THE UNIVERSITY OF MELBOURNE

Melbourne Institute Working Paper Series

Working Paper No. 17/08

The Australian Firearms Buyback and Its Effect on Gun Deaths

Wang-Sheng Lee and Sandy Suardi

MELBOURNE INSTITUTE
of Applied Economic and Social Research

The Australian Firearms Buyback and Its Effect on Gun Deaths*

Wang-Sheng Lee[†] and Sandy Suardi[‡]

[†] **Melbourne Institute of Applied Economic and Social Research**

The University of Melbourne

[‡] **Department of Economics and Finance, La Trobe University**

Melbourne Institute Working Paper No. 17/08

ISSN 1328-4991 (Print)

ISSN 1447-5863 (Online)

ISBN 978-0-7340-3285-0

August 2008

* Corresponding author: Sandy Suardi, s.suardi@latrobe.edu.au. Thanks to Christine Neill and Andrew Leigh for making the data used in their paper available, and to John Creedy for helpful comments.

Melbourne Institute of Applied Economic and Social Research

The University of Melbourne

Victoria 3010 Australia

***Telephone* (03) 8344 2100**

***Fax* (03) 8344 2111**

***Email* melb-inst@unimelb.edu.au**

***WWW Address* http://www.melbourneinstitute.com**

Abstract:

The 1996-97 National Firearms Agreement (NFA) in Australia introduced strict gun laws, primarily as a reaction to the mass shooting in Port Arthur, Tasmania in 1996, where 35 people were killed. Despite the fact that several researchers using the same data have examined the impact of the NFA on firearm deaths, a consensus does not appear to have been reached. In this paper, we re-analyze the same data on firearm deaths used in previous research, using tests for unknown structural breaks as a means to identifying impacts of the NFA. The results of these tests suggest that the NFA did not have any large effects on reducing firearm homicide or suicide rates.

1. Introduction

Australia's 1996-97 National Firearms Agreement (NFA), which involved the buyback and destruction of over 600,000 guns within a few months, is one of the most massive government adjustments to gun control regulations in the developed world in recent history. The extent of the buyback and accompanying swift nationwide change in the firearm regulatory environment following the enactment of the NFA (prohibition on certain types of firearms, registration requirements etc.) makes this policy a natural experiment in which the results of a methodical evaluation would be extremely interesting.[1] Indeed, this call has been heeded by several quantitative researchers (e.g., see Reuter and Mouzos 2003; Baker and McPhedran, 2006; Chapman et al., 2006; Neill and Leigh, 2007). Yet, more than a decade later, despite considerable research using homicide statistics from the same source – the Australian Bureau of Statistics (ABS) – there is still considerable rhetorical debate regarding the impact of the NFA on firearm homicide. In a typical time-series design, violence rates are analysed using time-series regression methods to see if there is a significant downward shift in homicides around the time a new gun law goes into effect. Much of the debate surrounding the effects of the NFA has revolved around interpretations of the results of quantitative analyses and their accompanying statistical tests. This paper aims to contribute to this debate by a re-analysis of the same data used by previous researchers, using an alternative time-series approach based on unknown structural breaks. As shown in Piehl et al. (2003), such tests for unknown structural breaks provide a useful framework for estimating potential treatment effects in an evaluation framework. In this paper, by employing a battery of structural break tests and analyzing the available data in a rigorous fashion, an attempt is made to resolve the debate surrounding the effects of the NFA.

A noteworthy feature of these studies that analyze the effects of the NFA is that they are all based on aggregate statistics for the whole of Australia. Unfortunately, the consistency of the approach taken by each state and territory government in Australia in the enactment of uniform firearm legislation has meant that an evaluation can only be conducted for the whole country and not for individual states or counties. In other words, unlike most quasi-experiments analyzed in the evaluation literature, there is no

[1] See, for example, Mouzos (1999) or Reuter and Mouzos (2003) for more background and details on the NFA.

appropriate comparison group as the entire population in Australia was targeted.[2] This makes the use of difference-in-difference or matching estimators that explicitly rely on comparison groups inappropriate. It also makes approaches often used in the US gun control literature that are based on state-level variation in gun laws and which implicitly rely on comparison groups infeasible.[3] Hence, from a policy perspective, although the NFA is a clean example of a natural experiment, from a statistical perspective, evaluation of the reform is not a straightforward exercise.

The main aim of this paper is to analyze the effect of the NFA on reducing firearm related deaths in Australia, with an emphasis on deaths by homicide and suicide. In addition, following the previous literature, we also examine homicides and suicides by methods other than a firearm to allow for the possibility of method substitution effects.

Several features of the Australian gun buyback make it of interest to policy makers in the US and other developed countries where gun control is an important issue on the political agenda. First, it has been argued that many gun laws fail because they are local, allowing guns from contiguous counties or states with more lenient laws to "leak" into the stricter jurisdictions. As Australia is geographically isolated with no domestic supply of the prohibited firearms, a nationwide implementation of the gun buyback implies that leakage would not be a serious issue. Second, the gun buyback scheme was large and better funded than comparable efforts in the US, helping to test the hypothesis of whether size of the buyback matters.[4] Third, in countries where the gun culture might not be as prevalent as in the US, where homicide rates are lower by international standards but where occasional mass murders occur (e.g., Britain), the impact of the NFA in Australia could be viewed as an indication of what the effect of a similar buyback scheme would be in such countries.

---

[2] Reuter and Mouzos (2003) make a passing reference to the possible use of New Zealand as a comparison group, as New Zealand participated in the Australasian Police Ministers Council which produced the proposal for the NFA, but chose not to implement the NFA. But it is clear that the Australian and New Zealand population and environment are quite different and that the New Zealand experience cannot represent a useful counterfactual.

[3] See, for example, Kleck and Patterson (1993).

[4] See Kennedy, Braga and Piehl (1996) and Sherman (2001) for reviews of the US gun buyback literature, where it was generally found that past efforts were on too small a scale or too poorly implemented. In addition, a recent critical review of the firearms and violence literature in the US by the National Academy of Sciences states that arguments for and against a market-based intervention such as gun buybacks appear to be largely based on speculation, not on evidence from research. The report states that it is at present not known if it is actually possible to shut down illegal pipelines of guns to criminals, and whether such access restrictions make useful public policy (Wellford, Pepper and Petrie, 2004).

The remainder of this paper is organized as follows. Section 2 provides a brief review of the literature on the effects of the NFA. Section 3 introduces the time-series data on homicides and suicides that we use for our empirical analysis. Section 4 discusses the econometric model and issues regarding model selection. In particular, we present results of unit root tests that check for the stationarity of the data, as well as discuss how we choose the appropriate lag structure for our time-series model. Section 5 discusses the structural break tests we employ and their results when used to analyze the data. Finally, section 6 concludes.

2. Background

This section provides a brief recap of some of the more prominent studies that have attempted to estimate the impact of the NFA.

