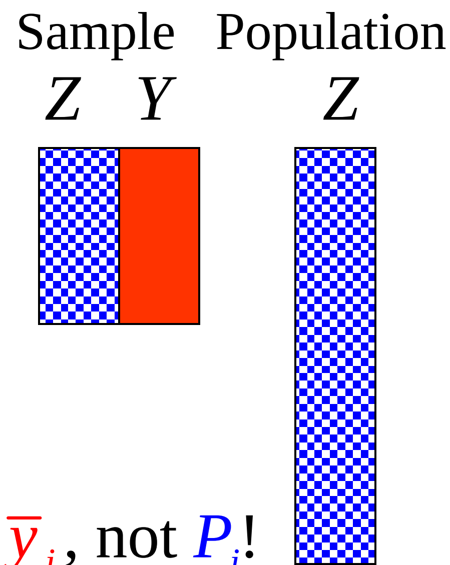
$\bar{y}_{\text{mod}} =$ prediction estimate for $y_{ji} \sim \text{Nor}(\mu_j, \sigma_j^2)$

$\quad =$ design-weighted estimate with weight $w_j$

Hence two approaches intersect ...

# One categorical post-stratifier $Z$

$$\bar{y}_{\text{mod}} = \bar{y}_{\text{wt}} = \sum_{j=1}^{J} P_j \bar{y}_j = \sum_{j=1}^{J} w_j n_j \bar{y}_j \Big/ \sum_{j=1}^{J} w_j n_j$$




1. Approaches differ in small samples:

Model replaces $\bar{y}_j$ by prediction $\hat{\mu}_j$ from model

E.g. $\mu_j \sim \text{Nor}(\mu, \tau^2)$ shrinks weight towards 1.

Design modifies weight $w_j$ -- but problem is with $\bar{y}_j$, not $P_j$!

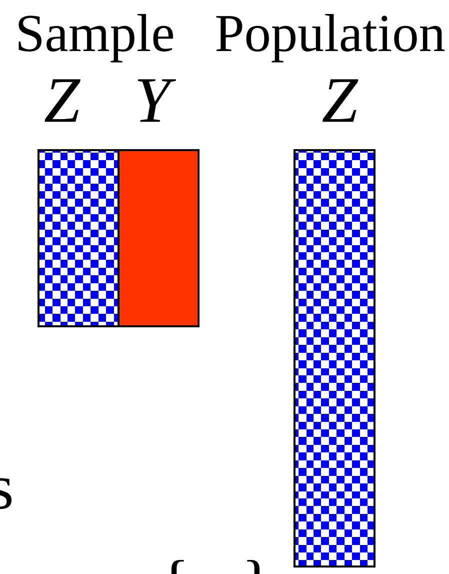
Changing $\bar{y}_j$ requires a model --

Modifications of weights often have an implicit model

Preferable to be explicit!

# One categorical post-stratifier $Z$

$$\bar{y}_{\text{mod}} = \bar{y}_{\text{wt}} = \sum_{j=1}^{J} P_j \bar{y}_j = \sum_{j=1}^{J} w_j n_j \bar{y}_j \Big/ \sum_{j=1}^{J} w_j n_j$$




2. Variance estimation:

Model automatically fixes the counts $\{n_j\}$;

Bayes with objective prior yields t-type corrections

Design approach is not clear on whether to condition on $\{n_j\}$

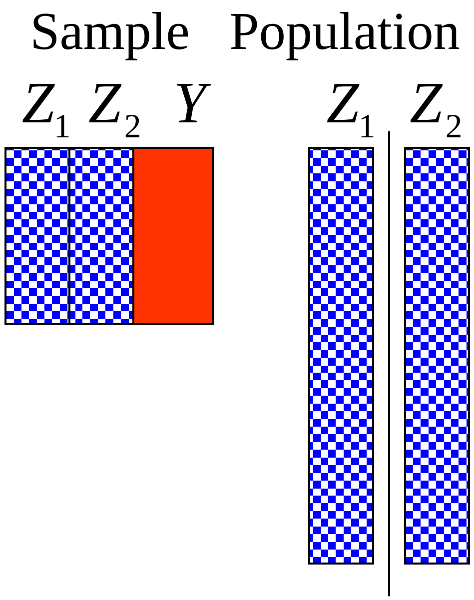
$\{n_j\}$ are not fixed in repeated sampling

But if allowed to vary, the sampling variance is not defined!

