Going Beyond Ideal Point Points: Modeling Measurement Model Measurement Error

AP Retreat

Bertrand Wilden

June 7, 2023

Overview

- Latent variables (e.g. political ideology) are often measured using statistical models
- When using these variables in downstream analysis, bias can occur if measurement uncertainty is ignored
- I develop a method for fixing this

Measurement Models

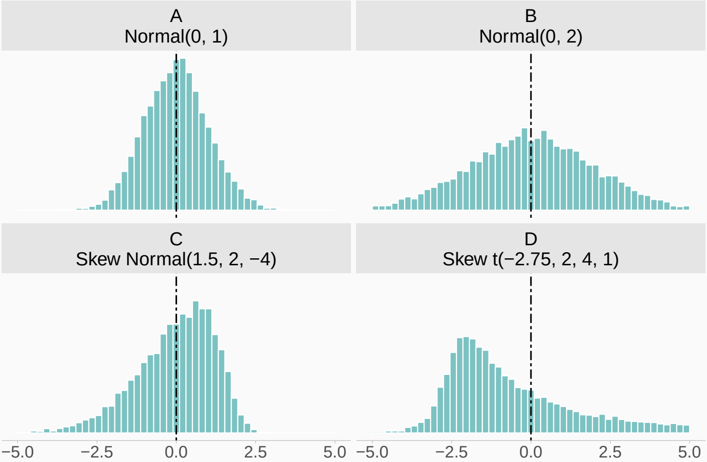
- Used to measure latent variables: *ideology, corruption, democracy, racial resentment*
- Example model: item response theory (IRT) ideal point model

Theory-Testing Models

- Testing a causal theory: *being an ideologically extreme candidate hurts one's reelection chances*
- Example model: regression with controls

Bayesian Measurement Models

- Output is a full posterior distribution of values for the latent variable according to their relative plausibility
- In practice, only a point estimate such as mean or median is used is subsequent analyses



Four different measurement model posterior distributions with mean zero

Three Problems With Ignoring Measurement Model Uncertainty

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

Three Problems With Ignoring Measurement Model Uncertainty

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

Three Problems With Ignoring Measurement Model Uncertainty

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

Solution: Joint Bayesian Measurement Error Model

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

- y, outcome in theory-testing model
- θ, latent variable
- y^*, training data in measurement model
- $f(\cdot)$, 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

Joint IRT Ideal Point Measurement Error Model

$$y_i \sim \text{Normal}(\beta_0 + \beta_1 \theta_i, \sigma^2)$$
$$y^*_{ij} \sim \text{Bernoulli}[\Phi(\gamma_j \theta_i + \xi_i)]$$

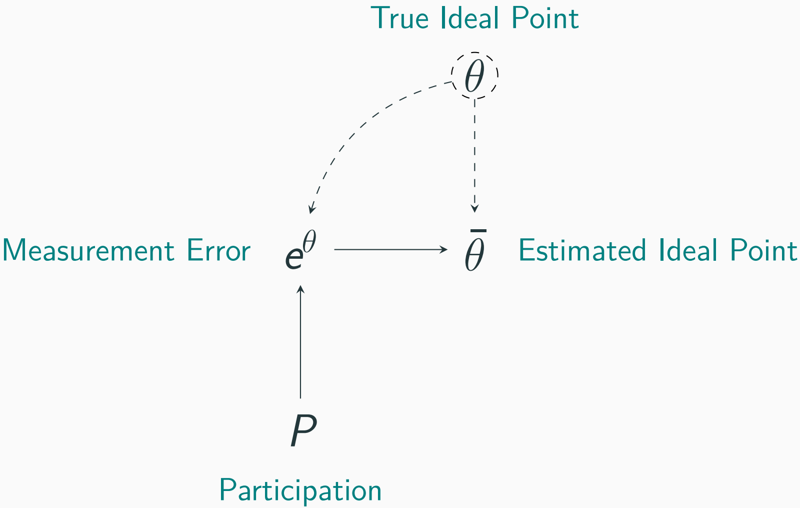
Is simplified as:

$$y_i \sim \text{Normal}(\beta_0 + \beta_1 \theta_i, \sigma^2)$$
$$\bar{\theta}_i \sim \text{Normal}(\theta_i, \sigma^2_{\theta[i]})$$