Based on data from 1980 to 1999, Reuter and Mouzos (2003) found that the Australian buyback that targeted low-risk weapons may have had a modest effect on homicides (see their Figure 4.2). They note, however, that an alternative outcome to consider is the occurrence of another mass shooting (like the “Port Arthur” incident in 1996 where 35 people were killed that precipitated the enactment of the NFA). They highlight that the fact that there were no mass murders committed with a firearm in the five years post-NFA can be viewed as a slightly more promising outcome. It is clear, however, that more follow-up time than was available to the authors would be necessary before a proper assessment of the effects of the NFA could be made.

With more follow-up data in hand, several researchers revisited the question addressed in Reuter and Mouzos (2003). Using data from 1979 to 2003 and comparing the trends in firearm deaths pre- and post-NFA, Chapman et al. (2006) found that post-NFA, there were accelerated declines in annual total gun deaths and firearm suicides and a non-significant accelerated decline in firearm homicides. They also make the point that swings in the data by 2003 are so obvious that if one were given the data and were asked to guess the date of a major firearm intervention, it would be clear that it happened between 1996 and 1998.

On the other hand, based on a slightly longer time series of data from 1979 to 2004, Baker and McPhedran (2006) found that an examination of the long-term trends indicated that the only category of sudden death that may have been influenced by the introduction of the NFA was firearm suicide. Homicide patterns (firearm and non-firearm) were not influenced by the NFA. They therefore concluded that the gun buyback and restrictive legislative changes had no influence on firearm homicide in Australia. As an explanation for the divergent findings, Baker and McPhedran (2007) critique the conclusions reached by Chapman et al. (2006) because it appears the outcomes of the statistical analyses in Chapman et al. were contrary to the results of the statistical tests undertaken.

More recently, Neill and Leigh (2007) re-analyze the results of Baker and McPhedran (2006). They find that re-analysing the results either with a longer time series or using the log of the death rate, however, strengthens the evidence against the null hypothesis that the NFA had no effect on firearm suicides or homicides. In particular, they find a statistically significant reduction in deaths due to both firearm homicides and suicides.

Summing up the extant literature, it therefore appears that despite the fact that the data come from the same source, ten years post-NFA, there is no consensus on whether the NFA had an impact on firearm homicides.

3. Data

The data used in this paper come from the ABS statistics on Causes of Death (COD). The data, which span the period 1915-2004, comprise of death rates (expressed as per 100,000 of total population) classified under firearm suicide rates (FS), firearm homicide rates (FH), non-firearm suicide rates (NFS) and non-firearm homicide rates (NFH). Although two alternative data sources are available for analysing homicide in Australia – the National Homicide Monitoring Program at the Australian Institute of Criminology, and the Recorded Crime Australia publication by the ABS – the COD statistics have the longest span.[5] The COD data were used by all the above mentioned studies; where they differ is the time period used. Baker and McPhedran (2006) and Chapman et al. (2006) possibly use data starting from 1979 because of revisions to the categories of causes of death, including a move from the use of ICD-8 to ICD-9 categorization in 1979.[6] However, as pointed out by Neill and Leigh (2007), this seems unlikely to have greatly affected estimates of suicide or homicide deaths and they prefer to base their analysis on the longer time series from 1915 to 2004. For the purposes of this paper, we therefore employ the same data used in Neill and Leigh (2007) spanning

[5] Mouzos (2003) provides more discussion of the differences between the three alternative data sources on homicide in Australia.
[6] Neither study provides an explanation why 1979 was chosen as the starting time point.

the period 1915 to 2004 in order to maximize the sample size for our time series analysis.[7]

Given that a time series model of FS, FH, NFS or NFH using the raw data would imply that deaths could be negative sometime in the future, before proceeding with any formal statistical analysis, we first perform the transformation used in Neill and Leigh (2007) to ensure that model predicted death rates are always positive and express all rates in natural logarithms. A plot of the transformed data is shown in Figure 1.

Focusing on gun related death rates, we can see from Panels 1 and 2 of Figure 1 that there is a downward trend in both suicide and homicide rates starting from around about 1985. Apart from the dip in gun related death rates towards the end of the sample period, there appears to be a decline in death rates around about the early 1940s, and a drastic fall in firearm homicide rates in 1950. A cursory look at the plot around 1996-1997 suggests that the NFA is unlikely to have had a large effect on gun deaths. However, the downward trend in death rates that started in 1985 could have masked any possible effect that the NFA might exert on both FS and FH rates.

In contrast to the FS and FH series, we observe that there has been a rising trend in the non-gun related suicide and homicide rates since 1985. The fall in non-firearm related deaths is also noticeable in the early part of 1940s.

[7] The data used in this paper is provided in Appendix A.

Figure 1: Plots of the Natural Logarithm of the Series

Notes: FS, FH, NFS and NFH are expressed as total number of deaths per 100,000 of the Australian population.

4. Methods

We model all four data series in our COD data using the Box and Jenkins (1976) methodology to obtain an appropriate Autoregressive and Integrated Moving Average (ARIMA) model. Although we start off our empirical analysis by employing the Box-Jenkins approach to modelling the homicide data that were also used in the studies by Baker and McPhedran (2006), Chapman et al. (2006) and Neill and Leigh (2007), our point of departure is in how we use the model to make inferences on the effect of the NFA. The Baker and McPhedran (2006) and Neill and Leigh (2007) studies estimate a regression relationship based on pre-NFA data and use the regression model to predict values post-NFA. A comparison is then made between the actual post-NFA homicide rates and the predicted post-NFA homicide rates to make inferences regarding the effects of the NFA. Put another way, these studies use forecast errors as the method for identifying the treatment effect.

Chapman et al. (2006) adopt a different strategy and estimate separate regressions pre- and post-NFA. Their idea is to compare the slopes of the two regressions and to test if they are significantly different. Our approach based on unknown structural break tests generalizes this strategy by avoiding the need to choose a particular date that defines the pre- and post-periods. As pointed out by Britt, Kleck and Bordua (1996) when re-evaluating the 1976 District of Columbia gun law, the effective date of the law is merely a legalism and makes no special claims about when the new law will actually begin to have an effect on the target variable. Put another way, it is possible that there could be announcement effects or lagged impacts that a dummy variable indicating the time a gun law was legally enforced would not capture.

4.1 Stationarity of the Data

As is typical of any time-series analysis, an important first step of the modelling exercise is to determine the stationarity property of the series. We first use two tests – the Augmented Dickey Fuller (ADF) test and the Phillips-Perron (PP) test – to determine whether the series possesses a unit root.[8] Both tests share a common null hypothesis of a unit root process but differ in the way they correct for the problem of serial correlation in the regression residuals. The results of both tests are reported in Table 1. It can be seen that for the two series FS and NFS, both the ADF and PP tests

[8] As the ADF and PP tests are commonly employed in empirical research, we do not elaborate on the way we perform these tests.

unanimously fail to reject the existence of a unit root at all conventional levels of significance. For the FH series, the ADF and PP tests both strongly reject the null of a unit root. However, the two tests provide conflicting evidence about the unit root property of the NFH series; the PP test suggests that the NFH series is stationary while the ADF test result implies otherwise.

