Measurement Error in Police Recorded Crime Rates and Sensitivity Analysis

Jose Pina-Sánchez

David Buil-Gil

Errors in Police Recorded Crime Rates

Errors in Police Recorded Crime Rates

- We hypothesise different measurement error mechanisms:

Errors in Police Recorded Crime Rates

- We hypothesise different measurement error mechanisms:
  - *systematic*, since not all crime is reported to the police
  - *random*, subject to variability across officers and forces, as a result of the different recording practices
  - *multiplicative*, errors proportional to the true extent of crime in the area
  - $Y^* = Y \cdot U$

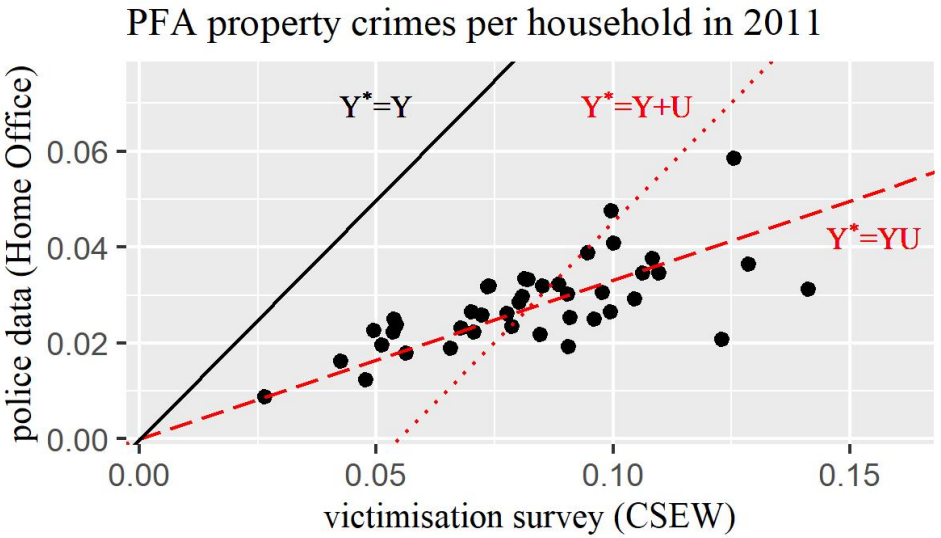
    $U \sim N((0,1), \sigma)$

Errors in Police Recorded Crime Rates

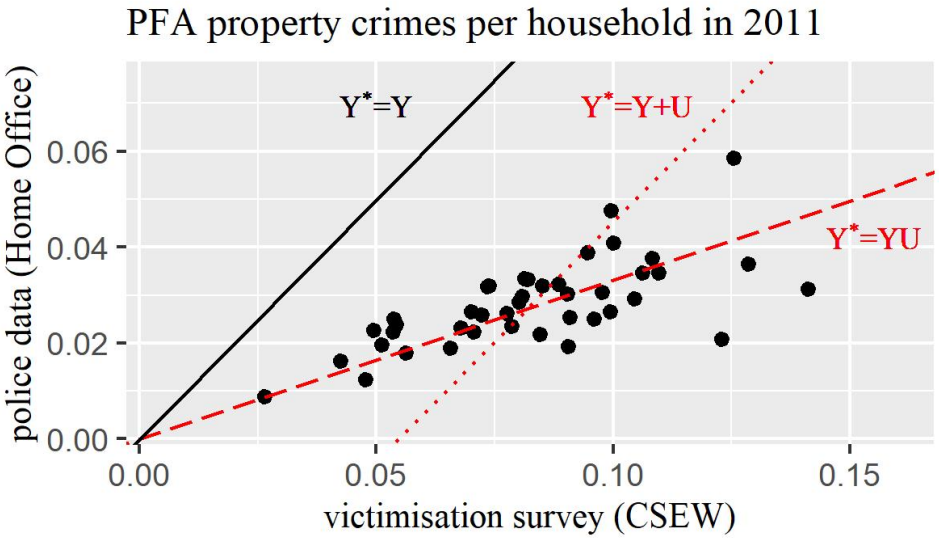
- We hypothesise different measurement error mechanisms:
  - *systematic*, since not all crime is reported to the police
  - *random*, subject to variability across officers and forces, as a result of the different recording practices
  - *multiplicative*, errors proportional to the true extent of crime in the area
  - $Y^* = Y \cdot U$