(Holt and Smith 1971, Little 1993)

# Ex 3. One stratifier $Z_1$, one post-stratifier $Z_2$

Design-based approaches




(A) Standard weighting is $w_i = w_{is} \times w_{ip}(w_{is})$

Notes: (1) $Z_1$ proportions are not matched!

(2) why not $w_i^* = w_{ip} \times w_{is}(w_{ip})$?

(B) Deville and Sarndal (1992) modifies sampling weights $\{w_{is}\}$ to adjusted weights $\{w_i\}$ that match poststratum margin, but are close to $\{w_{is}\}$ with respect to a distance measure $d(w_{is}, w_i)$.

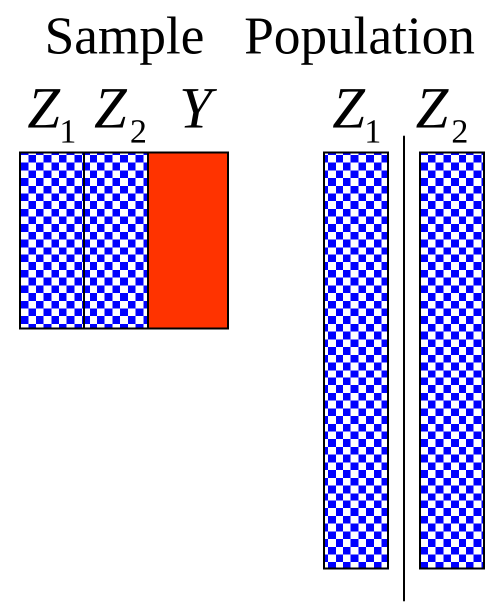
Questions:

What is the principle for choosing the distance measure?

Should the $\{w_i\}$ necessarily be close to $\{w_{is}\}$?

# Ex 3. One stratifier $Z_1$, one post-stratifier $Z_2$

<u>Model-based approach</u>




Saturated model: $\{n_{jk}\} \sim \text{MNOM}(n, \pi_{jk})$;

$$y_{jki} \sim \text{Nor}(\mu_{jk}, \sigma_{jk}^2)$$

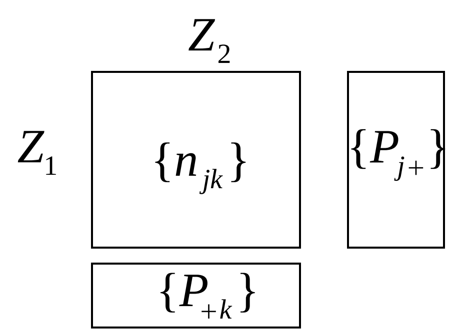
$$\bar{y}_{\text{mod}} = \sum_{j=1}^{J}\sum_{k=1}^{K} \hat{P}_{jk}\bar{y}_{jk} = \sum_{j=1}^{J}\sum_{k=1}^{K} w_{jk} n_{jk} \bar{y}_{jk} \Big/ \sum_{j=1}^{J}\sum_{k=1}^{K} w_{jk} n_{jk}$$

$n_{jk} =$ sample count, $\bar{y}_{jk} =$ sample mean of $Y$

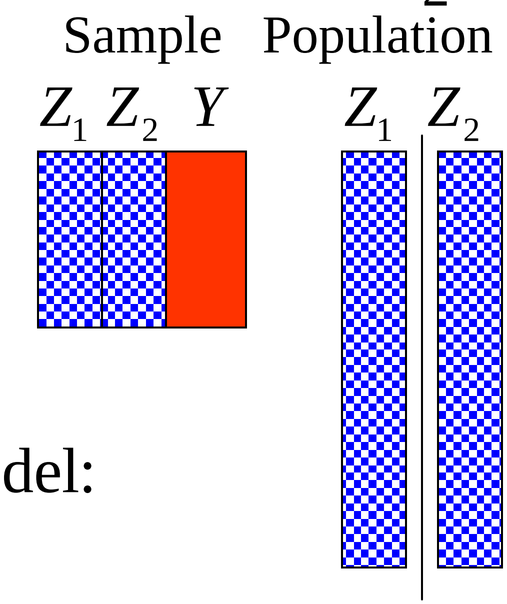
$\hat{P}_{jk} =$ proportion from raking (IPF) of $\{n_{jk}\}$

