On The Use of Empirical Bayes for Comparative Interrupted Time Series with an Application to Mandatory Helmet Legislation

Jake Olivier Joanna J.J. Wang Scott Walter Raphael Grzebieta

University of New South Wales

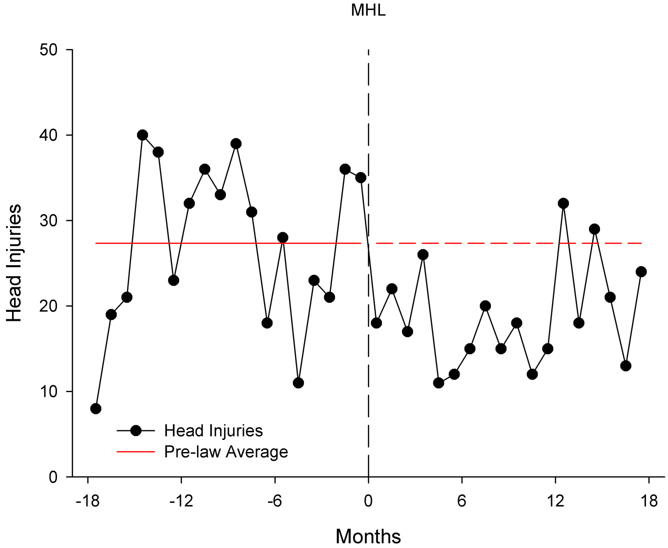
August 2013

Outline

Mandatory bicycle helmet legislation in NSW

- Intervention directed at increasing helmet wearing among cyclists
  - $\rightarrow$ Lower bicycle related head injuries
  - Not a panacea for all bicycle related injuries
- Applies to all age groups
- Went into effect in two stages
  - Adults ($>$16): 1 January 1991
  - Children: 1 July 1991
- Led to greater helmet wearing rates (~25% to ~80%)
- Associated with fewer bicycle related head injuries

Adult head injury hospitalisations in NSW



Criticisms of MHL

- MHL is very controversial
- Leads to reductions in cycling?
  - Fewer cyclists → fewer bicycle related head injuries?
- Leads to increased risk to cyclists
  - Risk compensation, rotational injuries, safety in numbers?
- Has a negative health economics impact?
  - Quit cycling → no exercise → more obesity?
  - Morbidity/mortality from obesity outweighs safety benefit of helmets?
- Loss of freedom?
- Debate rages on after 20+ years
- The anti-helmet advocacy group Bicycle Helmet Research Foundation is the main proponent of these criticisms[1]

[1] www.cyclehelmets.org

BIG QUESTIONS

Question 1

Is the drop in head injury associated solely, partly or not at all with the helmet law?

Question 2

Did the helmet law **CAUSE** the drop in head injury? (via increased helmet wearing)

Question 3

Did declines in cycling **CAUSE** the drop in head injury?

Causal Inference for Population-based Interventions

- Pre- and post-intervention periods are not randomised

$\Rightarrow$**Causal inference is difficult**

- Relevant data is often missing
  - cycling exposure, risk of injury
- Routinely collected data is probably best option for assessment
  - hospitalisation data, census data, police data (traffic, criminal reports)
- A rigorous analysis is paramount
  - There are many examples where different analyses result in different conclusions[2]
- What is the **best** analytic method/framework?
  - Interrupted time series is most common

[2] Ramsay et al. (2003) "Interrupted time series designs in health technology assessment: Lessons learned from two systematic reviews of behavior change strategies."

