Evaluating the effect of a law as a social intervention: Testing the long-run effect of an interrupted time series

Abstract

Data referring to laws, or economy, health, to population,… are published with open access, but they are rarely analysed. This paper highlights the importance of analysing this kind of data to test the effectiveness of a law, bearing in mind that a law can be seen as an intervention programme. Longitudinal data must be analysed in an appropriate way, by contrasting the long-run trend (LRT) of the data before and after the law.

As an example, the month-by-month cases of domestic violence murders (DVM) are analysed, as an indicator of effectiveness in the LRT of a Law against Gender Violence. It is found that: (a) the Law shows itself to be effective, because the change in trend is statistically significant; (b) before the Law came into force, the LRT of DVM was approximately 6 per month; and (c) after implementation of the Law, the LRT is 2.5 DVM per month.

Key words

Domestic gender violence; Murders; Interrupted time series; Programme evaluation; Intervention effect.

Websites that publish periodic official records (monthly, quarterly, annual, etc.) can be found on the Internet, but these data are rarely analysed to verify the effectiveness of the measures used to fit the data and their long-run trend (LRT). This research will provide details on the following points: (a) the importance of analysing the official data from before and after the implementation of a law in order to test its effectiveness; (b) the importance of analysing the longitudinal data by means of dynamic processes (time series, data trend, etc.), instead of analysing the data using cross-sectional data statistics; and (c) the advisability of bearing in mind the LRT of the time series both prior to and following the implementation of the law, contrasting the two trends statistically. As an example, we wish to check the effectiveness of a law against domestic violence by determining whether the number of monthly gender-based domestic violence murders (DVM) has decreased since its implementation.

DVM is a murder in which a man kills a woman with whom he is or has been involved in a relationship in which they were married, living together or partners. It is important to know the time process of DVM because it would allow us to establish: (a) a system of short-term forecasting, but also (b) the LRT of the frequency of DVM, thereby making it possible to determine whether it is a relatively stable trend or, conversely, rising or descending. This knowledge could be used to set up campaigns to promote social awareness and training by special groups in order to lower the incidence (Black, Basile, Breiding et al. 2011; Fleming, Gruskin, Rojo and Dworkin, 2015; Foshee, Reyes, Agnew-Bruce et al., 2014; Foshee, Reyes, Tharp et al., 2014; Hayton, 2015; Jakobsen, 2014; Jewkes, 2002; Langbein, 2015; McCollister, French, and Fang, 2010; Samarasekera and Horton, 2015).

Gender-based DVM are relatively infrequent in advanced western countries, but throughout the whole of Spain there are an average of 4.99 murders per month (from the year 2000 until 2018), which give rise to familial and social anguish as a result of the loss of human lives. As pointed out by the United Nations (2015), one of the possible causes (and the possible effects, we might add) of the inequality between men and women is violence against women, which leaves its mark on those who suffer it and in the children who witness that abuse. The worst kind of abuse is clearly that which results in murder.

The Importance of Periodic Evaluation of a Law

(a) A law as an intervention programme

The evaluation of public programmes is a complex interdisciplinary matter (Scriven, 2008, has called it ‘transdisciplinary’), involving a range of disciplines such as psychology, economics, sociology, public administration and statistics. In a similar vein, Leeuw (2010) stresses the importance of collaboration between legislators, public administrators, governors, etc. and experts in evaluation to verify the effectiveness of laws.

A social intervention programme can be considered as (i) a *planned series of actions* (sometimes in the form of a protocol), in which (ii) *a set of human, temporal and material means* are used, involving a cost that is to be funded by a (public or private) organisation, (iii) to *achieve improvement goals* in a particular area. In the following we will briefly expand on each of these items. The (i) *planned series of actions* is aimed at modifying the content of the goals, and may range from rules (in the case of a public body) to objective or affective information (information or advertising campaigns); (ii) the *human, temporal and/or material means* may range from direct instructions given by technicians to oral or written media (‘word of mouth’, written documents, radio, press, television, etc.), over a certain period of time, and (iii) the *improvement goals* may cover a particular social field (small group, community, private or public organisations, a territory, etc.), about a given content, such as aspects related to education, health or of a social nature, and about specific personal aspects, such as attitudes, habits, particular behaviours (Kaufmann, 2014; Monsen, 2018; Parker, 2018).