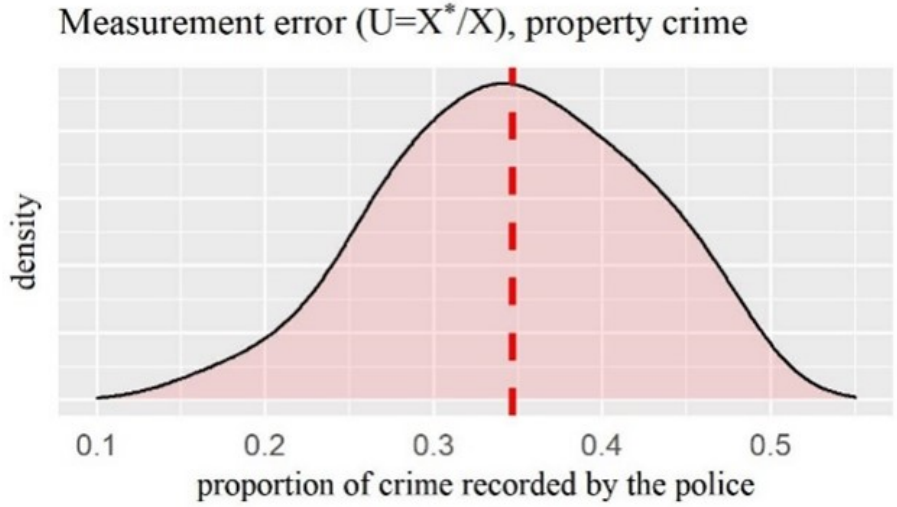
    $U \sim N((0,1), \sigma)$
- To examine this proposition we compare police data (data.police.uk) and crime survey (CSEW) estimates

Measurement Error in Crime Rates



Measurement Error in Crime Rates





Tractable Impact

- In some simple settings, we can anticipate its impact algebraically, and adjust accordingly
- e.g. a linear model exploring causes of crime

Tractable Impact

- In some simple settings, we can anticipate its impact algebraically, and adjust accordingly
- e.g. a linear model exploring causes of crime
  - $Y^* = \alpha + \beta_1 X_1 + \beta_2 X_2 + \epsilon$
  - where, $Y^* = Y \cdot U$, and $U \sim N((0,1), \sigma)$

Tractable Impact

- In some simple settings, we can anticipate its impact algebraically, and adjust accordingly
- e.g. a linear model exploring causes of crime
  - $Y^* = \alpha + \beta_1 X_1 + \beta_2 X_2 + \epsilon$
  - where, $Y^* = Y \cdot U$, and $U \sim N((0,1), \sigma)$
  - then, $Y = \dfrac{\alpha + \beta_1 X_1 + \beta_2 X_2 + \epsilon}{U}$
  - regression coefficients biased upwards proportionally to the under-recording rate, i.e. overdimensioned effect sizes

Hard to Anticipate Impact

- Much harder to predict when crime rates are used as a covariate
  - systematic and random errors can reinforce or oppose each other
  - how they are associated with other explanatory variables affect the direction and magnitude of the bias
  - Pina-Sánchez, J., Buil-Gil, D., Brunton-Smith, I., and Cernat, A. (2022). The impact of measurement error in models using police recorded crime rates. *Journal of Quantitative Criminology*, 39(4):975–1002.

