Don't Mis-Measure Measurement Model Measurement Error

OMG Workshop

Bertrand Wilden

March 7, 2023

Measurement Models

- Used to measure latent variables: *ideology, corruption, democracy, racial resentment*

Theory-Testing Models

- Testing a causal theory: *being an ideologically extreme candidate hurts one's reelection chances*

Using Measurement Model Outputs in Theory-Testing Models

- Because measurement models are statistical constructs they contain uncertainty estimates

Using Measurement Model Outputs in Theory-Testing Models

- Because measurement models are statistical constructs they contain uncertainty estimates
- In practice, almost everyone disregards this uncertainty and uses only the *maximum a posteriori* (MAP) values from measurement models for their theory-testing models

Using Measurement Model Outputs in Theory-Testing Models

- Because measurement models are statistical constructs they contain uncertainty estimates
- In practice, almost everyone disregards this uncertainty and uses only the *maximum a posteriori* (MAP) values from measurement models for their theory-testing models
    - Mean, median, mode estimates from the measurement model

Three Problems With Ignoring Measurement Model Uncertainty

1. Uncertainty is not properly propagated in theory-testing models using MAP values - ***Coefficient uncertainty is underestimated***

Three Problems With Ignoring Measurement Model Uncertainty

1. Uncertainty is not properly propagated in theory-testing models using MAP values - ***Coefficient uncertainty is underestimated***
2. There is too much random noise in the MAP values - ***Coefficients are biased towards zero***

Three Problems With Ignoring Measurement Model Uncertainty

1. Uncertainty is not properly propagated in theory-testing models using MAP values - ***Coefficient uncertainty is underestimated***
2. There is too much random noise in the MAP values - ***Coefficients are biased towards zero***
3. Possibility of backdoor confounding through the measurement process - ***Coefficients are biased in unpredictable ways***

Measurement Model Cases

- Bayesian Improved Surname Geocoding (**BSIG**)

Measurement Model Cases

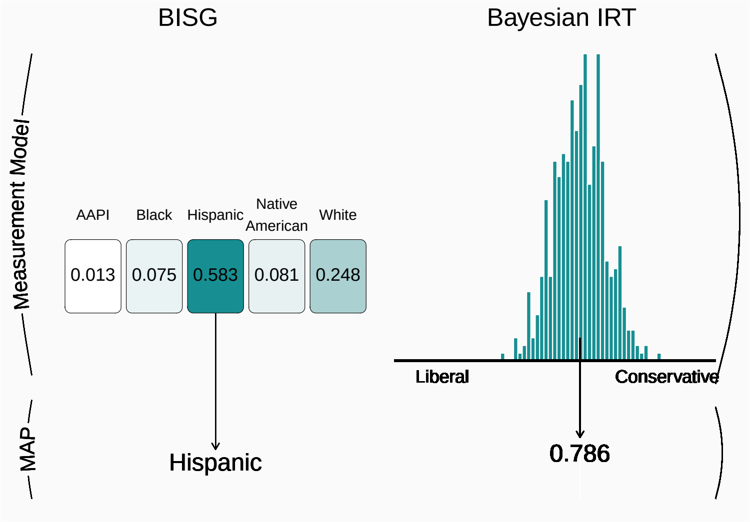
- Bayesian Improved Surname Geocoding (**BSIG**)
    - Used to measure race/ethnicity based on names and location. Measurement model is categorical

Measurement Model Cases

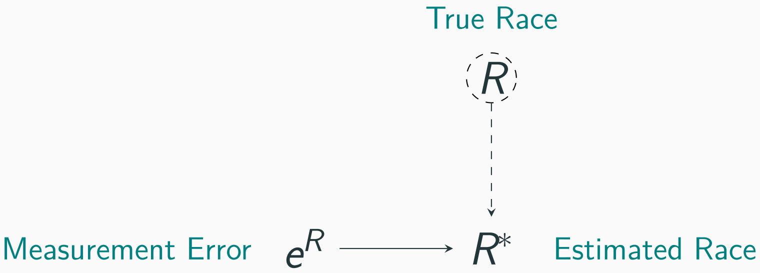
- Bayesian Improved Surname Geocoding (**BSIG**)
    - Used to measure race/ethnicity based on names and location. Measurement model is categorical
- Item-Response Theory Ideal Points (**IRT**)