Outline

1. Motivating Example
2. Interrupted Time Series
3. Empirical Bayes ITS
4. Results
5. Discussion

Interrupted Time Series (ITS)

- Type of quasi-experimental design
  - Participants are not randomised
- Estimates a time series before and after an intervention
  - Comparing single pre- and post-intervention effects can hide *history*
  - Multiple pre- and post-intervention observations avoids *regression to the mean*
- Important comparisons made between pre- and post-intervention time series
  - Change in level (immediate impact)
  - Change in slope (gradual impact)

Scary maths

Interrupted time series (basic structural model)

$$y_t = \mu_t + \gamma_t + \sum_{j=1}^{k} \delta_j x_{jt} + \lambda w_t + \varepsilon_t$$

μ_t := trend
γ_t := seasonal component
x_{jt} := jth explanatory variable
δ_j := coefficient for x_{jt}
λ := intervention effect
w_t := pre/post-law indicator
ε_t := irregular component

Not so scary maths

Effects are additive

- Outcome is comprised of

$$\left(\begin{array}{c}\text{basic}\\\text{pattern}\end{array}\right)+\left(\begin{array}{c}\text{cyclical}\\\text{effects}\end{array}\right)+\left(\begin{array}{c}\text{other}\\\text{effects}\end{array}\right)+\left(\begin{array}{c}\text{law}\\\text{effects}\end{array}\right)+\left(\begin{array}{c}\text{random}\\\text{noise}\end{array}\right)$$

Simple ITS

$$\log(y_T) = \beta_0 + \beta_1 T + \beta_2 I + \beta_3 TI + u_T$$

where

$T :=$ time

$I := \begin{cases} 0 & \text{pre-intervention} \\ 1 & \text{post-intervention} \end{cases}$

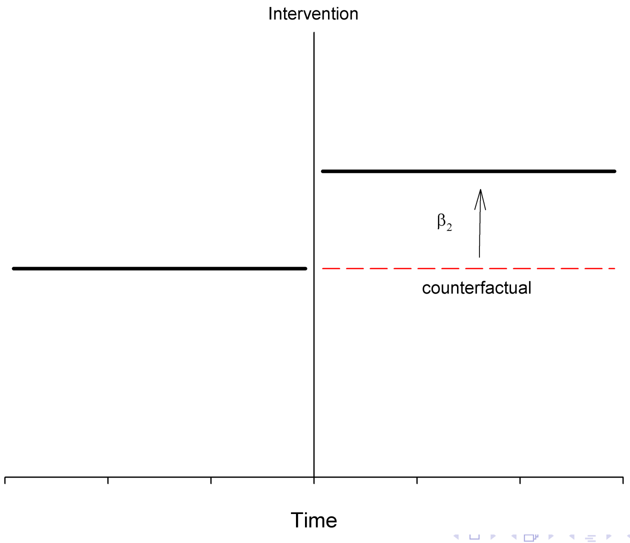
$u_T :=$ error process (time dependent?)

- Could also include cyclical effects or other (confounding) variables
- **Counterfactual** (or trajectory) is the estimated time series if the intervention had not occurred, for example

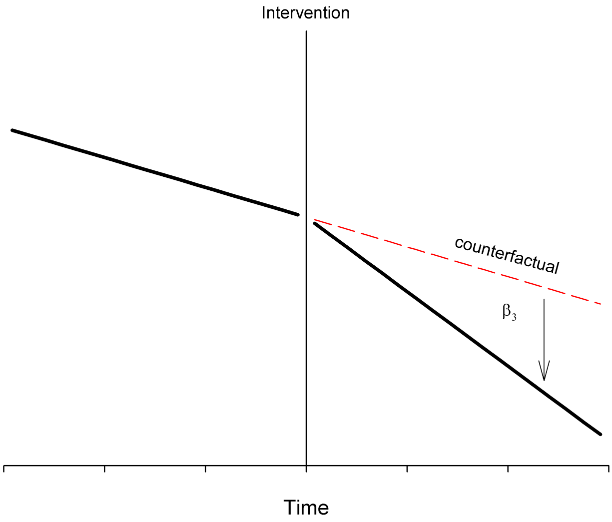
$$\log(\hat{y}_T) = \hat{\beta}_0 + \hat{\beta}_1 T$$

