CURRENT ISSUES IN
CRIMINAL JUSTICE

Journal of the Institute of Criminology

Current Issues in Criminal Justice is published three times per year, in July, November and March. This issue of *Current Issues in Criminal Justice* forms the second of three issues published in the twentieth volume. Volume 20 Number 3 will be published in March 2009. *Current Issues in Criminal Justice* Number 1 (October 1989) and Number 2 (March 1990) should be regarded as forming Volume 1, Numbers 1 and 2, respectively.

The views expressed in this issue are those of the individual authors and do not necessarily reflect those of the Institute of Criminology or the institution with which they are affiliated. Whilst every reasonable care has been taken in the preparation of this publication, no liability is assumed for any errors or omissions. *Current Issues in Criminal Justice* is a peer reviewed journal. The Editorial Board would like to thank the reviewers of submissions to this issue of the Journal.

Typeset by the Institute of Criminology.
Printed and bound by Ligare Printery, Sydney, Australia.

Subscriptions
> Subscription rates within Australia (gst incl.): $110 (1 year), $209 (2 years), $308 (3 years). Overseas subscriptions: $115 (1 year), $220 (2 years), $325 (3 years). Single copies of the current issue are $44. Back issues are available for $27.50. For subscriptions in North America, contact Gaunt, Inc, who are sole distributors of the journal in that country (Gaunt Building, 3011 Gulf Drive, Holmes Beach, Florida 34217-2199 USA).

Membership
> Membership to the Institute of Criminology includes a Journal subscription, entry to seminars and 10% discount on Institute publications. Contact the Institute for details.

Contact details
> Institute of Criminology, Sydney University Law School. 173–175 Phillip Street, Sydney, NSW 2000, Australia Ph: 61 2 9351 0239 Fax: 61 2 9351 0200
> e-mail: d.koester@usyd.edu.au <www.criminology.law.usyd.edu.au>

Current Issues in Criminal Justice is indexed in APAIS: Australian Public Affairs Information Service, produced by the National Library of Australia, and in CINCH, the Australian Criminology Database, produced by the Australian Institute of Criminology. Both databases are available on-line or in CD-ROM format. Contact: INFORMIT, phone: (03) 9341 3265; email: <info@rmitpublishing.com.au>.

Current Issues in Criminal Justice is published online at Informit: <http://search.informit.com.au>, and at HeinOnline: <http://heinonline.org>.

Lecturers
    T Anthony, BA (Hons) LLB (Hons) PhD (*Syd*)
    A Loughnan, BA (Hons) LLB (Hons) (*Syd*) LLM (*NYU*) PhD (*LSE*)
    K Walton, LLB (*Edin*) MA (*UCL*) PhD (*Edin*)
Publishing Officer
    D Koester, BA (*Syd*)
Administration Officer
    R Miller, BA (*Syd*) MCrim (*Syd*)
Publishing and Media Officer
    K Moore, BA Mass Comm (*Macq*) MCrim (*Syd*)

Management Committee

The Management Committee is comprised of the teaching and research staff and the Administrative staff of the Institute of Criminology. Additional members of the Committee are:
    Associate Professor E Baldry, University of New South Wales
    Associate Professor K Biber, University of Technology Sydney
    Professor D Brown, University of New South Wales
    Associate Professor M Carmody, University of Western Sydney
    Dr J Cashmore, University of Sydney
    Professor C Cunneen, University of New South Wales
    Professor G Triggs, Dean, Faculty of Law, University of Sydney

Advisory Committee

Chair
    The Honourable J Spigelman, AC, Chief Justice of the Supeme Court of New South Wales
Advisor to the Director
    The Honourable Justice M Kirby, AC CMG, High Court of Australia
Members
    Mr B Barbour, New South Wales Ombudsman
    The Honourable Justice RO Blanch, AM, Chief Judge, District Court, Sydney
    Professor J Braithwaite, Australian National University
    Mr P Byrne, SC, Barrister
    Profesor D Chappell, University of Sydney, Professorial Fellow, University of Wollongong
    Mr P Cloran, Deputy Chief Magistrate, Local Court of New South Wales
    Mr N Cowdery, AM, QC, Director of Public Prosecutions
    His Honour Judge R Dive, Senior Judge, NSW Drug Court
    Professor P Grabosky, Australian National University
    Mr L Grant, Assistant Commissioner, NSW Department of Corrective Services
    Mr A Haesler, SC, Barrister and NSW Public Defender
    Professor RW Harding, Inspector of Custodial Services, Western Australia
    The Honourable Justice PJ Hidden, AM, Supreme Court of New South Wales
    His Honour Judge M Marien SC, Judge of the Distict Court of New South Wales
    Ms P Musgrave, Director, Criminal Law Review Division, NSW Attorney-General's
       Department
    His Honour Judge SR Norrish, QC, Judge of the Distict Court of New South Wales
    The Honourable Justice RNJ Purvis, AM, QC,
    Ms J Sanders, Youth Justice Coalition, Shopfront
    The Honourable Justice PJ Staunton, AM, Industrial Relations Commission
    Mr B Thomas, Director, Crime Prevention Division, NSW Attorney-General's Department
    Mr PG Ward
    Dr D Weatherburn, Director, NSW Bureau of Crime Statistics and Research
    His Honour Judge GD Woods, QC, Judge of the Distict Court of New South Wales

CURRENT ISSUES IN CRIMINAL JUSTICE

CONTENTS

Do Gun Buy-backs Save Lives? Evidence from Time Series Variation

Christine Neill* and Andrew Leigh**

Abstract

Three recent papers have examined the effect of a national tightening of firearm legislation and gun buy-back in Australia in 1996-1997 on firearm and non-firearm death rates. Despite analysing almost the same data, the three papers reach rather different conclusions. In this article, we highlight key methodological concerns with the papers. We also make some judgments as to the evidence on the effectiveness of the Australian legislation. Drawing strong conclusions from simple time series analysis is not warranted, but to the extent that this evidence points anywhere, it is towards the firearms buy-back reducing gun deaths.

Introduction

Understanding the relationship between firearms availability and gun deaths is critical for policy makers around the world. Australia's 1996-1997 National Firearms Agreement (NFA), which tightened gun ownership and licensing requirements and removed 600,000 guns from a country with a population of 20 million people, offers a potentially useful policy experiment to analyse this relationship. A decade later, there are three papers that have examined the effects of the NFA: Ozanne-Smith et al (2004), Baker and McPhedran (2006), and Chapman et al (2006).

As its name suggests, the Australian National Firearms Agreement is a policy change that took place at the national level. In this sense, it is analogous to the 1994 US federal ban on assault weapons and large capacity magazines. In their study of that ban, Koper and Roth (2001a:33) found that 'the ban may have contributed to a reduction in gun homicides, but a statistical power analysis of our model indicated that any likely impact from the ban will be very difficult to detect statistically for several more years'. Kleck (2001:79), in a criticism of that paper, argued that 'longitudinal impact evaluations of unique macro-level interventions, such as a change in federal law, cannot be even minimally persuasive' and that for this and several other reasons, publication of the article was misguided.