- $\bar{\theta}_i$, posterior mean from measurement model
- $\sigma^2_{\theta[i]}$, posterior variance from measurement model

Measurement Error and Attenuation Bias

Measurement Model



Theory-testing Model

$$\begin{tikzcd}
& \theta \arrow[dl, dashed, bend right] \arrow[d, dashed] & \\
e^{\theta} \arrow[r] & \bar{\theta} \arrow[r] & \overset{\text{Outcome}}{Y} \\
P \arrow[u] & &
\end{tikzcd}$$

Simulation Study

1. Generate latent variable distributions from true θ_i

Simulation Study

1. Generate latent variable distributions from true θ_i
2. Generate outcome y_i as a function of true θ_i

Simulation Study

1. Generate latent variable distributions from true θ_i
2. Generate outcome y_i as a function of true θ_i
3. Use maximum likelihood to estimate distributional parameters (mean, variance, skew) from step 1 for use in simplified measurement model

Simulation Study

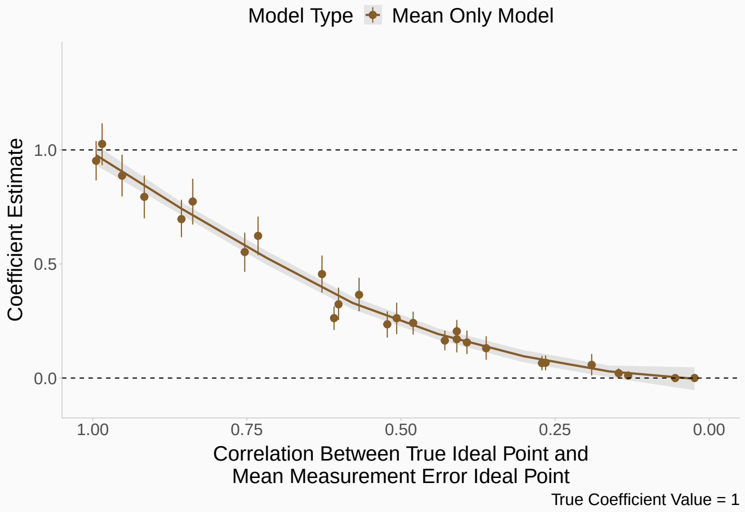
1. Generate latent variable distributions from true θ_i
2. Generate outcome y_i as a function of true θ_i
3. Use maximum likelihood to estimate distributional parameters (mean, variance, skew) from step 1 for use in simplified measurement model
4. Compare parameter recovery using a mean-only model and a joint measurement error model

Mean-Only Model

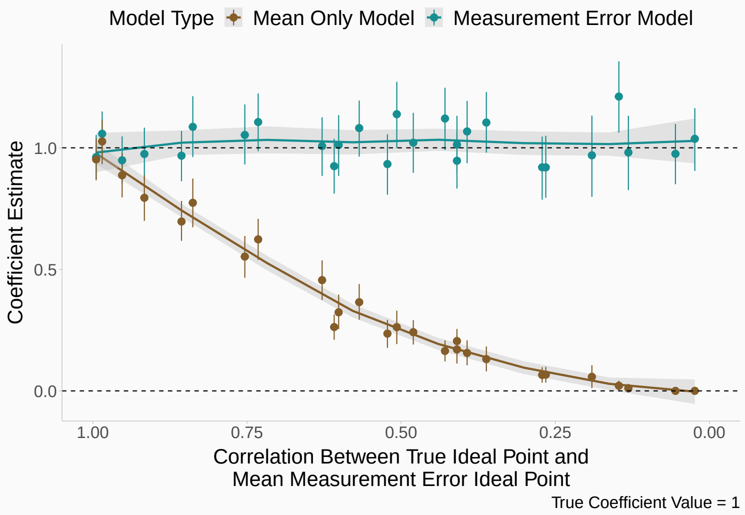
$$
\begin{aligned}
y_i &\sim \text{Normal}(\mu_i, \sigma^2) \\
\mu_i &= \beta_0 + \beta_1 \bar{\theta}_i \\
\beta_0, \beta_1 &\sim \text{Normal}(0, 2) \\
\sigma &\sim \text{Half Student t}(3, 0, 2)
\end{aligned}
$$