A law can be defined as a set of rules issued by a state body, usually the government or parliament, aimed at regulating conduct. The law is applied to regulate conduct and facts (Dworkin, 1986; Kantorowicz, 1958). The legislator is usually concerned about the ‘wording’ of the law, but also has to provide a series of human and material means to enforce that law.

A programme and a law have both common and differential aspects. In the following we will briefly address the ones they have in common: (a) a set of norms and actions that must be implemented by experts in their respective fields; thus, a programme can be implemented by specialists (experts in health, education or in different fields), while a law is conveyed by means of a legal text and is binding, although its application is ensured by experts in different fields from both public and private bodies: police, inspectors (labour, tax office, etc.), educators, healthcare workers, and so on; (b) implicitly, a programme for improving habits usually has a limited duration, determined by the resources available to the organisation, whereas a law has a longer, although undetermined, duration, until the law is abolished; (c) the coercive means of a programme are of a diffuse nature, being implicit in the goals (if you smoke less, your health improves), while the law may contain coercive means of an economic (fines) or even penal nature (prison); and (d) in a programme, the goal may be directed towards internal or external behaviours (attitudes, opinions, habits, etc.), whereas the law is directed towards changing manifest behaviours (Dunn, 2017; Leeuw and Schmeets, 2016).

From a broad point of view, a law is a special type of intervention programme with the peculiarity that the source of the law is a body with legislative power and whose scope of application is that of the territory of the legislative body. Consequently, one can, and must, verify how effective (in terms of the goals reached) and efficient (in terms of the economic benefits) a law is as an intervention programme. Proof of this is the fact that modern states, such as the European Union (DEVCO, 2018), the USA (DSS, 2018) and so forth, have developed bodies and standards to assess their own programmes or those funded by their ministerial bodies for implementation in other countries. At the same time, although there are public institutions that do so, it would be advisable for state-dependent bodies, for the sake of transparency, to periodically publish basic statistical data referring to specific contents of their economic, social, public, etc. activity.

In public management and programme evaluation the intervention of legislators and social professionals is largely guided by theory, while at the same time there are authors who are beginning to contemplate a type of criminology that is based on the data obtained (Lösel, 2018). There are, however, other disciplines such as medicine, education or clinical psychology that have their basic postulates grounded in 'evidence-based practice'. In these disciplines systematic reviews are mostly conducted by means of meta-analyses (Carr and McNulty, 2016; Claes, Van Loon, Vandevelde and Schalock, 2015; Ellsberg et al., 2015; Newton and Miah, 2017).

Despite the differences in laws, intervention programmes and adequate systems of analysis, over time progress can be expected to be made in legal practice and in the evaluation of programmes based on scientific evidence (Eller, Gerber and Robinson, 2018; Partington, 2012). In fact, there are private institutions (such as the Campbell Collaboration (https://www.campbellcollaboration.org/) that publish systematic reviews on research carried out on laws within a particular domain.

(b) The importance of analysing temporal data about a law by means of dynamic processes

It is important that data relating to the effects of the law are published on a regular basis and that these temporal data are analysed using dynamic procedures such as time series, trends, time-domain methods, etc. rather than as cross-sectional data (with t or F tests).