The issues raised in the exchange between Koper and Roth (2001a, 2001b) and Kleck (2001) are important ones, that have implications for the statistical analysis of policy experiments and their interpretation. The fact that firearm control legislation is controversial makes it particularly important that researchers in the area undertake their work with a clear understanding of the limitations of statistical analysis, and that the robustness of the results is checked.

---

* Corresponding author. Department of Economics, Wilfrid Laurier University, 75 University Ave West, Waterloo, ON, N2L 3C5, Canada. Email cneill@wlu.ca. Website: www.wlu.ca/sbe/cneill

** Research School of Social Sciences, Australian National University, Canberra, Australia. Email: andrew.leigh@anu.edu.au. Website: andrew.leigh@anu.edu.au

In this article, we discuss the results of the three Australian papers and their robustness to alternative specifications. Our comment focuses on four key issues: (1) whether and under what circumstances a pure time series study can identify the effects of national policy changes; (2) understanding the power of the tests used; (3) sensitivity to the model specification used; and (4) sensitivity to the time period used.

A Brief Description of the Australian Studies

In 1996, following the Port Arthur massacre, in which 35 people died, Australia's federal and state governments agreed to the standardisation of firearms legislation across Australian states. A key provision of the 1996 National Firearms Agreement (NFA) was that certain types of semi-automatic rifles, and semi-automatic and pump action shotguns, were declared illegal. These weapons were subject to a buy-back under which owners who turned in newly illegal weapons were paid market prices (see Reuter and Mouzos 2003 for a more complete description of the NFA). Around 600,000 guns were returned and destroyed by September 1997, around 20% of the stock of guns in Australia. The cost of the buy-back was around half a billion Australian dollars. Although it had fairly broad public support, the new legislation was nonetheless controversial, drawing heavy criticism from individuals and organizations involved in the sport of shooting and in hunting activities.

A decade after the NFA was implemented, there have been three studies that seek to evaluate whether it was successful in achieving its key aims of reducing firearm deaths. Each uses data from 1979 to the early 2000s. The starting date in two of the three (Ozanne-Smith et al 2004; Chapman et al 2006) appears to have been selected by data availability.[1] Baker and McPhedran (2006), on the other hand, graphed data going back as far as 1915, but their statistical analysis discarded all observations prior to 1979. Ozanne-Smith et al (2004) used state-level data, while Chapman et al (2006) and Baker and McPhedran (2006) used only national-level data on death rates. None of the studies included socio-economic control variables.

Of the three studies, Chapman et al (2006) and Baker and McPhedran (2006) are the most directly comparable. Although they used almost the same data set, they come to quite different conclusions as to the results of the policy.[2] To some extent, this derives from different empirical specifications.

Baker and McPhedran (2006) estimated ARIMA(1,1,1) models on data from 1979 to 1996 for firearm and non-firearm homicide, suicide, and accidental deaths, with the dependent variable being the number of deaths per 100,000 individuals (the death rate).[3] They then calculated mean projections for the years 1997 to 2004 and conducted a t-test to determine whether the projected series were statistically significantly different from the actual series. So far as we can determine, their test statistics did not account for the fact that the projections themselves are subject to uncertainty. However, they did perform a heuristic test that implicitly took this into account: graphing the 95% confidence interval around the

[1]   Data going back to 1979 at the national level are available from Australian Bureau of Statistics (ABS) publications, but earlier data must be purchased from ABS Consultancy Services.
[2]   Chapman et al (2006) exclude data after 2003 from their analysis, based on concerns over data reliability. The number of deaths identified in each year differs somewhat between Chapman et al (2006) and Baker and McPhedran (2006). It is unclear why this is the case.
[3]   ARIMA(1,1,1) refers to an auto-regressive integrated moving-average model, with integration of order 1 and first-order serial correlation and moving average components.

point estimates, and identifying statistically significant departures from the model as occurring if the actual series passed outside the confidence interval.

Baker and McPhedran (2006) found that there was a statistically significant drop in firearm suicides after 1997, and no statistically significant change in firearm homicides, or non-firearm suicides or homicides. They concluded that 'suicide rates in Australia were highly influenced by other societal changes, confounding the ability to discern any effect on firearm suicides that may have resulted from the NFA' and that '[h]omicide patterns (firearm and non-firearm) were not influenced by the NFA, the conclusion being that the gun buy-back and restrictive legislative changes had no influence on firearm homicide in Australia' (2006:463).

Chapman et al (2006) used a negative binomial regression, which ensured that they did not predict negative death rates (we return to this issue below). Unlike Baker and McPhedran, they did not allow for the possibility that there is serial correlation in the data, but they did include controls for pre-existing trends in death rates, and allowed for the NFA to affect the death rate in two ways: through a level shift, or by affecting its rate of change. They found what appear to be statistically significant downward movements in both firearm suicides and homicides, and a faster rate of decrease in those series after 1997 (although in the case of firearm homicides, this is not statistically significant). They also recognised the possibility of method substitution, and concluded that the fact that non-firearm deaths also decreased after 1997 suggested that method substitution did not occur. Unlike the other two studies, Chapman et al examined mass shootings, pointing out that while Australia averaged one mass shooting per year in the decade prior to 1997, there were no mass shootings in Australia during the decade 1997-2006. They therefore argued that the NFA was successful in its key aim of preventing further firearms massacres.[4]

Ozanne-Smith et al (2004) took a somewhat different approach, using sub-national variation. The authors noted that the state of Victoria tightened firearm legislation in 1988, and argued that the implementation of the NFA in 1996-97 meant that the other Australian states and territories 'caught up' with Victoria's tougher legislation.[5] They then estimated a Poisson model that compared the rates of decline in firearm deaths in Victoria relative to the rest of Australia after 1988 and then again after 1997. Because they used sub-national variation in policy, they were able to control for any national-level changes in firearm death rates by including a full set of year dummy variables, rather than relying on time trends. They found that there was a significant decline in firearm deaths in Victoria relative to the rest of Australia between 1988 and 1996, and that firearm deaths fell in the rest of Australia relative to Victoria after 1997, suggesting that the firearm legislation had significant impacts on deaths. The largest effect was found in suicides. They do not, however, consider the possibility of method substitution.

The three studies therefore agreed on several key points. First, firearm suicides dropped after 1997, and this drop was statistically significant and large in magnitude. Second, firearm homicides dropped substantially, although statistical tests may not find this drop to have been statistically significant. And third, although it cannot be ruled out, there does not appear to have been substantial method substitution, since non-firearm death rates also

---

[4]   Chapman et al do not estimate the statistical significance of this difference, but given the prior history of mass shootings in Australia, the probability that this change in the frequency of mass shootings was due to mere chance is well below 1%.

