

Time series robustness checks to test the effects of the 1996 Australian firearm law on cause-specific mortality

Benjamin Ukert¹  · Elena Andreyeva¹ · Charles C. Branas²

© Springer Science+Business Media B.V. 2017

Abstract

Objectives Many studies utilize time series methods to identify causal effects without accounting for an underlying time trend. We show that accounting for trends changes the conclusions in the study of Chapman et al. (*JAMA*, 316(3), 291–299, 2016), who evaluated the impact of the Australian firearm law in 1996. We also introduce a new empirical method that tests whether their empirical strategy can actually identify a causal effect that is also useful for panel analyses.

Methods We use national data from the Australian Bureau of Statistics, assembled in annual counts of: total firearm deaths, firearm suicides, and firearm homicides. These data are used in an independent re-analysis of the impact of the 1996 Australian firearm law that accounts for underlying stochastic trends. We then estimate a series of artificially created interruptions using interrupted times series analysis in a time frame before 1996, to test for changes in the slope of mortality across several years prior to the actual regulatory changes. This tests whether the empirical model produces effects in years other than the year of the intervention, thereby testing if the results can simply be replicated at random using other interruption years.

Results Controlling for stochastic trends produces less statistical evidence of the impact of the firearm law on firearm mortality than previously reported by Chapman et al. (*JAMA*, 316(3), 291–299, 2016). Introducing artificial interruptions in 1990 through 1995 produces statistically significant decreases in all firearm-related mortality

✉ Benjamin Ukert
bukert@wharton.upenn.edu

¹ Wharton, Leonard Davis Institute of Health Economics, and Perelman School of Medicine, University of Pennsylvania, 308 Colonial Penn Center, 3641 Locust Walk, Philadelphia, PA 19104-6218, USA

² Gelman Professor and Chair, Department of Epidemiology, Columbia University, Mailman School of Public Health, New York 10032 NY, USA

measures well above the expected type 1 error. Overall, 19 out of the 36 artificial interruption models we tested were found to be statistically significant, suggesting that the empirical model can be implemented in multiple non-intervention years with results similar to the true 1996 interruption year.

Conclusions Current evidence showing decreases in firearm mortality after the 1996 Australian national firearm law relies on an empirical model that may have limited ability to identify the true effects of the law.

Keywords Australia · Firearm regulation · Interrupted time series · Methods

Introduction

Many changes in law or policy are experienced by the entire population of a country or region. An important question for policymakers and scientists is the identification of the causal impact and magnitude of such events. Interrupted time series (ITS) methods are commonly used to identify the impact of an event experienced by an entire population. This method has a long history in criminology and the social sciences because it is straightforward to operationalize while also potentially capturing causal effects in the presence of no formal control population (Cook and Campbell 1979; Enders 2014; Greene 2003; McCleary et al. 2017). For example, D'Alessio and Stolzenberg (1995) estimated the impact of sentencing guidelines in Minnesota on jail sanctions, Pridemore et al. (2007, 2008) estimated the impact of unanticipated catastrophic events, such as the collapse of the Soviet Union and terror attacks in the USA on violence, and Messner et al. (2001) estimated the impact of economic deprivation on homicide arrests. Similarly, scientists have utilized ITS analyses to estimate the impact of firearm regulations in Australia, and the impact of macroeconomic shocks, such as the recession and austerity interruptions in Greece, on mortality (Branas et al. 2015; Chapman et al. 2006, 2016; Humphreys et al. 2017; Klieve et al. 2009; Laliotis et al. 2016).

In its simplest case, the ITS approach identifies potential causal effects by comparing the outcome variable's time trend (level) before the intervention to the time trend (level) in the post-intervention period. The difference in the pre and post time trend levels defines the causal impact of the intervention. In this approach, the null hypotheses states that the post period can be projected from the pre period, and that any change in the post period is, therefore, attributable to the causal effect. In other words, had the intervention not taken place, the trend would have continued uninterrupted. Several variations of time series models have been developed to accommodate different data generating processes and intervention types (Bernal et al. 2017; Dugan 2010; Shadish et al. 2002). For example, ITS methods account for seasonality with harmonic terms, and potential non-independence of observations (correlation) in the time series with autoregressive integrated moving average (ARIMA) models.

However, many ITS studies do not account for potential trends and seasonality that are unrelated to any intervention or event, thus violating the ITS assumption that the pre-intervention time trend can be projected in the post-intervention period (Chapman et al. 2006, 2016; Gagné et al. 2010; Spittal et al. 2012). For example, time series can experience a general upward (or downward) trend in certain time periods irrespective of any specific event. An empirical model ignoring such trends may lead to incorrect

inferences. In many cases, such a trend is a process that sometimes only depends on its previous level of observation. In this case, observations are not independent across time—a requirement for correctly identifying any causal effects of an intervention.