Table 1: Results for Unit Root Tests

	ADF	PP	ZA
With intercept			Break in Intercept
Firearm suicides	-0.4450 (0) [0.8957]	-0.1645 (1) [0.9380]	-3.0460 {1997}
Firearm homicides	-3.9163 (0) [0.0029]	-3.6860 (1) [0.0059]	-2.4714 {1993}
Non-firearm suicides	-2.2820 (0) [0.1800]	-2.4347 (6) [0.1353]	-3.3755 {1935}
Non-firearm homicides	-2.2211 (1) [0.2004]	-3.5398 (2) [0.0091]	-3.6688 {1970}
With trend and intercept			Break in Intercept and trend
Firearm suicides	-0.9416 (0) [0.9456]	-0.9416 (0) [0.9456]	-3.6539 {1990}
Firearm homicides	-3.9350 (0) [0.0145]	-3.7015 (1) [0.0273]	-2.8015 {1979}
Non-firearm suicides	-2.8774 (0) [0.1748]	-3.0436 (5) [0.1266]	-3.5805 {1931}
Non-firearm homicides	-2.7772 (1) [0.2096]	-4.0492 (2) [0.0105]	-4.0469 {1938}

Note: Figures in ( ) for the ADF and PP tests are the AIC-based selected lag length and the Newey-West selected bandwidth respectively. Figures in [ ] are p-values. Zivot and Andrew (ZA) (1992) test has a null hypothesis of a unit root with no break, and an alternative hypothesis of stationarity with a single break. The figure in {} under ZA test denotes break date. The 1%, 5% and 10% critical values for the ZA test with break(s) in intercept (intercept and trend) are -5.34 (-5.57), -4.80 (-5.08), and -4.58 (-4.82) respectively.

An alternative assessment of the unit root property of the data series requires that we account for a possible break in the data generating process. This is accomplished by employing the Zivot and Andrews (1992) (ZA) test that allows an endogenously determined breakpoint in the intercept, the trend function, or in both. This view is consistent with our objective of determining whether the national buy-back and tightened legislation in 1996-1997 had an effect on the death rates associated with firearm, and altered the dynamics of the FS and FH series. As argued by Perron (1989), failing to account for a structural break in the conventional unit root test may lead to a loss of power and wrongly infer the presence of a unit root when in fact the series is stationary around a one time structural break.

In its general form with breaks in both the intercept and the trend function, the test involves running the following regression for all potential breakpoints, T_B ($1 < T_B < T$),

$$\Delta y_t = \mu + \beta t + \theta_1 DU_t + \gamma_1 DT_t + \alpha y_{t-1} + \sum_{i=1}^{k} c_i \Delta y_{t-i} + \varepsilon_t \tag{1}$$

where DU_t and DT_t are break dummy variables that are defined as

$$DU_t = \begin{cases} 1 & if \quad t > T_B \\ 0 & \text{otherwise} \end{cases}$$

and

$$DT_t = \begin{cases} t - T_B & if \quad t > T_B \\ 0 & \text{otherwise} \end{cases}$$

where T is the sample size and k is the number of lags determined for each possible breakpoint by the Bayesian Information Criteria. Equation (1) is sequentially estimated and T_B is chosen so as to minimize the one-sided t-statistics of the unit root null hypothesis with no break (i.e. $H_0 : \alpha = 0$).

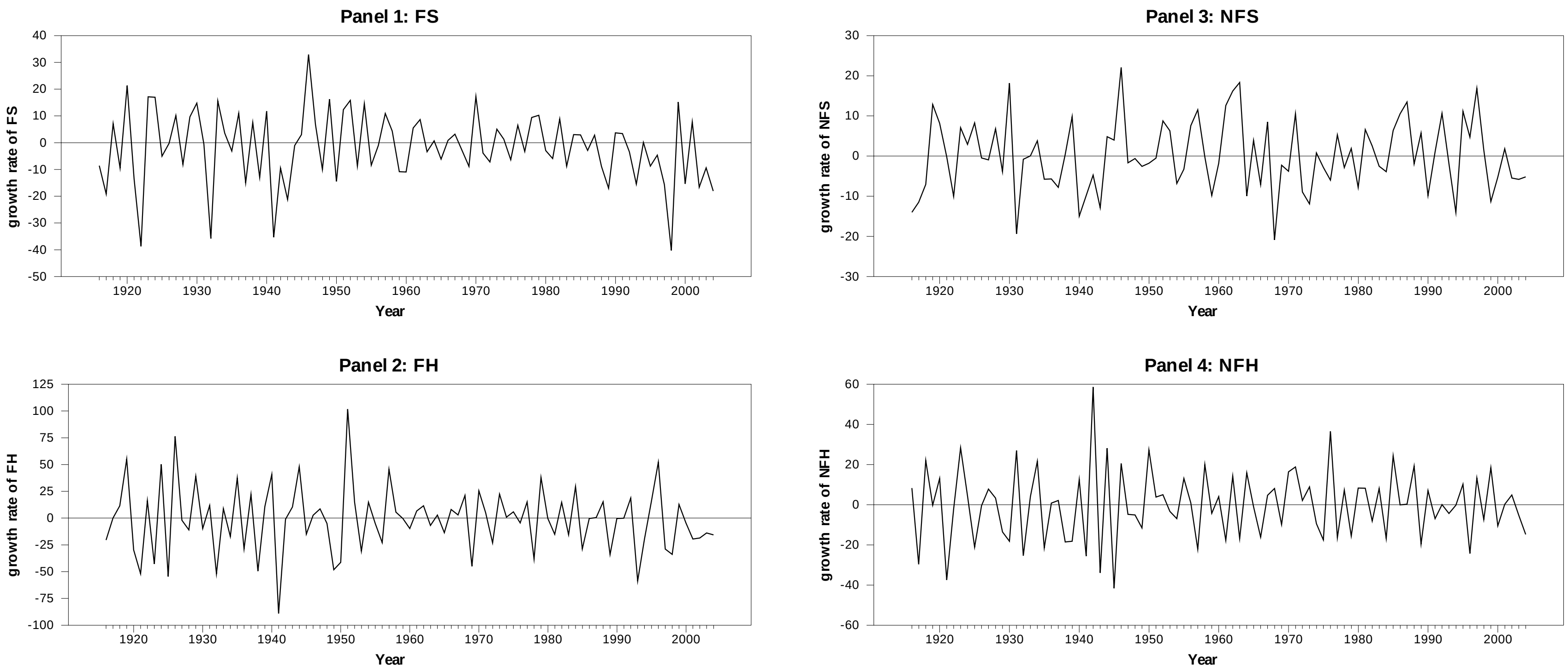
It is common to exclude the end-points of the sample when implementing the ZA unit root tests. This is due to the fact that the asymptotic distribution of their test statistics diverges to infinity when the end points are included. We report the results for 'trimming region' of the sample as suggested by Zivot and Andrews (1992) that is (0.15T, 0.85T). We also consider other trimming factors like 10% and 5% to accommodate the break date at around 1996-1997. Although not reported here, the results are largely consistent with those reported in Table 1, except that the break dates differ in some cases. Critical values at conventional levels of significance for the unit root tests are obtained from Zivot and Andrews (1992). For consistency with the reporting of unit root tests results, only the results of the ZA test for a structural break in the intercept, and a break in both the intercept and trend are reported.[9] Table 1 shows that once a potential break has been taken account of in the unit root test, the results for all series fail to reject the unit root null.