- β_2 and β_3 are comparisons between the counterfactual and the post-intervention model

Change in Level



Change in Slope



Threat to Internal Validity

- Unmeasured confounding is a major weakness of ITS
- The use of a control/comparator time series is often recommended[3]
  - Also affected by unmeasured confounding
  - Not subject to the intervention
  - Observations over the same time period
  - Could be a related observation from the same study unit
- Treatment and control are modelled simultaneously
  - Comparative interrupted time series (CITS)

---

[3] Shadish, Cook & Campbell (2002) *Experimental and Quasi-experimental designs for generalized causal inference.*

Comparative ITS

$$\log(y_T) = \beta_0 + \beta_1 T + \beta_2 I + \beta_3 C + \beta_4 TI + \beta_5 TC + \beta_6 IC + \beta TIC_7 + u_T$$

where

$$C := \begin{cases} 1 & \text{primary time series} \\ 0 & \text{comparative time series} \end{cases}$$

$u_T :=$ error process (time dependent?)

- The comparison of the two times series is

$$\log(y_T^p / y_T^c) = (\beta_3 + \beta_6 I) + (\beta_4 + \beta_7 I)\, T$$

- β_6 and β_7 are comparisons between the counterfactual and the post-intervention model **relative to the comparative time series**
- Assumes unmeasured confounding factors are identical for y_T^p and y_T^c and therefore cancel out

How to Choose a Comparative Time Series?

Question 4

How do we know whether a comparative time series has accounted for unmeasured confounding?

Question 5

Given multiple comparators, how do you choose the best one?

How to Choose a Comparative Time Series?

1. Linden and Adams (2011) recommend choosing a comparative time series that is *similar* to the primary time series **before** the intervention[4]
   - Only time varying component?
2. Walter et al. (2013) chose comparative time series based on highest within-time period correlation[5]
   - What if unmeasured confounders are not similar?

$$\phi = \frac{cov(\varepsilon_t^p, \varepsilon_t^c)}{\sqrt{var(\varepsilon_t^p) var(\varepsilon_t^c)}} \neq \frac{cov(\eta_t^p + \varepsilon_t^p, \eta_t^c + \varepsilon_t^c)}{\sqrt{var(\eta_t^p + \varepsilon_t^p) var(\eta_t^c + \varepsilon_t^c)}}$$

[4] Linden & Adams (2011) "Applying a propensity score-based weighting model to interrupted time series data: improving causal inference in programme evaluation"

[5] Walter, Olivier, Churches & Grzebieta (2013) "The impact of compulsory helmet legislation on cyclist head injuries in New South Wales, Australia: A response"

Outline

Empirical Bayes ITS

- Basic idea:
  1. Pre-intervention data is used to estimate a *prior* model
  2. This model is extrapolated over the post-intervention period (i.e., counterfactual)
  3. Post-intervention observations are analysed relative to the counterfactual (*posterior*)
- Pre-intervention model

$$E\left(\log(y_T^{EB})\right) = \alpha_0 + \alpha_1 T + \alpha_2 C + \alpha_3 TC, \quad T < 0$$

- Counterfactual residuals

$$\Delta_T = \log(y_T) - \log(\hat{y}_T^{EB}), \quad T > 0$$

- No intervention effect when $\bar{\Delta}_T = 0$
- Residuals will have systematic pattern if unmeasured confounders are not similar

Comparative Empirical Bayes ITS

- Including a comparative time series

$$\Delta_T^p - \Delta_T^c = \log(y_T^p / y_T^c) - \log(\hat{y}_T^{EB-p} / \hat{y}_T^{EB-c})$$

- No **relative** intervention effect when $\bar{\Delta}_T^p - \bar{\Delta}_T^c = 0$
- Residuals will have systematic pattern if unmeasured confounders are not similar

Outline

1. Motivating Example
2. Interrupted Time Series
3. Empirical Bayes ITS
4. Results
5. Discussion

NSW Data

- Hospital presentations from 1 July 1989 to 30 June 1992
- Cases identified from ICD-9-CM
- Primary outcome: bicycle-related head injury hospitalisations
- Possible comparators
  - Bicycle-related arm injury hospitalisations (no head injury)
  - Bicycle-related leg injury hospitalisations (no head injury)
  - Pedestrian-related head injury hospitalisations
  - Australian beer production (sensitivity analysis?)