[5]   In the case of the gun buy-back, this assumption is clearly violated. Reuter and Mouzos (2003:132) show that the number of firearms handed back in Victoria in 1997 was 4,300 per 100,000 people, a higher rate than for Australia as a whole (3,400 per 100,000 people).

decreased. Despite what would appear to be considerable agreement, however, the interpretation of the findings in the three papers was quite different, and the debate in the Australian media over the results has been quite heated. Ozanne-Smith et al (2004) and Chapman et al (2006) argued that the statistical evidence favours the conclusion that firearm deaths fell and the NFA was effective. Baker and McPhedran, on the other hand, have interpreted the evidence as showing that the NFA had no effect.[6]

We now turn to an analysis of key concerns we have with the interpretation of these studies, and of the methodology used. Our focus here will be on Baker and McPhedran (2006), although we will also comment on the other two papers throughout.

Use of Time Series Variation to Estimate the Effects of National Law Changes

Kleck (2001:79) argued that one can never make any claims as to the effect of a national law change because '[w]e just do not have the macro-level data to measure most crime-related variables at regular intervals between census years. This is the main reason why longitudinal impact evaluations of unique macro-level interventions, such as a change in a federal law, cannot be even minimally persuasive.'[7] This point is based on earlier work published in Britt et al (1996), which argues that an appropriate control needs to be identified in order to account for these unobserved determinants of firearm death rates.

Baker and McPhedran (2006) were clearly aware of the value of having a control group, but are confused about how to identify such a control group and how to use it in a statistical model. They stated that '[t]he inclusion of suicide and homicide by methods other than firearm provided a control against which the political, social and economic culture into which additional legislative requirements for civilian firearm ownership occurred could be evaluated, as well as determining the level of method substitution within homicide and suicide' (2006:457).

Britt et al (1996) argued against the use of non-firearm death rates as a control for firearm death rates, since the two may be determined by different underlying socio-economic factors. However, the comments by Baker and McPhedran raise another, perhaps more

---

6   This is broadly the conclusion in their published paper. Statements to the media, however, were much stronger. In a summary of the research released at the Sporting Shooters' Association of Australia website, the conclusions are:
- 'The reforms did not affect rates of firearm homicide in Australia.
- The reforms could not be shown to alter rates of firearm suicide, because rates of suicide using other methods also began to decline in the late 1990s.
- …
- It must be concluded that the gun buyback and restrictive legislative changes had no influence on firearm homicide in Australia.
- The lack of effect of a massive buyback and associated legislative changes in the requirements for obtaining a firearm licence or legally possessing a firearm has significant implications for public and justice policy, not only for Australia, but internationally.'
Source: 'Gun Laws and Sudden Death: Did the Australian Firearms Legislation of 1996 Make a Difference? Executive Summary'. Available online on 18 December 2007, from www.ic-wish.org/Executive%20Summary.pdf

7   Koper and Roth are clearly aware of these difficulties, and highlight them in their paper, noting that 'the law has not produced a clear impact on gun violence' (2001a:69), although they suggest that the ban likely 'contributed to a reduction in gun homicides' (2001a:33).

important concern – specifically, that the possibility of method substitution invalidates the use of the non-firearm death rate as a control for the firearm death rate. If the gun buy-back caused an increase in non-firearm homicides, the non-firearm homicide rate cannot be a good control for the firearm homicide rate. A formal discussion of the problems that arise when attempting to identify the effect of national policy changes using only time series data, or using non-firearm deaths both as a control group and to examine substitution effects, is set out in Appendix A.

It is unfortunate that such factors make it extremely difficult to draw conclusions on the effects of national-level policy changes using only time series data, given that such policy changes are often of high policy importance. Kleck (2001) appears to argue that these types of problems mean that such studies should not be undertaken at all.[8] On the other hand, such studies may be able to provide indicative evidence, even if it is not conclusive, and perhaps may point researchers to areas where more research is needed. We do think, however, that researchers need to be aware of the drawbacks of such studies.

These lessons do not appear to have been learned by Baker and McPhedran (2006). That paper's conclusions appear to draw opportunistically on either method substitution or underlying trend arguments to justify a conclusion that the NFA had no effect, even when the statistical tests they used suggested otherwise. For instance, they found a statistically significant decline in firearm suicides following the introduction of the NFA, but no statistically significant decline in non-firearm suicides. Yet they argued that they cannot say that the NFA had any effect on firearm suicides because non-firearm suicides began to decrease after 1999. In analysing whether there was method substitution in homicides, the logic becomes rather more twisted. Baker and McPhedran (2006:461) stated that, although there was no statistically significant change in either firearm or non-firearm homicides, and thus no evidence that there was method displacement, there was a 'theoretical possibility that displacement from firearm homicide to other methods may have occurred at an increasing rate throughout the entire time series, potentially contributing to the relatively stable rate of non-firearm homicide over time' (presumably countering what would otherwise have been a downward trend in non-firearm homicides), although they immediately state that they do not empirically assess this possibility.

Chapman et al (2006) took a fairly similar approach to Baker and McPhedran (2006), in that they separately estimated models of firearm and non-firearm death rates, and discussed the problem of method substitution. Their conclusions are therefore subject to the same concerns regarding the use of purely time-series techniques to analyse the effects of policy changes. The authors were clearly aware of this, stating that '[g]iven the observational nature of the data available … conclusions regarding the causality of the association must remain interpretive rather than definitive' (Chapman et al 2006:366).

Ozanne-Smith et al (2004) on the other hand, accounted for the problem of identifying an appropriate control group by using changes in firearm deaths in Victoria as a control for firearm death changes in the rest of Australia. Some concerns may remain, however. To the extent that the NFA did have some effect on Victorian firearm death rates, their estimates will understate the magnitude of the NFA's effect. On the other hand, not including variables that control for possible determinants of firearm death rates opens the possibility

---

[8]   Although much of Kleck (2001) focuses on the relatively small impact that the 1994 US federal ban on assault weapons and large capacity magazines had on overall gun ownership (a critique that is not applicable to the Australian NFA), Kleck concludes by saying 'there is no effective way to assess the impact on crime rates of a unique national policy change' (2001: 80). This critique is potentially applicable to the Australian reform.

that confounding factors could be responsible for the differential rates of decline in firearm deaths. It should be noted that none of the three papers attempts to include other socio-economic controls, however.

Statistical Significance and the Power of Tests

A common axiom in empirical research is that if a test fails to reject the null hypothesis, the researcher should not automatically *accept* the null hypothesis. If the sample size is small or the dependent variable is mis-specified, the test is commonly described as being 'weak'. In such an instance, there may well be a true effect, but the test lacks the statistical power to identify the effect at conventional levels of statistical significance. When statistical tests are weak, the proper approach is to focus attention on the point estimates and their associated standard errors (McCloskey & Ziliak 1996).

This point is well known in the literature on the effectiveness of firearms legislation. Indeed, Koper and Roth (2001b) emphasised this point in their response to Kleck (2001), noting that their statistical tests had extremely low power so that despite their finding of a drop in firearm deaths that was very large in magnitude, they were unable to reject the null hypothesis of no effect on statistical grounds.