Violation of the independence of observations across time will lead to a time series process that is correlated across time (i.e., the mean of the time series is not constant over time, also referred to as a non-stationary process) and generate estimates that attribute the correlation of observations across time to the causal effect. Even in the case of a stationary time series, the estimated model needs to be correctly specified, which pertains to whether we expect changes in the dependent variable to occur gradually, that is, a change in slope over time, an immediate change in the dependent variable, or a jump in the dependent variable with a potential change in slope over time. Similarly, one has to define a model that has a temporal ordering related to any actual policy change. For instance, a law passed in 2010 should, therefore, be captured with an empirical model that identifies the effect of the policy beginning in 2010 and not earlier. There are several methods to test for stationarity of the time series process. However, there is no commonly utilized method to test whether an empirical model is correctly specified, especially in panel data analyses with a difference-in-differences empirical model.

We contribute in two ways to the existing literature. First, we highlight the need to test for whether a time series is stationary, defined as having a constant mean over time in the dependent variable. This is something that has not been addressed in many prior studies, and we show how estimates differ when one adjusts the data to a stationarity process. Second, we propose a study design that can help identify whether an empirical specification in the time series model correctly captures the causal effect of interest. This empirical specification has its roots in a thought experiment that a correctly specified model intending to capture the causal effect of an intervention should not provide statistically significant estimates (beyond the typically expected level of type 1 error) in a (stationary time series) period without an intervention. In other words, the specification tests the underlying ITS assumption that the pre-intervention trend can be projected in the post-intervention, which should be the case in a period of a missing intervention. Our method tests the validity of an empirical model by performing artificial law implementations in a sample that excludes observations from the post-intervention time. These regressions test whether the ITS assumption of a continuous projection of the pre-intervention trend hold in a period when no actual intervention happened. We would expect to find no statistically significant results in independent variables of interest in time periods with no intervention.

To this end, we take Chapman et al. (2016) as a case study to test whether their empirical specification passes typical time series tests for stationarity, and whether their empirical model is correctly specified, that is, captures the causal impact of the firearm law on cause-specific mortality. Chapman et al. (2016) look at the causal effect of the 1996 Australian firearm law on firearm-related mortality. The 1996 law was a response to a gun-involved massacre in Tasmania, after which the Australian government implemented stricter national firearm regulations. These regulations were implemented between June 1996 and 1998 and included a ban on certain types of long guns, mandatory licensing for firearm owners, and gun registration. Moreover, a comprehensive gun buyback program was implemented in all states in January 1997 to allow current gun owners to sell their newly prohibited guns.

Several studies have evaluated the impact of the 1996 Australian gun regulations on firearm-related outcomes relying on ITS analysis (Chapman et al. 2006, 2016; Klieve

et al. 2009; Lee and Suardi 2010). However, different regressions and ITS specifications have led to inconclusive results. On the one hand, there is evidence that the regulatory changes reduced firearm-related mortality and prevented mass shootings in the decade after the introduction of the gun laws (Chapman et al. 2006, 2016). On the other hand, studies also found limited effects of the gun regulations on firearm-related outcomes (Klieve et al. 2009; Lee and Suardi 2010).

To extrapolate policy relevant conclusions from the Australian firearm law and any other policy, the use of the ITS approach needs to satisfy the underlying assumptions of a stationary time series and a correctly specified model that can capture the actual impact of the event. Current studies may not necessarily employ an estimation strategy that can appropriately isolate the causal effect of the gun regulations themselves. They may simply capture a decreasing trend, or an acceleration of a decreasing trend in the firearm-related mortality rate that would have occurred even without the set of nationally implemented gun regulations (see Fig. 1a). Moreover, the empirical model may be mis-specified even if one has a stationary time series.

Our results provide two salient findings. First, Chapman et al.'s model does not pass our proposed empirical checks, given that it does not account for non-stationarity of the time series. Artificial law implementations prior to 1996 still produce statistically significant results for decreases in firearm mortality. As such, it appears that the decrease in Australian firearm mortality following the 1996 passage of new firearm regulations was part of an existing downward trend, and was not necessarily caused by