Hard to Anticipate Impact

- Much harder to predict when crime rates are used as a covariate
  - systematic and random errors can reinforce or oppose each other
  - how they are associated with other explanatory variables affect the direction and magnitude of the bias
  - Pina-Sánchez, J., Buil-Gil, D., Brunton-Smith, I., and Cernat, A. (2022). The impact of measurement error in models using police recorded crime rates. *Journal of Quantitative Criminology*, 39(4):975–1002.
- Also, the measurement error itself is often correlated with the focal or outcome variable (non-differential errors)
  - e.g.1 higher reporting rates where police is seen as legitimate
  - e.g.2 lower recording in high-rise housing areas

Hard to Anticipate Impact

- Much harder to predict when crime rates are used as a covariate
  - systematic and random errors can reinforce or oppose each other
  - how they are associated with other explanatory variables affect the direction and magnitude of the bias
  - Pina-Sánchez, J., Buil-Gil, D., Brunton-Smith, I., and Cernat, A. (2022). The impact of measurement error in models using police recorded crime rates. *Journal of Quantitative Criminology*, 39(4):975–1002.
- Also, the measurement error itself is often correlated with the focal or outcome variable (non-differential errors)
  - e.g.1 higher reporting rates where police is seen as legitimate
  - e.g.2 lower recording in high-rise housing areas
- We must conduct sensitivity analysis

Recounting Crime with Measurement Error *(RCME)*

- An R package based on the simulation of measurement error
  - Pina-Sánchez, J., Brunton-Smith, I., Cernat, A., and Buil-Gil, D. (2023). Exploring the impact of measurement error in police recorded crime rates through sensitivity analysis. Crime Science, 12(14):1–8
- Easy to use and flexible (applicable to a range of outcome model and forms of measurement error)

Recounting Crime with Measurement Error *(RCME)*

- An R package based on the simulation of measurement error
  - Pina-Sánchez, J., Brunton-Smith, I., Cernat, A., and Buil-Gil, D. (2023). Exploring the impact of measurement error in police recorded crime rates through sensitivity analysis. Crime Science, 12(14):1–8
- Easy to use and flexible (applicable to a range of outcome model and forms of measurement error)
  - assumes multiplicative errors, but allows the user to define ranges of...
  - systematic errors (S), i.e. under-recording
  - and differential errors (D), i.e. the correlation between measurement error and the variable of interest

Example of Sensitivity Analysis

- A linear model examining the effect of collective efficacy on criminal damage rates, controlling for a few area characteristics
  - $log(Y^*) = \alpha + \beta_1 X + \beta_k Z_k + \epsilon$
  - population: London's 982 MSOAs

Example of Sensitivity Analysis

- A linear model examining the effect of collective efficacy on criminal damage rates, controlling for a few area characteristics
  - $log(Y^*) = \alpha + \beta_1 X + \beta_k Z_k + \epsilon$
  - population: London's 982 MSOAs
- We estimate it first using police data (naive model), and we re-estimate it using adjusted crime rates (sensitivity analysis)
  - the adjusted crime rates: Y^*/U
  - the measurement error term: $U = S + D \cdot X$
  - we need to consider values for S (systematic error) and D (differential errors)