The reasons for using a time series model, instead of a traditional system of comparisons of means, are: (i) substantively, social behaviour presents *regularity and continuity*, which is due to the inertia of each person, group or society, tending to periodically repeat what has been done previously; in short: past behaviour explains future behaviour; (ii) from a methodological point of view, the transversal models of data analysis (means comparisons, etc.) assume the serial independence of the data (which implies no correlation), but in longitudinal studies, this principle of independence is transgressed, since the longitudinal variable is very likely to be correlated with past values of itself, and to have an autoregressive statistical dependence, such that a value for a particular moment depends on the values immediately preceding it; (iii) if the residuals are autocorrelated, and in temporary data this is the most frequent, such a system is very likely to make type I errors or, stated in other words, to infer that there is an effect when in reality it does not exist (Kmenta, 1971; Rouanet and Lépine, 1970); and (iv) if the values of the lag variable are omitted as an independent variable, and this autoregressive variable is part of the explanatory model of the behaviour, the coefficients obtained are biased (they would not coincide with the true values of b_0, b_1,…) and they are inconsistent, that is to say, increasing the sample is not enough to obtain the correct coefficients of the model. This in turn leads to confidence intervals that are not correct for hypothesis testing, and the inferences drawn no longer have any substantive meaning (Draper and Smith, 2014; Gujarati, Porter and Gunasekar, 2013). In order to make this text easier to read, the equations used will be included in the Appendices, together with the most technical developments. In Appendix 1 we describe a quick representation of a time series applied to our data (Equation A1.1), and a functional formalisation (Equation A1.2) of our autoregressive hypothesis (Box, Jenkins, Reinsel and Ljung, 2015; Box-Steffensmeier, Freeman, Hitt et al., 2014; Gujarati, Porter and Gunasekar, 2013; Harvey, 1993; Hedeker and Gibbons, 2006; Vasileiadou and Vliegenthart, 2014; Wooldridge, 2010).

The effect of a social intervention in which the dependent variable is a time series can be formulated as an interrupted time series (McCleary, McDowall and Bartos, 2017; McDowall, McCleary, Meidinger and Hay, 1980; Shadish, Cook and Campbell, 2002) or intervention analysis (Box, Jenkins, Reinsel and Ljung, 2015), where the contrast of the time series is tested with a dummy variable (comparing effect before vs. effect later).

(c) The desirability of taking into account the LRT of the effects of the law

It is important for any statistical model to have a coherent *prospective* capacity (Freedman, 2009), that is, it must be capable of making steady forecasts in the long term (even beyond the time limits of the actual data). Empirically, in our case, the stability of the model indicates that it tends towards a reasonable value. A time model cannot tend to produce a countless number of DVM every month, in the same way that its expected value cannot be below zero. Normally, data analysts tend to fit their valid time models only within the data collection interval, but the most correct thing to do would be to check whether the trend of the data is stable, both in means and in variance.

In this research we will use the inverse function (*1/t*) as the LRT model, as it has a set of formal properties that are very interesting from a statistical-mathematical point of view. One very important property possessed by the inverse *t* function is that it has a monotonic rising or descending shape, with a maximum or minimum asymptotic value. In other words, it has a maximum value which it tends towards or, in the opposite case, it has a minimum value which it does not exceed, and thus the LRT will have a steady finite value; moreover, it meets the condition of parsimony, the data tending towards a constant mean (Box, Jenkins, Reinsel and Ljung, 2015; Freedman, 2009). Later we will discuss the importance of not differentiating data in a time series model, basically so as not to miss the long-term dynamics.

An example: The Law against gender violence

In the year 2004, Law 1/2004, on Comprehensive Protection Measures against Gender Violence (*Ley de Medidas de Protección Integral contra la Violencia de Género* – hereinafter, “the Law”) was enacted in Spain. This law contains a series of measures of different natures, including: (a) protective (school, the mass media, etc.); (b) legal (creation of courts with specialised judges, public prosecutors and police, etc.); (c) technical (professional counselling teams, funding for training courses, etc.); (d) judicial and police supervision (monitoring abuser compliance with restraining orders, i.e. by means of electronic bracelets in order to prevent the wearer from going near their possible victims, etc.); and (e) social, by offering protection for the victim (protection of battered women, relocation of abused women in protected housing, aid for social assistance, geographic mobility, reserving employment positions, etc.).

The Law included the creation and coordination of a series of specialised courts (435 courts for the whole of Spain, of which 14 were exclusively anti-violence and the remaining 421 were compatible with other kinds of family-based offences and juvenile delinquency). The exclusive anti-violence courts, however, first came into operation on 29 June 2005, and the number previously included in the Law was extended to 17. These figures have increased considerably. According to a report drawn up in 2015 by the Governing Body of the Spanish Judiciary (*Consejo General del Poder Judicial*), in that year there were 106 exclusive courts and 355 compatible courts, most of which were set up prior to the year 2008. Law 1/2004 was later complemented with Decrees introduced after the Law (RD 513/2005, RD 515/2005, RD 1452/2005, RD 1369/2006, and so on). Other measures included the creation of action bodies and commissions, although the effects of their introduction (training, implementation, coordination, etc.) may have started with about 36 months' delay with respect to the Law, according to the answers given in interviews held with specialists; that is, full practical implementation of the effects of the Law began around January 2008.