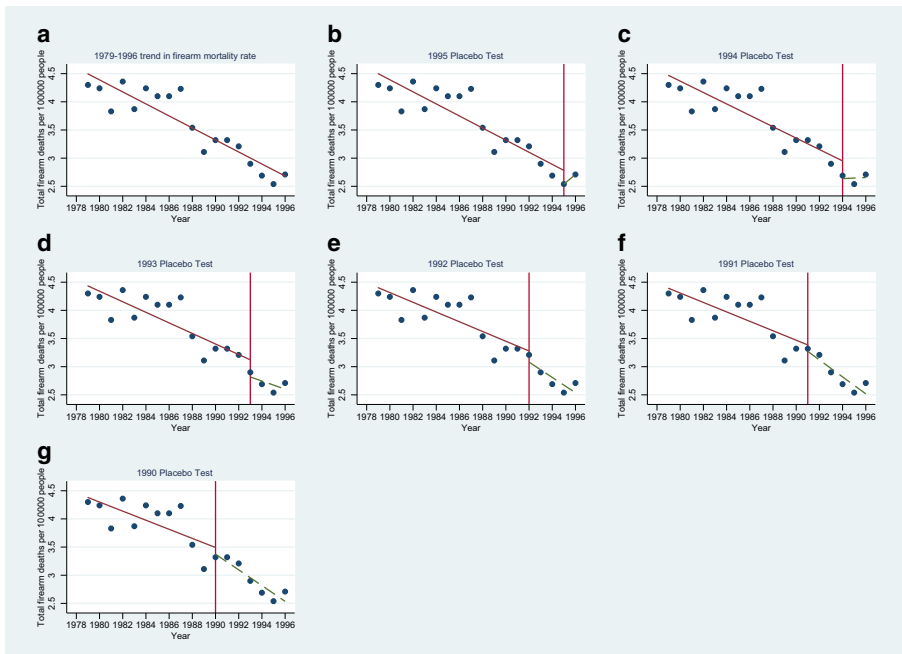


Fig. 1 Trends in firearm mortality and pre and post artificial law implementations. Each dot represents an observation, the solid vertical line indicates the artificial implementation year, the solid downward sloping line represents a trend in total firearm mortality rate before the corresponding artificial implementation year, and the dotted downward sloping line represents a trend in total firearm mortality rate after the corresponding artificial implementation year

the 1996 gun regulations. Thus, one cannot take their results as reliable evidence on the magnitude of the impact of the Australian firearm law. Once we account for non-stationarity using an ARIMA model, the coefficients on total firearm mortality, firearm suicides, and firearm homicides remain negative, but only total firearm and firearm suicide mortality remain statistically significant. These findings suggest some benefits of the law in terms of firearm-related mortality, though one should remain cautious given the low level of statistical significance. Additionally, we do not find any statistically significant estimates in the ARIMA model for all regressions that apply an artificial law implementation, suggesting that the ARIMA specification provides a correctly specified model to plausibly estimate the causal effects of the 1996 firearm regulations on firearm mortality.

Data and methods

Background on time series and stationarity of the time series

A large existing literature discusses techniques to identify causal effects in a time series (Campbell and Stanley 1966; Hansen 2001; McCleary et al. 2017; Shadish et al. 2002). Specifically, a time series model analyzes the impact of an intervention on a dependent variable that is repeatedly collected over time. The underlying assumption of the ITS approach is that the pre-intervention trend would have continued into the post-intervention period had the intervention not taken place. Thus, the causal impact of an intervention is captured by the difference in trends from the pre- to post-intervention periods.

Given that time series are disposed to having trends that are unrelated to interventions, these have to be properly described, so that one does not interpret existing random trends as the causal impact of an intervention. Two forms, deterministic and stochastic trends, can affect the level of a series over time. These need to be accounted for in a time series because the ITS approach compares the mean of the outcome variable before and after an intervention. For this to be a reasonable strategy, the time series before and after the intervention must possess a mean that is constant, that is, free of the influence of the trend. A series with a constant mean is stationary in its level, while a series that lacks a constant mean is non-stationary.

A common approach to test for non-stationarity is the use of the augmented Dickey–Fuller test and the Phillips–Perron test. While a deterministic trend can be accounted for with a linear continuous time variable, this does not address the stochastic trends—such as a random walk. A random walk drifts up and down in a non-consistent direction for an extended period and a standard method of removing a random walk is to convert a time series to first differences. One such empirical specification is the ARIMA model that allows for differencing the time series and also allows the direct specification of the number of (lags) past periods and residual terms that affect the time series.

Data and model description

We obtained annual data on firearm-related deaths in Australia from the Australian Bureau of Statistics (ABS) from 1979 to 2013. These data had been used in the

previous study (Chapman et al. 2016). We replicate Chapman et al.'s (2016) ITS model using the same data. Their empirical model employs a negative binomial regression estimation strategy to account for overdispersion in the dependent variable:

$$\ln(d_i) = \beta_{01}\ln(n_i) + \beta_{02} + \beta_{12}Year_i + \beta_{22}Law_i + \beta_{32}Year_i \times Law_i + \varepsilon_i \quad (1)$$

where Law_i , the variable of interest, is a dummy variable equal to 1 after the law implementation, $Year_i$ is a continuous variable of the calendar year and is coded as calendar year since the year of the gun law implementation, d_i indicates the total number of firearm-related deaths, firearm suicides, or firearm homicides in year i , n_i indicates the population at risk in year i , and ε_i represents the residual in year i . β_{22} represents the impact of the implementation of the firearm law in year i and measures the difference in the firearm-related mortality mean from before and after the law, that is, measuring the change in the form of an intercept movement. Finally, β_{32} measures the impact of the law on the firearm-related mortality as time progresses, that is, the change in trend after the implementation of the law.