Adjusting Crime Rates

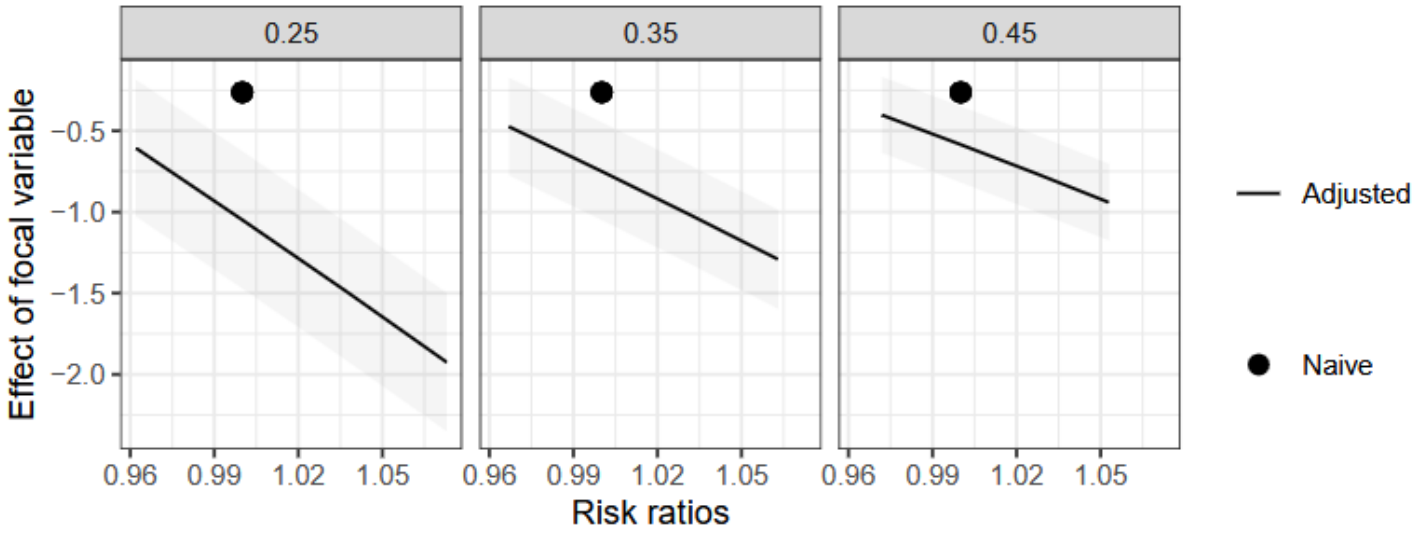
- We consider a range of parameters to simulate the measurement error
  - based on the CSEW and substantive knowledge of the subject
  - recording rates: $(0.25, 0.35, 0.45)$
  - differential errors (as ORs): $(0.95, 1, 1.05, 1.10)$

Adjusting Crime Rates

- We consider a range of parameters to simulate the measurement error
  - based on the CSEW and substantive knowledge of the subject
  - recording rates: $(0.25, 0.35, 0.45)$
  - differential errors (as ORs): $(0.95, 1, 1.05, 1.10)$

```
robust_collective = rcme_out(
  formula = "crime ~ collective + unemployment + white + age",
  data = data,
  focal_variable = "collective",
  R = c(0.25, 0.35, 0.45),
  D = c(0.95, 1, 1.05, 1.10))
```

The *Recounting Crime* Adjustment



- The effect size attributed to collective efficacy appears robust
  - the crime reduction effect of collective is severely under-estimated
  - the magnitude of that bias proportional to the extent of under-recording and the differential errors
  - even if we consider recording rates to be lower in areas where collective efficacy is higher, its effect remains underestimated

RCME

- If you want to know more:
  - https://github.com/RecountingCrime/rcme
  - RCME tutorial (Brunton-Smith et al., 2024)
  - Adjusting systematic and differential errors using simulations (Pina-Sánchez et al., 2024)

RCME

- If you want to know more:
  - https://github.com/RecountingCrime/rcme
  - RCME tutorial (Brunton-Smith et al., 2024)
  - Adjusting systematic and differential errors using simulations (Pina-Sánchez et al., 2024)
- *RCME*'s downsides
  - does not consider random errors
  - does not handle 0s properly

RCME

- If you want to know more:
  - https://github.com/RecountingCrime/rcme
  - RCME tutorial (Brunton-Smith et al., 2024)
  - Adjusting systematic and differential errors using simulations (Pina-Sánchez et al., 2024)
- *RCME*'s downsides
  - does not consider random errors
  - does not handle 0s properly
- With Bayesian statistics we can overcome those problems
  - can be used for any model, and adjust for any form of measurement error