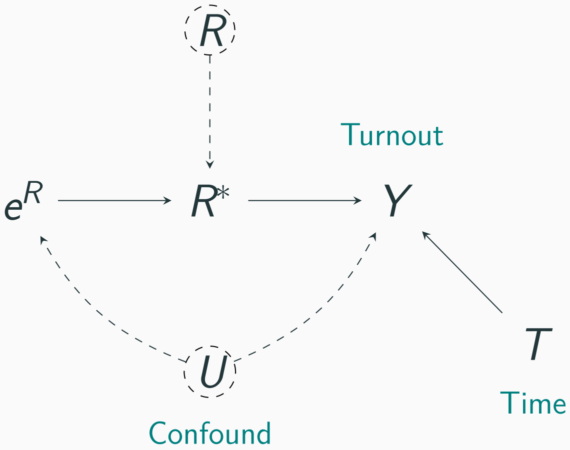
Measurement Model Cases

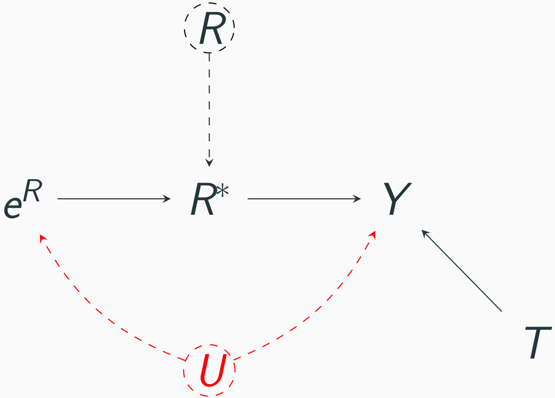
- Bayesian Improved Surname Geocoding (**BSIG**)
  - Used to measure race/ethnicity based on names and location. Measurement model is categorical
- Item-Response Theory Ideal Points (**IRT**)
  - Used to infer ideology of political actors. Measurement model is continuous


BISG
Bayesian IRT
Measurement Model
MAP
AAPI
Black
Hispanic
Native American
White
0.013
0.075
0.583
0.081
0.248
Hispanic
Liberal
Conservative
0.786

Measurement Error and Backdoor Confounding: BSIG Example


True Race
R
Measurement Error
e^R
R^*
Estimated Race


R
Turnout
e^R
R^*
Y
T
U
Confound
Time



R
e^R
R*
Y
T
U

Solution: Joint Bayesian Model

$$y_i \sim f(\pi_i)$$
$$\pi_i^* \sim g(\pi_i)$$

- y, outcome in theory-testing model
- π, latent variable
- π^*, MAP estimate of latent variable
- $f(\cdot)$, theory-testing model
- $g(\cdot)$, measurement model

Simplifications

- Applied researchers know a lot about $f(\cdot)$, but not $g(\cdot)$

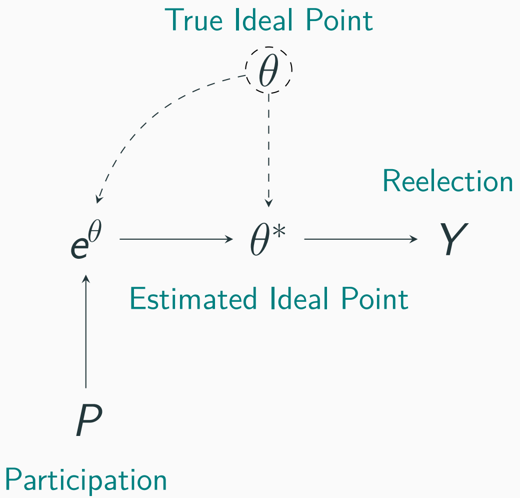
Simplifications

- Applied researchers know a lot about $f(\cdot)$, but not $g(\cdot)$
- The real $g(\cdot)$ can take a lot of computation time