Notice that this model cannot account for a trend described as a random walk. To account for a stochastic trend, we calculate the firearm-related mortality rates directly and test for stationarity with the augmented Dickey–Fuller test before specifying an empirical model, which reveals that the data is non-stationary and exhibits a random walk in the rate of firearm-related mortality. Thus, in addition to replicating Chapman et al.'s (2016) model using a negative binomial regression, we propose a model that takes the first difference of the data and estimates an ARIMA model of autoregressive order 1 based on autocorrelation charts and residual values.

Replication results

Table 1 illustrates the replication results from Eq. (1) for the negative binomial and the ARIMA models for total firearm deaths, firearm suicides, and firearm homicides. The first three rows (panel 1) represent the results from the ARIMA model, while rows four through six (panel 2) represent the results for the negative binomial model. In terms of the coefficients on the change in the intercept, Law_i , both are relatively similar in magnitude between panels 1 and 2. However, they are more statistically significant in the negative binomial specification. Specifically, we observe statistical significance at the 1% level for total deaths and suicides, and 10% for homicides in the negative binomial model. However, we only find statistical significance at the 10% level for total deaths and suicides, and lower statistical significance for homicides in the ARIMA model. Similarly, the coefficients capturing trend change after the law, $Law_i * Year_i$, are all significant at the 10% level in the negative binomial model, but are all insignificant in the ARIMA model.

These results show that accounting for trends in a time series can significantly alter the empirical conclusions as well as the magnitude of the causal impact of an intervention. The ARIMA model provides more marginal evidence that the firearm law decreased the total and suicide firearm rates, while also suggesting that firearm homicides were not affected. Additionally, the ARIMA model does not provide any evidence of a change in trend after the intervention, contrary to the findings in the

Table 1 The impact of actual law implementations on cause-specific mortality. Autoregressive integrated moving average (ARIMA) and negative binomial replication results

Firearm mortality rate	Panel 1: ARIMA			Panel 2: negative binomial		
	Total	Suicides	Homicides	Total	Suicides	Homicides
Law	-0.655* (0.396)	-0.547* (0.261)	-0.107 (0.146)	-0.402*** (0.065)	-0.427*** (0.058)	-0.262* (0.136)
Law*Year	0.0304 (0.0765)	0.033 (0.0476)	-0.0025 (0.0295)	-0.0196*** (0.006)	-0.019*** (0.006)	-0.0254* (0.014)
Sample size	34	34	34	35	35	35

Coefficients and standard errors in parentheses for the ARIMA model are presented in columns 1–3 and estimates for the negative binomial model are presented in columns 4–6. The coefficient of the change in intercept is shown in the row labeled Law and the coefficient of the change in the trend is represented in the row labeled Law*Year. Heteroskedasticity robust standard errors clustered at the state level are in parentheses.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

negative binomial model. Thus, we conclude that the law implementation did not impact the trend in firearm-related mortality and did not necessarily change the overall mean of firearm-related mortalities.

Robustness checks

Next, we perform robustness checks for both the negative binomial and ARIMA models from Table 1. Here, we continue to rely on the same ABS data used by Chapman et al. (2016). However, unlike Chapman et al., we utilize a subsample of the data that only includes pre-1997 observations, that is, the years from 1979 to 1996, thus, the period before the actual firearm law took effect.

Recall that the goal of the specification checks is to test whether the empirical model captures a causal effect in periods of no intervention. Thus, we test the assumption of the ITS model as to whether one would observe a continuous trend in the absence of an intervention in a period of no intervention. In other words, we test whether Chapman et al.'s empirical model is actually able to capture the causal impact of the firearm law on cause-specific mortality in Australia—setting aside the potential non-stationarity problem. If their estimation strategy is correctly specified, we would expect to not find statistically significant results in a pre period of an artificial law implementation (at most, we would expect to capture a stochastic trend). On the contrary, if we find consistent statistically significant estimates, then the proposed model in Chapman et al. (2016) captures trend changes in a period where no firearm law was enacted, therefore, questioning whether the trend changes during the intervention period can be fully attributed to the intervention.

We perform this test by implementing artificial law interruptions in the years from 1990 to 1995 in separate regressions. For example, an artificial implementation in 1990 suggests a pre artificial law period from 1979 to 1989 and a post artificial law time period from 1990 to 1996. Again, we first emulate Chapman et al.'s ITS analysis

using a negative binomial regression estimation strategy to account for overdispersion:

$$\ln(d_i) = \beta_{01}\ln(n_i) + \beta_{02} + \beta_{12}Year_i + \beta_{22}Law_i + \beta_{32}Year_i \times Law_i + \varepsilon_i \quad (2)$$

where all variables are the same as previously described in Eq. (1), except that the Law_i variable is equal to 1 in an earlier time period (i.e., 1990 and onwards for the artificial 1990 intervention, etc.).

We also estimate our preferred ARIMA model with the artificial law interruptions for all firearm outcome measures. Overall, we conduct separate analyses for total firearm suicide, total firearm homicide, and total firearm deaths. These artificial interventions lead to six models for each dependent variable, leading to 18 estimates, and 36 coefficients of interest in all negative binomial models and, also 36 coefficients of interest in the ARIMA model.