Although the ADF, PP and ZA tests lead to different conclusions regarding stationarity of the data, we favor the results of the ZA test because it is more consistent with our objective of determining whether the gun buyback and tightened legislation in 1996-1997 had an effect on the death rates associated with firearms, and altered the dynamics of the FS and FH series. Moreover, a preliminary plot of the data in Figure 1 also suggests that there could have been a change in the underlying process governing the series.

Having confirmed that all series are integrated with order one, we take the first difference of all series. Figure 2 shows plots of the data expressed in growth rates (i.e. the first difference of the logarithmic of the data multiplied by a hundred). Based on eye-balling the data, it is difficult to assess whether any breaks have taken place in the growth rate of the series. In section 5, after specifying an appropriate ARIMA model, we employ formal structural break tests to determine if any breaks do exist.

[9] We also performed the test for a break in the trend and the results, although not reported here, are consistent with the finding in Table 1. Results for this set of test are available from the authors upon request.

Figure 2: Plots of the Series Growth Rates



Notes: FS, FH, NFS and NFH are expressed as total number of deaths per 100,000 of the Australian population.

4.2 Order of the ARIMA model

In practice, to apply the unknown structural break point technique in a program evaluation setting, one first needs to define the regression relationship of interest. In other words, we need to have the correct specification of the regression model under the null hypothesis of no break. For example, if a linear trend belongs in the model, then we would need to include it in the analysis in order to have the correct inference on the break in mean. However, if there is no trend in the true relationship and we include a trend variable, this will obscure inference on the break in mean as inclusion of a trend could absorb some of the change in mean.

Although attempting to explain a series using only information about the history of the series (i.e., univariate time series analysis) appears to ignore information in related series, such an approach is not necessarily inferior to a structural approach. This is because a pure univariate time series model can be useful as a summary device when it is not clear what the appropriate structural model should be. Alternatively, even if economic theory is useful in suggesting an economic structural relationship, it is often the case that there is an equivalent ARIMA representation of that model. We discuss in more detail how our econometric model is specified below using the Box-Jenkins approach.

The optimal order of the ARIMA(p,1,q) process is determined by minimizing the value of either the Akaike Information Criteria (AIC) or the Bayesian Information Criteria (BIC) from estimating different ARIMA models for p=q=10. The results in Table 2 show that the BIC provides a more parsimonious model than the AIC. We also performed the Ljung-Box (1978) test to ensure that the residual of the regression is free from serial correlation.

When applying the structural break tests in the next section, for each of the four data series, we employ the specification suggested by AIC and assume that is the correct specification of the regression model under the null hypothesis of there being no structural breaks. The results also show that the chosen models for all series are free from various orders of serial correlation.

Table 2: ARIMA Model Selection Results

Variable	ARIMA model based on AIC	ARIMA model based on BIC
Firearm suicides (FS)	(0,1,1) 710.8814	(0,1,0) 711.2031
Firearm homicides (FH)	(4,1,2) 842.2294	(0,1,1) 847.1406
Non-firearm suicides (NFS)	(0,1,3) 636.4619	(0,1,0) 638.6316
Non-firearm homicides (NFH)	(0,1,1) 723.8419	(0,1,1) 726.3305

$$\Delta y_t = \mu + \alpha u_{t-1} + u_t$$

	FS	FH	NFS	NFH
μ	-1.7675 (1.4020)	-1.0873 (1.2740)	0.0508 (0.9325)	-0.0091 (0.5309)
α	-	-0.5733* (0.0908)	-	-0.6504* (0.0820)
Q(5)	2.646 [0.7543]	2.634 [0.7562]	1.539 [0.9085]	4.783 [0.4429]
Q(10)	13.617 [0.1912]	12.061 [0.2810]	4.151 [0.9402]	10.859 [0.3686]
Q(15)	19.785 [0.1803]	18.466 [0.2390]	15.598 [0.4092]	12.088 [0.6724]

Note: Figures in (p,d,q) indicate the lag order for AR, the order of integration, and the lag order for MA respectively. Marginal significance levels are displayed in [.]. Figures in (.) are robust standard errors. Q is the Ljung-Box test for k-th order serial correlation in the residual u_t, for k=5, 10 and 15. * denotes significance at the 1% level.

5. Structural Break Tests and Empirical Results

In this section, to examine whether the NFA has an influence at reducing the suicide and homicide growth rates that are related to firearms, we adopt a statistical approach similar to the one employed in Piehl et al. (2003), who suggest that tests for unknown structural breaks can be used as a policy evaluation tool. This approach involves identifying any unknown structural break in the time series for the homicide data. In the context of evaluation, a finding of a structural break that coincides with the implementation of the NFA can be interpreted as evidence supporting an effect of the NFA. On the other hand, a finding of no structural break or a structural break at an alternative date would be evidence against there being an effect of the NFA.

As in the case of the ZA test, one decision that needs to be made when applying these structural break tests is the choice of the “trimming” value. When one searches over all possible locations for a break in some parameters, it is important to specify how far into the sample one starts looking for a break and how close to the end of the sample one stops looking. The reason for not looking from the first observation to the last is that there must be a sufficient number of observations on either side of the point under consideration to estimate the regression relationship both before and after the break point. With 90 observations in our sample, the standard 15 percent trimming factor (i.e., the amount recommended by Andrews (1993)) does not concur with our interest in searching for a break at around about 1996-1997 when the NFA program was implemented. We therefore consider a trimming factor of 5 percent which provides about 5 data observations for the first and last breakpoint that lie close to the beginning and end of the sample.