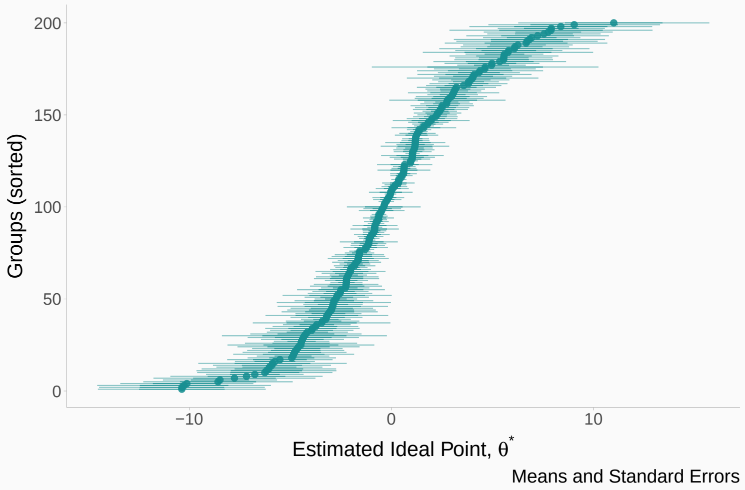
Simplifications

- Applied researchers know a lot about $f(\cdot)$, but not $g(\cdot)$
- The real $g(\cdot)$ can take a lot of computation time
- My solution is to use an approximation of $g(\cdot)$ that fits the measurement error posterior distribution