A correctly specified model capturing the impact of the firearm law suggests that the artificial interruptions should not find statistically significant estimates above the statistically expected type 1 error level. If the above regressions find statistical significance in quantities larger than expected by chance, that is, 5% for a p -value of 0.05, then we can suspect that the actual impact of the law in 1996 is unclear in the currently specified empirical model. Statistically significant impacts on the coefficients of interest in the pre-1997 period may imply that Chapman et al. (2016) only capture time-varying impacts that are unrelated to the actual passage of the law, such as a trend that is present in a non-stationary time series. On the other hand, if all of our artificial law implementations show limited statistically significant results, then this should provide strong support that their estimation strategy is internally valid and presents actual causal impacts of the law.¹

Results

Figure 1a presents the overall trend in total firearm mortality from 1978 to 1996. Panels b–g illustrate the before and after trends in the total firearm mortality rate with breaks in the time series for artificial laws from 1990 to 1995. All panels reveal that a simple descriptive comparison of before and after trends suggests some sort of discontinuity in firearm mortality rates, and sometimes a potential change in trend after the artificial law. Given that there is no actual law implementation over this period, these results suggest potential challenges to the identification strategies of other studies relying on variation in firearm mortality after the implementation year of the law under study, that is, in 1997.

In each column of Table 2, we present regression results of the artificial law implementations on cause-specific mortality in a different year. Panel 1 of Table 2 presents estimates and standard errors in parentheses on the impact of artificial implementations of gun laws on total firearm mortality. Artificial law implementations

¹ We also stratified the pre-treatment period into two unique periods, 1979–1986 and 1986–1996, and estimated artificial interruptions in 1982 and 1993, respectively. We similarly find statistically significant impacts of the law and the interaction in Chapman et al.'s (2016) model well above expected levels. Such subsample analysis speaks to the multiple comparison problem outlined in Hansen (2001), which may be a potential concern in the results of the five artificial interruptions in the same time series.

Table 2 The impact of artificial law implementations in 1990 to 1995 on cause-specific mortality: Chapman et al. (2016) negative binomial model

Panel 1		Total firearm mortality count		1991		1992		1993		1994		1995	
Artificial law in year	1990	-0.505** (0.201)	-0.575*** (0.201)	-0.564*** (0.200)	-0.482*** (0.174)	-0.143 (0.137)	0.016 (0.109)						
Law	1990	-0.087** (0.040)	-0.108*** (0.041)	-0.103** (0.042)	-0.065 (0.040)	0.020 (0.052)	0.101 (0.103)						
Law*Year	18	18	18	18	18	18	18						
Sample size	18	18	18	18	18	18	18						
Panel 2		Firearm suicide count		1991		1992		1993		1994		1995	
Artificial law in year	1990	-0.555*** (0.155)	-0.622*** (0.154)	-0.614*** (0.154)	-0.496*** (0.143)	-0.199* (0.116)	-0.062 (0.095)						
Law	1990	-0.099*** (0.031)	-0.119*** (0.031)	-0.119*** (0.033)	-0.085*** (0.034)	-0.019 (0.045)	0.018 (0.091)						
Law*Year	18	18	18	18	18	18	18						
Sample size	18	18	18	18	18	18	18						
Panel 3		Firearm homicide count		1991		1992		1993		1994		1995	
Artificial law in year	1990	-0.305 (0.502)	-0.424 (0.517)	-0.408 (0.506)	-0.473 (0.402)	0.104 (0.293)	0.359* (0.213)						
Law	1990	-0.031 (0.100)	-0.069 (0.106)	-0.044 (0.107)	0.026 (0.093)	0.210* (0.111)	0.489*** (0.205)						
Law*Year	18	18	18	18	18	18	18						
Sample size	18	18	18	18	18	18	18						

Table 2 presents coefficients and standard errors from Chapman et al.'s model with artificial law implementations in a dataset limited to 1979–1996. The coefficient of the change in intercept is shown in the row labeled Law and the coefficient of the change in the trend is represented in the row labeled Law*Year. Heteroskedasticity robust standard errors clustered at the state level are in parentheses.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

in 1990, 1991, 1992, and 1993 lead to negative and significant effects on total firearm mortality. However, in reality, there is no law implementation during these years, suggesting, therefore, that the regression model only captures a preexisting downward trend in firearm mortality.

Similarly, panel 2 shows statistically significant artificial law coefficients in 1990–1993 on firearm suicide mortality. Thus, the artificial laws capture a decreasing suicide mortality trend in all regressions, rather than a causal effect. The only exceptions occur in panel 3, where all artificial law implementations for the years 1990–1994 present insignificant coefficients, but do show statistically significant effects in 1994 and 1995. Panel 3 estimates are closest to passing the robustness checks with limited rejections of the null hypothesis.

