• Lecture 1 — Introduction, Examples, Effects of Measurement Error in Linear Models

• Lecture 2 — Data Types, Nondifferential Error, Estimating Attenuation, Exact Predictors, Berkson/Classical

• Lecture 3 — Distinguishing Berkson from Classical, Structural/Functional, Regression Calibration, Non-Additive Errors

• Lecture 4 — SIMEX and Instrumental Variables

• Lecture 5 — Bayesian Methods

• Lecture 6 — Nonparametric Regression with Measurement Error
LECTURE 1: INTRODUCTION, EXAMPLES AND LINEAR MEASUREMENT ERROR MODELS

OUTLINE

• Why We Need Special Methods For Measurement Errors

• Measurement Error Examples

• Structure of a Measurement Error Problem

• Classical Error Model in Linear Regression
Why We Need Special Methods To Handle Measurement Errors

- These lectures are about strategies for regression when some predictors are measured with error.
- Remember your introductory regression text . . .
  - Snedecor and Cochran (1967), “Thus far we have assumed that $X$-variable in regression is measured without error. Since no measuring instrument is perfect this assumption is often unrealistic.”
  - Steele and Torrie (1980), “. . . if the $X$’s are also measured with error, . . . an alternative computing procedure should be used . . .”
  - Neter and Wasserman (1974), “Unfortunately, a different situation holds if the independent variable $X$ is known only with measurement error:”
• **Measurement error in outcome:** If

\[ Y = \beta_0 + \beta_1 X + \epsilon \]

and we observe \( Y^* = Y + U \), then

\[ Y^* = \beta_0 + \beta_1 X + (\epsilon + U) \]

* All that has happened is that the error variance is bigger

* **Standard regression applies**
**Why Special Methods Are Needed — Cont.**

- **Consider now:**

  \[ Y = \beta_0 + \beta_1 X + \epsilon \]
  \[ W = X + U \]

  and

  only \( W \) and \( Y \) are observed.

  Then

  \[ Y = \beta_0 + \beta_1 W + (\epsilon - \beta_1 U) \]

  * This *seems like a standard regression problem* with additional error.

  - Can we analyze it that way?
WHY SPECIAL METHODS ARE NEEDED — CONT.

- From previous page: $X = W - U$ implies
  \[ Y = \beta_0 + \beta_1 W + (\epsilon - \beta_1 U) \]

- **Problem:** $W$ is correlated with the “error” $\epsilon - \beta_1 U$

- Better to replace $W$ by a function of $W$, namely $E(X|W)$, since
  \[ X = E(X|W) + V, \quad \text{where } E(X|W) \text{ and } V \text{ are uncorrelated} \]
  and then we have the model
  \[ Y = \beta_0 + \beta_1 E(X|W) + (\epsilon + \beta_1 V) \]

  * special case of “regression calibration”
  * leads to unbiased estimates
EXAMPLES OF MEASUREMENT ERROR PROBLEMS

- Different measures of nutrient intake, e.g., by a food frequency questionnaire (FFQ)
- Systolic Blood Pressure
- Radiation Dosimetry
- Exposure to arsenic in drinking water, dust in the workplace, radon gas in the home, and other environmental hazards
MEASURES OF NUTRIENT INTAKE

- $Y =$ average daily % calories from fat by a FFQ.
- $X =$ true long-term average daily percentage of calories from fat
- Assume $Y = \beta_0 + \beta_x X + \epsilon$
  * we do not assume that the FFQ is unbiased
- $X$ is never observable. It is measured with error:
  * Along with the FFQ, on 6 days over the course of a year women are interviewed by phone and asked to recall their food intake over the past year (24–hour recalls). Their average is recorded and denoted by $W$.
    - $W$ is an “alloyed gold standard”
  * The analysis of 24–hour recall introduces some error $\Rightarrow$ analysis error
  * Measurement error = sampling error + analysis error
  * Classical measurement error model:
    $$W_i = X_i + U_i,$$
    $U_i$ are measurement errors
**Heart Disease Vs Systolic Blood Pressure**

- $Y =$ indicator of Coronary Heart Disease (CHD)
- $X =$ true long-term average systolic blood pressure (SBP) (maybe transformed)
- Assume $P(Y = 1) = H(\beta_0 + \beta_x X)$, $H$ is the logistic or probit function
- Data are CHD indicators and determinations of systolic blood pressure for $n = 1600$ in Framingham Heart Study
**Heart Disease vs Systolic Blood Pressure — Cont.**

- $X$ measured with error:
  - SBP measured at two exams (and averaged) $\Rightarrow$ sampling error
  - The determination of SBP is subject to machine and reader variability $\Rightarrow$ analysis error
  - Measurement error = sampling error + analysis error
  - Measurement error model
    $$W_i = X_i + U_i, \quad U_i \text{ are measurement errors}$$
General Structure Of A Measurement Error Problem

- \( Y = \) response, \( Z = \) error-free predictor, \( X = \) error-prone predictor

- \( E(Y|Z, X) = f(Z, X, \beta) \) (outcome model)

- Observed data: \((Y_i, Z_i, W_i), i = 1, \ldots, n\)

  - \( * \ E(Y|Z, W) \neq f(Z, W, \beta) \) (source of our worries)

- Error model relating \( W_i \) and \( X_i \) (measurement model)

  - \( * \ W_i = X_i + U_i \) (classical error model)

  - \( * \ W_i = \gamma_{0,em} + \gamma_{x,em}^{t}X_i + \gamma_{z,em}^{t}Z_i + U_i \) (error calibration model)
A Classical Error Model

- $W_i = X_i + U_i$ (additive)

- $U_i$ are:
  - independent of all $Y_i, Z_i$ and $X_i$
  - IID$(0, \sigma_u^2)$

- In addition, we may have a model for the distribution of $(X, Z)$ (exposure model)
  - called a structural model
SIMPLE LINEAR REGRESSION WITH A CLASSICAL ERROR MODEL

- \( Y = \text{response, } X = \text{error-prone predictor} \)
- \( Y = \beta_0 + \beta_x X + \epsilon \)
- Observed data: \((Y_i, W_i), \ i = 1, \ldots, n\)
- \( W_i = X_i + U_i \) (additive)
- \( U_i \) are:
  * independent of all \( Y_i, Z_i \) and \( X_i \)
  * IID\((0, \sigma_u^2)\)

What are the effects of measurement error on the usual analysis?
**Simulation Study**

- Generate $X_1, \ldots, X_{50}$, IID $N(0, 1)$

- Generate $Y_i = \beta_0 + \beta_x X_i + \epsilon_i$
  
  * $\epsilon_i$ IID $N(0, 1/9)$
  * $\beta_0 = 0$
  * $\beta_x = 1$

- Generate $U_1, \ldots, U_{50}$, IID $N(0, 1)$

- Set $W_i = X_i + U_i$

- Regress $Y$ on $X$ and $Y$ on $W$ and compare
Effects of Measurement Error

True Data Without Measurement Error
Observed Data With Measurement Error
THEORY BEHIND THE PICTURES: THE Naive ANALYSIS

- Least Squares Estimate of Slope:

\[
\hat{\beta}_x = \frac{S_{y,w}}{S_w^2} = \frac{n^{-1} \sum (Y - \bar{Y})(W - \bar{W})}{n^{-1} \sum (W - \bar{W})^2}
\]

where

\[S_{y,w} \rightarrow \text{Cov}(Y, W) = \text{Cov}(Y, X + U) = \text{Cov}(Y, X) = \sigma_{y,x}\]

\[S_w^2 \rightarrow \text{Var}(W) = \text{Var}(X + U) = \sigma_x^2 + \sigma_u^2\]

So

\[
\hat{\beta}_x \rightarrow \frac{\sigma_{y,x}}{\sigma_x^2 + \sigma_u^2} = \left(\frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2}\right) \frac{\sigma_{y,x}}{\sigma_x^2} = \lambda \beta_x
\]

where

\[
\lambda = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2} = \frac{\sigma_x^2}{\sigma_u^2} = \text{attenuation factor} = \text{reliability ratio}
\]

* It is the relative size of the error that matters
THEORY BEHIND THE PICTURES: THE Naive ANALYSIS

- Least Squares Estimate of Intercept:

\[
\hat{\beta}_0 = \bar{Y} - \hat{\beta}_x \bar{W} \\
\rightarrow \mu_y - \lambda \beta_x \mu_x \\
= \beta_0 + (1 - \lambda) \beta_x \mu_x
\]

- Estimate of Residual Variance:

\[
\text{MSE} \rightarrow \sigma^2_\epsilon + (1 - \lambda) \beta_x^2 \sigma^2_x
\]
MORE THEORY: JOINT NORMALITY

- $Y, X, W$ jointly normal $\implies$
  
  $\ast \ Y \mid W \sim \text{Normal}$
  
  $\ast \ E(Y \mid W) = \beta_0 + (1 - \lambda)\beta_x \mu_x + \lambda \beta_x W$
  
  $\ast \ Var(Y \mid W) = \sigma^2_\epsilon + (1 - \lambda)\beta_x^2 \sigma^2_x$

- Intercept is shifted by $(1 - \lambda)\beta_x \mu_x$

- Slope is attenuated by the factor $\lambda$

- Residual variance is inflated by $(1 - \lambda)\beta_x^2 \sigma^2_x$
MORE THEORY: IMPLICATIONS FOR TESTING HYPOTHESES

- Because

\[ \beta_x = 0 \quad \text{iff} \quad \lambda \beta_x = 0 \]

it follows that

\[ [H_0 : \beta_x = 0] \equiv [H_0 : \lambda \beta_x = 0] \]

so the naive test of \( \beta_x = 0 \) is valid (correct Type I error rate).

- The naive test of \( H_0 : \beta_x = 0 \) is asymptotically efficient when \( E(X \mid W) \) is linear in \( W \).

- The discussion of naive tests when there are multiple predictors measured with error, or error-free predictors, is more complicated.
Sample Size for 80% Power. True slope $\beta_x = 0.75$. Variances $\sigma^2_x = \sigma^2_\epsilon = 1$. 
Multiple Linear Regression With Error

- Model

\[ Y = \beta_0 + \beta_z^t Z + \beta_x^t X + \epsilon \]
\[ W = X + U \]

- Regressing \( Y \) on \( Z \) and \( W \) estimates

\[
\begin{pmatrix}
\beta_{z*} \\
\beta_{x*}
\end{pmatrix} = \Lambda \begin{pmatrix}
\beta_z \\
\beta_x
\end{pmatrix} \neq \begin{pmatrix}
\beta_z \\
\beta_x
\end{pmatrix}
\]

- \( \Lambda \) is the attenuation matrix or reliability matrix

\[
\Lambda = \begin{pmatrix}
\sigma_{zz} & \sigma_{zx} \\
\sigma_{xz} & \sigma_{xx} + \sigma_{uu}
\end{pmatrix}^{-1} \begin{pmatrix}
\sigma_{zz} & \sigma_{zx} \\
\sigma_{xz} & \sigma_{xx}
\end{pmatrix}
\]

- Biases in components of \( \beta_{x*} \) and \( \beta_{z*} \) can be multiplicative or additive

  * Naive test of \( H_0 : \beta_x = 0, \beta_z = 0 \) is valid
From previous page:

\[
\begin{pmatrix}
\beta_{z*} \\
\beta_{x*}
\end{pmatrix} = \Lambda \begin{pmatrix}
\beta_z \\
\beta_x
\end{pmatrix}
\]

\[
\Lambda = \begin{pmatrix}
\sigma_{zz} & \sigma_{zx} \\
\sigma_{xz} & \sigma_{xx} + \sigma_{uu}
\end{pmatrix}^{-1} \begin{pmatrix}
\sigma_{zz} & \sigma_{zx} \\
\sigma_{xz} & \sigma_{xx}
\end{pmatrix}
\]

* Naive test of \( H_0 : \beta_x = 0 \) is valid

* Naive test of \( H_0 : \beta_{x,1} = 0 \) is typically not valid (\( \beta_{x,1} \) denotes a subvector of \( \beta_x \))

* Naive test of \( H_0 : \beta_z = 0 \) is typically not valid (same is true for subvectors)
**Multiple Linear Regression With Error**

- For $X$ scalar, attenuation factor in $\beta_{x*}$ is

$$\lambda_1 = \frac{\sigma^2_{x|z}}{\sigma^2_{x|z} + \sigma^2_u}$$

  * $\sigma^2_{x|z} = \text{residual variance in regression of } X \text{ on } Z$
  * $\sigma^2_{x|z} \leq \sigma^2_x \implies$

$$\lambda_1 = \frac{\sigma^2_{x|z}}{\sigma^2_{x|z} + \sigma^2_u} \leq \frac{\sigma^2_x}{\sigma^2_x + \sigma^2_u} = \lambda$$

  * $\implies\text{Collinearity accentuates attenuation}$

- Biased estimates of $\beta_z$:

$$\beta_{z*} = \beta_z + (1 - \lambda_1)\beta_x \Gamma_z,$$

  * $\Gamma_z$ is from $E(X \mid Z) = \Gamma_1 + \Gamma_z^t Z$
**Analysis Of Covariance**

- These results have implications for the two group ANCOVA.
  