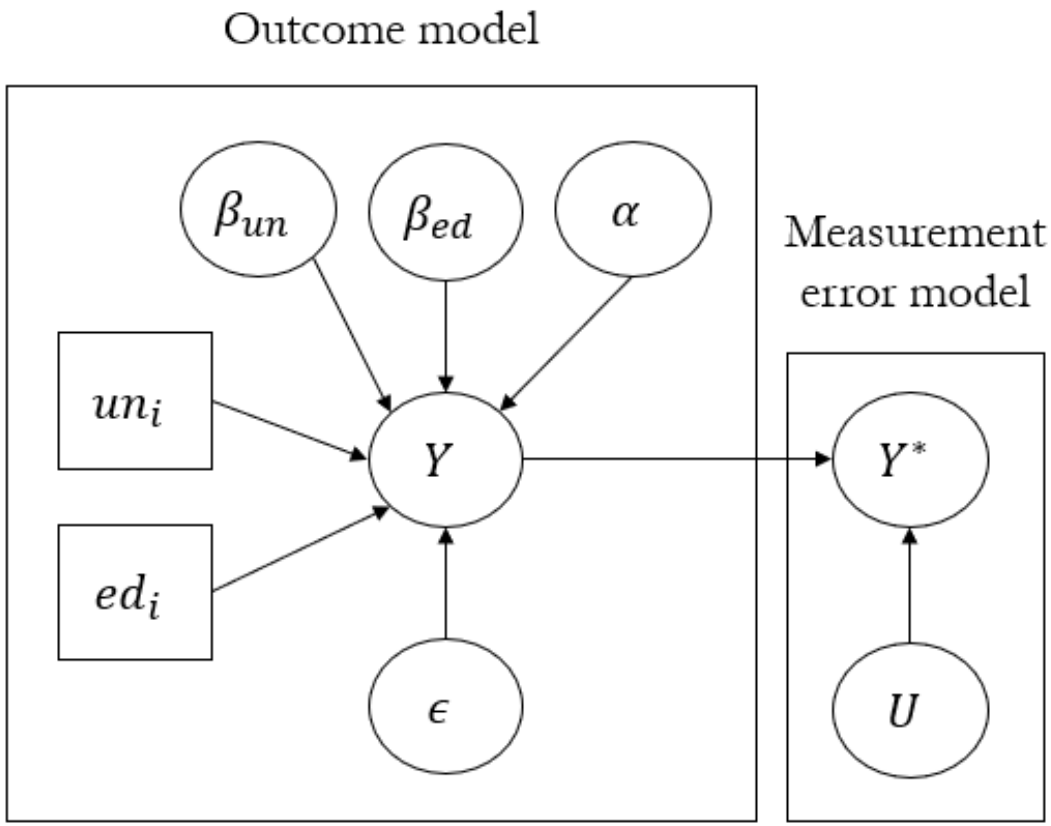
Bayesian Adjustments

- We specify an outcome and a measurement model
  - the former reflects the substantive relationship that we want to estimate
  - the latter any form of measurement error that might be affecting any of our variables