Chapman et al (2006), alone of the three studies, note the likely low power of the tests given the volatility of the data and the relatively short time series available. Baker and McPhedran (2006) do not discuss the power of their tests at all, moving almost seamlessly from findings that there was not a statistically significant decrease in firearm homicides to stating that '[h]omicide patterns (firearm and non-firearm) were not influenced by the NFA, the conclusion being that the gun buy-back and restrictive legislative changes had no influence on firearm homicide in Australia' (2001:463). They nowhere attempt to interpret the magnitude of the effect, using statistical significance as the sole criterion for assessing the effectiveness of the NFA.

Model Specification and Robustness of the Results

Particularly in the case where estimates are based on a potentially weak source of identification, it is useful to examine the robustness of the results to alternative model specifications. Britt et al (1996) discuss the need to test the sensitivity of estimates of firearm law changes to changes in model specification. Unfortunately, none of the three papers considered here presents the results of more than a single regression for each type of death examined, and little attention is paid to alternative specifications. In this paper, we show that even small changes to the model can result in important differences in the conclusions. We illustrate in relation to Baker and McPhedran (2006), whose modelling strategy we consider, for reasons explained in the text, to be particularly questionable. We vary their specification in only two ways: changing the time period used in the analysis and the functional form of the estimating equation.

Length of the Time Period Used

Britt et al (1996:371) highlight the possibility that results may be highly sensitive to changes in the time period used in the analysis of policy changes: 'Since the evaluation of a law's effectiveness may vary with the specific time series used, there is the potential for research

outcomes to be manipulated merely by the timing of the study.' They argue that checking robustness to the time series used is an important part of serious empirical research.

None of the three studies of the Australian NFA assess the sensitivity of their results to changes in the time period under examination. All begin their sample in 1979 and end with the most recently available data.[9] As mentioned earlier, in two of the three cases, this appears to have been due to data limitations.

The exception is Baker and McPhedran (2006), who show graphs of Australian homicide and suicide rates from 1915 onwards. Despite the fact that they have 95 years of data, their empirical analysis uses only 26 years of data (1979-2004), and their ARIMA model is estimated on data from 1979 to 1996 only.

There is no discussion in Baker and McPhedran as to the reasons for restricting their sample to the shorter time period. We can think of two possible explanations. First, there may have been a structural break in the series from 1978 to 1979. Baker and McPhedran do not attempt to make that argument, however, and we can see no obvious reason why there should have been a structural break in 1978-79. The second possible reason is that the data may have come from different sources, or have been calculated differently. The ABS provides data on deaths due to assault and deaths due to self-harm in its Cause of Death publications going back to 1915. While there have been revisions to the categories of causes of death, including a move from the use of ICD-8 to ICD-9 categorisation in 1979, this seems unlikely to have greatly affected estimates of suicide or homicide deaths. As Kriesfeld & Harrison (2005) show in the case of the shift from ICD-9 to ICD-10, the recategorisation of ICD codes tends to have virtually no impact on the well-established categories of homicide and suicide.

The problems with using a short time series are compounded by Baker and McPhedran's use of an ARIMA(1,1,1) model. In the late 1970s and early 1980s, firearm homicide and suicide rates were very high, relative to historical averages. Estimating a time-series model with a linear time trend from a local maximum will bias the results to finding a long-term declining trend, resulting in predictions of firearm death rates that are relatively low. Thus, such models will almost certainly underestimate any downward shift in firearm death rates post-NFA. Although a similar criticism could be applied to Chapman et al (2006), in their case the use of a negative binomial specification alleviated some of the problems associated with choosing a starting point of 1979. That said, using the same modelling technique as Chapman et al but extending the period back in time also increases the estimated effect of the NFA in reducing firearm homicides and suicides. Ozanne-Smith et al (2004) also used 1979 as their starting point, but their estimates are not biased by that choice because they used sub-national variation and include Australia-wide year fixed effects in their model.

How much difference does the choice of starting point actually make to the results? We begin by attempting to replicate the results in Baker and McPhedran (2006), and then extend the time series under consideration.[10] Table 1 shows how this affects the results as reported

---

[9]   An exception is Chapman et al (2006), who exclude data from 2004 that was considered to be of dubious quality. No results are reported in that paper from different time periods, however.

[10] We could not obtain the full set of data underlying Figures 1 and 2 in Baker and McPhedran (2006) from the corresponding author. Our results are therefore based upon Table 1 in Baker and McPhedran (2006), supplemented with data kindly provided by Jenny Mouzos of the Australian Institute of Criminology, and the most recent population data from the Australian Bureau of Statistics. We first used precisely the data laid out in Baker and McPhedran (2006) and estimated ARIMA(1,1,1) models for the period 1979 to 1996. We did not have access to the same statistical package used by Baker and McPhedran, and initially used STATA. STATA

by Baker and McPhedran (2006). Details of the parameter estimates in the ARIMA(1,1,1) model are in Appendix Table B1.

Table 1: **Comparison of predicted and observed rates of firearm homicide and suicide, ARIMA(1,1,1) model (dependent variable is the number of deaths per 100,000 people)**

	BM	Replication	Full Sample (1915-2004)	Extend by a Decade (1969-2004)
Panel A: Dep Var is Firearm Homicide				
Predicted rate	0.28	0.303	0.450	0.400
Observed rate	0.27	0.269	0.269	0.269
Difference	-0.01	-0.034	-0.181	-0.131
P (Predicted≤Observed)	0.14	0.0248	0.0002	0.0011
Lives Saved per Year	2	7	36	26
Panel B: Dep Var is Firearm Suicide				
Predicted rate	1.85	1.854	2.465	2.020
Observed rate	1.22	1.230	1.230	1.230
Difference	-0.630	-0.624	-1.235	-0.790
P (Predicted≤Observed)	0.001	0.0000	0.0000	0.0000
Lives Saved per Year	126	125	247	158

Note: Figures in the column headed 'BM' are taken directly from the text in Baker and McPhedran (2006). Replication shows our best attempts to produce the same results, using the same time period (1979-1996). Predicted rate is based on estimating an ARIMA model using data up to 1996, and forecasting out to 2004. Observed rate is the average in the data from 1996 to 2004. Lives saved per year is calculated by multiplying the difference (change in the rate per 100,000 people) by 200 (since the population of Australia is 20 million). The model estimates and predictions are from R.

Several points are notable. First, as expected, predicted firearm suicides and homicides after 1997 are considerably smaller when data from 1979 onwards is used, than when the full data set is used. The models estimated from the longer time series suggest that there were on average 250 fewer firearm deaths per year after the implementation of the NFA

---

did not yield estimates similar to those reported by Baker and McPhedran. For instance, the firearm homicide model did not converge. Simply changing the population figures to more recent ABS estimates helped somewhat. Switching to the statistical package R brought our estimates considerably closer to those in Baker and McPhedran; the results reported here are therefore from R. We consider the estimates from R to be more reliable, but the sensitivity of the estimates to the statistical package used is somewhat concerning, and itself suggests that the ARIMA(1,1,1) specification is not ideal. Note, however, that regardless of the statistical package used, and despite coming quite close on firearm suicides, we could not replicate Baker and McPhedran's predictions of the average firearm homicide rate after 1997 in the absence of the NFA. We also note that the statistical 'test' used by Baker and McPhedran is more heuristic than formal. McDowall and Loftin (2005) discuss more formal tests for structural breaks in an I(1) model. Here, we nonetheless use the methods described in Baker and McPhedran, in order to focus attention on the robustness of the results to modelling changes. All data used in this paper and the programs used to obtain the statistical results (in STATA and R), as well as details of those results in numerical and graphical form, are available at www.wlu.ca/sbe/cneill.

than would have been expected based on the predictions from the ARIMA model – close to double the numbers implied by Baker and McPhedran's estimates.