Joint Measurement Error Model

$$
\begin{aligned}
y_i &\sim \text{Normal}(\mu_i, \sigma^2) \\
\mu_i &= \beta_0 + \beta_1 \theta_i \\
\bar{\theta}_i &\sim \text{Normal}(\theta_i, \sigma^2_{\theta[i]}) \\
\theta_i &\sim \text{Normal}(\pi, \tau) \\
\beta_0, \beta_1 &\sim \text{Normal}(0, 2) \\
\sigma &\sim \text{Half Student t}(3, 0, 2) \\
\tau &\sim \text{Half Student t}(3, 0, 2) \\
\pi &\sim \text{Normal}(0, 1)
\end{aligned}
$$



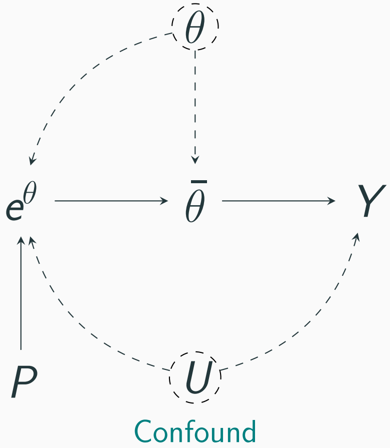
Model Type
Mean Only Model
Coefficient Estimate
1.0
0.5
0.0
1.00
0.75
0.50
0.25
0.00
Correlation Between True Ideal Point and
Mean Measurement Error Ideal Point
True Coefficient Value = 1



Model Type
Mean Only Model
Measurement Error Model
Coefficient Estimate
1.0
0.5
0.0
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

Measurement Error and Confounding Bias

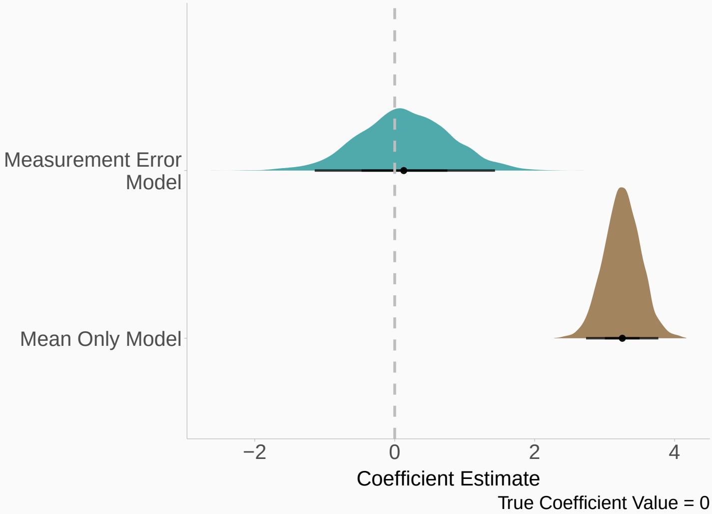
Theory-testing Model



Skew-Normal Joint Measurement Error Model

$$
\begin{aligned}
y_i &\sim \text{Normal}(\mu_i, \sigma^2) \\
\mu_i &= \beta_0 + \beta_1 \theta_i \\
\bar{\theta}_i &\sim \text{Skew Normal}(\theta_i, \omega_{\theta[i]}, \alpha_{\theta[i]}) \\
\theta_i &\sim \text{Normal}(\pi, \tau) \\
\beta_0, \beta_1 &\sim \text{Normal}(0, 2) \\
\sigma &\sim \text{Half Student t}(3, 0, 2) \\
\tau &\sim \text{Half Student t}(3, 0, 2) \\
\pi &\sim \text{Normal}(0, 1)
\end{aligned}
$$