Bayesian Adjustments

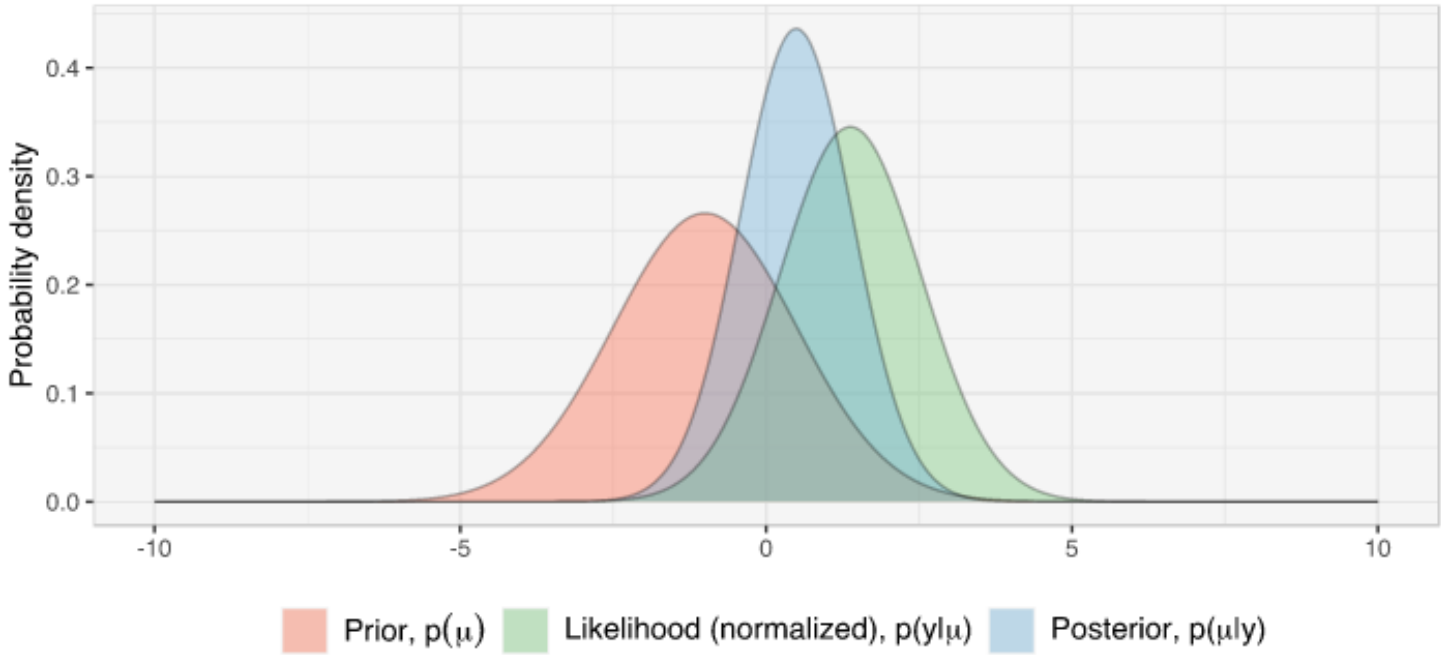
- We specify an outcome and a measurement model
  - the former reflects the substantive relationship that we want to estimate
  - the latter any form of measurement error that might be affecting any of our variables
- These two models are estimated simultaneously
  - using Markov chain Monte Carlo (MCMC) estimation methods
  - we obtain a 'posterior distribution' for each estimate included in our models
  - reflecting the probability distribution of an estimate given: i) the models we are using, ii) the data that we observe, and iii) any prior knowledge we might want to include

Bayesian Adjustments



Informative Priors

The *posterior distribution* as a combination of the *likelihood function* and *prior distribution*



Source: Bolstad (2018)

Bayesian Adjustments: Final Thoughts

- Main downsides:
  - the method with the steepest learning curve
  - computationally intensive

Bayesian Adjustments: Final Thoughts

- Main downsides:
  - the method with the steepest learning curve
  - computationally intensive
- However,
  - software becoming faster and more user friendly
  - I use the *rjags* package to call JAGS from R

Bayesian Adjustments: Final Thoughts

- Main downsides:
  - the method with the steepest learning curve
  - computationally intensive
- However,
  - software becoming faster and more user friendly
  - I use the *rjags* package to call JAGS from R
- To know more:
  - Gustafson, P. (2003). Measurement error and misclassification in statistics and epidemiology: impacts and Bayesian adjustments. Chapman and Hall/CRC.
  - Pina-Sánchez, J., & Varela, A. (2024). Adjusting for Measurement Error in Police Recorded Crime Counts Using Bayesian Statistics. CrimRxiv

In Conclusion

In Conclusion

- The type of measurement error observed in police data can lead to strong biases when used in regression models
  - the validity of much of the literature is under question

In Conclusion

- The type of measurement error observed in police data can lead to strong biases when used in regression models
  - the validity of much of the literature is under question
- Not always possible to anticipate the specific impact

In Conclusion

- The type of measurement error observed in police data can lead to strong biases when used in regression models
  - the validity of much of the literature is under question
- Not always possible to anticipate the specific impact
- Sensitivity analysis should be widely employed
  - the *RCME* package can help with that
  - for everything else, try Bayesian adjustments