Nevertheless, in total, we would expect a rejection of the null hypothesis by chance in as much as 10% of all cases, which, when rounded up, would allow for an acceptable rejection of four times ($10\% \times 36 = 3.6$). However, we reject the null hypothesis 19 times for the 36 coefficients across Table 1. This leads to a rejection level of over 50%, and questions the interrupted series conclusions presented by Chapman et al. (2016) to identify the causal effect of the 1996 firearm law.

Table 3 is constructed in a similar way as Table 2, but illustrates regression results of the artificial law implementations on cause-specific mortality using the ARIMA model. Panels 1, 2, and 3 present results for the impact of artificial implementations of gun laws on total firearm mortality rate, firearm suicide rate, and firearm homicide rate, respectively. In all columns, artificial law implementations in 1990 to 1995 lead to statistically insignificant effects on all firearm mortality types. These results are in stark contrast to those presented in Table 2 and suggest that the ARIMA model passes the specification checks, thus yielding causal estimates on the impact of the firearm law. Specifically, contrary to Table 2, these results suggest that changes in the total firearm mortality rate, reported in panel 1 of Table 1, indeed occurred due to the new firearm law and are not just a part of the existing downward trend.

Discussion

Concerns remain in regards to the time series estimation strategy utilized by Chapman et al. (2016) to evaluate the effect of the Australian national firearm law on firearm mortality, especially when one does not test for an underlying trend in the time series. In addition, the law was not immediately implemented after its passage in May of 1996 as each Australian state had to ratify the law, which some states did not do until mid 1997. Therefore, the empirical model capturing an immediate drop in firearm-related mortality might not be the best approach to identify a potentially gradual impact of the law. To this end, we used a new methodological approach to test whether the empirical model is potentially mis-specified. Our results indicate that the model passes the robustness checks when one relies on an ARIMA model, implying that there was, indeed, an immediate drop in total firearm-related mortality. However, Chapman et al.'s (2016) model does not pass the specification check with artificial law implementations, highlighting the need to adjust their model by accounting for a potential non-stationary time series with either an ARIMA or another type of model. For example, one could adjust the empirical specification by directly controlling for a stochastic trend in the form of harmonic terms (Bernal et al. 2017).

Table 3 The impact of artificial law implementations in 1990 to 1995 on cause-specific mortality: ARIMA model

Panel 1		Firearm mortality rate				
Artificial law in year	1990	1991	1992	1993	1994	1995
Law	0.235 (1.291)	- 0.023 (6.657)	- 0.271 (1.048)	- 0.257 (1.230)	- 0.024 (1.232)	0.208 (1.071e + 07)
Law*Year	- 0.004 (0.155)	- 0.050 (0.159)	- 0.053 (0.183)	0.009 (0.218)	0.115 (0.564)	0.292 (1.074e + 07)
Sample size	17	17	17	17	17	17
Panel 2		Firearm suicide rate				
Artificial law in year	1990	1991	1992	1993	1994	1995
Law	- 0.034 (1.058)	- 0.254 (8.504)	- 0.394 (1.226)	- 0.347 (1.063)	- 0.185 (2.874)	- 0.019 (8.144e + 06)
Law*Year	- 0.0417 (0.152)	- 0.0767 (0.161)	- 0.0825 (0.180)	- 0.0536 (0.268)	0.000322 (1.204)	0.0570 (5.166e + 06)
Sample size	17	17	17	17	17	17
Panel 3		Firearm homicide rate				
Artificial law in year	1990	1991	1992	1993	1994	1995
Law	0.262 (0.447)	0.219 (1.733)	0.112 (0.249)	0.080 (0.356)	0.148 (0.228)	0.228 (400,235)
Law*Year	0.035 (0.041)	0.024 (0.044)	0.026 (0.052)	0.059 (0.049)	0.113 (0.118)	0.229 (408,355)
Sample size	17	17	17	17	17	17

Table 3 presents coefficients and standard errors from the ARIMA model with artificial law implementations in a dataset limited to 1979–1996. The coefficient of the change in intercept is shown in the row labeled Law and the coefficient of the change in the trend is represented in the row labeled Law*Year. Heteroskedasticity robust standard errors clustered at the state level are in parentheses.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Overall, we provide an empirical test that can help identify potential concerns in a time series model, which does not correctly control for underlying trends in the data. This approach can also be applied to panel and repeated cross-sectional analyses to test the appropriateness of a proposed model. Additionally, we believe that even in cases where an ARIMA model may not pass the proposed robustness checks, such violation can be corrected with an adjustment of the empirical model in terms of how it captures the causal effect, because the violations can occur due to a mis-specified empirical model. A recent study provides a detailed explanation of the types of regression models applicable to ITS designs, and we believe that a more careful model selection will result in more robust estimates that pass our robustness checks (Bernal et al. 2017).