to known margins $\{P_{j+}\}, \{P_{+k}\}$

$w_{jk} = n\hat{P}_{jk} / n_{jk}$ = model weight



# Ex 3. One stratifier $Z_1$, one post-stratifier $Z_2$




<u>Model-based approach</u>

$$\bar{y}_{\text{st}} = \sum_{j=1}^{J}\sum_{k=1}^{K}\hat{P}_{jk}\bar{y}_{jk} = \sum_{j=1}^{J}\sum_{k=1}^{K} w_{jk} n_{jk}\bar{y}_{jk} \Big/ \sum_{j=1}^{J}\sum_{k=1}^{K} w_{jk} n_{jk}$$

What to do when $n_{jk}$ is small?

Model: replace $\bar{y}_{jk}$ by prediction from modified model:

e.g. $y_{jki} \sim \text{Nor}(\mu + \alpha_j + \beta_k + \gamma_{jk}, \sigma^2_{jk})$,

$\sum_{j=1}^{J}\alpha_j = \sum_{k=1}^{K}\beta_k = 0, \gamma_{jk} \sim \text{Nor}(0, \tau^2)$ (Gelman 2007)

Setting $\tau^2 = 0$ yields additive model,

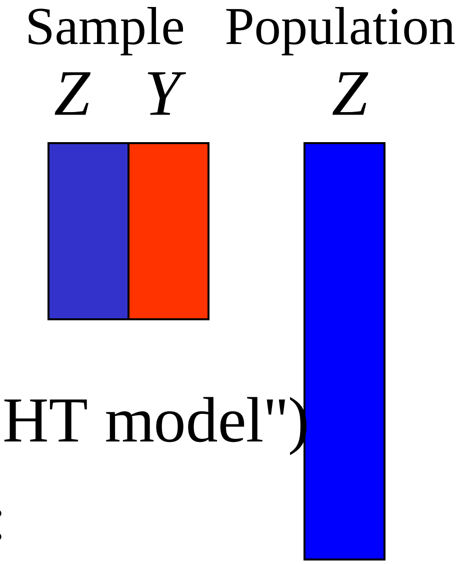
otherwise shrinks towards additive model

Design: arbitrary collapsing, ad-hoc modification of weight

# Ex 4. One continuous (post)stratifier $Z$

Consider PPS sampling, $Z$ = measure of size

<u>Design</u>: HT or Generalized Regression




$$\bar{y}_{\text{wt}} = \frac{1}{N}\left( \sum_{i=1}^{n} y_i / \pi_i \right); \pi_i = \text{ selection prob (HT)}$$

$$\bar{y}_{\text{wt}} \approx \text{ prediction estimate for } y_i \sim \text{Nor}(\beta\pi_i, \sigma^2\pi_i^2) \text{ ("HT model")}$$