  * $X$ = true covariate
  
  * $Z$ = dummy indicator of group 1, say

- We are interested in estimating $\beta_z$, the group effect. Biased estimates of $\beta_z$:

  $$\beta_{z*} = \beta_z + (1 - \lambda_1)\beta_x \Gamma_z,$$

  * $\Gamma_z$ is from $E(X \mid Z) = \Gamma_1 + \Gamma_z^t Z$

  * $\Gamma_z$ is the difference in the mean of $X$ among the two groups.

  * Thus, biased unless $X$ and $Z$ are unrelated.

  ➤ Use a randomized Study!!!
If we wish to predict $Y$ based on $W$, then the regression of $Y$ on $W$ is the correct model to use.

* Though this assumes that the distribution of $X$ and the measurement error distribution will be the same in the future as in the study.

since $\lambda = \frac{\sigma_x^2}{\sigma^2_x + \sigma_u^2}$

In public health, interventions will change the true $X$, not $W$, so we are interested in the regression of $Y$ on $X$. 
SUMMARY OF THE EFFECTS OF MEASUREMENT ERROR IN SIMPLE LINEAR REGRESSION

- **Regression Model**

\[ Y = \beta_0 + \beta_x X + \epsilon \]
\[ W = X + U \]

- **Attenuation Factor (Reliability Ratio)**

\[ \lambda = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2} \]

\[ 0 < \lambda \leq 1 \]

\[ \lambda = 1 \iff \sigma_u^2 = 0 \]

<table>
<thead>
<tr>
<th>Regression</th>
<th>Intercept</th>
<th>Slope</th>
<th>Residual Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Y ) on ( X )</td>
<td>( \beta_0 )</td>
<td>( \beta_x )</td>
<td>( \sigma_\epsilon^2 )</td>
</tr>
<tr>
<td>( Y ) on ( W )</td>
<td>( \beta_0 + (1 - \lambda)\beta_x \mu_x )</td>
<td>( \lambda \beta_x )</td>
<td>( \sigma_\epsilon^2 + (1 - \lambda)\beta_x^2 \sigma_x^2 )</td>
</tr>
</tbody>
</table>
END OF LECTURE 1
LECTURE 2: DATA TYPES, NONDIFFERENTIAL ERROR, ESTIMATING ATTENUATION, EXACT PREDICTORS, BERKSON MODEL

OUTLINE

- Nondifferential measurement error
- Estimating the attenuation
- Replication and validation data
- Internal and external subsets
- Transportability across data sets
- Is there an “exact” predictor?
- Berkson and classical measurement error
A response $Y$

Predictors $X$ measured with error (unobserved).

Predictors $Z$ measured without error.

A major proxy $W$ for $X$.

Sometime, a second proxy $T$ for $X$. 

---

**THE BASIC DATA**

---

- A response $Y$
- Predictors $X$ measured with error (unobserved).
- Predictors $Z$ measured without error.
- A major proxy $W$ for $X$.
- Sometime, a second proxy $T$ for $X$. 

---
Roughly speaking, the error is said to be *nondifferential* if $W$ and $T$ would not be measured if one could have measured $X$.

More formally, $(W, T)$ are *conditionally independent* of $Y$ given $(X, Z)$.

$$\Rightarrow (W, T) \text{ would provide no additional information about } Y \text{ if } X \text{ were observed}$$

This often makes sense, but it may be fairly subtle in each application.
Nondifferential Error

• Many crucial theoretical calculations revolve around nondifferential error.

• Consider simple linear regression: \( Y = \beta_0 + \beta_x X + \epsilon, \ \epsilon \) independent of \( X \).

\[
E(Y|W) = E \left[ \left\{ E(Y|X, W) \right\} | W \right] = E \left[ \left\{ E(Y|X) \right\} | W \right] = \beta_0 + \beta_x E(X|W).
\]

* This reduces the problem to estimating \( E(X|W) \). For example,

\[
E(X|W) = \lambda W + (1 - \lambda) \mu_x \quad \text{(under joint normality and classical error)}
\]

where \( \lambda = \frac{\sigma_x^2}{\sigma_w^2} \).

• If the error is differential, then (1) fails, and no simplification is possible.
HEart Disease Vs Systolic Blood Pressure

- $Y =$ indicator of Coronary heart Disease (CHD)
- $X =$ true long-term average systolic blood pressure (SBP) (maybe transformed)
- Assume $P(Y = 1) = H(\beta_0 + \beta_xX)$
- Data are CHD indicators and determinations of systolic blood pressure for $n = 1600$ in Framingham Heart Study
- $X$ measured with error:
  - SBP measured at two exams (and averaged) $\Rightarrow$ sampling error
  - The determination of SBP is subject to machine and reader variability
- It is hard to believe that the short term average of two days carries any additional information about the subject’s chance of CHD over and above true SBP.
- Hence, nondifferential
**Is This Nondifferential?**

- From Tosteson et al. (1989).

- \( Y = I\{\text{wheeze}\} \).

- \( X \) is personal exposure to \( \text{NO}_2 \).

- \( W = (\text{NO}_2 \text{ in kitchen, } \text{NO}_2 \text{ in bedroom}) \) is observed in the primary study.
Is This Nondifferential?

- From Küchenhoff & Carroll

  \[ Y = I\{\text{lung irritation}\}. \]

- \( X \) is actual personal long–term dust exposure

- \( W = \) is dust exposure as measured by occupational epidemiology techniques.

  * The sampled the plant for dust.

  * Then they tried to match the person to where he/she worked.
What Is Necessary To Do An Analysis?

- In linear regression with classical additive error $W = X + U$, one needs:
  - Nondifferential error
  - An estimate of the error variance $\text{var}(U)$ — since $\lambda = \frac{\sigma_w^2 - \sigma_u^2}{\sigma_w^2}$.

- How do we get the latter information?

- The best way is to get a subsample of the study in which $X$ is observed. This is called \textit{validation}.
  - In many applications not possible.

- Another method is to do replications of the process, often called \textit{calibration}.

- A third way is to get the value from another similar study.
In a replication study, for some of the study participants you measure more than one $W$.

The standard model is

$$W_{ij} = X_i + U_{ij}, \quad j = 1, \ldots, m_i.$$ 

This is a one-factor ANOVA with mean squared error $\sigma_u^2$ estimated by

$$\hat{\sigma}_u^2 = \frac{\sum_{i=1}^n \sum_{j=1}^{m_i} (W_{ij} - \bar{W}_i.)^2}{\sum_{i=1}^n (m_i - 1)}.$$ 

As the proxy for $X_i$ one would use $\bar{W}_{i.}$ where

$$\bar{W}_{i.} = X_i + \bar{U}_{i.}$$

$$\text{var}(\bar{U}_{i.}) = \frac{\sigma_u^2}{m_i}.$$
Replication

- Replication allows you to test whether your model is additive with constant error variance.
- If $W_{ij} = X_i + U_{ij}$ with $U_{ij}$ symmetrically distributed about zero and independent of $X_i$, we have a major fact:
  - The sample mean and sample standard deviation are uncorrelated.
    - Eckert, Carroll, and Wang (1997) use this fact to find a transformation to additivity
- Also, if $U_{ij}$ are normally distributed, then so too are differences $W_{i1} - W_{i2} = U_{i1} - U_{i2}$.
- Graphical diagnostics can be implemented easily in any package.
REPLICATION: WISH

- The WISH study measured caloric intake using a 24–hour recall.
  - There were 6 replicates per woman in the study.

- A plot of the caloric intake data showed that $W$ was nowhere close to being normally distributed in the population.
  - If additive, then either $X$ or $U$ is not normal.

- When plotting standard deviation versus the mean, one often uses the rule that the method “passes” the test if the max-to-min is less than 2.0.
  - A little bit of non–constant variance does not seem to hurt. See Carroll & Ruppert (1988)
  - **Caveat:** Carroll & Ruppert studied effects of heteroscedasticity on efficiency, not on bias due to measurement error.
WISH, Caloric Intake, Q–Q plot of \( W \).
WISH, Caloric Intake, Q–Q plot of Differenced $W$s.
WISH, Caloric Intake, plot for additivity, loess and OLS.
**REPLICATION: WISH**

- Taking logarithm of $W$ improves all the plots.
WISH, Log Caloric Intake, Q–Q Plot of Observed $W$. 

WISH Log-Calories, FFQ
Normal QQ--Plot
WISH, Log Caloric Intake, Q–Q Plot of Differenced $W$s.
WISH, Log Caloric Intake, Plot For Additivity, Loess and OLS.

- Perhaps a bit over-transformed — square-root might be better.
TRANSPORTABILITY OF ESTIMATES

- In linear regression, we only require need the measurement error variance (after checking for semi-constant variance, additivity, normality).

- In general though, more is needed. Let’s remember that if we observe $W$ instead of $X$, then the observed data have a regression of $Y$ on $W$ that effectively acts as if

$$E(Y|W) = \beta_0 + \beta_x E(X|W) \approx \beta_0 + \beta_x \{\lambda W + (1 - \lambda)\mu_x\}.$$ 

- As we will see later, in general problems we can do a likelihood analysis if we know the distribution of $X$ given $W$. 
TRANSPORTABILITY OF ESTIMATES

- It is tempting to try to use outside data and transport an estimate of $\lambda$ to your problem.

  * **Bad idea!!!**

    $\lambda = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2}$

  * Note how this depends on the distribution of $X$.

  * It is rarely the case that two populations have the same $X$ distribution, even when the same instrument is used.

    - the instrument effects only $\sigma_u^2$, not $\sigma_x^2$

- Maybe one should transport the estimate of $\sigma_u^2$ and use

  $\hat{\lambda} = \frac{\hat{\sigma}_w^2 - \hat{\sigma}_{u,\text{transported}}^2}{\hat{\sigma}_w^2}$
EXTERNAL DATA AND TRANSPORTABILITY

- We say that a model is transportable across studies if the model holds with the same parameters in the two studies.

  ∗ Internal data is ideal since there is no question about transportability.

- With external data, transportability back to the primary study cannot be taken for granted.

  ∗ Sometimes transportability clearly will not hold. Then the value of the external data is, at best, questionable.

  ∗ Even is transportability seems to be a reasonable assumption, it is still just that, an assumption.
EXTERNAL DATA AND TRANSPORTABILITY

- As an illustration, consider two nutrition data sets which use exactly the same FFQ
- Nurses Health Study
  - Nurses in the Boston Area
- American Cancer Society
  - National sample
- Since the same instrument is used, error properties should be about the same.
  - But maybe note the entire distribution!!
  - var(differences, NHS = 47)
  - var(differences, ACS = 45)
  - var(sum, NHS = 152)
  - var(sum, ACS = 296)
FFQ Histograms in NHS and ACS
IS THERE AN “EXACT” PREDICTOR?