A second key concern with the time period selected by Baker and McPhedran is that estimating an ARIMA(1,1,1) model using data only from 1979-1996 is dubious, especially for firearm homicides. The point estimates over the shorter time period are sensitive to the specification, and even to the statistical package used. This partly reflects the well-known difficulties associated with estimating an ARIMA model with such a short time series. However, it also reflects the fact that the ARIMA(1,1,1) model appears to be inappropriate in this case – simple statistical tests reject the hypothesis that Australian homicide rates follow a non-stationary process.[11] Baker and McPhedran present no statistical tests to show whether this model is appropriate. Augmented Dickey-Fuller tests, which are a simple way to determine whether a series is integrated (non-stationary), are shown in Appendix Table B2. They strongly reject the null hypothesis that the series are integrated in the case of firearm and non-firearm homicides.

Third, these results are not driven by some distant historical episode. The final column of Table 1 shows that even if we extend the sample period by only a decade, the probability that firearm deaths were not lower than would be expected based on a simple ARIMA(1,1,1) model is well below 1% for both homicides and suicides.

Model Specification: The Resurrection Problem

In modelling death rates, particularly where the absolute number of deaths is rather small, it is important to carefully consider whether the model used is appropriate to the task. Most empirical models of death rates consequently use an empirical specification appropriate to count data (Poisson or negative binomial) or take other steps to ensure that predicted death rates do not fall below zero (using the natural logarithm of the death rate as the dependent variable, or explicitly allowing for zero observations through a Tobit model, for instance).

As the dependent variable, Baker and McPhedran (2006) use the number of deaths of a particular type per 100,000 people. If the firearm death rate were high and stable, this might not present a problem. However, since the firearm death rate is low and volatile, the estimates in Baker and McPhedran place a non-zero probability on the death rate for accidental firearm deaths and firearm homicides falling below zero. As is shown graphically in Figure 1 (which merely repeats figures 4A and 5 from Baker and McPhedran 2006), it cannot be rejected at the 95% confidence level that there will be negative deaths after 1994 for accidental firearm deaths, and after 2004 for firearm homicides. Projecting out further, the models predict that by 2010, deaths attributable to assault with a firearm or an accidental firearm incident would be negative. This is concerning: an effective modelling strategy should place a zero probability on the occurrence of a logically impossible event. Furthermore, it points clearly to the issue raised earlier of the low power of the tests. If, in order for Baker and McPhedran to be satisfied that firearm homicides fell after 1997, there must be a negative number of firearm homicides, then the test clearly has no statistical power.

---

[11]   McDowall (2002) notes that to date little attention has been paid to the possibility that crime rates can be described by non-stationary processes. Using a time series from 1925 to 2000, described in McDowall and Loftin (2005) as 'short', he finds that US homicide rates are well-described by an I(1) model with first-order serial correlation.
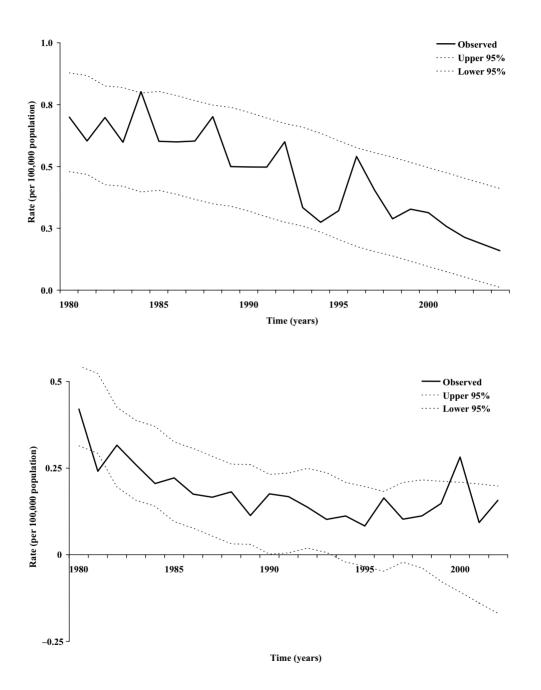
Figure 1: Death Rates Cannot Fall Below Zero

Note: Panels are from Baker and McPhedran (2006). First panel shows their Figure 4A (firearm homicide rate per 100,000 people) and the second panel shows their Figure 5 (accidental firearm death rate per 100,000 people).

There are straightforward and well-known solutions to this problem. In order to ensure that predicted death rates are always positive, researchers can use a Poisson model, or take the log of the rate, rather than the rate itself. Such approaches are common in studies of the impact of gun laws on deaths (see e.g. Ozanne-Smith et al 2004; Duggan 2001; Beautrais et al 2006; Chapman et al 2006). Indeed, we have been unable to find another study examining firearm-related deaths that simply uses the death rate as a dependent variable with no further specification checks.

We check the sensitivity of Baker and McPhedran's results to the use of a log rather than a levels specification. Table 2 shows estimates of the post-1996 reduction in deaths, taken from ARIMA(1,1,1) models with the dependent variable as the log of the death rate, both for the period 1979-2004 and for the full period 1915-2004. The coefficient estimates are converted into level terms to facilitate comparison. The results show an increase in the predicted death rate in the log model compared with the levels models shown in Table 1. More importantly, though, there is a very clear reduction in the probability that the observed series is greater than the predicted series for firearm homicides in the model estimated from 1979 to 1996. The estimated probability that the predicted series is smaller than the actual series drops from 2.5% to 0.26% (compare the second columns of Table 1 and Table 2). Thus, there is extremely strong evidence in this model that the observed firearm homicide rate was lower after 1997 than would have been expected based on an ARIMA(1,1,1) log model estimated from 1979 to 1996.

Table 2: Comparison of predicted and observed rates of firearm homicide and suicide, ARIMA(1,1,1) model (dependent variable is the death rate)

	BM 1979-1996 Dep Var is rate	Replication 1979-1996 Dep Var is log rate	Full Sample 1915-1996 Dep var is log rate
Panel A: Dep Var is Firearm Homicide			
Predicted rate	0.28	0.338	0.446
Observed rate	0.27	0.269	0.269
Difference	-0.01	-0.069	-0.177
P (Predicted≤Observed)	0.14	0.0026	0.0002
Lives Saved per Year	2	14	35
Panel B: Dep Var is Firearm Suicide			
Predicted rate	1.85	1.941	2.396
Observed rate	1.22	1.230	1.230
Difference	-0.63	-0.712	-1.166
P (Predicted≤Observed)	0.001	0.0000	0.0000
Lives Saved per Year	126	142	233

Note: Figures in the column headed 'BM' are taken directly from the text in Baker and McPhedran (2006). The other columns use the log of the death rate, rather than the death rate itself, as the dependent variable in the model estimated to obtain forecasts. These forecasts are then converted back to a death rate, to ensure comparability with Table 1. Model estimates and predictions are from R.