5.1 The Quandt (1960) Test

Given a stationary time series, following Piehl et al. (2003), we first perform the Quandt (1960) endogenous structural change test which searches for the largest Chow (1960) statistic over all possible break dates. This test involves splitting the sample into two sub-periods and estimating the parameters for each sub-period. A Wald-statistic is then employed to test the equality of the two sets of parameters. Although the idea underlying the use of the Chow test bears resemblance to the approach used in Chapman et al. (2006) where pre- and post-NFA regressions were estimated and their slopes compared, an important difference is that it does not assume that the break date is known *a priori*. Even when the date of a policy change is known, tests for unknown breaks can be useful because of possible anticipation effects and lag effects. The use of tests for unknown breaks circumvents the need to choose the appropriate lag or lead to capture potential policy effects. In addition, in such instances, the resulting statistical tests would not have the correct critical values.

To compute the Chow statistics for a particular break date, we split the sample at that break date and estimate the model parameters separately on each sub-sample, as well as for the whole sample. The respective residual sum of squares (RSS) are computed and used to calculate the Wald-statistic as follows

$$W = \frac{RSS - (RSS_1 + RSS_2)/k}{\left(RSS_1 + RSS_2\right)/T - 2k} \tag{2}$$

where RSS is the residual sum of squares for the whole sample; the subscripts 1 and 2 denote the first and second sub-samples; T is the number of observations, and k is the number of regressors in the sub-sample regression. Thus the test is one of how much the RSS for the whole sample is bigger than the sum of the RSS for the two sub-samples. If the coefficients do not change much between the samples, the RSS will not rise much upon imposing the constancy parameter restriction across the two sub-samples. Having computed a sequence of the Chow statistics as a function of candidate break dates, we then plot in Figure 3 these values on the y-axis against the candidate break dates on the x-axis.

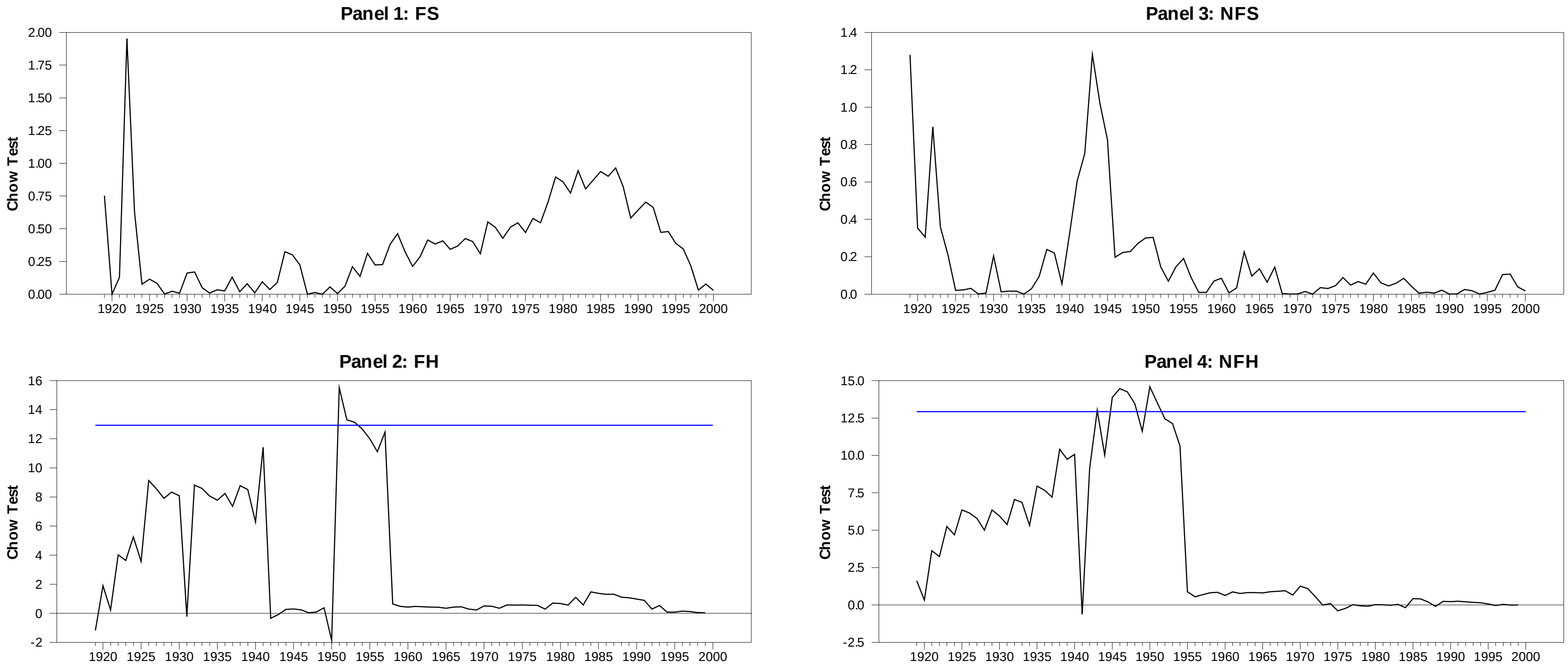
In panel 1 of Figure 3, we can see significant variation of the Chow test sequence across candidate break dates, with the test statistic reaching a high of 1.95 in 1922. Given that the break date is not known *a priori*, we cannot rely on standard Chi-square distribution for inference. The asymptotic critical values for a class of tests, including the Wald test, where the break date is only identified under the alternative hypothesis is computed by Andrews (1993). The 5 percent level critical values for a trimming factor of 0.05 and with one restriction is 9.84, and 12.93 with two restrictions. These asymptotic critical values are considerably larger than the comparable chi-square critical values.[10]

Visual inspection of the sequence of Chow test for the FS series suggests that in not a single candidate break date do we observe the test statistic above the 9.84 critical value. This implies that we cannot reject the null hypothesis of no structural break. Hence, this suggests that the FS time series do not possess a structural break. In other words, we fail to find any evidence that show the NFA has an effect at reducing the firearm suicide growth rates. The same result also holds for the NFS growth series.

In contrast, for the FH and NFH growth series, we find at the 5 percent significance level that there is evidence for the existence of structural breaks. For the FH growth rate, the test statistic is above the 12.93 critical value in the period 1951-53; for the NFH growth rate, there appears to be breaks at various points in time: 1943, 1946-48 and 1950-51.

[10] The 5% significance level chi-square critical values for one degree of freedom is 3.84, and 5.99 for two degrees of freedom.

Figure 3: Chow Test Sequence as a Function of Break Dates



In summary, our implementation of the Quandt test fails to find any evidence that a structural break occurred around the time of the NFA in any of the four data series examined. In the next two sub-sections, we employ alternative structural break tests as robustness checks.