• One can distinguish between two concepts of $X$:
  
  * the actual exact predictor, which may never be observable under any circumstances.
  
  * the operationally-defined exact predictor which could be observed, albeit at great cost and effort.

• Example: $X$ is long-term caloric intake.
  
  * actual $X$ is long-term intake over lifetime.
  
  * operationally-defined $X$ is long-term intake since inception of the study, as measured by the best available instrument.
One could probably construct hypothetical examples where the effects of actual $X$ and of operationally-defined $X$ are quite different, even opposite in sign.

Obviously, a policy to change $X$, say to reduce intake of saturated fat or of calories, affects the actual $X$, not the operationally defined $X$.

Nonetheless, the operationally-defined $X$ is all we can work with.

“Gold standard” generally means the operationally-defined $X$.

* We will take “exact predictor” and “gold standard” to both refer to operationally-defined $X$.

* However, for some authors, “gold standard” mean the actual exact $X$ even if not operationally-defined.
The Berkson Model

- The classical Berkson model says that
  \[ \text{True Exposure} = \text{Observed Exposure} + \text{Mean Zero Error} \]
  \[ X = W + U_b \] (or \[ X = W \times U_b \]),

- It is assumed that \( W \) and \( U_b \) are indep. and \( E(U) = 0 \) (additive error) or \( E(U) = 1 \) (multiplicative error) so that
  \[ E(X|W) = W \]

- Compare with classical measurement error model where
  \[ W = X + U \]
  and
  \[ E(X|W) = \lambda W + (1 - \lambda)\mu_x. \]
THE BERKSON MODEL — CONT.

- From previous page

\[ E(X|W) = W \]

- In the linear regression model,
  
  * Ignoring error still leads to unbiased intercept and slope estimates,
  * but the error about the line is increased.

  ➤ To see this, note that

  \[ Y = \beta_0 + \beta_1(W + U_b) + \epsilon = \beta_0 + \beta_1 W + (\beta_1 U_b + \epsilon) \]

- In the logistic model with normally distributed measurement error, ignoring error

  * Leads to a bias in slope (and intercept):

  \[
  \text{Est. Slope} \approx \frac{\text{True Slope}}{\sqrt{1 + \beta_1^2 \sigma_{u,b}^2 / 2.9}}.
  \]

  * For many problems, this attenuation is only minor.
LECTURE 3: BERKSON/CLASSICAL, STRUCTURAL/FUNCTIONAL, REGRESSION CALIBRATION

OUTLINE

• What is Berkson? What is classical?

• Functional versus structural modeling
  • Classical and flexible structural modeling

• Regression calibration

• Multiplicative error
WHAT’ S BERKSON? WHAT’ S CLASSICAL?

• **Berkson**: \( X = W + U_b \) and \( W \) and \( U_b \) are independent

• **Classical**: \( W = X + U_c \) and \( X \) and \( U_c \) are independent

• In practice, it may be hard to distinguish between the classical and the Berkson error models.

  * In some instances, neither holds exactly.

  * In some complex situations, errors may have both Berkson and classical components, e.g., when the observed predictor is a combination of 2 or more error–prone predictors.

• **Berkson model**: a nominal value is assigned.

  * Direct measures cannot be taken, nor can replicates.

• **Classical error structure**: direct *individual* measurements are taken, and can be replicated but with variability.
The reason why people have trouble distinguishing the two is that in real life, it takes hard thought to do so.
LET’S PLAY STUMP THE EXPERTS!

- Framingham Heart Study

  * Predictor is systolic blood pressure

- All workers with the same job classification and age are assigned the same exposure based on job exposure studies.

- Using a phantom, all persons of a given height and weight with a given recorded dose are assigned the same radiation exposure.

- The radon gas concentration in houses that have been torn down is estimated by the average concentration in a sample of nearby houses
Berkson Error In Nonlinear Models

- Suppose that

\[ Y = f(X, \beta) + \epsilon \]

- Then

\[ Y = f(W, \beta) + \frac{1}{2} f''(W, \beta) \sigma_b^2 + \left\{ f'(W, \beta) U_b + \frac{1}{2} f''(W, \beta) (U_b^2 - \sigma_b^2) + \epsilon \right\} \]

\* Here

\[ f'(W, \beta) = \frac{\partial}{\partial W} f(W, \beta) \]

and so forth

\* The nature of the bias (in red) depends on the sign of \( f''(W, \beta) \)

\* In the linear case \( f''(W, \beta) = 0 \) and we get the previous result: no bias
**Mixtures Of Berkson And Classical Error**

- Let $L$ be long-term average radon concentration for all houses in a neighborhood
- $X$ is long-term average radon concentration in a specific house in this neighborhood
- $W$ is average measured radon concentration for a sample of houses in this neighborhood
- Model of Tosteson and Tsiatis (1988)

\[
X = L + U_b
\]
\[
W = L + U_c
\]

$L, U_b, U_c$ are independent

*Used by Reeves, Cox, Darby, and Whitley (1998) and Mallick, Hoffman, and Carroll (2002)*
MIXTURES OF BERKSON AND CLASSICAL ERROR

• From previous page: Model of Tosteson and Tsiatis (1988)

\[ X = L + U_b \]

\[ W = L + U_c \]

• \( U_c = 0 \Rightarrow X = W + U_b \Rightarrow \text{Berkson} \)

• \( U_b = 0 \Rightarrow W = X + U_c \Rightarrow \text{Classical} \)

• More generally,

\[ W = X + U_c - U_b \]

and \( U_c \) is independent of \( X \) while \( U_b \) is independent of \( W \)
**Functional and Structural Modeling Classical Error Models**

- The common linear regression texts make distinction:
  - **Functional:** $X$’s are **fixed** constants
  - **Structural:** $X$’s are **random** variables

- If you pretend that the $X$’s are fixed constants, it seems plausible to try to estimate them as well as all the other model parameters.

- This is the functional maximum likelihood estimator.
  - Every textbook has the linear regression functional maximum likelihood estimator.

- Unfortunately, the functional MLE in nonlinear problems has two defects.
  - It’s really nasty to compute.
  - It’s a lousy estimator (badly inconsistent).

- Structural $\Rightarrow$ empirical Bayes type analysis
More useful distinction:

- **Functional modeling:** No assumptions made about the $X$’s (could be random or fixed)
- **Classical structural modeling:** Strong parametric assumptions made about the distribution of $X$. Generally normal, lognormal or gamma.
- **Flexible structural modeling:** Structural, but flexible parametric family. Tries to get the best of both worlds.

Flexible is similar in spirit to **Quasi-structural modeling** (Pierce et al., 1992)

- Assume the that distribution of $X$ is the (unknown) empirical distribution of the $X$ values in the sample
Choosing Between Functional or Structural Models

- **Key questions:**
  - How sensitive are inferences to an assumed **STRUCTURAL** model?
  - How much does it “cost” to be functional?

- **FUNCTIONAL**: No need to perform extensive sensitivity analyses.

- Many functional methods are simple to implement (and some are computed using little more than standard software).

- Functionality focuses emphasis on the error model.

- Because of “latent-model” robustness, a functional analysis serves as a useful check on a parametric structural model.
Choosing Between Functional or Structural Models

- **Structural** models can be viewed as an *empirical Bayes* method of dealing with a large number of nuisance parameters (the true covariate values) (Whittemore, 1989)

- Best known *functional* methods can be *inefficient* (missing data, thresholds, non-parametric regression).
Choosing Between Functional Or Structural Models

• One should consider the distribution of the true X’s

  ✪ Example: A-bomb survivor study (Pierce et al., 1992):

    ▶ All subjects survived the acute effects
    ▶ This provides some information
    ▶ A high radiation measurement is likely to be due to a high measurement error

  ✪ This information can, perhaps, be captured by during structural modeling

    ▶ The distribution of true doses is different among survivors than among all exposed to the attacks
Choosing Between Functional Or Structural Models

- My opinion is that one should use structural models
  - But care is needed when modeling the distribution of $X$
**Some Functional Methods**

- Regression Calibration/Substitution
  
  * Replaces true exposure $X$ by an estimate of it *based only on covariates* but not on the response.
  
  * In linear model with additive errors, this is the classical *correction for attenuation*.
  
  * In Berkson model, one simply ignores the measurement error.

- SIMEX is a fairly general functional method.
  
  * It assumes only that you have an error model and that you can “add on” measurement error to make the problem worse.
**Regression Calibration—Basic Ideas**

- **Key idea:** replace the unknown $X$ by $E(X|Z,W)$ which depends only on the known $(Z,W)$.
  
  * This provides an approximate model for $Y$ in terms of $(Z,W)$.
  * Called the “conditional expectation approach” by Lyles and Kupper (1997)

- Generally applicable.
  
  * Depends on the measurement error being “not too large” in order for the approximation to be sufficiently accurate.
  * For some models, if $\text{var}(X|Z,W)$ is constant then the only bias due to the regression calibration approximation is in the intercept parameter.
Why does regression calibration work?

- \( X = E(X|Z, W) + U \) where \( U \) is uncorrelated with any function of \((Z, W)\)

- If

\[
Y = X^T \beta_x + Z^T \beta_z
\]

then

\[
Y = E(X|Z, W)^T \beta_x + Z^T \beta_z + (U^T \beta_x + \epsilon)
\]

where \((U^T \beta_x + \epsilon)\) is uncorrelated with the regressors \(Z\) and \(E(X|Z, W)\)

- Therefore, regression of \(Y\) on \(E(X|Z, W)\) and \(Z\) gives unbiased estimates

  - \(E(X|Z, W)\) could be replaced by the linear regression (best linear predictor) of \(X\) using \((Z, W)\)
The general algorithm is:

* Using replication, validation, or instrumental data, develop a model $E(X | Z, W) = m(Z, W, \gamma_{cm})$ and estimate $\gamma_{cm}$.

* Replace $X$ by $m(Z, W, \hat{\gamma}_{cm})$ and run your favorite analysis.

* Obtain standard errors by the bootstrap or the “sandwich method.”
  
  ▶ Can automatically correct for uncertainty about $\gamma_{cm}$

In linear regression, regression calibration is equivalent to the “correction for attenuation.”
The Regression Calibration Algorithm

- Easily adjusts for different amount of replication
- If $\overline{W}_i$ is the average of $m_i$ replicates then

$$\lambda_i = \frac{\sigma^2_x}{\sigma^2_x + \sigma^2_u/m_i}$$

$$E(X_i|\overline{W}_i) = \lambda_i \overline{W}_i + (1 - \lambda_i) \mu_x$$
**Logistic Regression, Normal X**

- Consider the logistic regression model
  \[ \Pr(Y = 1|X) = \{1 + \exp(-\beta_0 - \beta_x X)\}^{-1} = H(\beta_0 + \beta_x X). \]

- Suppose that
  - \( X \) and \( U \) are normally distributed.
  - Then \( X \) given \( W \) is normal with mean \( E(X|W) = \mu_x(1-\lambda)+\lambda W \) and variance \( \lambda \sigma_u^2 \).

- If the event is “rare”, then \( H(\beta_0 + \beta_x X) \approx \exp(\beta_0 + \beta_x X) \).