1st and 2nd Criteria

- Results from CITS models for each comparator

Comparator	Pre-law similarity $\hat{\beta}_5$ (SE)	Within-time correlation $\hat{\phi}$
Arm	-0.008 (0.015)	0.026
Leg	0.023 (0.021)	0.096
Head-Peds	-0.008 (0.020)	-0.063
Beer	**0.003 (0.015)**	**0.185**

- Australian beer production is the "best" comparator using these criteria

Empirical Bayes Criterion

- Models were fit to pre-intervention data using each potential comparator
- Linear models fit to counterfactual residuals

Comparator	Intercept	Slope
Arm	-0.263 (0.138)	0.010 (0.013)
Leg	-0.263 (0.157)	-0.025 (0.015)
Head-Peds	-0.383 (0.190)	0.001 (0.018)
Beer	-0.494 (0.165)	0.010 (0.016)

- All slope estimates are statistically non-significant and "small"
- All intercept estimates are statistically significant (or nearly so)

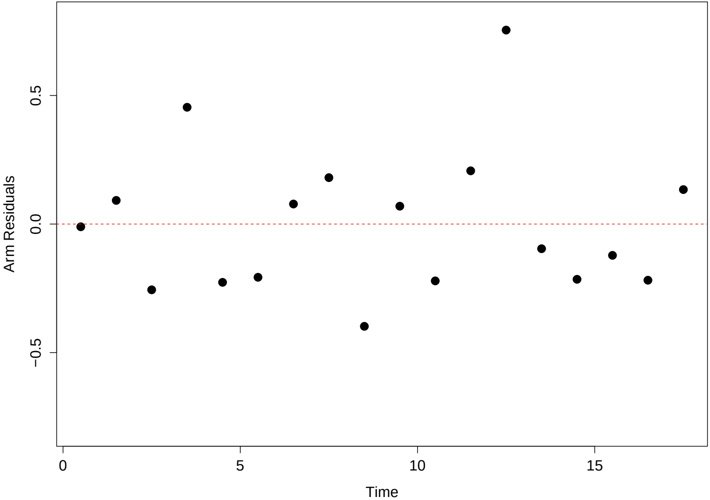
Empirical Bayes Criterion

- Head injuries had the greatest relative decline compared to Australian beer production

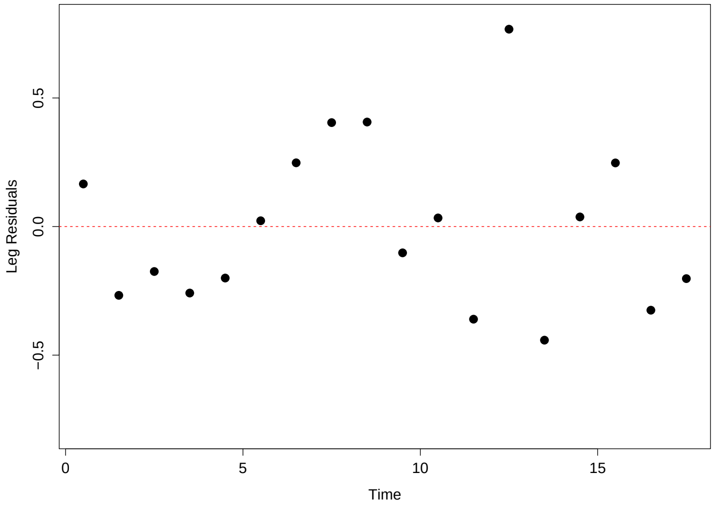
$$\exp\left(-\mathbf{0.494}\right) - \mathbf{1} = -\mathbf{39}\%$$