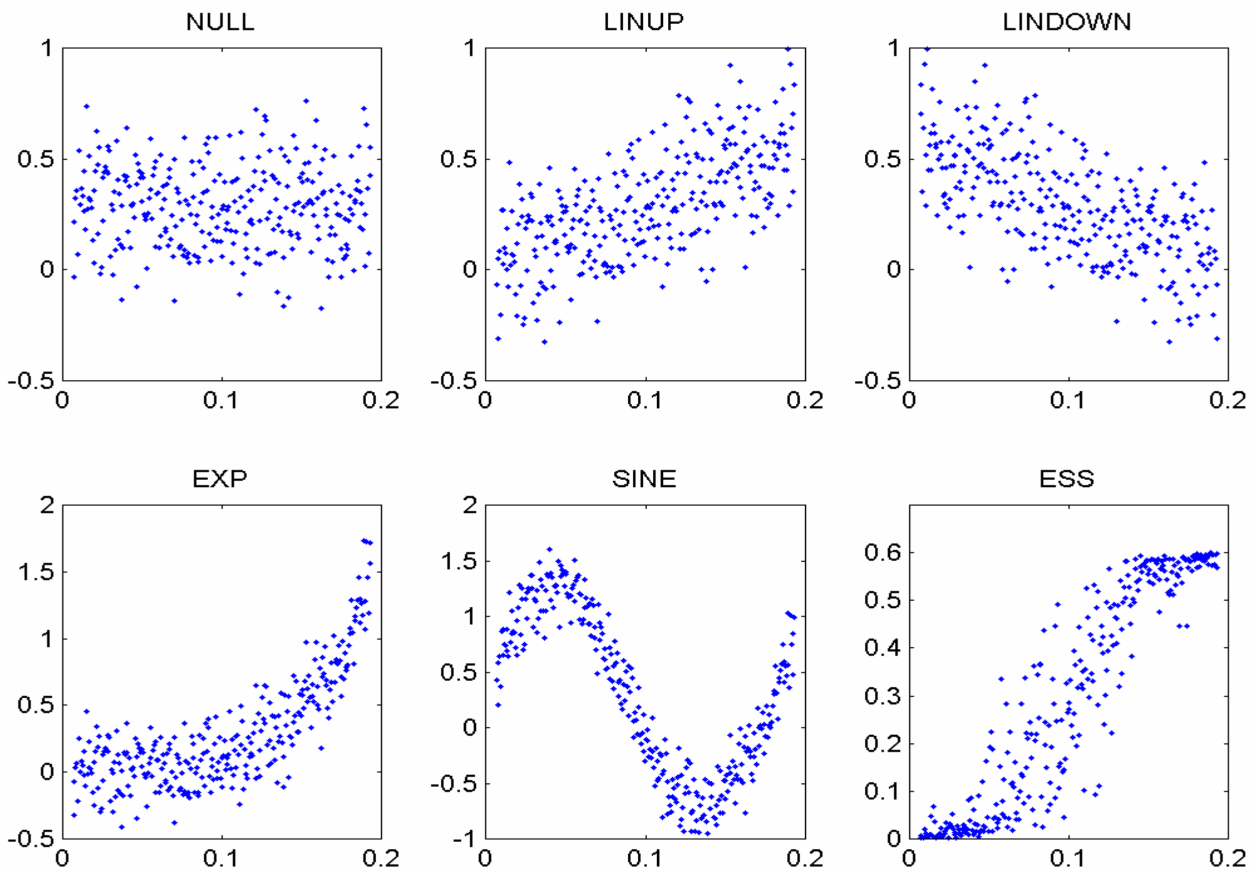
This motivates following robust modeling approach:

$$\bar{y}_{\text{mod}} = \frac{1}{N}\left( \sum_{i=1}^{n} y_i + \sum_{i=n+1}^{N} \hat{y}_i \right),\ \hat{y}_i \text{ predictions from:}$$

$$y_i \sim \text{Nor}(S(\pi_i), \sigma^2\pi_i^k); S(\pi_i) = \text{penalized spline of } Y \text{ on } Z$$

(Zheng and Little 2003, 2005)

# Simulation: PPS sampling in 6 populations



# Estimated RMSE of four estimators for N=1000, n=100

| Population | | model | wt | gr |
|---|---|---|---|---|
| NULL | Normal | **20** | 33 | 21 |
| | Lognormal | 32 | 44 | **31** |
| LINUP | Normal | **23** | 24 | 25 |
| | Lognormal | **25** | 30 | 30 |
| LINDOWN | Normal | 30 | 66 | **29** |
| | Lognormal | **24** | 65 | 28 |
| SINE | Normal | **35** | 134 | 90 |
| | Lognormal | **53** | 130 | 84 |
| EXP | Normal | **26** | 32 | 57 |
| | Lognormal | **40** | 41 | 58 |

# 95% CI coverages: HT

| Population | V1 | V3 | V4 | V5 |
|---|---|---|---|---|
| NULL | 90.2 | 91.4 | 90.0 | 90.4 |
| LINUP | 94.0 | 95.0 | 95.0 | 95.0 |
| LINDOWN | 89.0 | 89.8 | 90.0 | 90.6 |
| SINE | 93.2 | 93.4 | 93.0 | 93.0 |
| EXP | 93.6 | 94.6 | 95.0 | 95.0 |
| ESS | 95.0 | 95.6 | 95.4 | 95.2 |

V1 Yates-Grundy, Hartley-Rao for joint inclusion probs.
V3 Treating sample as if it were drawn with replacement
V4 Pairing consecutive strata
V5 Estimation using consecutive differences