- Then, by using moment generating functions, the observed data follow
  \[
  \Pr(Y = 1|W) = E \left\{ \Pr(Y = 1|X, W) \right\} |W = \begin{cases} 
  E \left\{ \Pr(Y = 1|X) \right\} |W 
  \approx E \{ \exp(\beta_0 + \beta_x X) |W \} = \exp \{ \beta_0 + (1/2)\beta_x^2 \lambda \sigma_u^2 + \beta_x E(X|W) \} 
  \approx H \{ \beta_0 + (1/2)\beta_x^2 \lambda \sigma_u^2 + \beta_x E(X|W) \}. 
  \end{cases}
  \]

- Typically, the regression calibration approximation works fine in this case.
A different approximation uses the fact that $H(x) \approx \Phi(kx)$ with $k = 0.588 \approx 1/1.70$

\[ \text{error} = \{ \Phi(kx) - H(x) \} \quad \text{and relative error} = \frac{\text{Error}}{H(x)}. \]
**Logistic Regression, Normal X**

\[
Pr(Y = 1|X) = \Phi\{k(\beta_0 + \beta_x X)\} = Pr\{U \leq k(\beta_0 + \beta_x X)\}
\]

Therefore

\[
Pr(Y = 1|W) = E[Pr\{U \leq k(\beta_0 + \beta_x X)\}|W]
\]

\[
= Pr[U - \beta_x kU_b \leq k\{\beta_0 + \beta_x E(X|W)\}]
\]

\[
= \Phi \left[ \frac{k\{\beta_0 + \beta_x E(X|W)\}}{(1 + \beta_x^2 k^2 \sigma^2_{X|W})^{1/2}} \right] \approx H \left\{ \frac{\beta_0 + \beta_x E(X|W)}{(1 + \beta_x^2 k^2 \sigma^2_{X|W})^{1/2}} \right\}
\]

Here

\[
U_b \sim N(0, \sigma^2_{X|W}) = N(0, \lambda\sigma^2_u)
\]
A third approximation is

\[
\Pr(Y = 1|W) = H \left[ \beta_0 + \beta_x \{ E(X|W) + U_b \} \right]|W
\]

\[
\approx H \{ \beta_0 + \beta_x E(X|W) \} + \frac{1}{2} \beta_x^2 H'' \{ \beta_0 + \beta_x E(X|W) \} \sigma^2_{X|W}
\]

Note that

\[
H''(x) > 0 \text{ if } x < 0
\]

\[
H''(x) < 0 \text{ if } x > 0
\]

- This approximation is applicable to most models, not just logistic regression
Estimating The Calibration Function

- Need to estimate $E(X|Z, W)$.
  - How this is done depends, of course, on the type of auxiliary data available.
- Easy case: validation data
  - Suppose one has internal, validation data.
  - Then one can simply regress $X$ on $(Z, W)$ and transports the model to the non-validation data.
  - Of course, for the validation data one regresses $Y$ on $(Z, X)$, and this estimate must be combined with the one from the non-validation data.
- Same approach can be used for external validation data, but with the usual concern for non-transportability.
Estimating the Calibration Function: Instrumental Data: Rosner’s Method

- Internal unbiased instrumental data:
  * Suppose \( E(T|Z,X) = E(T|Z,X,W) = X \) so that \( T \) is an unbiased instrument.
  * If \( T \) is expensive to measure, then \( T \) might be available for only a subset of the study. \( W \) will generally be available for all subjects.
  * Then
    \[
    \]
- Thus, \( T \) regressed on \((Z,W)\) follows the same model as \( X \) regressed on \((Z,W)\), although with greater variance.
  * So one regresses \( T \) on \((Z,W)\) to estimate the parameters in the regression of \( X \) on \((Z,W)\).
More formally, suppose we have the model $W_i = \gamma_{0,\text{em}} + \gamma_{x,\text{em}}^t X_i + \gamma_{z,\text{em}}^t Z_i + U_i$

(error calibration model)

Under joint normality, this implies a calibration model

$$X_i = \gamma_{0,\text{cm}} + \gamma_{w,\text{cm}}^t W_i + \gamma_{z,\text{cm}}^t Z_i + U_{\text{cm},i}$$

* Sometimes one can use validation data to regress $X$ on $Z$ and $W$

* Suppose we only have an “alloyed gold standard” $T_i = X_i + U_{T,i}$ on a “validation sample”

  ▶ the regress $T$ on $Z$ and $W$

  ▶ This works! — think of $T_i = X_i + U_i$ as outcome measurement error
ESTIMATING THE CALIBRATION FUNCTION: REPLICATION DATA

• Suppose that one has unbiased internal replicate data:
  ∗ $W_{ij} = X_i + U_{ij}$, $i = 1, \ldots, n$ and $j = 1, \ldots, k_i$, where $E(U_{ij} | Z_i, X_i) = 0$.
  ∗ $\bar{W}_i := \frac{1}{k_i} \sum_j W_{ij}$.
  ∗ Notation: $\mu_z$ is $E(Z)$, $\Sigma_{xz}$ is the covariance (matrix) between $X$ and $Z$, etc.

• Use standard least squares theory to get the best linear unbiased predictor of $X$ from $(W, Z)$:

$$E(X | Z, \bar{W}) \approx \mu_x + (\Sigma_{xx} \Sigma_{xz}^t) \left\{ \Sigma_{xx} + \frac{\Sigma_{uu}}{k} \Sigma_{xz} \right\}^{-1} \left( \bar{W} - \mu_w \right)$$

(best linear approximation = exact conditional expectation under joint normality).

• Need to estimate the unknown $\mu$’s and $\Sigma$’s.

• Essentially a flexible structure method since existence of $\mu_x$ and $\Sigma_{xx}$ is assumed.
Estimating The Calibration Function: Replication Data, Continued

- $\hat{\mu}_z$ and $\hat{\Sigma}_{zz}$ are the “usual” estimates since the $Z$’s are observed.

- $\hat{\mu}_x = \hat{\mu}_w = \sum_{i=1}^{n} k_i \overline{W}_i / \sum_{i=1}^{n} k_i$.

- $\hat{\Sigma}_{xz} = \sum_{i=1}^{n} k_i (\overline{W}_i - \hat{\mu}_w)(Z_i - \hat{\mu}_z)^t / \nu$
  where $\nu = \sum k_i - \sum k_i^2 / \sum k_i$.

- $\hat{\Sigma}_{uu} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{k_i} (W_{ij} - \overline{W}_i)(W_{ij} - \overline{W}_i)^t}{\sum_{i=1}^{n} (k_i - 1)}$.

- $\hat{\Sigma}_{xx} = \left[ \left\{ \sum_{i=1}^{n} k_i (\overline{W}_i - \hat{\mu}_w)(\overline{W}_i - \hat{\mu}_w)^t \right\} - (n - 1)\hat{\Sigma}_{uu} \right] / \nu$. 
Regression Calibration Versus Correcting The Naive Estimator For Attenuation

• If

\[ g\{E(Y|X)\} = \beta_0 + \beta_1 X \quad \text{(GLM)} \]

and

\[ E(X|W) = \alpha_0 + \lambda W \]

then these two slope estimators are equivalent:

* Regress \( Y \) on \( E(X|W) \) (e.g., Rosner, Spiegelman, Willett, 1989)
* Regress \( Y \) on \( W \) and divide the slope by \( \lambda \) (e.g., Carroll, Ruppert, Stefanski, 1995)

• Equivalence does not hold in general, e.g., for multiplicative error, when \( E(X|W) \) is nonlinear.
The multiplicative lognormal error model is

\[ W = X U, \quad \log(U) \sim N(\mu_u, \sigma_u^2) \]  

(1)

- If \( \log(X) \) is normal then we can convert (1) to a model for \( X \) given \( W \)

\[ \log(X) = \alpha + \lambda \log(W) + U_b \]

where \( U_b \) is \( N(0, \sigma_b^2) \) and independent of \( W \)

- If \( Y \) were linear in \( \log(X) \) then we would work with \( \log(X) \) and have an additive error model

  \* But typically it is assumed that \( Y \) is linear in \( X \), not \( \log(X) \)
The multiplicative error model with outcome linear in exposure is well-supported by empirical work:

- Lyles and Kupper (1997, Biometrics):
  - “there is much evidence for this model”
  - “the more biologically relevant predictor is true mean exposure on the original scale”
MULTIPLICATIVE ERROR

* Pierce, Stram, Vaeth, Schafer (1992, JASA)
  - data from Radiation Effects Research Foundation (RERF) in Hiroshima
  - “On both radiobiological and empirical grounds, focus is on models where the expected response is linear or quadratic in dose rather than linear on logistic and logarithmic scales.”
  - “It is accepted that radiation dose-estimation errors are more homogeneous on a multiplicative than on an additive scale”
  - “The distribution of true doses is extremely nonnormal”
  - However, it seemed unlikely that a lognormal model for the true doses would fit as well as the Weibull model that they used
MULTIPLICATIVE ERROR

* Seychelles study (preliminary analysis of Thurston)
  - validation data (brain versus maternal hair MeHg): error appears to be multiplicative and lognormal
  - researchers use MeHg concentration, not log concentration, as the exposure
MULTIPLICATIVE ERROR

• from previous page:

\[
\log(X) = \alpha + \lambda \log(W) + U_b
\]  \hspace{1cm} (2)

• From (2)

\[
X = W^\lambda \exp(U_b)
\]

and

\[
E(X|W) = W^\lambda \exp \left( \alpha + \sigma^2_b / 2 \right)
\]

Not linear in \(W\)
MULTIPLICATIVE ERROR

- Therefore if $E(Y|X) = \beta_0 + \beta_1 X$ then
  
  $$E(Y|W) = \beta_0 + \{\beta_1 \exp (\alpha + \sigma_B^2/2)\} W^\lambda$$

- Therefore, regress $Y$ on $W^\lambda$ and divide the slope estimate by $\exp (\alpha + \sigma_B^2/2)$
  
  * This is regression calibration again

- Note that the regression of $Y$ on $W$ is not linear even though the regression of $Y$ on $X$ is linear
  
  * This is because the regression of $X$ on $W$ is not linear
**Multiplicative Error**

Simulation: \( W = XU, \log(X) = N(0, 1/4), \log(U) = N(-1/8, 1/4) \)

\[ Y = X/2 + N(0, 0.04) \]
**MULTIPLICATIVE ERROR**

\[ W = XU, \log(X) = N(0,1/4), \log(U) = N(-1/8,1/4) \]

\[ Y = X/2 + N(0,0.04) \]
\[ W = XU, \log(X) = N(0, 1/4), \log(U) = N(-1/8, 1/4) \]
\[ Y = X/2 + N(0, 0.04) \quad \textbf{Note: } \lambda = 1/2 \]
MULTIPLICATIVE ERROR

Other work

• Hwang (1986) regresses $Y$ on $W$ (not $W^\lambda$) and then corrects

  * This is the “correction method” that is equivalent to regression calibration only when $E(X|W)$ is linear in $W$ – which is not the case here

  * Consistent but badly biased in simulations of Lyle and Kupper

  * Bias is still noticeable for $n = 10,000$

    * Previous plots suggest why

  * These results are consistent with those of Iturria, Carroll, and Firth (1999)
MULTIPLICATIVE ERROR

Other work

• Lyles and Kupper (1999) propose a method based on quasi-likelihood

  * somewhat superior to regression calibration in a simulation study

    ▶ especially when measurement error is large relative to equation error (plus measurement error in $Y$)

  * it is a weighted version of regression calibration

    ▶ more efficient than regression calibration because it weights according to inverse conditional variances

  * a special case of expanded regression calibration (Carroll, Ruppert, Stefanski, 1995)
**MULTIPLICATIVE ERROR**

**Other work**

- Iturria, Carroll, and Firth (1999) study polynomial regression with multiplicative error
  
  * One of their general methods is a special case of Hwang’s
  
  * Their “partial regression” estimator assumes lognormality and generalizes the estimator discussed earlier
  
  * In a simulation with lognormal $X$ and $U$, the partial regression estimator is superior to the ones that make less assumptions
    
    ▶ this is similar to the results of Lyles and Kupper
**Transformations To Additive Error**

- Model is

  \[ h(W) = h(X) + U \]

  * Includes additive and multiplicative error as a special case

- Eckert, Carroll, and Wang (1997) consider parametric (power) and nonparametric (monotonic cubic spline) models for \( h \)

  * can transform to a specific error distribution
  * could, instead, transform to constant error variance
  * if model holds, then both methods have the same target transformation
  * it could be more efficient to use both sources of information (Ruppert and Aldershof, 1989)

- Nusser et al. (1996) use transformations to estimate the distribution of \( X \) where \( X \) is the daily intake of a dietary component
END OF LECTURE 3
LECTURE 4: SIMEX AND INSTRUMENTAL VARIABLES

OUTLINE

- The Key Idea Behind SIMEX (Simulation/Extrapolation)
- An Empirical Version of SIMEX
- Details of Simulation Extrapolation Algorithm
- Example: Measurement Error in SBP in Framingham Study
- IV (Instrumental Variables): Rationale
- IV via prediction and the IV algorithm
- Example: Calibrating a FFQ
- Example: CHD and Cholesterol
ABOUT SIMULATION EXTRAPOLATION

• A functional method
  ∗ no assumptions about the true $X$ values

• For bias reduction and variance estimation
  ∗ like bootstrap and jackknife

• Not model dependent
  ∗ like bootstrap and jackknife

• Handles complicated problems

• Computer intensive

• Approximate, less efficient for certain problems
The effects of measurement error on an estimator can be studied with a simulation experiment in which additional measurement error is added to the measured data and the estimator recalculated.

