Crime Surveys User Conference 2020

The impact of measurement error in police crime records

Recounting Crime

Jose Pina-Sánchez (University of Leeds)
Ian Brunton-Smith (University of Surrey)
David Buil-Gil (University of Manchester)
Alexandru Cernat (University of Manchester)

The Problem

- We are all aware that police recorded crime is deeply flawed
  - Under-reporting/under-detection of crime
  - Recording inconsistencies across forces

The Problem

- We are all aware that police recorded crime is deeply flawed
  - Under-reporting/under-detection of crime
  - Recording inconsistencies across forces
- Yet, we still use police recorded crime rates in our research
  - Key variable to assess the causes and consequences of crime
  - Multivariate models based on police data will likely be biased

The Problem

- We are all aware that police recorded crime is deeply flawed
  - Under-reporting/under-detection of crime
  - Recording inconsistencies across forces
- Yet, we still use police recorded crime rates in our research
  - Key variable to assess the causes and consequences of crime
  - Multivariate models based on police data will likely be biased
- In this paper (and project) we seek to tackle this problem
  - Identify the nature and prevalence of measurement error in police data
  - Illustrate the impact it has when used in regression models
  - Suggest methods for its adjustment

Measurement Error in Police Data

- We focus on crime counts/rates
  - We compare police data (data.police.uk) and CSEW estimates of acquisitive crime at the Police Force Area level per year
  - Assuming the latter is a gold standard

Measurement Error in Police Data

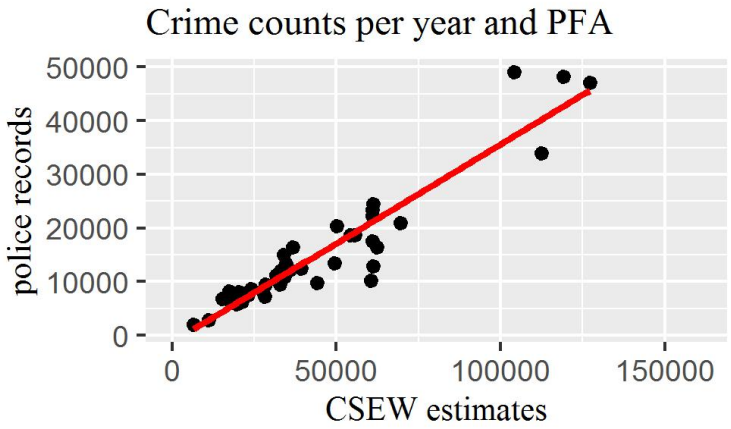
- We focus on crime counts/rates
  - We compare police data (data.police.uk) and CSEW estimates of acquisitive crime at the Police Force Area level per year
  - Assuming the latter is a gold standard
- To estimate the prevalence of the measurement error first we need to consider its form
  - We can anticipate systematic (under-reporting/under-detection) and random (inconsistencies across forces) errors
  - And that these errors are multiplicative (proportional to the true value)

Measurement Error in Police Data

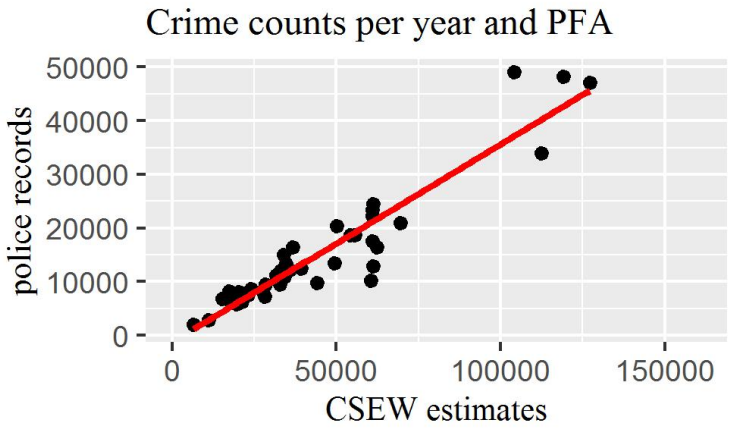
- We focus on crime counts/rates
  - We compare police data (data.police.uk) and CSEW estimates of acquisitive crime at the Police Force Area level per year
  - Assuming the latter is a gold standard
- To estimate the prevalence of the measurement error first we need to consider its form
  - We can anticipate systematic (under-reporting/under-detection) and random (inconsistencies across forces) errors
  - And that these errors are multiplicative (proportional to the true value)