Hypothesis

Our hypothesis is that the incidence of DVM follows a relatively regular pattern, which depends on: (a) previous DVM values – in other words, it follows an autoregressive pattern; (b) the effect of Law 1/2004; (c) the trend of the series; and (d) the change in trend of the series before and after the Law, which is statistically formalised as the interaction (product) between the Law effect and the trend.

Appendix 2, Equation A2.1, shows the functional form of the hypothesis, and its statistical formulation can be seen in Equation A2.2. Each element of the hypothesis will be outlined in the section on variables.

Method and data analysis

Variables

DVM

In this work we will use the number of DVM per month perpetrated in Spain from January 2000 to October 2018 as the reference variable. Table 1 shows the official data for DVM, by months, in Spain over that period. The figures for the years 2000 to 2002 were collected from the Spanish Ministry of Equality (2009, p. 310) and from the year 2003 until the end of the period they were obtained from the Ministry of Health, Social Services and Equality, of the Government of Spain (2018). The fact that the body in charge of monitoring DVMs has changed does not affect the nature of the data, as the second of the institutions responsible for publishing the official data, the Spanish Ministry of Health, Social Services and Equality, is in fact the result of a more recent merging of the former Ministry of Equality with several others. Furthermore, in both cases the data are taken from the official National Police records, that is to say, the information comes from the same source. Data, input and output SPSS files can be obtained from the website: http://repositori.uji.es/XXXXX/XXXXX/. [Note.- We are managing the site name]

Insert Table 1 here

Autoregressive pattern

A monthly time series, such as DVM, has its own trend, with immediate memory (one month before) and cyclical memory. Thus, it can be expected that what happened in the same month in the previous year (one year before) will tend to be repeated in any given month. In our case we suggest that the variable DVM (Y_t) is a function of the number of DVM that occurred a month earlier (Y_{t-1}) and twelve months before (Y_{t-12}).

The effect of Law 1/2004

The *Law Effect* will be identified by means of a dummy variable (*DLE*), with a value of 0 ($DLE = 0$) in the months prior to the effect of Law 01/2004, that is to say, from January 2000 until December 2007 (or from $t = 1$ to $t = 96$), and a value of 1 ($DLE = 1$) from January 2008 until the end of data collection, in October 2018 (from t = 97 until t = 226), as in McDowall et al. (1980), Hardy (1993) and Langbein (2015). Statistically, however, a dummy variable reflects a difference in level between the situation before and after the social intervention. In our case, besides a dummy variable, we will also add a trend to the time series, through an inverse function of time, and test the difference in the LRT before versus after by means of the interaction between the inverse function and the dummy variable.

The LRT of DVM

If we take the order in which the data were recorded as the value of *t* (from $t = 1$ to $t = 226$), an easy representation of an asymptotic function is that of the inverse function of *t* ($1/t$), that is, $Trend = 1/t$. Have in mind that in the long run, a series must tend towards a maximum or minimum limit (asymptotic limit), which it cannot exceed (Freedman, 2009), and the inverse function fulfils this condition.

Our hypothesis is that the series displayed a rising trend before the implementation of the Law, and that one of the effects of the Law will be a change in this trend (Morrison, 2012), which is likely to be descending. The change in trend between the two periods, before and after the Law Effect, is represented in Equation A2.2 by the interaction *Law Effect·Trend* (Jaccard, Turrisi, and Wan, 2003; Jose, 2013;

Hayes, 2013), that is, the product between *Law Effect* and *1/t* (*DLE·(1/t)*). For the properties and the calculation of the LRT, please see Appendix 3, Equations A3.1.