# 95% CI coverages: B-spline

| Population | V1 | V2 | V3 |
|---|---|---|---|
| NULL | 95.4 | 95.8 | 95.8 |
| LINUP | 94.8 | 97.0 | 94.6 |
| LINDOWN | 94.2 | 94.2 | 94.6 |
| SINE | **88.0** | 92.6 | 97.4 |
| EXP | 94.4 | 95.2 | 95.6 |
| ESS | 97.4 | 95.4 | 95.8 |

Fixed with more knots

V1 Model-based (information matrix)
V2 Jackknife
V3 BRR

# Why does model do better?

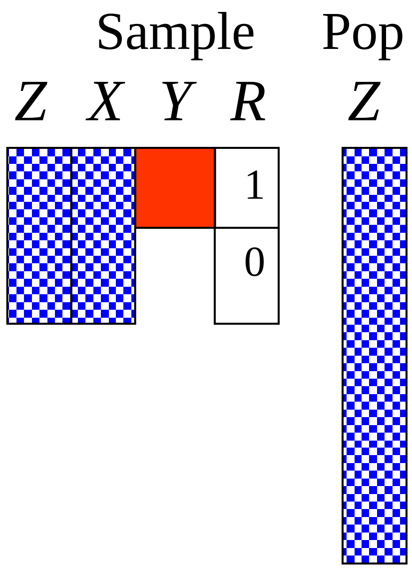
- Assumes smooth relationship – HT weights can "bounce around"
- Predictions use sizes of the non-sampled cases
  - HT estimator does not use these
  - Often not provided to users (although they could be)
- Little & Zheng (2007) also show gains for model when sizes of non-sampled units are not known
  - Predicted using a Bayesian Bootstrap (BB) model
  - BB is a form of stochastic weighting

# Outline of talk

1. Big picture: design vs. model-based inference, weighting vs. prediction
2. Comparisons of weighting and prediction
3. Weighting and prediction for nonresponse
4. Robust modeling strategies
5. Variance estimation and inference

# Ex 5. Unit nonresponse

- Predict nonrespondents by regression on design variables $Z$ and any observed survey variables $X$
- For bias reduction, predictors should be related to propensity to respond $R$ <u>and</u> outcome $Y$
- In choosing from a set of predictors, good prediction of $Y$ is more important than good prediction of $R$



# Impact of weighting for nonresponse

$\text{corr}^2(X,Y)$

| $\text{corr}^2(X,R)$ | Low | High |
|---|---|---|
| Low | --- | var $\Downarrow$ |
| High | var $\uparrow$ | var $\Downarrow$ bias $\Downarrow$ |

Too often adjustments do this?

- Standard "rule of thumb" $\text{Var}(\bar{y}_{\text{w}}) = \text{Var}(\bar{y}_{\text{u}})(1+\text{cv}(w))$ fails to reflect that nonresponse weighting can <u>reduce</u> variance
- Little & Vartivarian (2005) propose refinements

# Weighting squared?

- Nonresponse weights are often computed using units weighted by their sampling weights $\{w_{1i}\}$

$$w_{2j}^{-1} = \left(\sum_{r_i=1, x_i=j} w_{1i}\right) / \left(\sum_{r_i=1, x_i=j} w_{1i} + \sum_{r_i=0, x_i=j} w_{1i}\right)$$

  - Gives unbiased estimate of response rate in each adjustment cell defined by *X*
  - Not correct from a prediction perspective
  - For nonresponse bias, need to condition on *X* <u>and</u> *Z*, not just *X*
  - Does not correct for bias when *Z* is associated with *R* and *Y*
  - Need to condition on design variables involved in sampling weight (as in predictive inference)

# Simulation Study

- Simulations to provide insight into the variance and bias of estimators of weighted and unweighted rates and alternative estimators, under a variety of population structures and nonresponse mechanisms. (Little & Vartivarian 2003)
- Categorical outcome, to avoid distributional assumptions such as normality.
- 25 populations to cover all combinations of models for $Y$ and $R$ given $Z$ and $X$