- “Response variable” is the estimator under study
- “Independent factor” is the measurement error variance
  - “Factor levels” are the variances of the added measurement errors
- Objective is to study how the estimator depends on the variance of the measurement error
OUTLINE OF THE ALGORITHM

• Add measurement error to variable measured with error
  * \( \Lambda \) controls amount of added measurement error
  * \( \sigma_u^2 \) increased to \( (1 + \Lambda)\sigma_u^2 \)
  * Average over many simulations to remove Monte Carlo variation

• Recalculate estimates — called pseudo-estimates

• Plot Monte Carlo average pseudo-estimates versus \( \Lambda \)

• Extrapolate to \( \Lambda = -1 \)
  * \( \Lambda = -1 \) corresponds to case of no measurement error
Your estimate when you ignore measurement error.
What happens to your estimate when you have more error, which you add on by simulation, but you still ignore the error.
What statistician can resist fitting a curve?.
Now extrapolate to the case of no measurement error.
AN EMPIRICAL VERSION OF SIMEX: FRAMINGHAM DATA EXAMPLE

• Data

  * $Y$ = indicator of CHD
  * $W_k = \text{SBP at Exam } k, \ k = 1, 2$
  * $X = \text{“true” SBP}$

  Data:

  $$(Y_j, W_{1,j}, W_{2,j}), \ j = 1, \ldots, 1660$$

• Model Assumptions

  * $W_1, W_2 \mid X \text{ iid } N(X, \sigma^2_u)$
  * $\Pr(Y = 1 \mid X) = H(\alpha + \beta X), \ H \text{ logistic}$
Framingham Data Example: Three Naive Analyses:

- Regress $Y$ on $\overline{W} \mapsto \hat{\beta}_A$
- Regress $Y$ on $W_1 \mapsto \hat{\beta}_1$
- Regress $Y$ on $W_2 \mapsto \hat{\beta}_2$

<table>
<thead>
<tr>
<th>$\Lambda$</th>
<th>Measurement Error Variance $= (1 + \Lambda)\sigma_u^2/2$</th>
<th>Slope Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-1$</td>
<td>0</td>
<td>?</td>
</tr>
<tr>
<td>$0$</td>
<td>$\sigma_u^2/2$</td>
<td>$\hat{\beta}_A$</td>
</tr>
<tr>
<td>$1$</td>
<td>$\sigma_u^2$</td>
<td>$\hat{\beta}_1, \hat{\beta}_2$</td>
</tr>
</tbody>
</table>
Logistic regression fits in Framingham using first replicate, second replicate and average of both
A SIMEX–type plot for the Framingham data, where the errors are not computer–generated.
A SIMEX–type extrapolation for the Framingham data, where the errors are not computer–generated.
Simulation And Extrapolation Steps: Adding Measurement Error

- Framingham Example:
  \[
  \begin{align*}
  &\bar{W}_1 \leftarrow W_1 = \bar{W}_1 + (W_1 - W_2)/2 \\
  &\bar{W}_2 \leftarrow W_2 = \bar{W}_1 + (W_2 - W_1)/2 
  \end{align*}
  \]

- In General:
  \[
  \begin{align*}
  &\text{Best Estimator(data)} \leftarrow \text{Best Estimator(data} + \sqrt{\Lambda} \{\text{Independent } N(0, \sigma_u^2) \text{ Error} \} \\
  &\Lambda \text{ controls amount (variance) of added measurement error}
  \end{align*}
  \]
SIMULATION AND EXTRAPOLATION ALGORITHM: ADDING MEASUREMENT ERROR (DETAILS)

- For $\Lambda \in \{\Lambda_1, \ldots, \Lambda_M\}$

- For $b = 1 \ (1) \ B$, compute:

  * $b$th pseudo data set

    $$W_{b,i}(\Lambda) = W_i + \sqrt{\Lambda} \ \text{Normal} \ (0, \ \sigma^2_u)_{b,i}$$

  * $b$th pseudo estimate

    $$\hat{\theta}_b(\Lambda) = \hat{\theta} \left( \{Y_i, W_{b,i}(\Lambda)\}_1^n \right)$$

  * the average of the pseudo estimates

    $$\hat{\theta}(\Lambda) = B^{-1} \sum_{b=1}^{B} \hat{\theta}_b(\Lambda) \approx E \left( \hat{\theta}_b(\Lambda) \mid \{Y_j, X_j\}_1^n \right)$$
SIMULATION AND EXTRAPOLATION STEPS: EXTRAPOLATION

- Framingham Example: (two points $\Lambda = 0, 1$)
  - Linear Extrapolation — $a + b\Lambda$

- In General: (multiple $\Lambda$ points)
  - Linear — $a + b\Lambda$
  - Quadratic — $a + b\Lambda + c\Lambda^2$
  - Rational Linear — $(a + b\Lambda)/(c + \Lambda)$ (exact for linear regression)
SIMULATION AND EXTRAPOLATION ALGORITHM: EXTRAPOLATION (DETAILS)

- Plot $\hat{\theta}(\Lambda)$ vs $\Lambda$ ($\Lambda > 0$)
- Extrapolate to $\Lambda = -1$ to get $\hat{\theta}(-1) = \hat{\theta}_{\text{SIMEX}}$
- For certain models and assuming the extrapolant function is chosen correctly
- $E \left\{ \hat{\theta}(-1) \mid \text{True Data} \right\} = \hat{\theta}_{\text{TRUE}}$
**Example: Measurement Error in Systolic Blood Pressure**

- Framingham Data:

  \[
  \left( Y_j, \text{Age}_j, \text{Smoke}_j, \text{Chol}_j, W_{A,j} \right), \quad j = 1, \ldots, 1615
  \]

- *Y* = indicator of CHD
- *Age* (at Exam 2)
- *Smoking Status* (at Exam 1)
- *Serum Cholesterol* (at Exam 3)
- *Transformed SBP*

  \[
  W_A = (W_1 + W_2) / 2,
  \]

  \[
  W_k = \ln (\text{SBP} - 50) \text{ at Exam } k
  \]
EXAMPLE: PARAMETER ESTIMATION

- Consider logistic regression of $Y$ on Age, Smoke, Chol and SBP with transformed SBP measured with error
  - The plots on the following page illustrate the simulation extrapolation method for estimating the parameters in the logistic regression model
Age

Lambda Coefficient (x 1e+2)

-1.0 0.0 1.0 2.0

4.36 4.95 5.54 6.13 6.72

Smoking

Lambda Coefficient (x 1e+1)

-1.0 0.0 1.0 2.0

3.43 4.68 5.93 7.18 8.43

Cholesterol

Lambda Coefficient (x 1e+3)

-1.0 0.0 1.0 2.0

5.76 6.82 7.87 8.93 9.98

Log(SBP-50)

Lambda Coefficient (x 1e+0)

-1.0 0.0 1.0 2.0

1.29 1.50 1.71 1.92 2.12
INSTRUMENTAL VARIABLES: RATIONALE

- Remember, $W = X + U$, $U \sim \text{Normal}(0, \sigma_u^2)$.

- The most direct and efficient way to get information about $\sigma_u^2$ is to observe $X$ on a subset of the data.

- The next best way is via replication, namely to take $\geq 2$ independent replicates
  
  * $W_1 = X + U_1$ and $W_2 = X + U_2$.
  * If these are indeed replicates, then we can estimate $\sigma_u^2$ via a components of variance analysis.

- The third method is to use Instrumental Variables.
  
  * Sometimes $X$ cannot be observed and replicates cannot be taken.
  * Then IV’s can help.
What Is An Instrumental Variable?

\[ Y = \beta_0 + \beta_x X + \epsilon; \]
\[ W = X + U; \]
\[ U \sim \text{Normal}(0, \sigma_u^2). \]

- In linear regression, an instrumental variable \( T \) is a random variable which has three properties:
  - \( T \) is independent of \( \epsilon \)
  - \( T \) is independent of \( U \)
  - \( T \) is related to \( X \).
  - You only measure \( T \) to get information about measurement error: it is not part of the model.
  - In our parlance, \( T \) is a surrogate for \( X \)!
- Whether \( T \) qualifies as an instrumental variable can be a difficult question.
**A N E X A M P L E: C A L I B R A T I N G A Q U E S T I O N N A I R E**

\[ X = \text{usual (long–term) average intake of Fat (log scale)}; \]
\[ Y = \text{Fat as measured by a questionnaire}; \]
\[ W = \text{Fat as measured by 6 days of 24–hour recalls} \]
\[ T = \text{Fat as measured by a diary record} \]

- In this example, the time ordering was:
  
  \* Questionnaire
  
  \* Then one year later, the recalls were done fairly close together in time.
  
  \* Then 6 months later, the diaries were measured.

- One could think of the recalls as replicates, but some researchers have worried that substantial correlations exist, i.e., they are not *independent* replicates.

- The 6–month gap with the recalls and the 18–month gap with the questionnaire makes the diary records a good candidate for an instrument.
Using Instrumental Variables: Motivation

- In what follows, we will use underscores to denote what coefficients go where.

- For example, $\beta_{Y|1X}$ is the coefficient for $X$ in the regression of $Y$ on $X$.

- Let’s do a little algebra:

  \[
  Y = \beta_{Y|1X} + \beta_{Y|1X}X + \epsilon; \\
  W = X + U; \\
  (\epsilon, U) = \text{independent of } T.
  \]

- This means

  \[
  E(Y \mid T) = \beta_{Y|1T} + \beta_{Y|1T}T = \beta_{Y|1X} + \beta_{Y|1X}E(X \mid T) \\
  = \beta_{Y|1X} + \beta_{Y|1X}E(W \mid T) = \beta_{Y|1T} + \beta_{Y|1X}\beta_{W|1T}T.
  \]
From previous page

\[ E(Y \mid T) = \beta_{Y|1T} + \beta_{Y|1X} \beta_{W|1T}T. \]

We want to estimate \( \beta_{Y|1X} \)

* From above

\[ \beta_{Y|1T} = \beta_{Y|1X} \beta_{W|1T}. \]

* Equivalently,

\[ \beta_{Y|1X} = \frac{\beta_{Y|1T}}{\beta_{W|1T}}. \]

* Regress \( Y \) on \( T \) and divide its slope by the slope of the regression of \( W \) on \( T \)!
THE DANGERS OF A WEAK INSTRUMENT

• Remember that we get the IV estimate using the relationship

\[
\beta_{Y|1X} = \frac{\beta_{Y|1T}}{\beta_{W|1T}}.
\]

• The division causes increased variability.

  ✴ If the instrument is very weak, the slope \( \beta_{W|1T} \) will be near zero.

  ✴ This will make the IV estimate very unstable.
Here’s another way to do the same thing and get the IV estimator without doing division explicitly:

* Regress $W$ on $T$.

* Form predicted values: $\beta_{W|1T} + \beta_{W|1T}T$.

* Regress $Y$ on these predicted values

* The slope is the instrumental variables estimate.
**LOGISTIC REGRESSION EXAMPLE**

- These data come from a paper by Satten and Kupper.

  
  \[
  \begin{align*}
  Y &= \text{Evidence of Coronary Heart Disease (binary)} \\
  W &= \text{Cholesterol Level} \\
  T &= \text{LDL} \\
  Z &= \text{Age and Smoking Status}
  \end{align*}
  \]

- It’s not particularly clear that \(T\) is really an instrument. While it may well be uncorrelated with the error in \(W\) and the random variation in \(Y\), it might also be included as part of the model itself!