The most important term in Equation A2.2 is the interaction between the trend and the dummy variable, that is, *[DLE·(1/t)]*, because it reflects the difference in the trends between before and after the Law, and is the main object of this research. We would like to point out that, as far as we are aware, this is the first time the inverse function is used to model LRT in a time series and its interaction with a dummy variable is employed to test the effect of an intervention. Box, Jenkins, Reinsel and Ljung (2015) do not include the inverse function in their chapter on transfer functions, but it has the advantage of being very easy to apply and interpret, and can be a good representation of social processes that increase or decrease.

Residual analysis (e_t variable)

The term e_t in Equation A4.1 is the forecasting error of the equation (or residuals of the model), which must be: (a) normal, we will check the normality with the Shapiro-Wilk normality test; (b) white noise (or in other words, with statistically non-significant simple and partial autocorrelations, we will apply the Ljung-Box test, in agreement with Box, Jenkins, Reinsel and Ljung, 2015); and (c) the variance of the errors must be equal before and after the implementation of the Law, so the fit will be the same for the two sections of the time series (before and after the Law Effect). To check this property, we will apply Levene's test of equality of variances.

Descriptive analysis

Table 1 shows raw values, the means and the standard deviations of DVM by months and years. The data analyses were performed with SPSS (2016), the level of significance for all the analyses was 5%, the values of the mean (M) and the standard deviation (SD) of the whole series being 4.99 and 2.27, respectively. It can be seen that January, July and August are the months with the highest mean number of murders (M > 5.30), while the lowest means (M < 4.75) occurs in February, March, April and November. By years, 2008 and 2010 were the ones that presented the highest incidence, with a monthly mean above 6.00. The lowest incidence was found for the years 2016, 2001, 2017 and 2018 (M < 4.30). The M of DVM before the Law Effect (t = 1 to t = 96) is 5.28 (SD = 2.10), whereas after the Law Effect (t = 97 to t = 226) it is 4.78 (SD = 2.38). Because it is a dynamic, time-dependent process, however, we are interested in the LRT, rather than the difference in means, and, moreover, the correct way to analyse our variable Y_t is through a time series model.

Results

Model

Table 2 shows the results obtained in the regression equation developed in Equation A4.1, in agreement with the hypothesis that was proposed. The estimation procedure used was ordinary least squares (OLS) regression, because it is the *best unbiased linear estimator* under conditions of normality and non-autocorrelation of residuals (Dunteman, 1984; Gujarati, Porter and Gunasekar, 2013).

Insert Table 2 here

The multiple correlation coefficient between DVM and the IVs is *.300* (*5*, *208*), $p = .001$, which indicates that 9.0% of the variance of DVM is explained by the IVs. In Table 2 it can be seen how the overall model is significant, as are each of its coefficients, except the one corresponding to Y_{t-1} ($\beta = -.006$, $p = .926$).

Insert Figure 1 here

Figure 1 shows the real values and the predicted values of DVM according to the results in Table 2 and Equation A2.5. This latter can be broken down into two equations (one for each dummy value of the variable Law Effect): for before the Law in Equation A2.6, and for after the Law in Equation A2.7.

We also performed an analysis of residuals (e_t) of the model in Table 2, the value obtained in the Shapiro-Wilk normality test being .988 (214 df), p = .062, and hence we can accept the normality of the residuals. The Box and Ljung test for 12 lags yields 13.735 (12 df), p = .318, the non-autocorrelation of the residuals thus being accepted. We also conducted the Levene equality of variances test of e_t before and after the Law, the result being a value of 1.902 (1, 212), p = .169, and hence the variances are not statistically different.

The main result to be noted in Table 2 is that there is a statistically significant interaction between the variable Law Effect and the series trend, where the interaction is represented by means of *[(1/t)·DLE]*), this term representing the trend change. Consequently, the LRT of the series, according to the results in Appendix 3 (Equations and results of A3.2 and A3.3), reveals that prior to the implementation of the Law, the LRT is around *6.00 DVM*, while after its implementation, the LRT is about *2.48 DVM* per month, these trends being significantly different.

Discussion

The general hypothesis put forward here has been confirmed, as we have pointed out (except the partial hypothesis for the lag 1 of Y_t). The main result to be taken into account is that there has been a change of trend after the implementation of the Law, as compared to the trend prior to the implementation of this Law. Prior to the implementation, the DVM had an increasing LRT, thus gradually rising (asymptotically) to 6.00 per month. Since the implementation of this Law, however, this trend has changed, that is, DVMs are decreasing to 2.48 per month.