# RMSE's of three methods

| Y model | **XZ** | **XZ** | XZ | **XZ** | XZ | **add** | **add** | add | **add** | add | **Z** | **Z** | Z | **Z** | Z | **Ave*** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| R model | **XZ** | **add** | X | **Z** | Ø | **XZ** | **add** | X | **Z** | Ø | **XZ** | **add** | X | **Z** | Ø | **Ave*** |
| Pred(xz) | **40** | **41** | 42 | **36** | 36 | **38** | **39** | 39 | **37** | 33 | **36** | **38** | 38 | **35** | 34 | **38.2** |
| urr(x) | **49** | **53** | 42 | **51** | 37 | **54** | **71** | 40 | **59** | 34 | **58** | **76** | 39 | **63** | 35 | **47.1** |
| wrr(x) | **51** | **58** | 42 | **53** | 37 | **56** | **77** | 41 | **61** | 35 | **56** | **71** | 39 | **62** | 35 | **47.1** |

(*Models for *Y* that exclude *Z* are omitted to save space – methods are all similar for these cases)

**urr(x) is biased when both *Y* and *R* depend on *Z*.**

**wrr(x) does not generally correct the bias in these situations: similar to urr[x] overall**

Prediction based on model for *X* and *Z* is best

# Item nonresponse

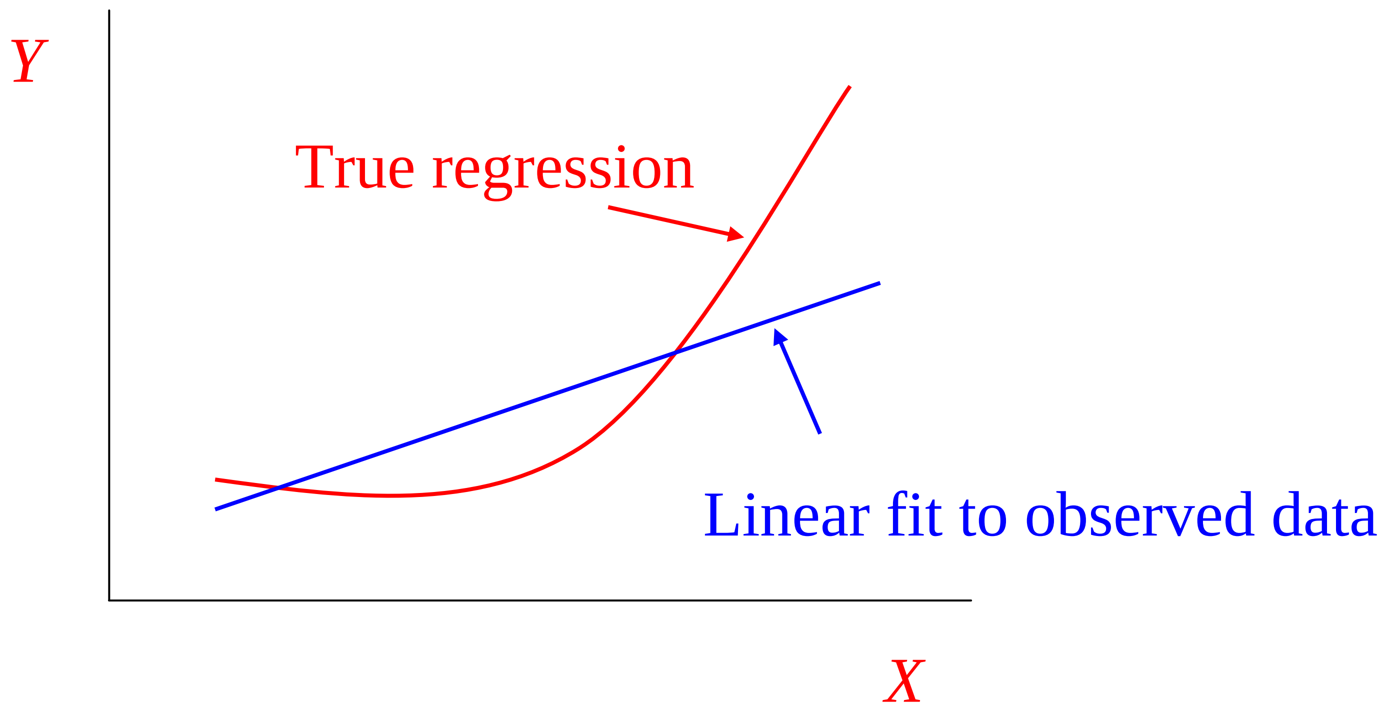
- Item nonresponse generally has complex "swiss-cheese" pattern
- Weighting methods are possible when the data have a monotone pattern, but are very difficult to develop for a general pattern
- Model-based multiple imputation methods are available for this situation (Little & Rubin 2002)
  - By conditioning fully on all observed data, these methods weaken MAR assumption