Note that a part of our concern with the model specification in Baker and McPhedran (2006) is the combination of the ARIMA(1,1,1) model – and in particular the assumption that the series are integrated –, the 1979 starting date, and the use of levels rather than logs. We present results from the ARIMA(1,1,1) models here, in order to show how sensitive the results are to even small specification changes. (The results from AR(1) models or from simple linear regression models that incorporate time trends typically yield point estimates of a similar or larger magnitude to those shown here.[12])

Here again, our criticism on model specification is unique to Baker and McPhedran's analysis. Both Ozanne-Smith et al (2004) and Chapman et al (2006) use techniques (Poisson and negative binomial regression respectively) that rule out the possibility of negative deaths. In addition, the specification used by Chapman et al (2006) (allowing for both trend breaks and breaks in the level of the series) appears, if anything, to work against finding any effect of the NFA.

Conclusion

This article has reviewed the available evidence on the effects of Australia's National Firearms Agreement on homicide and suicide rates. While we mainly focus on general methodological points, it is useful to consider whether we can confidently draw any conclusions from the three studies currently available.[13]

Although we can point to flaws in all three papers, we believe that Baker and McPhedran (2006) contains too many statistical and interpretive deficiencies, several of which are outlined above, to enable objective readers to rely on it to any extent. Chapman et al (2006) provide useful summary evidence on the trends in firearm and non-firearm deaths, and uniquely among the three, also some evidence on mass shooting events. Their empirical model is reasonable, although some testing of alternative specifications and time periods would have been helpful to the paper. If anything, their empirical strategy and use of 1979 as a starting point bias their results against finding a downward shift in firearm death rates coincident with the NFA. Ultimately, however, as the authors themselves acknowledge, they cannot be sure that any decline in firearm death rates after 1997 is causally related to the NFA.

Of the three, Ozanne-Smith et al (2004) have the most satisfying identification strategy, relying on an earlier policy change in Victoria and cross-state differences in firearm homicide and suicide rates to identify a plausibly causal effect of tighter firearm regulations. They find a 14% drop in death rates in the rest of Australia relative to Victoria after 1997, mostly due to lower firearm suicide. It is likely that this identification strategy too would have underestimated the effect of the NFA, since there is a reasonable possibility that the NFA had some effects in Victoria (a substantial number of guns were handed in by Victorians under the buy-back). A key weakness of the paper is that it does not consider whether there is evidence of method substitution. Further, recent advances in techniques for dealing with policy experiment studies with small numbers of policy changes (Bertrand et al 2004) suggest that there may be some concerns with the calculation of standard errors in that paper.

---

[12]   Results available on request.
[13]   In principle, a variety of statistical approaches could be used to model the effect of the NFA on firearm deaths. Exploring the full gamut of approaches may be a useful exercise, but it is beyond the scope of this paper.

Given that the time series and other available evidence to date suggest a substantial fall in firearm deaths, can we say there was (likely) a decline in overall homicides and suicides following the NFA? Making such an inference would require a conclusion to be drawn on method substitution. Here, Ozanne-Smith et al (2004) provide no guidance; the best evidence we have to rely on is the time series evidence. In our view – and the views expressed in Baker and McPhedran (2006) and Chapman et al (2006) – the lack of any marked, sustained increase in non-firearm suicide or homicide rates after 1997 suggests there is little reason to suspect any long-run method substitution effect. It is, however, difficult to be certain of this without first having in place a model that satisfactorily explains movements in total deaths, and simple time-series models are not sufficient in this regard. Further, it would likely be difficult to identify method substitution if it occurred, given that in Australia, firearm deaths are small relative to total numbers of deaths and to the volatility in overall deaths. Despite these difficulties, and although we cannot rule out the possibility that non-firearm suicide and homicide would have fallen faster in the absence of the NFA, the fact that overall violent deaths have fallen since 1996 suggests there has not been substantial method substitution.

As we point out in our re-analysis of some of the findings in Baker and McPhedran (2006), the high degree of variability in the underlying data and the fragility of the estimated results with respect to different specifications and even statistical packages used, suggest that time series analysis alone cannot conclusively identify the effect of a national law change on death rates. However, to the extent that the available evidence points anywhere, it is towards the conclusion that the NFA reduced gun deaths.

The main thrust of this article, however, has been to raise some methodological concerns with the existing studies – in particular, regarding robustness to small changes in the time period and the model specification.

Aside from the question of the robustness of the results, a critical issue is whether time series analysis alone can ever be definitive in drawing conclusions about the effects of national policy changes. In the appendix to this article we show that it cannot, except under certain rather stringent assumptions. We do not believe that this means that such analysis is not worth undertaking or publishing. However, authors do need to be careful in interpreting their results.

Finally, it would be simple for researchers to make publicly available their statistical programs and data, allowing others to replicate their work. In two of the three papers we examine, researchers provide the full data set used in their empirical work in their paper, but even then it is difficult to replicate the results described in published papers because of uncertainties regarding the precise specification used.[14] The additional cost of making the data available in a package ready for statistical analysis and posting the statistical programs used to generate the final results (which in each case appears to have involved a single regression for each type of death examined) would be small. This would assist other researchers in satisfying themselves as to the validity of the results.[15]

---

[14] The term 'replicate' has been the subject of some recent controversy. Here, we use it in its most limited sense, to mean recovering the same estimates in a published paper using the same data set. We see such 'replication' as a first, rather than a final, step in assessing a paper's conclusions.

[15] All data and statistical programs used in this paper are available at: www.wlu.ca/sbe/~cneill

Acknowledgements

References

Baker J & McPhedran S 2006 'Gun Laws and Sudden Death: Did the Australian Firearms Legislation of 1996 Make a Difference?' *British Journal of Criminology* Advance Access published on 18 October 2006 (doi:10.1093/bjc/azl084)

Beautrais AL, Ferguson DM & Horwood LJ 2006 'Firearms legislation and reductions in firearm-related suicide deaths in New Zealand' *Australian and New Zealand Journal of Psychiatry* vol 40 pp 253-259

Bertrand M, Duflo E & Mullainathan S 2004 'How much should we trust difference-in-differences estimators?' *Quarterly Journal of Economics* vol 119 no 1 pp 249-275

Britt CL, Kleck G & Bordua DJ 1996 'A reassessment of the DC gun law: some cautionary notes on the use of interrupted time series designs for policy impact assessment' *Law & Society Review* vol 30 no 2 pp 361-380

Chapman S, Alpers P, Agho K & Jones M 2006 'Australia's 1996 gun law reforms: faster fall in firearm deaths, firearm suicides and a decade without mass shootings' *Injury Prevention* vol 12 no 6 pp 365-372