It should be borne in mind that the autoregressive coefficients of Y_{t-1} ($\beta = -.01$) and Y_{t-12} ($\beta = -.15$) are not significant and significant, respectively, thus indicating the existence of a seasonal series every 12 months. It should also be noted that there is a slight oscillation of the values, so that if a particular month shows a slight rise in relation to the mean in the DVM value, 12 months later the expected value tends to drop in relation to the mean, and the same happens in the opposite case. Another important property of autoregressive coefficients is that the impact of a moment tends to disappear with time, which is a short-run fit system, as we will see later.

Moreover, in Equations A2.6 and A2.7, the coefficients of *1/t* are linked with the shape of the trend, so that before the Law (–35.05) indicates that the shape is ascendant (negative sign) and relatively quick (the absolute value is not very large), but after the effect of the Law (coefficient: 408.12) it is descendant (positive sign) and with a slower decrease (its absolute value is relatively large).

It can be seen that the series rises slightly before the Law, and that if it continued to increase according to its own inertia, indicated by the result of Equation A3.2, it would reach a value of approximately 6.00. After the implementation of the Law, however, the values tend slowly towards a value of 2.48 (result of Equation A3.2). This is due to the fact that the terms that contain the values *(1/t)* gradually diminish over time, as the higher the value of t is, the lower the value of the term will be. Figure 2 shows the forecasted values for each value of Y_t in accordance with Equations A2.5, A2.6 and A2.7.

Insert Figure 2 here

Earlier we pointed out that the autoregressive coefficients tend to disappear over time. For example, the differential effect of Y_{t-1} on Y_t ($\beta_1 = -.01$) will be $-.01 \cdot Y_{t-1}$, and the effect of Y_t on Y_{t+1} will result in: $Y_{t+1} = -.01 \cdot Y_t = -.01(-.01 \cdot Y_{t-1}) = .0001 \cdot Y_{t-1}$, and the effect of Y_{t-1} on Y_{t+2} has a value of almost zero. The same occurs with the effect of Y_{t-12} ($\beta_2 = -.15$) on Y_t, which tends to disappear ($Y_t = -.15 \cdot Y_{t-12}$; $Y_{t+12} = -.15 \cdot Y_t = -.15 \cdot (-.15 \cdot Y_{t-12}) = .0225 \cdot Y_{t-12}$, or ; $Y_{t+24} = .0225 \cdot Y_t$; $Y_{t+36} = .0034 \cdot Y_t$; so, the long effect is practically ‘zero’). This is because the autoregressive variables (Y_{t-1} and Y_{t-12}) are a short-term fit process (Kitagawa, 2010; Kirchgässner, 2013), but, on the other hand, the system of interaction between the trend and the effect of the Law ($(1/t) \cdot DLE$) implies a long-term fit process. This can be seen in Figure 2, in which, after having removed the real values as of the effect of the intervention (month 97, or January 2008), the autoregressive forecasted values of Y_t gradually diminish, leaving only the effect of the trend (thin discontinuous line, LRT: 6), and after the intervention (total data, month 226, or October 2018), the trend smoothes out (thick discontinuous line, LRT: approximately 2.5).

An analyst would have probably recommended differentiating the series before beginning the autoregressive and trend analyses. For example, in Figure 2 it can be seen that there is a slight upward trend until December 2008, which then slowly decreases from that moment on. In such a case some analysts would recommend differentiating the series (Box, Jenkins, Reinsel and Ljung, 2015), but this recommendation is usually made in the field of time series with data from industry, where the important thing is the short-term prediction. In the social sciences, however, such a strategy is not recommended and there are authors who have criticised this system of analysis of differentiation, because statistically the long-term dynamics is missed. Consequently, differentiation is not altogether empirically, analytically or theoretically correct, as will be detailed in the following points:

(a) If it is assumed that the series has to be differentiated, then this is equivalent to saying that the variable DVM does not have a stable mean; yet, in a normal society, any variable can be expected to move within certain maximum and minimum values (from the trend of the data alone, it is most unlikely that DVM will increase indefinitely or decrease below a DVM value of zero), which makes it unnecessary to differentiate the series.