5.2 The Bai (1997) Sequential Multiple Breaks Test

The Quandt test discussed in the previous sub-section only considers a single structural break and it does not allow for more than one break. In this sub-section, we perform Bai's (1997) sequential multiple structural breaks test to accommodate for more than a single breakpoint in the firearm and non-firearm related homicide growth rates. The idea of sequential estimation in Bai's test is to consider one break at a time rather than simultaneous estimation of multiple breaks. Because the model is treated as if there were only one break point at each stage of the analysis, this method is not only computationally less intensive but is also robust to misspecification in the number of breaks. An important insight of this method is that in the presence of multiple breakpoints, the sum of squared residuals (as a function of the break date) can have a local minimum around each break date. As such, the global minimum can be used as a break date estimator while the other local minima can also be regarded as candidate break date estimators. Accordingly, the sample is split at the break date estimate and the Quandt test is performed on the sub-samples to test for other breaks. Bai (1997) shows that important improvements can be obtained from re-estimation of break dates based on refined samples. In a relative long data series, testing for sequential breaks conditional on earlier breaks is important as there could be later break points in the series that could well coincide with the NFA date.

The Bai (1997) multiple break test proceeds by estimating the sum of squared residuals of the two sub-samples for a given break date T_s. For ease of exposition, suppose we consider a mean shift in a linear process. We can estimate a single break point for the mean shift by minimizing the sum of squared residuals among all possible sample splits. Let us denote the mean of the first T_s observations by $\bar{Y}_{T_s}$ and the mean of the last $T - T_s$ observations by $\bar{Y}_{T-T_s}$. Then the sum of squared errors for the whole sample is:

$$S_T(T_s) = \sum_{t=1}^{T_s} (Y_t - \overline{Y}_{T_s})^2 + \sum_{t=T_s+1}^{T} (Y_t - \overline{Y}_{T-T_s})^2 \quad (3)$$

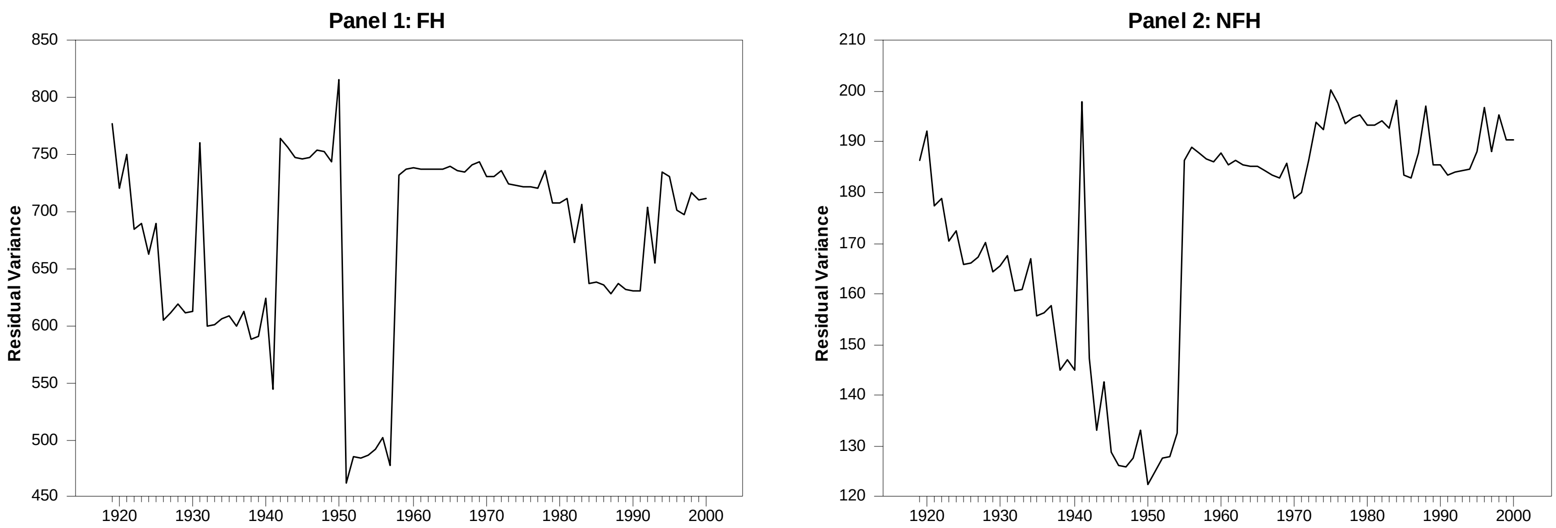
Bai defines a break point estimator as $\hat{T}_s = \arg\min_{1 \le T_s \le T-1} S_T(T_s)$. We first estimate the ARIMA(0,1,1) model with intercept for both the FH and NFH growth series on the sub-samples at various break dates across the whole sample. We then obtain the sum of squared residuals for the whole sample for each break date as given in (3).

We only apply the Bai test to the FH and NFH series since the Quandt test in the previous section found evidence for breaks in them. In other words, we do not perform the test on both FS and NFS due to the lack of evidence in support of a structural break in their series. In all of our analysis, we do not consider the possibility of a break in the trend given that the trend function is not statistically significant in all four ARIMA models.[11] In Figure 3, the residual variances for FH (panel 1) and NFH (panel 2) are plotted on the vertical axis against the break dates on the horizontal axis. In these figures, a random and erratic variation of the sub-sample estimates (and hence the sum of squared errors) would imply that the true parameters are constant. However, if there is a structural break, then there will be systematic variations in the sub-sample estimates across possible break dates, and there will be a well-defined minimum for the sum of squared residuals near the true break date.

Turning to panel 1 of Figure 4, we observe three well-defined minima for the residual variance of the FH growth series: a global minimum in 1951 and two local minima in 1941 and 1987. The visual evidence indicates that two or three structural breaks are possible in the sample period. We split the sample into two sub-samples at the break date given by the global minimum (i.e. 1951) and test for structural breaks on the two sub-samples. For the period 1952-2004, we find the Quandt test (test statistic equals 33.22) rejects the null hypothesis of parameters constancy and identifies 1987 as the structural break date. As for the period 1915-51, we fail to find evidence for a break.

[11] The plot of growth data in Figure 1(b) also does not reveal any evidence of a trend in any of the series.

Figure 4: Least Squares Break Date Estimation – Residual Variance as a Function of Break Dates



Next, we divide the sample in 1987 and conduct tests for structural breaks in the periods 1915-87 and 1988-2004. No structural breaks are identified in the latter period although for the former period, the Quandt test has a test statistic of 20.16 and rejects the null of parameter constancy at the 5 percent level in 1951. Finally, we re-consider the sample period 1951-87 in which we have failed to find any evidence of a structural break so far.

It is evident from a sequence of Bai's structural break test that two structural breaks appeared to have occurred in 1951 and 1987. None of these dates, however, coincide with the date when the NFA was implemented.

Repeating the analysis for NFH series, we find that there is only a single break date in 1950 (see Panel 2 of Figure 4). Again, this date does not coincide with the NFA date.