Duggan M 2001 'More Guns, More Crime' *Journal of Political Economy* vol 109 no 5 pp 1086-1114

Kleck G 2001 'Impossible policy evaluation and impossible conclusions – a comment on Koper and Roth' *Journal of Quantitative Criminology* vol 17 no 1 pp 75-80

Koper CS & Roth JA 2001a 'The impact of the 1994 federal assault weapon ban on gun violence outcomes: an assessment of multiple outcome measures and some lessons for policy evaluation' *Journal of Quantitative Criminology* vol 17 no 1 pp 33-74

Koper CS & Roth JA 2001b 'A priori assertions versus empirical inquiry: a reply to Kleck' *Journal of Quantitative Criminology* vol 17 no 1 pp 81-88

Kriesfeld R & Harrison J 2005 'Injury Deaths, Australia, 1999' *Injury Research and Statistics Series* Number 24 (AIHW cat. no. INJCAT 67) AIHW Adelaide

McCloskey DN & Ziliak ST 1996 'The Standard Error of Regressions' *Journal of Economic Literature* vol 34 no 1 pp 97-114

McDowall D & Loftin C 2005 'Are U.S. crime rate trends historically contingent?' *Journal of Research in Crime and Delinquency* vol 42 no 4 pp 359-383

McDowall D 2002 'Tests of nonlinear dynamics in U.S. homicide time series, and their implications' *Criminology* vol 40 no 3 pp 711-735

Ozanne-Smith J, Ashby K, Newstead S, Stathakis VZ & Clapperton A 2004 'Firearm related deaths: the impact of regulatory reform' *Injury Prevention* vol 10 pp 280-286

Reuter P & Mouzos J 2003 'Australia: A Massive Buyback of Low-Risk Guns' in Ludwig J & Cook PJ (eds) *Evaluating Gun Policy: Effects on Crime and Violence* (pp 121-156) Washington, DC

Appendix A. A Simple Model of Controls Versus Substitutes

The following discussion is framed in terms of suicide, where method of choice has been more commonly studied, but the analysis could be applied equally to murder. It is described in terms of firearm and hanging, simply to make the model more concrete and relevant to the issues discussed in this paper. Suppose these are the only two possible forms of suicide, and that numbers of suicides per year in a particular country are determined by:

$$Firearm_t = \alpha_1 + \beta_1 X_t + \rho_1 guns_t + \varepsilon_t \tag{1}$$
$$Hanging_t \quad = \alpha_2 + \beta_2 X_t + \rho_2 guns_t + \upsilon_t \tag{2}$$

We assume that there are sufficient numbers of deaths that the errors are approximately normally distributed, and that there is no serial correlation in the error terms. We also assume that $\text{cov}(\varepsilon_t \upsilon_t) = 0$.

In this simple model, there are thus two reasons why firearm and hanging deaths may be related to each other. First, movements in X_t cause changes in deaths. This variable is intended as a simple representation of what Baker and McPhedran (2006) refer to as the 'political, social and economic culture', which may affect the decision to suicide. Suppose, for instance, that increases in the unemployment rate increase the numbers of suicides – then X_t is the unemployment rate and β_1 and β_2 would be positive. For the purposes of evaluating the effect of changes in the gun laws, we are not concerned about these coefficients, except that we want to ensure that we do not mis-attribute falls in firearm-related suicides to general social changes.

The variable $guns_t$ represents the legal environment around gun ownership. The model above allows that changes in gun laws may have led to changes in both firearm and hanging suicides. The simple intuition that drives tightening of gun laws is typically that ρ_1 would be negative: a tightening of gun laws would reduce firearm-related suicides. However, there is a possibility that method substitution may occur, so that changes in gun laws that reduce firearm-related suicides may actually *increase* suicides by hanging.

Given this model, total suicides are:

$$
\begin{aligned}
Suicide_t \quad &= Firearm_t + Hanging_t \\
&= (\alpha_1 + \alpha_2) + (\beta_1 + \beta_2)X_t + (\rho_1 + \rho_2)guns_t + (\varepsilon_t + \upsilon_t) \\
&= (\alpha_1 + \alpha_2) + (\beta_1 + \beta_2)X_t + (\rho_1 + \rho_2)guns_t + \mu_t
\end{aligned}
\tag{3}
$$

The central questions for this paper are:
(1)   is ρ_1 negative? (did firearm-related deaths fall after the change in laws?);
(2)   is ρ_2 positive? (was there method substitution?); and
(3)   is $(\rho_1 + \rho_2)$ negative? (did overall suicides fall after the change in laws?)

A way of determining this would be to estimate a stacked model:

$$
\begin{aligned}
&Firearm_t \\
&\qquad = \alpha_1 + \beta_1 X_t + \rho_1 guns_t + \alpha_2 H_t + \beta_2(H_t X_t) + \rho_2(H_t guns_t) + \mu_t \\
&Hanging_t
\end{aligned}
\tag{3}
$$

Suppose, however, that we have no information on X available to us. Can we answer these questions? There are two possibilities. First, if $\beta_1 = \beta_2$, then there is a potentially simple solution: subtract hangings from firearm deaths, so that we have:

$$Firearm_t - Hanging_t \quad = (\alpha_1 - \alpha_2) + (\beta_1 - \beta_2)\,X_t + (\rho_1 - \rho_2)guns_t + (\varepsilon_t + \upsilon_t)$$
$$= (\alpha_1 - \alpha_2) + (\rho_1 - \rho_2)guns_t + (\varepsilon_t + \upsilon_t) \qquad (4)$$

In this case, however, we will only be able to test the third of the propositions – we will not be able to say anything about the impact on firearm-related suicides, or on method substitution – and that only implicitly by assuming that $\rho_1 < 0$ and that $\rho_2 >= 0$. The assumption that firearm and hanging deaths respond in the same way to the same stimuli is also unsatisfactory.[16] For example, if unemployment rates affect suicides in low-income families more than in high-income families, but firearms tend to be owned by high-income families, then $\beta_1 \neq \beta_2$.

What if we simply omit X_t and perform the stacked regression analysis described in (3)? Then we have:

$$Firearm_t$$
$$= \alpha_1 + \rho_1 guns_t + \alpha_2 H_t + \rho_2(H_t guns_t) + \zeta_t \qquad (5)$$
$$Hanging_t$$

where $\zeta_t = \beta_1 X_t + \beta_2(H_t X_t) + \mu_t$. Estimates of ρ_1 and ρ_2 will be biased and inconsistent if there is any correlation between X_t and $guns_t$. Given the particular nature of this model, where there is only time series variation, and where $guns_t$ is a dummy equal to one after 1997, this is inevitable if there is any time-series variation in X_t. This is roughly similar to the problem of using a difference-in-difference estimator with time series variation and serial correlation as described in Bertrand et al (2004), except that in this case the lack of a reasonable control group exacerbates the difficulties. Suppose, for instance, that a drought tends to lead to higher firearm suicides, but not higher suicides by hanging (X_t is drought, $\beta_1 > 0$, $\beta_2 = 0$). Then the Australian drought of the mid-2000s should have led to more firearm suicides, but not hangings. Failing to account for this in estimating (5) would see all those excess suicides attributed to the NFA; that is, the estimate of ρ_1 would be biased in a positive direction.