  ∗ There are no replicates here to compare with.

- Note here that we have added variables, \(Z\), which are measured without error.

- The algorithm does not change: all regressions are done with \(Z\) as an extra covariate.
The IV algorithm in linear regression is:

**STEP 1:** Regress $W$ on $T$ and $Z$ (may be a multivariate regression)

**STEP 2:** Form the predicted values of this regression

**STEP 3:** Regress $Y$ on the predicted values and $Z$.

**STEP 4:** The regression coefficients are the IV estimates.

Only Step 3 changes if you do not have linear regression but instead have logistic regression or a GLM.

* Then the “regression” is logistic or GLM.
* Very simple to compute.
* Easily bootstrapped.

This method is “valid” in GLM’s to the extent that regression calibration is valid.
Logistic Regression Example: Results

- We used LDL/100, Cholesterol/100, Age/100

- The naive logistic regression leads to the following analysis:
  - Slope and bootstrap s.e. for Cholesterol: 0.65 and 0.29
  - Slope and bootstrap s.e. for Smoking: 0.065 and 0.260
  - Slope and bootstrap s.e. for Age: 7.82 and 4.26

- The IV logistic regression leads to the following analysis:
  - Slope and bootstrap s.e. for Cholesterol: 0.91 and 0.33
  - Slope and bootstrap s.e. for Smoking: 0.056 and 0.259
  - Slope and bootstrap s.e. for Age: 7.79 and 4.31

- Note here how only the coefficient for cholesterol is affected by the measurement error.
END OF LECTURE 4
LECTURE 5: BAYESIAN METHODS

OUTLINE

• Likelihood: Outcome, measurement, and exposure models

• Priors and posteriors

• MCMC

• Example: linear regression

• Example: Mixtures of Berkson and classical error and application to thyroid cancer and exposure to fallout
**Outcome model:** (true disease model)

\[ f(Y|X, Z, \theta_O) \]

Could be a GLM or a “nonparametric” spline model

**Measurement model:** (measurement error model)

\[ f(W|Y, X, Z, \theta_M) = f(W|X, Z, \theta_M) \] if nondifferential (which is assumed here)

Could be a classical Gaussian measurement error model

**Exposure model:** (predictor distribution model)

\[ f(X|Z, \theta_X)f(Z|\theta_Z) \] (usually \( f(Z|\theta_Z) \) is ignored)

Could be a low-dimensional parametric model or a more flexible model (e.g., normal mixture model)
\begin{align*}
\pi(\theta_O), \; \pi(\theta_M), \; \pi(\theta_X), \; \pi(\theta_Z) \quad (\text{usually } \pi(\theta_Z) \text{ is ignored})
\end{align*}
Likelihood

\[
f(Y, W|\theta_O, \theta_M, \theta_X, Z) = \int f(Y|X, Z, \theta_O) f(W|X, Z, \theta_M) f(X|Z, \theta_X) \, dX
\]

**Major problem:** computing the integral

**We will go in another direction — Bayesian analysis by MCMC.**
POSTERIOR AND MCMC

\[ f(\theta_O, \theta_M, \theta_X, X|Y, W, Z) \propto f(Y, W, Z, X, \theta_O, \theta_M \theta_X|Z) \]

\[ = f(Y|X, Z, \theta_O) f(W|X, Z, \theta_M) f(X|Z, \theta_X) \pi(\theta_O) \pi(\theta_M) \pi(\theta_X) \]

MCMC — For example we might sample successively from:

1. \([\theta_O|\text{others}]\]
2. \([\theta_M|\text{others}]\]
3. \([\theta_X|\text{others}]\]
4. \([X|\text{others}]\]
MCMC: Linear Regression With Replicate Measurements

\[ Y_i \mid X_i, Z_i, \beta, \sigma_\epsilon \sim N(\mathbf{Z}_i^T \beta_z + X_i \beta_x, \sigma_\epsilon^2) \]  (outcome model)

\[ W_{i,j} \mid X \sim IN(X_i, \sigma_u^2), \ j = 1, \ldots, J_i \]  (measurement model)

\[ [X_i \mid Z_i] = [X_i] \sim N(\mu_x, \sigma_x^2) \]  (exposure model)

\[ [\beta_x] = N(0, \sigma_{\beta_x}^2), \ [\beta_z] = N(0, \sigma_{\beta_z}^2 I), \ [\mu_x] = N(0, \sigma_{\mu_x}^2) \]

\[ [\sigma_\epsilon^2] = IG(\delta, \delta), \ [\sigma_u^2] = IG(\delta, \delta), \ [\sigma_x^2] = IG(\delta, \delta), \]

\( \sigma_{\beta} \) and \( \sigma_{\mu} \) are “large” and \( \delta \) is “small”

**unknowns:** \((\beta_x, \beta_z, \sigma_\epsilon), (\sigma_u), (X_1, \ldots, X_N), (\mu_x, \sigma_x)\)
MCMC: Linear Regression — Likelihood

Let

\[ C_i = \begin{pmatrix} Z_i \\ X_i \end{pmatrix}, \quad Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_N \end{pmatrix}, \quad \text{and} \quad \beta = \begin{pmatrix} \beta_z \\ \beta_x \end{pmatrix} \]

Likelihood:

\[
[Y_i, W_{i,1}, \ldots, W_{i,J_i}|\text{others}] \propto \frac{1}{\sigma_\epsilon \sigma_u} \exp \left\{ -\frac{1}{2} \left( \frac{(Y_i - C_i^T \beta)^2}{\sigma_\epsilon^2} + \sum_{j=1}^{J_i} \frac{(W_{i,j} - X_i)^2}{\sigma_u^2} \right) \right\}
\]
MCMC: Linear Regression — Posterior For $\beta$

$$[\beta|\text{others}] \propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{N} (Y_i - C_i^T \beta)^2 - \frac{1}{2\sigma^2_\beta} \|\beta\|^2 \right\}$$

Let $C$ have $i$th row $C_i^T$ and let $\lambda = \sigma^2_\epsilon / \sigma^2_\beta$. Then

$$[\beta|\text{others}] = N \left( \{C^T C + \lambda I\}^{-1} C^T Y, \sigma^2_\epsilon (C^T C + \lambda I)^{-1} \right)$$

- Exactly what we get for linear regression without measurement error.

  * Except now $C$ will vary on each iteration of MCMC, since it contains imputed $X$s.
MCMC: Linear Regression — Posterior For $\mu_x$

$$[\mu_x|\text{others}] \propto \exp \left\{ -\sum_{i=1}^{N} \frac{(X_i - \mu_x)^2}{2\sigma_x^2} - \frac{\mu_x^2}{2\sigma^2} \right\}$$

Let $\sigma^2_X = \sigma_x^2 / N$. Then

$$[\mu_x|\text{others}] \sim N \left\{ \left( \frac{1}{\sigma^2_X} + \frac{1}{\sigma^2_\mu} \right)^{-1} \left( \frac{\bar{X}}{\sigma^2_X} + \frac{0}{\sigma^2_\mu} \right), \left( \frac{1}{\sigma^2_X} + \frac{1}{\sigma^2_\mu} \right)^{-1} \right\},$$

$$\approx N (\bar{X}, \sigma^2_X) \text{ if } \sigma_\mu \text{ is large}$$
Define
\[ \sigma^2_{w,i} = \frac{\sigma^2_u}{J_i} \]

Then
\[
[X_i|\text{others}] \propto \exp \left[ -\frac{1}{2} \left\{ \frac{(Y_i - X_i/\beta_x - Z_i^T\beta_z)^2}{\sigma^2_{\epsilon}} + \frac{(X_i - \mu_x)^2}{\sigma^2_x} + \frac{(W_i - X_i)^2}{\sigma^2_{w,i}} \right\} \right]
\]

After some algebra, \([X_i|\text{others}]\) is normal with mean
\[
\frac{\{ (Y_i - Z_i^T\beta_z)/\beta_x \}(\beta^2_x/\sigma^2_{\epsilon}) + (\mu_x)(1/\sigma^2_x) + (W_i)(1/\sigma^2_{w,i})}{(\beta^2_x/\sigma^2_{\epsilon}) + (1/\sigma^2_x) + (1/\sigma^2_{w,i})}
\]
and variance
\[
\left\{ (\beta^2_x/\sigma^2_{\epsilon}) + (1/\sigma^2_x) + (1/\sigma^2_{w,i}) \right\}^{-1}
\]
MCMC: Linear Regression — Posterior For $\sigma_\epsilon^2$

Prior:

$$[\sigma_\epsilon^2] = IG(\delta, \delta)$$

$IG(\alpha, \beta)$:

mean = $\alpha/\beta$ and variance = $\alpha/\beta^2$ and density $\propto x^{-(\alpha+1)} \exp \left( -\frac{\beta}{x} \right)$

Posterior:

$$[\sigma_\epsilon^2 | \text{others}] \propto \frac{1}{(\sigma_\epsilon^2)^{\delta+N/2+1}} \exp \left\{ \delta + \frac{1}{2} \sum_{i=1}^{N} \left( \frac{(Y_i - X_i/\beta_x - Z_i^T \beta_z)^2}{\sigma_\epsilon^2} \right) \right\}$$

$$\Rightarrow [\sigma_\epsilon^2 | \text{others}] = IG \left\{ \delta + \frac{N}{2}, \delta + \frac{1}{2} \sum_{i=1}^{N} (Y_i - X_i/\beta_x - Z_i^T \beta_z)^2 \right\}$$
MCMC: Linear Regression Posterior For $\sigma_x^2$

\[
[\sigma_x^2 | \text{others}] \propto \frac{1}{(\sigma_x^2)^{\delta+\frac{N}{2}+1}} \exp \left\{ \frac{\delta + \frac{1}{2} \sum_{i=1}^{N} (X_i - \mu_x)^2}{\sigma_x^2} \right\}
\]

\[
\Rightarrow [\sigma_x^2 | \text{others}] = IG \left\{ \delta + \frac{N}{2}, \delta + \frac{1}{2} \sum_{i=1}^{N} (X_i - \mu_x)^2 \right\}
\]
MCMC: Linear Regression — Posterior For $\sigma_u^2$

\[
[\sigma_u^2 | \text{others}] \propto \frac{1}{(\sigma_u^2)^{\delta + \sum_{i=1}^{N} J_i / 2 + 1}} \exp \left\{ \delta + \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{J_i} (W_{i,j} - X_i)^2 \right\} \\
\Rightarrow [\sigma_u^2 | \text{others}] = IG \left\{ \delta + \frac{\sum_{i=1}^{N} J_i}{2}, \delta + \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{J_i} (W_{i,j} - X_i)^2 \right\}
\]
Mallick, Hoffman, and Carroll (2002) study models where both Berkson and classical measurement errors are present.