Another potential weakness of the time series approach, not addressed in this study, is that other unobservable factors may be correlated with the timing of the intervention that will influence the magnitude and level of significance of the results. For example,

macroeconomic fluctuations correlated with the passage of the firearm law, such as changes in the unemployment rate and fluctuation in the cost of guns, could have led to lower firearm-related mortality after 1996 (Carmichael and Ward 2001; Lee and Holoviak 2006; Tang and Lean 2007). Not controlling for these factors can lead to potentially overstated coefficients and spurious statistical significance of the impact of the firearm law on firearm-related mortality. A potential way to control for such factors is to add them to the model as covariates.

Similarly, the underlying ITS assumption of a projection of the pre trend in the post period had the intervention not taken place may, nevertheless, be too strong. The evaluation of the Australian gun regulations could be strengthened by introducing a comparison group, which would allow controlling for any common phenomenon that could have affected changes in the firearm mortality. To this end, surrounding countries with similar demographics may be used as control states, such as New Zealand, or estimates may rely on the synthetic control method (Abadie et al. 2010).

Other promising avenues include the use of monthly state-level data, which would allow researchers to identify the exact timing of the law implementation for each Australian state, increase sample size, precision, and generate more accurate variation attributable to the law implementation. One approach taken by Andreyeva and Ukert (2017) estimates the impact of the Australian firearm law that relies on state variation in firearm-related mortality and assumes that the impact of the law is proportional to the states' firearm-related mortality rate prior to the law implementation. Such an approach allows controlling for common trends unrelated to the intervention.

Conclusion

This paper contributes by highlighting the need to conduct time series stationarity tests before moving to data analyses. It also provides a new set of robustness checks to assess the validity of any empirical model. First, we replicate an interrupted time series (ITS) regression model from Chapman et al. (2016) using a negative binomial specification to account for overdispersion. Second, we test for non-stationarity using the augmented Dickey–Fuller test and the Phillips–Perron test, which reveal that the data is non-stationary, that is, exhibits a stochastic random walk. Thus, we employ the autoregressive integrated moving average (ARIMA) model to account for the trends and their relation to past observations. The results are similar in magnitude to the negative binomial model, but are less significant (10% level relative to 1% in the negative binomial model).

Third, we introduce an empirical strategy that tests whether the model from Chapman et al. (2016) is appropriately specified by testing for statistically significant estimates in a time period that excludes the intervention and post-intervention period. The model estimates artificial law implementations in five prior years, 1990 through 1995, yielding different artificial post-intervention lengths. Overall, we find that all three measures of firearm mortality (total mortality, suicides, and homicides) fail to pass the robustness checks in the negative binomial specification proposed by Chapman et al. (2016), but pass the specification checks in our ARIMA model.

We conclude that policymakers can rely only on select aspects of previous studies that show statistically significant impacts on cause-specific mortality and that future studies should better account for potential time series non-stationarity

and perform some of the robustness checks we incorporate here. As a conservative rule, we also propose that the use of our artificial interventions test should not result in statistically significant coefficients in an amount larger than 5% of the total artificial law coefficients of interest. Beyond stationarity concerns in time series analyses, our robustness checks can yield useful insight as to whether an appropriate conclusion has been reached, and can, thus, strengthen findings and the subsequent policy implications.

References

- Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, *105*(490), 493–505.
- Andreyeva, E., & Ukert, B. (2017). Do firearm regulations work? Evidence from the Australian National Firearms Agreement.
- Bernal, J. L., Cummins, S., & Gasparrini, A. (2017). Interrupted time series regression for the evaluation of public health interventions: A tutorial. *International Journal of Epidemiology*, *46*(1), 348–355.
- Branas, C. C., Kastanaki, A. E., Michalodimitrakis, M., Tzougas, J., Kranioti, E. F., Theodorakis, P. N., Carr, B. G., & Wiebe, D. J. (2015). The impact of economic austerity and prosperity events on suicide in Greece: A 30-year interrupted time-series analysis. *British Medical Journal Open*, *5*(1), 1–11.
- Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Skokie, IL: Rand McNally.
- Carmichael, F., & Ward, R. (2001). Male unemployment and crime in England and Wales. *Economics Letters*, *73*(1), 111–115.
- Chapman, S., Alpers, P., Agho, K., & Jones, M. (2006). Australia's 1996 gun law reforms: Faster falls in firearm deaths, firearm suicides, and a decade without mass shootings. *Injury Prevention*, *12*(6), 365–372.
- Chapman, S., Alpers, P., & Jones, M. (2016). Association between gun law reforms and intentional firearm deaths in Australia, 1979–2013. *JAMA*, *316*(3), 291–299.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston, MA: Houghton Mifflin.
- D'Alessio, S. J., & Stolzenberg, L. (1995). The impact of sentencing guidelines on jail incarceration in Minnesota. *Criminology*, *33*(2), 283–302.
- Dugan, L. (2010). Estimating effects over time for single and multiple units. In A. Piquero & D. Weisburd (Eds.), *Handbook of quantitative criminology* (pp. 741–763). New York, NY: Springer.
- Enders, W. (2014). *Applied econometric time series* (4th ed.). New York, NY: Wiley.
- Gagné, M., Robitaille, Y., Hamel, D., & St-Laurent, D. (2010). Firearm regulation and declining rates of male suicide in Quebec. *Injury Prevention*, ip-2009.
- Greene, W. H. (2003). *Econometric analysis*. Delhi, India: Pearson Education.
- Hansen, B. E. (2001). The new econometrics of structural change: Dating breaks in U.S. labor productivity. *The Journal of Economic Perspectives*, *15*, 117–128.
- Humphreys, D. K., Gasparrini, A., & Wiebe, D. J. (2017). Evaluating the impact of Florida's "stand your ground" self-defense law on homicide and suicide by firearm: An interrupted time series study. *JAMA Internal Medicine*, *177*(1), 44–50.
- Klieve, H., Barnes, M., & De Leo, D. (2009). Controlling firearms use in Australia: Has the 1996 gun law reform produced the decrease in rates of suicide with this method? *Social Psychiatry and Psychiatric Epidemiology*, *44*(4), 285.
- Laliotis, I., Ioannidis, J. P., & Stavropoulou, C. (2016). Total and cause-specific mortality before and after the onset of the Greek economic crisis: An interrupted time-series analysis. *The Lancet Public Health*, *1*(2), e56–e65.
- Lee, D. Y., & Holoviak, S. J. (2006). Unemployment and crime: An empirical investigation. *Applied Economics Letters*, *13*(12), 805–810.
- Lee, W.-S., & Suardi, S. (2010). The Australian firearms buyback and its effect on gun deaths. *Contemporary Economic Policy*, *28*(1), 65–79.
- McCleary, R., McDowall, D., & Bartos, B. J. (2017). *Design and analysis of time series experiments*. New York, NY: Oxford University Press.