In short, time series variation in and of itself is unable to recover plausible estimates of the effect of the Australian gun laws on deaths, except in the presence of extremely restrictive (and probably inaccurate) assumptions on the determinants of the numbers of deaths. Without having some credible control group – or at a minimum a model that introduces time-varying factors that affect homicide and suicide rates, such as personal income growth, unemployment rates, or health and social program changes – we can draw no definitive conclusions on either the extent of method substitution, or on the underlying direction of overall homicide and suicide rates.

---

16   This is the concern that Britt et al (1996) have with the use of non-firearm deaths as a control for firearm deaths.

Appendix B. Details of Regression Results

Table B1: Estimates from ARIMA (1,1,1) models of firearm and non-firearm suicides and homicides

	Firearm suicide			Firearm homicide			Non-firearm suicide			Non-firearm homicide		
	1979-96	1969-96	1915-96	1979-96	1969-96	1915-96	1979-96	1969-96	1915-96	1979-96	1969-96	1915-96
Level												
Trend*	**-0.087** (0.0154)	-0.024 (0.0363)	-0.006 (0.0063)	-0.020 (0.0044)	-0.004 (0.0086)	-0.001 (0.0043)	**0.196** (0.0445)	0.058 (0.1278)	0.023 (0.0193)	0.009 (0.0175)	0.015 (0.0138)	0.000 (0.0064)
AR	**0.562** (0.2338)	0.156 (0.5928)	**0.771** (0.0827)	0.001 (0.2703)	-0.149 (0.2782)	0.138 (0.1813)	**-0.444** (0.2345)	**-1.000** (0.0050)	**0.847** (0.0662)	-0.551 (0.3001)	-0.340 (0.3611)	-0.183 (0.1580)
MA	-1.000 (0.1985)	-0.319 (0.5388)	-1.000 (0.0404)	-1.000 (0.1832)	-0.554 (0.1966)	-0.736 (0.1296)	-1.000 (0.1634)	0.990 (0.0474)	-1.000 (0.0338)	-0.192 (0.3163)	-0.375 (0.4111)	-0.548 (0.1294)
AIC	-0.08	5.5	62.87	-21.09	-35.07	-102.01	41.67	63.8	196.68	-11.7	-18.01	-70.61
Observed mean		1.230			0.269			11.312			1.327	
Predicted mean	1.854	2.020	2.465	0.303	0.400	0.450	11.819	11.253	10.409	1.393	1.393	1.326
P (Predicted ≤ Observed)	0.0000	0.0000	0.0000	0.0248	0.0011	0.0002	0.2146	0.5464	0.9768 / 0.0232	0.2092	0.0692	0.5060
Log												
Trend*	**-0.031** (0.0064)	-0.010 (0.0128)	-0.002 (0.0025)	**-0.038** (0.0115)	-0.008 (0.0174)	-0.001 (0.0090)	**0.021** (0.0051)	0.006 (0.0135)	-0.001 (0.0090)	0.007 (0.0129)	**0.012** (0.0032)	0.000 (0.0064)
AR	**0.668** (0.2321)	0.158 (0.6782)	**0.799** (0.0778)	0.213 (0.2777)	0.119 (0.3536)	0.227 (0.1790)	**0.487** (0.2315)	**-1.000** (0.0030)	0.227 (0.1790)	-0.548 (0.3081)	0.146 (0.2475)	-0.217 (0.1677)
MA	-1.000 (0.1900)	-0.285 (0.6297)	-1.000 (0.0379)	-1.000 (0.1712)	-0.662 (0.2401)	-0.782 (0.1218)	-1.000 (0.1623)	0.991 (0.0383)	-0.782 (0.1218)	-0.197 (0.3281)	-1.000 (0.1552)	-0.513 (0.1499)
AIC	-36.46	-53.69	-105.53	3.89	2.23	28.34	-34.55	-57.34	28.34	-21.85	-28.25	-77.62
Observed mean		1.230			0.269			11.312			1.327	
*Predicted mean** *	1.941	2.031	2.396	0.338	0.405	0.446	12.125	11.296	10.396	1.351	1.556	1.315
P (Predicted ≤ Observed)	0.0000	0.0000	0.0000	0.003	0.0005	0.0002	0.1294	0.5119	0.9785	0.2590	0.0007	0.6212

* The estimated trend is the constant in the first-differenced model.

** For ease of interpretation, predictions for the log models were converted to levels for calculating the means and conducting the t-tests.

Coefficients in bold are significant at the 5% level.

Note: The estimates on the MA terms clearly indicate that the ARIMA(1,1,1) specification is not able to be cleanly estimated. These results are from R, but STATA's Arima command performed no better. Since our main concern here is to test the sensitivity of Baker and McPhedran's results to simple specification changes, we chose to maintain the ARIMA(1,1,1) model here despite these clear problems to make the specifications as comparable to Baker and McPhedran's as possible. We are unable to compare all of these estimates directly to those of Baker and McPhedran, since they used a different statistical package and do not report the full set of parameter estimates in their paper. Models that use an AR(1) specification or which use a simple linear regression, which we consider to be more appropriate to the data, yield point estimates not very different to those shown here, and are available on request.

Table B2: Augmented Dickey Fuller test results

	Level		Log	
	ADF statistic	p-value	ADF statistic	p-value
1979-2004				
No trend:				
Firearm suicides	0.44	0.9817	1.16	0.9933
Firearm homicides	-1.30	0.6296	-0.47	0.8990
Non-firearm suicides	-1.62	0.4712	-1.60	0.4845
Non-firearm homicides	**-3.76**	**0.0033**	**-3.75**	**0.0035**
With trend:				
Firearm suicides	-3.05	0.1200	-1.79	0.7106
Firearm homicides	**-5.44**	**0.0000**	**-3.66**	**0.0253**
Non-firearm suicides	-1.43	0.8510	-1.35	0.8734
Non-firearm homicides	**-3.69**	**0.0230**	**-3.70**	**0.0227**
1915-2004				
No trend:				
Firearm suicides	-1.98	0.2963	-0.45	0.9037
Firearm homicides	**-4.54**	**0.0002**	**-3.92**	**0.0019**
Non-firearm suicides	-2.38	0.1471	-2.28	0.1778
Non-firearm homicides	**-3.86**	**0.0024**	**-3.92**	**0.0019**
With trend:				
Firearm suicides	-2.29	0.4389	-0.94	0.9516
Firearm homicides	**-4.53**	**0.0014**	**-3.94**	**0.0109**
Non-firearm suicides	-2.97	0.1398	-2.88	0.1696
Non-firearm homicides	**-4.33**	**0.0028**	**-4.37**	**0.0025**

Note: Coefficients in bold are significant at the 5% level. This indicates that there is substantial evidence against the null hypothesis that the series is integrated. That is, it indicates that an I(1) model is inappropriate for the particular series. Typically, the preferred ADF test statistic is one based on a model that allows for a time trend.