- $W = \text{observed dose of radiation from fallout due to nuclear testing in Nevada}$
- $W = C \times DCF \times I \times TD \times FP$
  - $C = \text{time-integrated radioiodine concentration of milk}$
    - specific to producer, not individual (Berkson)
    - but one component is deposition rate of $I^{131}$ across regions of study (Classical)
Mixtures of Berkson and Classical Measurement Error

- $DCF =$ ingestion dose conversion factor
  - specific to age and isotope (Berkson) (Classical)
- $I =$ individual milk intake rate in liters per day, measured by a food frequency questionnaire (FFQ) (Berkson)
- $TD =$ time delay (depends on source of milk — from FFQ) (Berkson)
- $FP =$ frequency of purchase correction factor (from FFQ) (Berkson)
**Latent Variable Model**

- \( W \) = measured dose
- \( X \) = true dose
- \( L \) = latent intermediate variable

Berkson and classical errors are combined in:

\[
\log(X) = \log(L) + U_b \\
\log(W) = \log(L) + U_c
\]

\[U_b = 0 \Rightarrow \log(X) = \log(L) \Rightarrow \log(W) = \log(X) + U_c \text{ (Classical)}\]

\[U_c = 0 \Rightarrow \log(W) = \log(L) \Rightarrow \log(X) = \log(W) + U_b \text{ (Berkson)}\]
LATENT VARIABLE MODEL

• $\sigma^2_b + \sigma^2_c$ is known

• $p = \sigma^2_b / (\sigma^2_b + \sigma^2_c)$ is unknown

  * informative priors:
    - $p = 1$ (Berkson)
    - $p = 0$ (Classical)
    - $p$ is uniform on $[0.2, 0.8]$ (Mixture)
    - compare results from these three priors (sensitivity analysis)

• $Z =$ age at exposure, sex (known exactly)

• state = state of residence (Utah, Nevada, Arizona) (known exactly)
Exposure and Outcome Models

• Exposure model:

\[
[ \log(L) | Z, \text{state} = s ] \sim N( \alpha_s 0 + Z^T \alpha_s 1, \sigma_s^2 )
\]

• Outcome model:

* \( Y = \) indicator of thyroid cancer

* \( \logit\{\text{pr}(Y = 1|Z, X)\} = \beta_0 + Z^T \beta_1 + \log(1 + \theta X) \)
Approximate Model For $[Y|W, Z]$

- **Outcome model (from previous page):**
  
  * $Y$ = indicator of thyroid cancer
  * $\text{logit} \{\text{pr}(Y = 1|Z, X)\} = \beta_0 + Z^T \beta_1 + \log(1 + \theta X)$

- **Outcome model based on $W$:**
  
  * $\text{logit}\{\text{pr}(Y = 1|Z, W)\} \approx \beta_0 + Z^T \beta_1 + \log \left(1 + \theta \gamma W^{\lambda_{x|w,t}}\right)$
  
  where
  $$\gamma = \exp \left\{\left(1 - \lambda_{x|w,t}\right) \left(\mu_{L|Z} + \sigma_{L|Z}^2/2\right) + \sigma_b^2/2\right\}$$

  - $\mu_{L|Z}$ and $\sigma_{L|Z}^2$ are the mean and variance of $\log(L)$ given $Z$

  and

  $$\lambda_{x|w,t} = \frac{\sigma_{L|Z}^2}{\sigma_{L|Z}^2 + \sigma_c^2}$$
Berkson Case

- $\lambda_{x|w,t} = 1$
- $\gamma = \exp(\frac{\sigma^2_b}{2})$

and model on previous page simplifies to

$$\text{logit}\{\text{pr}(Y = 1|Z, W)\} \approx \beta_0 + Z^T\beta_1 + \log(1 + \theta \gamma W)$$

* $\gamma > 1$ so effect of Berkson error is for the naive analysis to **overestimate** the effect of dose

* In the regression, replace $W$ by $\gamma W$ to correct for bias (regression calibration)
Several parametric and semiparametric models were used

- nonparametric model for dose is constrained to be monotonically increasing

- nonparametric model for the distribution of $\log(L)$ uses a Pólya-tree prior (Lavine, 1992)
### Bayesian Analysis: Some Results for Parametric Models

<table>
<thead>
<tr>
<th>Error model</th>
<th>Post. mean of $\theta$</th>
<th>RR at 1 Gy</th>
<th>95% credible interval for RR</th>
</tr>
</thead>
<tbody>
<tr>
<td>No error</td>
<td>38.9</td>
<td>9.4</td>
<td>(4.5, 13.8)</td>
</tr>
<tr>
<td>Classical</td>
<td>74.1</td>
<td>17.1</td>
<td>(8.5, 24.6)</td>
</tr>
<tr>
<td>Berkson</td>
<td>31.9</td>
<td>7.9</td>
<td>(3.8, 11.4)</td>
</tr>
<tr>
<td>Mixture</td>
<td>56.1</td>
<td>13.2</td>
<td>(5.0, 23.1)</td>
</tr>
</tbody>
</table>

- Effects on a naive analysis:
  
  * Berkson error $\Rightarrow$ overestimate effect
  
  * Classical error $\Rightarrow$ underestimate effect
  
  * Mixture of errors $\Rightarrow$ underestimate effect (but less than if only classical error)
**Bayesian Analysis: Semiparametric Models**

- The semiparametric models have better fit by DIC.
- The semiparametric model for \( \log(L) \):
  - Leads to slightly smaller estimated posterior means of \( \theta \) and RR.
  - Credible intervals also widen.
- The semiparametric outcome model:
  - Leads to slightly higher estimated posterior means of \( \theta \) and RR.
  - Credible intervals also widen.
- Overall effect of both semiparametric models on mixture of errors model:
  - Raise RR from 13.2 to 14.2.
  - Change CI from (5.0, 23.1) to (1.7, 33.6).
**Bayesian Analysis: Assumption Of Independent Error**

- almost certainly false
- $C$ (radioiodine concentration) includes
  - deposition of $I^{131}$ by region
  - mass interception on vegetation
  - consumption of vegetation by cows
  - effective half-life of $I^{131}$ in vegetation
  - milk transfer coefficient (MTC)
- Utah study generated a distribution of log-normally distributed MTC’s with estimated mean and variance
  - if parameters known, then error is Berkson
  - parameters estimated from historical data and literature review
    $\Rightarrow$ shared classical errors
Sensitivity Analysis

- For the six gender/state combinations, Berkson errors for individuals in a group had a common correlation of $\rho$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>0.0</th>
<th>0.2</th>
<th>0.4</th>
<th>0.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E(\theta</td>
<td>\text{data})$</td>
<td>56.1</td>
<td>65.9</td>
<td>84.1</td>
</tr>
<tr>
<td>CI</td>
<td>(18.6, 102)</td>
<td>(21.4, 120)</td>
<td>(30.4, 143)</td>
<td>(38.2, 152)</td>
</tr>
</tbody>
</table>
END OF LECTURE 5
LECTURE 6: NONPARAMETRIC REGRESSION WITH MEASUREMENT ERROR

OUTLINE

- Earlier approaches: deconvolution kernels, SIMEX, regression calibration
- New Bayesian spline approach (Berry, Carroll, and Ruppert, 2002, JASA)
- Simulation results
- Example: A clinical trial
- Example: TSP and mortality, a sensitivity study
THE PROBLEM OF MEASUREMENT ERROR — ILLUSTRATION
THE PROBLEM OF MEASUREMENT ERROR — ILLUSTRATION
The Problem Of Measurement Error

- The regression model is
  \[ Y = m(X) + \epsilon \]
  where \( m \) is only known to be smooth.

- Observe
  \[ Y \text{ and } W = X + U, \text{ where} \]
  
  - \( E(U|X) = 0 \) and \( \text{var}(U|X) = \sigma^2_u \)
  
  - \( U|X \) normally distributed.

- Measurement error variance \( \sigma^2_u \) is estimated from internal replicate data. (Observe \( W_{ij}, j = 1, \ldots, J_i \).)
<table>
<thead>
<tr>
<th>Available Methods</th>
</tr>
</thead>
</table>

- Deconvolution kernels
- SIMEX
- Structural spline
- Bayesian spline
Deconvolution Kernels

- Globally consistent nonparametric regression by deconvolution kernels (Fan and Truong, 1993, *Annals*)
  - does not work so well
    - Fan & Truong show very poor asymptotic rates of convergence
    - simulations show poor finite-sample behavior
  - no methodology for bandwidth selection or inference
The SIMEX method is due to Cook & Stefanski (1995, *JASA*).

SIMEX has been previously applied to parametric problems.

Makes no assumptions about the true $X$’s. *(Functional)*

Results in estimators which are *approximately* consistent, i.e., consistent at least to order $O(\sigma_u^6)$. 
**SIMEX**

- Carroll, Maca, Ruppert (2001, *Biometrika*) (CMR) applied the SIMEX to non-parametric regression.
- **CMR** have asymptotic theory in the local polynomial regression (LPR) context.
  - The estimators have the usual rates of convergence.
  - They are approximately consistent, to order $O(\sigma_u^6)$.
- An asymptotic theory with rates seems very difficult for splines
  - but, simulations in **CMR** indicate that SIMEX/splines works a little *better* than SIMEX/kernel
  - problem seems due to undersmoothing
  - With better bandwidth selection, SIMEX/LPR is competitive with other methods.
The regression of $Y$ on the observed $W$ is

$$E(Y|W) = E \{m(X)|W\} = \int m(x)f(x|W)dx.$$ 

Suppose that we had:

* convenient flexible form for $m(x; \beta)$
* convenient flexible distribution for $f(x|W)$.

Then we could estimate $m(X; \beta)$ by minimizing over the data

$$\sum_{i=1}^{n} \left\{ Y_i - \int m(x; \beta)f(x|W_i) dx \right\}^2.$$
Regression Splines — “Plus Function” Basis

- Model

\[ E(Y|X) = m(X; \beta) := \sum_{j=0}^{J} \beta_j X^j + \sum_{k=1}^{K} \beta_{k+J}(X - \xi_k)^J_+ \]
**10-Knot Quadratic Spline Approximations**

- $\sin(15x)$
- $\sin\{5(1+x+x^2)\}$
- $\Phi\{18(x-.05)\}$
- $0.8x+\exp\{-35(x-.5)\}$

“blue” = function, “red” = spline approximation
**Regression Splines**

- **Model**

\[
E(Y|X) = m(X; \beta) := \sum_{j=0}^{J} \beta_j X^j + \sum_{k=1}^{K} \beta_{k+J} (X - \xi_k)^j_+ + \epsilon
\]

- \(X^j\) is replaced by \(E(X^j|W)\)

- \((X - \xi_k)^j_+\) is replaced by \(E\{(X - \xi_k)^j_+|W\}\)

- The key remaining issue: the joint distribution of \(X\) and \(U\).

- **CMR** used a mixtures of normals for \([X]\) and Gibbs sampling to estimate the parameters. **(Flexible structural)**

  - This is an extension to measurement error of an idea of Roeder & Wasserman (JASA, 1997).
**FULLY BAYESIAN MODEL**

**What’s New?**

**Answer:** Fully Bayesian MCMC method in Berry, Carroll, and Ruppert (2002, *JASA*) (BCR)

- Uses splines
  - smoothing or penalized
  - P-splines in this lecture

- Structural
  - $X_i$ are iid normal
  - but seems robust to violations of normality
**Fully Bayesian Model**

- Smoothing parameter is automatic
- Inference adjusts for the data-based smoothing parameter and for measurement error
- Allow global confidence bands
FULLY BAYESIAN MODEL — PARAMETERS

- **Regression Model**
  \[ Y_i = m(x_i; \beta) + \epsilon_i \]
  - \( m(x_i; \beta) \) is a P-spline
  - \( \epsilon_i \) iid \( N(0, \sigma^2_\epsilon) \)

- **Measurement Error Model**
  \[ W_{ij} = X_i + U_{ij} \] where \( U_{ij} \) iid \( N(0, \sigma^2_u) \)

- **Structural Model**
  \( X_i \) iid \( N(\mu_x, \sigma^2_x) \)

- **Parameters:** \( \beta, \sigma^2_e, \sigma^2_u, \mu_x, \sigma^2_x \)
FULLY BAYESIAN MODEL — PARAMETERS

• Priors

  – $\beta$ is $N(0, (\gamma K)^{-1})$ where $K$ is known. [$\alpha := \gamma \sigma^2_e$ is the smoothing parameter.]
  – $\gamma$ is Gamma($A_\gamma$, $B_\gamma$)
  – $\sigma^2_e$ is Inv-Gamma($A_e$, $B_e$)
  – $\sigma^2_u$ is Inv-Gamma($A_u$, $B_u$)
  – $\mu_x$ is $N(d_x, t_x^2)$
  – $\sigma^2_x$ is Inv-Gamma($A_x$, $B_x$)

• Hyperparameters: $A_e, B_e, A_u, B_u, A_x, B_x, d_x, t_x^2, A_\gamma, B_\gamma$

  – all fixed at values making the priors noninformative

  * E.g., $t_x^2 = 10^6$. 
**FULLY BAYESIAN MODEL — SMOOTHING**

- **Model**
  \[ E(Y|X) = m(X; \beta) := \sum_{j=0}^{J} \beta_j X^j + \sum_{k=1}^{K} \beta_{k+J}(X - \xi_k)^+ \]

- \( \beta \) is \( N(0, (\gamma K)^{-1}) \) where \( K \) is known. [\( \alpha := \gamma \sigma^2_e \) is the smoothing parameter.]