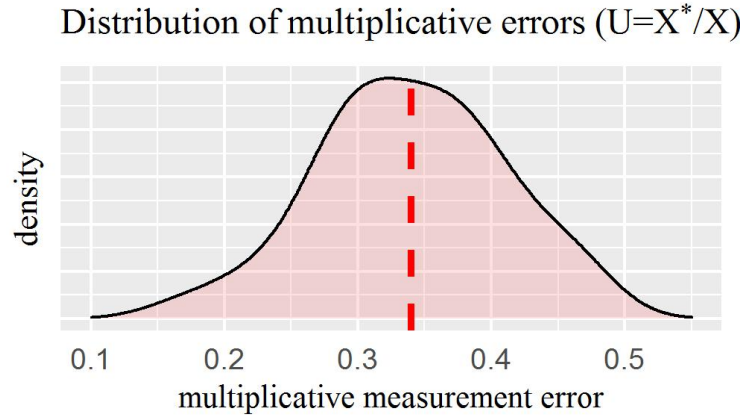
    multiplicative error model: $X^* = X \cdot U$, where $U \sim N(\in (0,1), \sigma)$

Measurement Error in Police Data



Measurement Error in Police Data





Impact of Measurement Error

- Let's consider a simple linear model, where the explanatory variable is affected by measurement error

$$Y = \alpha + \beta X^* + \epsilon$$

Impact of Measurement Error

- Let's consider a simple linear model, where the explanatory variable is affected by measurement error

$$Y = \alpha + \beta X^* + \epsilon$$

- Using OLS we can estimate α and β solving the following system of equations

$$\hat{\alpha}^* = \bar{Y} - \hat{\beta}\bar{X}^*$$

$$\hat{\beta}^* = \frac{cov(X^*Y)}{var(X^*)}$$

Impact on the Slope

- Let's focus on the slope since this is often what we are after

$$\hat{\beta}^* = \frac{cov(X^*Y)}{var(X^*)}$$

Impact on the Slope

- Let's focus on the slope since this is often what we are after

$$\hat{\beta}^* = \frac{cov(X^*Y)}{var(X^*)}$$

1. Random noise in X^* doesn't affect cov, but increases $var \rightarrow$ attenuates the slope

Impact on the Slope

- Let's focus on the slope since this is often what we are after

$$\hat{\beta}^* = \frac{cov(X^*Y)}{var(X^*)}$$

1. Random noise in X^* doesn't affect cov, but increases $var \rightarrow$ attenuates the slope
2. Systematic multiplicative error will lead to a change of scale, which affects the cov and especially $var \rightarrow$ the slope will be biased; augmented if the errors are negative

Impact on the Slope

- Let's focus on the slope since this is often what we are after

$$\hat{\beta}^* = \frac{cov(X^*Y)}{var(X^*)}$$

1. Random noise in X^* doesn't affect cov, but increases $var \rightarrow$ attenuates the slope
2. Systematic multiplicative error will lead to a change of scale, which affects the cov and especially $var \rightarrow$ the slope will be biased; augmented if the errors are negative

- Under-reporting will bias the slope upwards, while inconsistencies across PFAs will push it downwards

Adjustments

- Things become more complicated when we move away from simple linear regression
  - The errors in crime rates could affect bias the slopes of other variables included in the model
  - Harder to trace out if using non-linear models
  - And we have not even consider how it affects measures of uncertainty too