5.3 The Bai and Perron (1998, 2003) Test

Another test that we perform is based on the approach suggested by Bai and Perron (1998, 2003). In what follows we adhere to the practical recommendations made by Bai and Perron (2003).[12] We start by computing the double maximum test statistics $UD_{max}F_T(M,q)$ and $WD_{\max}F_T(M,q)$. These two tests have the null hypothesis of no structural break against an unknown number of breaks given some upper bound *M*. The $UD_{max}F_T(M,q)$ test applies equal weights to the individual tests while the $WD_{\max}F_T(M,q)$ test applies different weights to the individual tests such that the marginal p-values are equal across the number of breaks, *m* for $1 \leq m \leq M$.[13] The break points are obtained using the global minimization of the sum of squared residuals. Conditional on both test statistics being significant at the 5% level, which indicates the presence of at least one break, we then decide on the number of breaks by sequentially examining the $\sup - F(l+1|l)$ test statistics starting from $l=1$. We also set the maximum number of structural changes to M=4. The results are reported in Table 3.

Contrary to the results based on the Quandt (1960) and Bai (1997) tests, we find evidence of a single structural break in the FS series, with 1987 as the break date. However, once again, this date is different from the NFA implementation date

[12] See Section 5.5 on 'Summary and Practical Recommendations' in Bai and Perron (2003).

[13] For details on the application of weights to the individual tests in $WD_{\max}F_T(M,q)$ see Bai and Perron (1998).

suggesting that the agreement has no effect in influencing the firearm related suicide rate. The mean of the firearm related suicide growth rate appears to have decreased significantly after 1987, with the rate falling by more than 7 per cent. For the remaining series, we fail to identify the presence of any structural breaks.

Table 3: Tests for Multiple Breaks at Unknown Points in the Sample in the Mean Based on Bai and Perron (1998)

	Double maximum tests			
Series	UD*max*			WD*max*
FS	10.45			10.68
FH	3.97			4.72
NFS	3.10			5.33
NFH	4.10			5.36
	$\sup - F(l+1 \mid l)$ *test statistics*			
	F(2|1)	F(3|2)	F(4|3)	Estimated break dates and 90% confidence intervals
FS	3.15	4.07	4.07	1987 [1978; 2001]

	Estimated mean of growth rates (in percentage) by sub-sample	
	1916-1986	1987-2004
FS	-0.23	-7.80

Note: The critical values of UDmax and WDmax tests at the 5% level are 8.88 and 9.91, respectively. The critical values of the $\sup - F(l+1 \mid l)$ test at the 5% (10%) level are 10.13 (8.51), 11.14 (9.41) and 11.83 (10.04) for l = 1, 2 and 3 respectively.

6. Conclusion

This paper takes a closer look at the effects of the National Firearms Agreement on gun deaths. Using a battery of structural break tests, there is little evidence to suggest that it had any significant effects on firearm homicides and suicides. In addition, there also does not appear to be any substitution effects – that reduced access to firearms may have led those bent on committing homicide or suicide to use alternative methods.

Since the 1996 Port Arthur massacre, two other shooting incidents have attracted much media attention in Australia. An incident on 21 October 2002 at Monash University, in which a gunman killed two people and wounded five, prompted the National Handgun Buyback Act of 2003. Under this scheme that ran from July to December 2003, 70,000 handguns were removed from the community at a cost of approximately A$69 million. Another shooting on 18 June 2007, in which a lone gunman killed a man who had come to the aid of an assault victim and seriously wounded two others in Melbourne’s central business district during morning rush hour, renewed calls for tougher gun controls. Although gun buybacks appear to be a logical and sensible policy that helps to placate the public’s fears, the evidence so far suggests that in the Australian context, the high expenditure incurred to fund the 1996 gun buyback has not translated into any tangible reductions in terms of firearm deaths.

Appendix A: ABS Causes of Death Data 1915 to 2004

Year	Rate of firearm suicide	Rate of firearm homicide	Rate of non-firearm suicide	Rate of non-firearm homicide
1915	4.031636	0.5215052	9.166456	1.303763
1916	3.702076	0.4248284	7.970589	1.416094
1917	3.056176	0.4250311	7.104091	1.052458
1918	3.280707	0.4771938	6.621064	1.312283
1919	2.984728	0.8280212	7.529215	1.309429
1920	3.693712	0.6156186	8.170938	1.492409
1921	3.226317	0.366627	8.15745	1.026556
1922	2.190349	0.4308883	7.378962	1.005406
1923	2.599515	0.2810287	7.921496	1.334886
1924	3.080288	0.4646245	8.15674	1.393873
1925	2.929672	0.2693951	8.856365	1.128092
1926	2.922548	0.5779049	8.817177	1.122787
1927	3.234943	0.5661151	8.734347	1.213104
1928	2.983081	0.5077584	9.345927	1.253528
1929	3.284389	0.7507175	8.99297	1.094796
1930	3.806511	0.6808395	10.78512	0.9129438
1931	3.784579	0.7661091	8.886867	1.19513
1932	2.645654	0.4561473	8.818847	0.9274994
1933	3.092081	0.4977496	8.823744	0.9653326
1934	3.204859	0.4193273	9.165298	1.198078
1935	3.107225	0.6095514	8.652657	0.966362
1936	3.466909	0.457337	8.173054	0.9736851
1937	2.984404	0.5705478	7.563416	0.9948013
1938	3.218072	0.3478996	7.595809	0.8262616
1939	2.82731	0.3874993	8.381467	0.6888877
1940	3.182049	0.5824285	7.216432	0.7813066
1941	2.236319	0.2391033	6.540178	0.6047907
1942	2.033218	0.2367445	6.238914	1.08624
1943	1.644804	0.2626158	5.487288	0.7740255
1944	1.627971	0.4240934	5.759462	1.026032
1945	1.677559	0.365275	5.993215	0.6764351
1946	2.330828	0.3750758	7.474726	0.830525
1947	2.493615	0.4090056	7.348907	0.7916238
1948	2.257172	0.3891676	7.303379	0.7523907
1949	2.655517	0.240261	7.119314	0.6702018
1950	2.298655	0.1589496	6.99378	0.8803359
1951	2.600402	0.4393373	6.958153	0.9142966
1952	3.04523	0.5094681	7.595706	0.9610421
1953	2.790583	0.3743465	8.088153	0.9301944
1954	3.227052	0.4339828	7.555753	0.8679657
1955	2.967479	0.4130557	7.315433	0.9891596
1956	2.938817	0.3288928	7.893427	0.9972879
1957	3.277961	0.5186648	8.858794	0.7987437
1958	3.423985	0.5486504	8.839367	0.9753785
1959	3.072646	0.5469111	8.014733	0.9347208
1960	2.754253	0.4963494	7.873465	0.9732341
1961	2.910431	0.5308929	8.930377	0.8152998
1962	3.174369	0.5957761	10.50055	0.9402092
1963	3.068387	0.5570584	12.61144	0.794493
1964	3.089543	0.5731325	11.41787	0.9313403
1965	2.906654	0.5005416	11.87249	0.9220503
1966	2.931161	0.5431269	11.06082	0.7845167
1967	3.02566	0.5593657	12.04331	0.8220981
1968	2.939551	0.6911693	9.776299	0.8910255