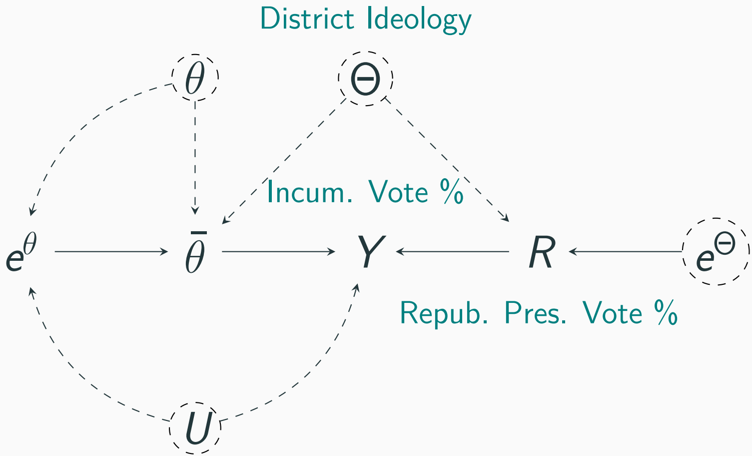
Skew-Normal Measurement Error Model Comparison



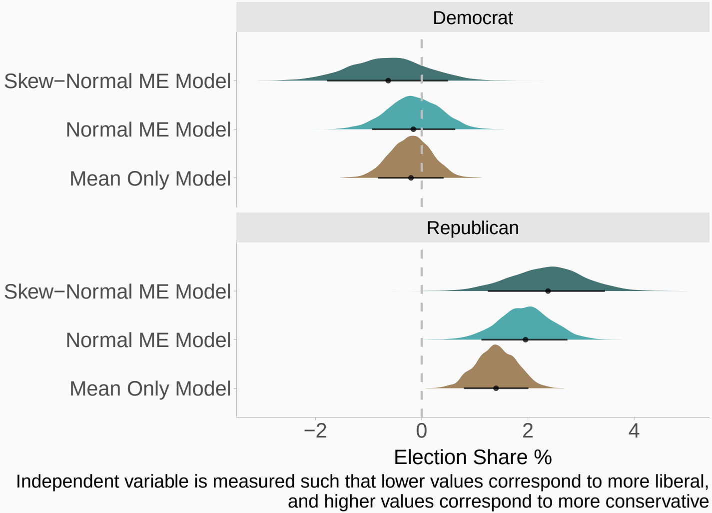
Case Study: Ideological Extremism and Electoral Success

Research Question: are ideologically extreme US House incumbents punished electorally?

- DV: General election vote share
- IV: Ideology estimated from previous Congress roll-call votes
- Data from 1990 to 2016

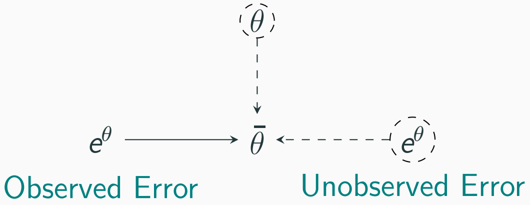

District Ideology
θ
Θ
Incum. Vote %
e^{θ}
$\bar{\theta}$
Y
R
e^{Θ}
Repub. Pres. Vote %
U

Effect of Incumbent Ideology on General Election Vote Share



Concluding Thoughts

- Uncertainty estimation matters
  - Better to use measurement models that provide uncertainty estimates than those that don't (eg NOMINATE)
  - Better to use state-of-the-art posterior sampling methods such as Hamiltonian Monte Carlo, rather than traditional Gibbs samplers for measurement models



Thank you!