# Outline of talk

1. Big picture: design vs. model-based inference, weighting vs. prediction.
2. Comparisons of weighting and prediction
3. Weighting and prediction for nonresponse
4. Robust modeling strategies
5. Variance estimation and inference

# Making predictions more robust

- Model predictions of missing values are potentially sensitive to model misspecification, particularly if data are not MCAR



# Relaxing Linearity: one *X*

- A simple way is to categorize $X_1$ and predict within classes -- link with *weighting* methods
- For continuous $X_1$ and sufficient sample size, a spline provides one useful alternative (cf. Breidt & Opsomer 2000) . We use a P-Spline approach:

$$(Y_1 \mid X_1, \phi) \sim \mathrm{Nor}\left( s_1(X_1, \phi), \sigma^2 \right)$$

$$s_1\left( X_1, \phi \right) = \phi_0 + \sum_{j=1}^{q} \phi_j X_1^{\,j} + \sum_{k=1}^{K} \phi_{q+k} \left( X_1 - \tau_k \right)_+^q,$$

$$\left( x \right)_+^q = x^q I\left( x \geq 0 \right),$$

$\tau_1 < \cdots < \tau_K$ are selected fixed knots

$\phi_{\mathrm{q+1}}, ..., \phi_{q+K}$ are random effects, shrink to zero

# More than one covariate

- When we model the relationship with many covariates by smoothing, we have to deal with the "curse of dimensionality".
  - One approach is to "calibrate" the model by adding weighted residuals (e.g. Scharfstein & Izzarry 2004, Bang & Robins 2005).
  - Strongly related to generalized regression approach in surveys (Särndal, Swensson & Wretman 1992)
  - Little & An (2004) achieve both robustness and dimension reduction with many covariates, using the conceptually simple model-based approach.

# Penalized Spline of Propensity Prediction (PSPP)

- Focus on a particular function of the covariates most sensitive to model misspecification, the response propensity score.
- Important to get relationship between *Y* and response propensity correct, since misspecification of this leads to bias (Rubin 1985, Rizzo 1992)
- Other *X*'s balanced over respondents and nonrespondents, conditional on propensity scores (Rosenbaum & Rubin 1983); so misspecification of regression of these is less important (loss of precision, not bias).

# The PSPP method

Define: $Y^{*}$=logit (Pr($R$ = 1|$X_1$,…,$X_p$ )) (Need to estimate)

$$(Y \mid Y^{*}, X_1, \ldots, X_p; \beta) \sim$$

$$\mathrm{Nor}(s(Y^{*}) + g(Y^{*}, X_2, \ldots, X_p; \beta), \sigma^2)$$

- Nonparametric part
- Need to be correctly specified
- We choose penalized spline

- Parametric part
- Misspecification does not lead to bias
- Increase precision
- $X_1$ excluded to prevent multicollinearity

# Double Robustness Property

- The PSPP method yields a consistent estimator for the marginal mean of $Y$, if:

  (a) the mean of $Y$ given $X_1,\ldots,X_p$ is correctly specified,

OR

  (b1) the propensity is correctly specified, and

  (b2) relationship of outcome with propensity is correctly specified

  ***Note: in (b), parametric part of model does not have to be correctly specified!***

- PSPP can be extended to handle general patterns of missing data
- Applies to other problems of selection, e.g. sampling (Little & An 2004)

# Role of Models in Classical Approach

- Models are often used to motivate the choice of estimator. For example:
  - Regression model $\longrightarrow$ regression estimator
  - Ratio model $\longrightarrow$ ratio estimator
  - Generalized Regression estimation: model estimates adjusted to protect against misspecification, e.g. HT estimation applied to residuals from the regression estimator (e.g. Särndal, Swensson & Wretman 1992).
- Estimates of standard error are then based on the randomization distribution
- This approach is design-based, model-assisted