- Messner, S. F., Raffalovich, L. E., & McMillan, R. (2001). Economic deprivation and changes in homicide arrest rates for white and black youths, 1967–1998: A national time-series analysis. *Criminology*, *39*(3), 591–614.
- Pridemore, W. A., Chamlin, M. B., & Cochran, J. K. (2007). An interrupted time-series analysis of Durkheim's social deregulation thesis: The case of the Russian Federation. *Justice Quarterly*, *24*(2), 271–290.
- Pridemore, W. A., Chamlin, M. B., & Trahan, A. (2008). A test of competing hypotheses about homicide following terrorist attacks: An interrupted time series analysis of September 11 and Oklahoma City. *Journal of Quantitative Criminology*, *24*(4), 381–396.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. New York, NY: Houghton Mifflin.
- Spittal, M. J., Pirkis, J., Miller, M., & Studdert, D. M. (2012). Declines in the lethality of suicide attempts explain the decline in suicide deaths in Australia. *PLoS One*, *7*(9), e44565.
- Tang, C. F., & Lean, H. H. (2007). Will inflation increase crime rate? New evidence from bounds and modified Wald tests. *Global Crime*, *8*(4), 311–323.

Dr. Benjamin Ukert is a postdoctoral researcher at the Leonard Davis Institute of Health Economics and the Injury Science Center at the Perelman School of Medicine, University of Pennsylvania. His research investigates how interventions and policy changes impact crime, health and health behaviors. To this end several of his studies investigated how access to healthcare as part of the Affordable Care Act impact healthcare access and risky health behaviors. Other work of his focused on the impacts of regulatory changes on crime and firearm related mortality.

Dr. Elena Andreyeva is a postdoctoral researcher at the Leonard Davis Institute of Health Economics and the Injury Science Center at the Perelman School of Medicine, University of Pennsylvania. Her research agenda is to combine her knowledge of health and urban economics and her spatial econometrics training to answer policy-relevant questions about population access to various services and amenities. She is particularly interested in analyzing (a) geographic access to healthcare services, (b) the effect of public policies on access to local amenities, such as crime, housing, and health choices, and (c) operational choices in various healthcare systems. Elena was awarded a PhD in Economics by Georgia State University's Andrew Young School of Policy Studies in summer of 2016.

Dr. Charles Branas has conducted research that extends from urban and rural areas in the US to communities across the globe, incorporating place-based interventions and human geography. He has led win-win science that generates new knowledge while simultaneously creating positive, real-world changes and providing health enhancing resources for local communities. His pioneering work on geographic access to medical care has changed the healthcare landscape, leading to the designation of new hospitals and a series of national scientific replications in the US and other countries for many conditions: trauma, cancer, stroke, etc. His research on the geography and factors underpinning gun violence has been cited by landmark Supreme Court decisions, Congress, and the NIH Director. Dr. Branas has also led large-scale scientific work to transform thousands of vacant lots, abandoned buildings and other blighted spaces in improving the health and safety of entire communities. These are the first citywide randomized controlled trials of urban blight remediation and have shown this intervention to be a highly cost effective solution to persistent urban health problems like gun violence. He has worked internationally on four continents and led multi-national efforts, producing extensive cohorts of developing nation scientists, national health metrics, and worldwide press coverage.