- Is Australian beer production the "best" comparator to cycling head injury hospitalisations?
- Residual analysis suggests cycling arm injuries are affected by similar unmeasured confounding

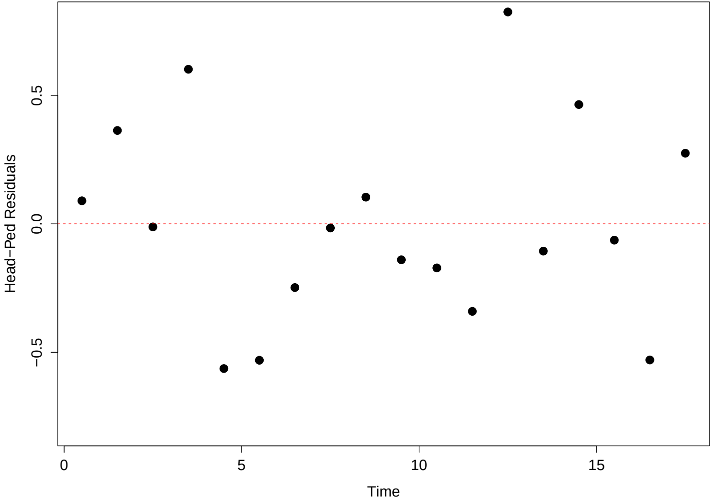
Arm Residuals



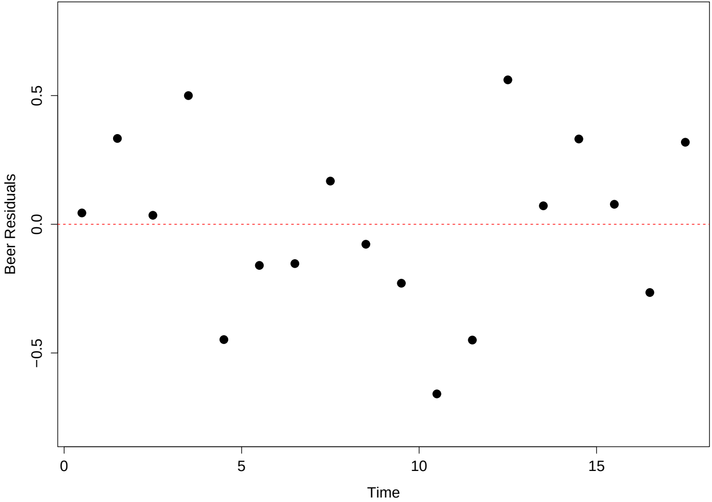
Leg Residuals



Pedestrian-Head Residuals



Beer Residuals



Outline

Which Comparator is Best?

- Linden and Adams criterion
  - All do not differ significantly in pre-law period (Beer production better than others)
- Walter et al. criterion
  - Beer production exhibits largest within-month correlation
- Empirical Bayes (residual analysis) criterion
  - Arm injury residuals appear random
  - Systematic pattern for others $\longrightarrow$ invalid statistical inference?
- Estimated intervention effect is smallest relative to arm injuries
  - Most conservative estimate

Discussion

- Causal inference for population-based interventions is difficult
- Interrupted time series is likely the best analytic approach
  - Threats to internal validity (due to lack of randomisation)
- The use of a comparative time series is promising
  - An analytic framework for choosing "best" comparator is needed

Acknowledgements

- School of Mathematics and Statistics, UNSW
  - Jake Olivier, Joanna J.J. Wang
- Transport and Road Safety (TARS) Research, UNSW
  - Joanna J.J. Wang, Raphael Grzebieta
- Centre for Health Systems and Safety Research, UNSW
  - Scott Walter
- NSW Department of Health

Thank You!

Questions?