# Comments

- Calibration approach yields double robustness
- However, relatively easy to achieve double robustness in the direct prediction approach, using methods like PSPP (see Firth & Bennett 1998)
- Calibration estimates can be questionable from a modeling viewpoint
- If model is robust, calibration is unnecessary and adds noise
  - Recent simulations by Guanyu Zhang support this

# Outline of talk

1. Big picture: design vs. model-based inference, weighting vs. prediction.
2. Comparisons of weighting and prediction
3. Weighting and prediction for nonresponse
4. Model-assisted estimation, double robustness
5. Variance estimation and inference

# Standard errors, inference

- Should be more emphasis on confidence coverage, less on estimating the standard error
- Model-based inferences
  - Need to model variance structure carefully
  - Bayes: good for small samples
- Sample reuse methods (bootstrap, jackknife, BRR)
  - More acceptable to practitioners
  - Large sample robustness (compare sandwich estimation)
  - Inferentially not quite as pure, but practically useful

# Summary

- Compared design-based and model-based approaches to survey weights
- Design-based: “VW beetle” (slow, reliable)
- Model-based: “T-bird” (more racy, needs tuning)
- Personal view: model approach is attractive because of flexibility, inferential clarity
- Advocate survey inference under “weak models”

# Acknowledgments


- JPSM for invitation to speak
- Don Rubin for inspiration and guidance
- Trivellore Raghunathan for being a great colleague
- Current and past graduate students for all their ideas and work
  - Di An, Hyonggin An, Michael Elliott, Laura Lazzaroni, Hui Zheng, Sonya Vartivarian, Mei-Miau Wu, Ying Yuan, Guangyu Zhang

# References


Bang, H. & Robins, J.M. (2005). *Biometrics* 61, 962-972.

Basu, D. (1971), p203-242, *Foundations of Statistical Inference*, Holt, Rinehart & Winston: Toronto.

Breidt, F.J. & Opsomer, J.D. (2000). *Annals Statist.* 28, 1026-53.

Deville, J-C and Sarndal, C-E. JASA 87, 376-382

Firth, D. & Bennett, K.E. (1998). *JRSS* B 60, 3-21.

Gelman, A. (2007). To appear in *Stat. Science* (with discussion)

Hansen, MH, Madow, WG & Tepping, BJ (1983) *JASA* 78, 776-793.

Holt, D., & Smith, T.M.F. (1979). *JRSS*A, 142, 33-46.

Horvitz, D.G., & Thompson, D.J. (1952). *JASA*, 47, 663-685.

Kish, L., & Frankel, M. R. (1974). *JRSS* B, 36, 1–37.

Little, R.J.A. (1991). *JOS*, 7, 405-424.

______ (1993). *JASA*, 88, 1001-12.

______ (2004). *JASA*, 99, 546-556.

______ (2006). *Am. Statist.*, 60, 3, 213-223.

______ & An, H. (2004). *Statistica Sinica*, 14, 949-968.

______ & Rubin, D.B. (2002). *Statistical Analysis with Missing Data*, 2nd Ed. New York: Wiley.

______ & Vartivarian, S. (2003). *Stat.Med.* 22, 1589-1599.

______ & Vartivarian, S. (2005). *Survey Meth.*, 31, 161-168.

______ & Zheng, H. (2007). To appear in *Bayesian Statistics* 8, J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith & M. West (Eds.)

Rizzo, L. (1992). *JASA* 87, 1166-1173.

Rosenbaum, P. R. & Rubin, D. B. (1983). *Biometrika*, 70, 41-55

Rubin, D.B. (1985). pp. 463-472 in *Bayesian Statistic*s, eds. I.M. Bernado, M.H. DeGroot, D.V. Lindley, & A.F.M. Smith, Amsterdam: Elsevier,.

Särndal, C.-E., Swensson, B. & Wretman, J.H. (1992), *Model Assisted Survey Sampling*, Springer: New York.

Scharfstein, D. & Irizarry, R. (2003). *Biometrics* **59**, 601-613.

Yuan, Y. & Little, R.J. (2007a). To appear in *Applied Statistics*.

______ (2007b). To appear in *Biometrics*.

Zhang, G. & Little, R.J. (2005). *American Statistical Association, Proc. Joint Stat.Meetings*.

Zheng, H. & Little, R.J. (2003). *JOS* 19, 2, 99-117.

______ (2005) *JOS* 21, 1-20.