Measurement Error and Attenuation Bias: IRT Ideal Point Example

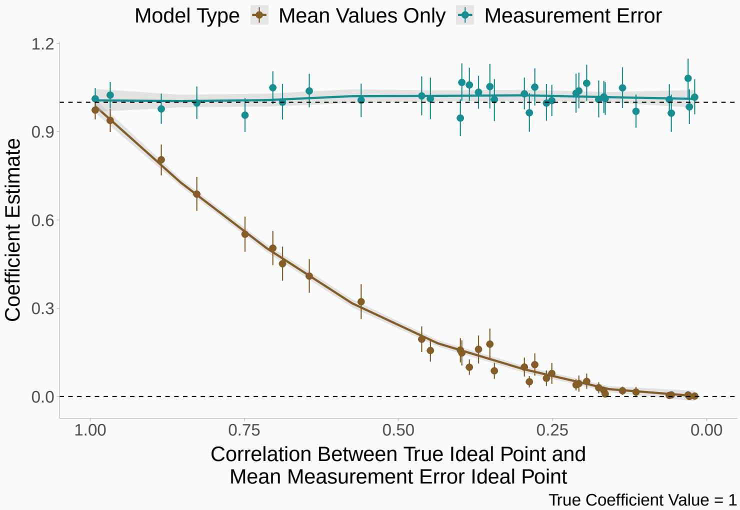

True Ideal Point
θ
Reelection
e^{θ}
θ^*
Y
Estimated Ideal Point
P
Participation

Simulated Random Normal Error



Joint Model for Normal Measurement Error

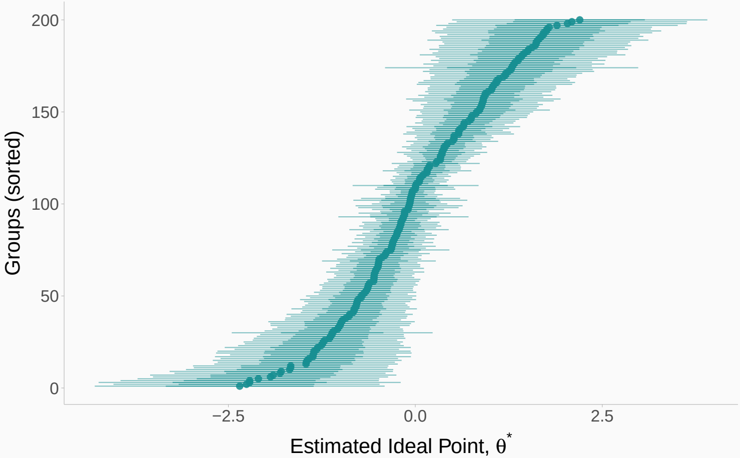
$$
\begin{aligned}
y_i &\sim \text{Normal}(\mu_i, \sigma) \\
\mu_i &= \beta_0 + \beta_1 X_{\text{TRUE,i}} \\
X_{\text{MEAS,i}} &\sim \text{Normal}(X_{\text{TRUE,i}}, X_{\text{SE,i}}) \\
X_{\text{TRUE,i}} &\sim \text{Normal}(\bar{X}_{\text{TRUE}}, \tau) \\
\beta_0, \beta_1 &\sim \text{Normal}(0, 2) \\
\sigma &\sim \text{Half Student t}(3, 0, 2) \\
\bar{X}_{\text{TRUE}} &\sim \text{Normal}(0, 1) \\
\tau &\sim \text{Half Student t}(3, 0, 2)
\end{aligned}
$$



Model Type
Mean Values Only
Measurement Error
1.2
0.9
0.6
0.3
0.0
Coefficient Estimate
1.00
0.75
0.50
0.25
0.00
Correlation Between True Ideal Point and
Mean Measurement Error Ideal Point
True Coefficient Value = 1

Skewed IRT Example

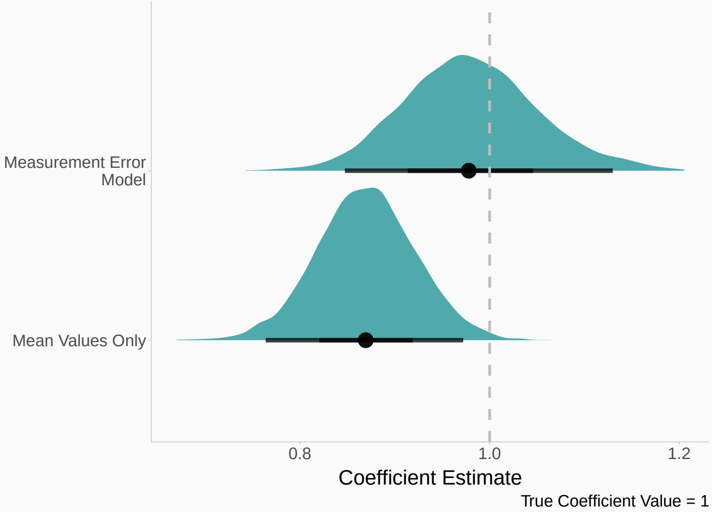
Simulated IRT Results



Means and Standard Errors

Joint Model for Skewed Measurement Error

$$
\begin{aligned}
y_i &\sim \text{Normal}(\mu_i, \sigma) \\
\mu_i &= \beta_0 + \beta_1 X_{\text{TRUE,i}} \\
X_{\text{MEAS,i}} &\sim \text{Skew Normal}(X_{\text{TRUE,i}}, X_{\text{SE,i}}, \alpha_i) \\
X_{\text{TRUE,i}} &\sim \text{Normal}(\bar{X}_{\text{TRUE}}, \tau_1) \\
\bar{X}_{\text{MEAS}} - X_{\text{MEAS,i}} &\sim \text{Normal}(\alpha_i, \tau_2) \\
\beta_0, \beta_1 &\sim \text{Normal}(0, 2) \\
\sigma &\sim \text{Half Student t}(3, 0, 2) \\
\bar{X}_{\text{TRUE}} &\sim \text{Normal}(0, 1) \\
\tau_1 &\sim \text{Half Student t}(3, 0, 2) \\
\alpha &\sim \text{Normal}(0, 1) \\
\tau_2 &\sim \text{Half Student t}(3, 0, 2)
\end{aligned}
$$



- Coefficient estimate attenuation of about 15% in realistic IRT example

Thank you!