Adjustments

- Things become more complicated when we move away from simple linear regression
  - The errors in crime rates could affect bias the slopes of other variables included in the model
  - Harder to trace out if using non-linear models
  - And we have not even consider how it affects measures of uncertainty too
- We need to employ adjustment methods
  - Bayesian adjustments are a good option - very flexible
  - We specify our outcome model of interest
  - And a measurement model for the variable affected by measurement error
  - Based on the estimated validity and reliability of that variable

Example of Impact and Adjustment

- We model the effect of acquisitive crime (X_1) and population density (X_2) on the % of white population (Y)
  - We do that in three steps

Example of Impact and Adjustment

- We model the effect of acquisitive crime (X_1) and population density (X_2) on the % of white population (Y)
  - We do that in three steps
- The - assumed - true model, using CSEW data
  - $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \epsilon$

Example of Impact and Adjustment

- We model the effect of acquisitive crime (X_1) and population density (X_2) on the % of white population (Y)
  - We do that in three steps
- The - assumed - true model, using CSEW data
  - $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \epsilon$
- The naive model, using police data
  - $Y = \alpha + \beta_1 X_1^* + \beta_2 X_2 + \epsilon$

Example of Impact and Adjustment

- We model the effect of acquisitive crime (X_1) and population density (X_2) on the % of white population (Y)
  - We do that in three steps
- The - assumed - true model, using CSEW data
  - $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \epsilon$
- The naive model, using police data
  - $Y = \alpha + \beta_1 X_1^* + \beta_2 X_2 + \epsilon$
- The Bayesian adjustment, using police data and what we know about the validity and reliability of the measurement error
  - $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \epsilon$
    $X_1 = X_1^* \cdot U$, where $U \sim N(1/0.34, 1/0.07)$

Example Using CSEW and Police Data

	outcome variable: % white in the area		
	true model	naive model	adjusted model
constant	0.977		
	(0.013)		
acquisitive crime	−0.073		
	(0.029)		
population density	−0.003		
	(0.002)		
observations	40	40	40

Example Using CSEW and Police Data

	outcome variable: % white in the area		
	true model	naive model	adjusted model
constant	0.977	0.976	
	(0.013)	(0.012)	
acquisitive crime	−0.073	−0.276	
	(0.029)	(0.095)	
population density	−0.003	−0.003	
	(0.002)	(0.002)	
observations	40	40	40

Example Using CSEW and Police Data

	outcome variable: % white in the area		
	true model	naive model	adjusted model
constant	0.977	0.976	0.976
	(0.013)	(0.012)	(0.013)
acquisitive crime	−0.073	−0.276	−0.081
	(0.029)	(0.095)	(0.030)
population density	−0.003	−0.003	−0.002
	(0.002)	(0.002)	(0.002)
observations	40	40	40

Discussion

- The type of measurement error observed in crime rates can be defined as
  - Multiplicative, with a strong negative systematic component, normally distributed across police forces

Discussion

- The type of measurement error observed in crime rates can be defined as
  - Multiplicative, with a strong negative systematic component, normally distributed across police forces
- These type of errors can lead to strong biases when used in regression models
  - The validity of much of the literature relying on such data is under question

Discussion

- The type of measurement error observed in crime rates can be defined as
  - Multiplicative, with a strong negative systematic component, normally distributed across police forces
- These type of errors can lead to strong biases when used in regression models
  - The validity of much of the literature relying on such data is under question
- Bayesian adjustments or similar can help
  - Very flexible, can be used with any kind of outcome model, and form of measurement error
  - Can be used as a sensitivity tool when all we have is an educated guess of the validity and reliability of police data

Discussion

- The type of measurement error observed in crime rates can be defined as
  - Multiplicative, with a strong negative systematic component, normally distributed across police forces
- These type of errors can lead to strong biases when used in regression models
  - The validity of much of the literature relying on such data is under question
- Bayesian adjustments or similar can help
  - Very flexible, can be used with any kind of outcome model, and form of measurement error
  - Can be used as a sensitivity tool when all we have is an educated guess of the validity and reliability of police data
- Next steps
  - Small area estimation, synthetic data, multi-trait multi-method models
  - `http://recountingcrime.com`