1969	2.691019	0.4403485	9.557194	0.8073056
1970	3.19812	0.5676662	9.20259	0.9514406
1971	3.07639	0.5969114	10.22402	1.147907
1972	2.863873	0.4735537	9.350807	1.172609
1973	3.013802	0.5923935	8.300913	1.281051
1974	3.053364	0.5975556	8.365779	1.165962
1975	2.864753	0.6334127	8.133595	0.9789106
1976	3.057062	0.6057115	7.66047	1.410951
1977	2.959365	0.7046107	8.074839	1.197838
1978	3.252258	0.4805263	7.848597	1.288368
1979	3.602988	0.702686	7.99822	1.102253
1980	3.497704	0.7009017	7.396895	1.197657
1981	3.296867	0.6030854	7.900419	1.299984
1982	3.602417	0.6980919	8.1005	1.198611
1983	3.3001	0.597656	7.899453	1.299252
1984	3.40193	0.8023421	7.599783	1.097604
1985	3.502591	0.6017109	8.100929	1.39977
1986	3.402348	0.5993127	9.002175	1.398396
1987	3.498551	0.6025625	10.2989	1.40188
1988	3.199823	0.7016625	10.10152	1.699717
1989	2.700064	0.4995714	10.69915	1.39761
1990	2.801034	0.4980918	9.69814	1.500135
1991	2.898628	0.497569	9.800951	1.400136
1992	2.800854	0.600183	10.90047	1.400427
1993	2.399942	0.3339542	10.69785	1.341477
1994	2.402724	0.274437	9.297252	1.33858
1995	2.202331	0.3209428	10.39744	1.482977
1996	2.102594	0.5406671	10.90072	1.163253
1997	1.798293	0.4050209	12.90127	1.328468
1998	1.202484	0.2885961	13.09906	1.23455
1999	1.400201	0.3275942	11.69828	1.484741
2000	1.200832	0.3132606	11.09987	1.336579
2001	1.298083	0.2575562	11.30157	1.339292
2002	1.099742	0.2138386	10.70211	1.405225
2003	1.001377	0.1861856	10.09931	1.333491
2004	0.8361743	0.1592713	9.591119	1.14974

Notes: Rates of homicide and suicide are expressed as total number of deaths per 100,000 of the Australian population.

References

Andrews, D. (1993). "Test for Parameter Instability and Structural Changes with Unknown Change Point" *Econometrica*, 61, pp.821-856.

Andrews, D. and W. Ploberger. (1994). "Optimal Tests When a Nuisance Parameter is Present Only under the Alternative," *Econometrica*, 62, pp.1383-1414.

Bai, J. and P. Perron. (1998). "Estimating and Testing Linear Models with Multiple Structural Changes" *Econometrica*, 66, pp.47-78.

Bai, J. and P. Perron. (2003). "Computation and Analysis of Multiple Structural Change Models" *Journal of Applied Econometrics*, 18, pp.1-22.

Baker, J. and S. McPhedran. (2006). "Gun Laws and Sudden Death: Did the Australian Firearms Legislation of 1996 Make a Difference?" *British Journal of Criminology*, Advance Access published on October 18, 2006.

Baker, J. and S. McPhedran. (2007). "Review and Critique of Chapman, S., P. Alpers, K. Agho and M. Jones. (2006)." Manuscript.

Box, G. and G. Jenkins. (1976). *Time Series Analysis, Forecasting and Control*. San Francisco: Holden Day.

Britt, C., G. Kleck and D. Bordua. (1996). "A Reassessment of the D.C. Gun Law: Some Cautionary Notes on the Use of Interrupted Time Series Designs for Policy Impact Assessment." *Law & Society Review*, 30, pp. 361-380.

Chapman, S., P. Alpers, K. Agho and M. Jones. (2006) "Australia's 1996 Gun Law Reforms: Faster Fall in Firearm Deaths, Firearm Suicides and a Decade without Mass Shootings," *Injury Prevention*, 12, pp. 365-372.

Hansen, B. (1997). "Approximate Asymptotic *p* Values for Structural-Change Tests." *Journal of Business & Economic Statistics*, 15, pp. 60-67.

Kennedy, D., A. Braga and A. Piehl. (1996). "Gun Buy-Backs: Where Do We Stand and Where Do We Go?" in *Under Fire: Gun Buy-Backs, Exchanges and Amnesty Programs*, edited by M. Plotkin, pp. 141-174. Washington: Police Executive Research Forum.

Kleck, G. and E. Patterson. (1993). "The Impact of Gun Control and Gun Ownership Levels on Violence Rates." *Journal of Quantitative Criminology*, 9, pp. 249-287.

Ljung, G. and G. Box. (1978). "On a Measure of a Lack of Fit in Time Series Models." *Biometrika*, 65, pp. 297-303.

Mouzos, J. (1999). "Firearm-Related Violence: The Impact of the Nationwide Agreement on Firearms." *Trends and Issues in Crime and Criminal Justice*, 116. Canberra: Australian Institute of Criminology.

Mouzos, J. (2003). "Australian Homicide Rates: a Comparison of Three Data Sources." *Trends and Issues in Crime and Criminal Justice* No. 261.

Neill, C. and A. Leigh. (2007). "Weak Tests and Strong Conclusions: A Re-Analysis of Gun Deaths and the Australian Firearms Buyback." Australian National University Discussion Paper No. 555.

Perron, P. (1989). "The Great Crash, the Oil Price Shock and the Unit Root Hypothesis." *Econometrica*, 57, pp. 1361-1401.

Piehl, A., S. Cooper, A. Braga and D. Kennedy. (2003). "Testing for Structural Breaks in the Evaluation of Programs." *Review of Economics and Statistics*, 85, pp. 550-558.

Quandt, R. (1960). "Tests for the Hypothesis That a Linear Regression System Obeys Two Separate Regimes." *Journal of the American Statistical Association*, 55, pp. 324-330.

Reuter, P. and J. Mouzos. (2003). "Australia: A Massive Buyback of Low-Risk Guns," in J. Ludwig and P. J. Cook, eds., *Evaluating Gun Policy: Effects on Crime and Violence.* Washington, DC: Brookings Institution Press.

Sherman, L. (2001). "Reducing Gun Violence: What Works, What Doesn't, What's Promising." *Criminal Justice*, 1, pp. 11-25.

Wellford, C., J. Pepper and C. Petrie, eds. (2004). *Firearms and Violence: A Critical Review*, Committee on Improving Research Information and Data on Firearms, National Academy of Sciences.

Zivot, E. and D. Andrews. (1992). "Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis." *Journal of Business & Economic Statistics*, 10, pp. 251-270.