- **Let**
  \[ K = \text{diag}(\Delta, \ldots, \Delta, 1, \ldots, 1) \]

  - \( \Delta \) is “small” — essentially 0
  - \( J + 1 \Delta \)’s followed by \( K \) ones
  - The prior on the jumps at the knots is that they are iid.
  - The prior on the polynomial coefficients is “diffuse,” i.e., “non-informative”.
Gibbs Sampling

- Iterate through \( \beta, \sigma_e^2, \sigma_u^2, \sigma_x^2, \mu_x, \gamma, X_1, \ldots, X_n \).
- All steps except one are easy, either gamma, inverse-gamma, or normal
  - E.g.,

\[
[\beta | \text{other parameters, } Y, W] \sim \text{Normal}
\]

\[
\text{Mean} = (X^T X + \gamma K)^{-1} X^T Y
\]

\[
\text{Cov} = \sigma_e^2 (X^T X + \gamma K)^{-1}.
\]

- Here \( X \) is one of the “other parameters”
- Essentially we’re fitting a spline to the imputed \( X \)’s and the observed \( Y \)’s
**GIBBS SAMPLING**

* Estimate of $\beta$, call it $\hat{\beta}$, is

$$(X^T X + \gamma K)^{-1} X^T Y$$

averaged over $\gamma$ and $X$. 

Gibbs Sampling

- The exception to the sampling being quick and easy is that a Metropolis-Hastings step is needed for $X_1, \ldots, X_n$.

$$[X_i|\mu_x, \sigma_x^2, \beta, \sigma_e^2, \sigma_u^2, Y, W]$$

$$\propto \exp \left( \frac{-1}{2\sigma_u^2} \sum_{j=1}^{m_i} (W_{ij} - X_i)^2 - \frac{1}{2\sigma_e^2} \{Y_i - m(X_i; \beta)\}^2 - \frac{1}{2\sigma_x^2} (X_i - \mu_x)^2 \right).$$
BAYESIAN INFERENCE

• Let $\tilde{X}$ be the spline basis function evaluated on a fine grid over some interval, $[a, b]$.

• $\tilde{X}\beta$ is the curve on $[a, b]$.

• $\tilde{X}\hat{\beta}$ is the estimated curve.

• Let $K_\alpha$ be the $(1 - \alpha)$ MCMC sample quantile of

$$\max_{\text{grid}} \left\{ \frac{\tilde{X}(\beta - \hat{\beta})}{\text{SD}(\tilde{X}\beta)} \right\}.$$

• Then,

$$\tilde{X}\hat{\beta} \pm K_\alpha \text{SD}(\tilde{X}\beta)$$

is a $100(1 - \alpha)\%$ simultaneous confidence band for the curve on $[a, b]$. 
**Bayesian Inference For Derivatives**

- Let \( \tilde{X}' \) be derivatives of the spline basis function evaluated on a fine grid over \([a, b]\).
- \( \tilde{X}' \beta \) is the curve’s derivative on \([a, b]\).
- \( \tilde{X}' \hat{\beta} \) is the estimated derivative.
- Let \( K'_\alpha \) be the \((1 - \alpha)\) MCMC sample quantile of
  \[
  \max_{\text{grid}} \left\{ \frac{\tilde{X}'(\beta - \hat{\beta})}{\text{SD}(\tilde{X}'\beta)} \right\}.
  \]
- Then,
  \[
  \tilde{X}' \hat{\beta} \pm K'_{.95} \text{SD}(\tilde{X}'\beta)
  \]
  is a \(100(1 - \alpha)\)% simultaneous confidence band for the derivative on \([a, b]\).
SIMULATIONS

The six cases were considered. $n_i \equiv 2$ in each case.
SIMULATIONS

• [Case 1]

The regression function is

\[ m(x) = \frac{\sin(\pi x/2)}{1 + 2x^2 \{\text{sign}(x) + 1\}}. \]

with \( n = 100, \sigma^2_\epsilon = 0.3^2, \sigma^2_u = 0.8^2, \mu_x = 0 \) and \( \sigma^2_x = 1. \)
SIMULATIONS

• [Case 2] Same as Case 1 except $n = 200$.

• [Case 3] A modification of Case 1 above except that $n = 500$.

• [Case 4]

Case 1 of CMR so that

$$m(x) = 1000x_+^3(1 - x)_+^3,$$

$$x_+ = xI(x > 0), \text{ with } n = 200, \sigma^2_\epsilon = 0.0015^2, \sigma^2_u = (3/7)\sigma^2_x, \mu_x = 0.5 \text{ and } \sigma^2_x = 0.25^2.$$
SIMULATIONS

- **[Case 5]**

A modification of Case 4 of CMR so that

\[ m(x) = 10 \sin(4\pi x), \]

with \( n = 500, \sigma^2_\epsilon = 0.05^2, \sigma^2_u = 0.141^2, \mu_x = 0.5 \) and \( \sigma^2_x = 0.25^2. \)

- **[Case 6]**

The same as Case 1 above except that \( X \) is a normalized chi-square(4) random variable. (Tests robustness against violation of the structural assumptions.)
## SIMULATIONS

### Mean Squared Bias $\times 10^2$

<table>
<thead>
<tr>
<th>Method</th>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 3</th>
<th>Case 4</th>
<th>Case 5</th>
<th>Case 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naive</td>
<td>5.59</td>
<td>4.92</td>
<td>5.21</td>
<td>1,108</td>
<td>3,733</td>
<td>4.83</td>
</tr>
<tr>
<td><strong>Bayes</strong></td>
<td><strong>0.78</strong></td>
<td><strong>0.38</strong></td>
<td>1.04</td>
<td>17.4</td>
<td>468</td>
<td>1.74</td>
</tr>
<tr>
<td>Flex. Structural, 5 knots</td>
<td>1.38</td>
<td>0.62</td>
<td><strong>0.46</strong></td>
<td>3.7</td>
<td>838</td>
<td><strong>1.47</strong></td>
</tr>
<tr>
<td>Flex. Structural, 15 knots</td>
<td>1.44</td>
<td>0.60</td>
<td>0.66</td>
<td><strong>3.3</strong></td>
<td><strong>226</strong></td>
<td>1.75</td>
</tr>
</tbody>
</table>

### Mean Squared Error $\times 10^2$

<table>
<thead>
<tr>
<th>Method</th>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 3</th>
<th>Case 4</th>
<th>Case 5</th>
<th>Case 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naive</td>
<td>6.91</td>
<td>5.57</td>
<td>5.38</td>
<td>1,155</td>
<td>3,793</td>
<td>5.77</td>
</tr>
<tr>
<td><strong>Bayes</strong></td>
<td><strong>2.84</strong></td>
<td><strong>1.56</strong></td>
<td><strong>1.47</strong></td>
<td><strong>195</strong></td>
<td>1,031</td>
<td><strong>2.69</strong></td>
</tr>
<tr>
<td>Flex. Structural, 5 knots</td>
<td>8.17</td>
<td>3.82</td>
<td>1.73</td>
<td>217</td>
<td>2,032</td>
<td>7.27</td>
</tr>
<tr>
<td>Flex. Structural, 15 knots</td>
<td>9.90</td>
<td>5.40</td>
<td>1.85</td>
<td>237</td>
<td><strong>799</strong></td>
<td>6.94</td>
</tr>
</tbody>
</table>

Results based on 200 Monte Carlo simulations for each case. SIMEX was not included in the table — it was not among the best estimators.
EXAMPLE — SIMULATED

- $Y = \sin(2X) + \epsilon$
- $X$ is $N(1, 1)$
- $\sigma_u = 1$
- $\sigma_e = 0.15$
- $n = 201$
- $n_i = 2$ for all $i$
- 15 knot quadratic P-splines
- 2,000 iterations of Gibbs. First 667 deleted as burn-in.
Results of Gibbs Sampling. Every twentieth iteration.

Note: $X(1) = -1.45$ and $\bar{W}(1) = -0.8$. Also, $\log(\sigma_e) = -1.9$. 
What does the Bayes approach work so well? Here’s my explanation:

Bayes uses all possible information to estimate $X$ and, especially, $m(X)$.

- $\|m(X) - E\{m(X)|W, Y, \text{ other param.}\}\|$
  $\approx \|m(X) - \text{ave}\{m(\hat{X})\}\| = 2.47$

- $\|m(X) - m(E\{X|W, Y, \text{ other param.}\})\|$
  $\approx \|m(X) - m(\text{ave}\{\hat{X}\})\| = 4.67$

- $\|m(X) - m(E(X|W))\| = 10.25$

- $\|m(X) - m(W)\| = 12.36$
EXAMPLE — CLINICAL TRIAL

• Study of a psychiatric medication.

• Treatment and control group.

• Evaluation at baseline ($W$) and at end of study ($Y$).
  
  – smaller values $\rightarrow$ more severe disease
  
  – scale is a combination of self-report and clinical interview so there is considerable measurement error
  
  – it is believed that $\sigma_u^2 \approx 0.35$.

• We are interested in $\Delta(X) := m(X) - X = E(Y - W|X)$.

• Preliminary Wilcoxon test found a highly significant treatment effect.

• Question: How does the treatment effect depend upon the baseline value?
Lecture 6

True Baseline Score

Change

-1 0 1 2
0.0
0.5
1.0
1.5
2.0
2.5
3.0

Treatment
Control

- use the same model as Berry, Carroll, and Ruppert (2002)
- compute MLEs using van Dyk’s (2002) nested Monte Carlo EM algorithm
- their estimator works well
  - this suggests that the success of the Berry, Carroll, and Ruppert method is due to using the likelihood rather than from the information in the prior
- Ganguli, Staudenmayer, and Wand extend their method to additive models
**EXAMPLE — A SENSITIVITY ANALYSIS**

Original study from Zanobetti, Wand, Schwartz, and Ryan (2000).

This analysis from Ganguli, Staudenmayer, and Wand (2005, to appear) — also see Ruppert, Wand, Carroll (2003, *Semiparametric Regression*)

- **Outcome** = log(natural mortality)
- **Exposure**
  - TSP = total suspended particles
- **Confounders**
  - DAY = days since beginning of study
  - Mean daily temperature
  - Relative humidity
EXAMPLE — A SENSITIVITY ANALYSIS

reliability ratio = \frac{\text{var } \{\log(TSP_i)\}}{\text{var } \{\log(TSP_i)\} + \sigma_{u}^{2}}.

**Problem:** No replicates so the measurement error variance $\sigma_{u}^{2}$ cannot be estimated.
Example — A Sensitivity Analysis

- No Measurement Error
- rel. ratio = 0.9
- rel. ratio = 0.8
- rel. ratio = 0.7

Additive Model Term for log(TSP)

Additive Model Term for Day

Additive Model Term for Temperature

Additive Model Term for Humidity
**Discussion**

- With the work of **CMR** and **BCR** we now have reasonably efficient estimators for nonparametric regression with measurement error.
  - SIMEX (LPR and splines) — in **CMR**
  - (Flexible) Structural splines — in **CMR**
  - Fully Bayesian (hardcore structural) — in **BCR**
With **BCR** we have a methodology that

- automatically selects the amount of smoothing
- estimates the unknown $X$’s
- allows inference that takes account of the effects of smoothing parameter selection and measurement error

Most efficient methods appear to be structural, though SIMEX may be competitive

- hardcore structural methods seem reasonably robust
END OF LECTURE 6

THANK YOU FOR YOUR ATTENTION!!!