(b) On analysing the data correctly, it is found that all the parameters in Equation 6 are significant. Moreover, the residuals are seen to be ‘white noise’ without heteroscedasticity, which is further evidence that it is not necessary to differentiate the series.

(c) According to Engle and Granger’s (1987) cointegration model, if a series is differentiated, the model will only be valid for that interval, but it is not valid for a long-term prediction. In other words, if the series is not differentiated, the model thus obtained describes the whole series (and future values) well, regardless of the trend of the sample interval. The model is therefore valid whether the series has a growing, descending or horizontal sample trend, because the important goal is its long-run behaviour.

One epistemologically and methodologically important aspect to bear in mind is that in variables of a social nature, like DVM, it is essential to achieve an LRT value that is stable over time, since if it were not stable, the DVM would be assumed to grow or decrease to the point of exceeding realistic values. Hence, we could have used a linear function and so, since it was ascending before the Law Effect, in the long run the forecast would have been plus infinity (in agreement with Equation A3.1(b)), and after the implementation of the Law its LRT, as it is decreasing, would be minus infinity. Alternatively, a square potential function could have been used, such as trend: $a \cdot t^2 + b \cdot t + c$, performing its corresponding interaction with the variable Law Effect. In this case, if the value of the coefficient *a* were positive ($+a \cdot t^2$), the series would grow in the long run towards plus infinity, and if there were a negative value of *a* ($-a \cdot t^2$), the LRT of the series would decrease until minus infinity, both values being absurd. The same approach could have been adopted with respect to a cubic potential function, and so on (fourth, fifth,… powers), but it would make no sense for the same reason. We may also find ourselves before the paradoxical situation in which the real fit (R^2 of the whole regression equation that was obtained) of one potential function (quadratic, cubic,…) within the real values were more accurate. However, a statistical time model is validated by its internal fit and by its LRT; that is, both the time interval in which the data have been gathered and the long-run forecasted values should be realistic (Huckfeldt, Kohfeld and Likens, 1982; Freedman, 2009).

This research has some shortcomings, i.e. (a) it has not been possible to disaggregate the possible effect of each of the different components of the Law (creation of courts, monitoring abuser compliance with restraining orders, relocation of abused women in protected housing, geographic mobility, training and specialisation of experts: policemen, judges, prosecutors, technical teams, etc.), and instead here we have taken all the components of the Law together as a set; (b) we have also tried to localise the budget devoted to the Law (month by month or year by year) as another IV, but the Central Government, each Autonomous Government inside their corresponding geographical and political areas, the respective provinces, the municipalities and so on are all involved in the total final budget. It should be noted that there are more than

8000 municipalities in Spain, and so to obtain these data by month and place is a very arduous, almost impossible, task.

To sum up: (a) we have confirmed the efficacy of the publication of official statistics with the aim of allowing statistical analyses to be performed in order to determine the effectiveness of the measures taken; (b) a Law can be considered an intervention programme that can be evaluated on the basis of its tangible results, and we have found that the Law Effect has been positive in Spanish society as a whole; (c) it has been shown that it is advisable to analyse longitudinal data by means of time series models in order to avoid making fundamental errors in the estimation of parameters and, hence, to draw correct inferences; (d) in temporal processes, attention must be paid to the long-term dynamics of the series, including the estimation of the LRT in each of the different phases (intervention A, intervention B, C,…) of the series; (e) in this study we have only tested DVM, but the Law has also, almost certainly, had other beneficial effects on other variables, such as the women who managed to survive episodes of gender violence (avoidance of psychological or physical abuse, for example), their closest relatives (above all their children, as they witness fewer arguments and violence) and the social and work context of the victims (better social relationships of those possibly affected, less time off work due to the consequences of abuse, fewer visits to the doctor's surgery, etc.); (f) the results found here are not to be taken as a ‘snapshot’ or a conclusive situation, but must be reviewed as time passes and more research is devoted to determining the effectiveness of the measures taken; and (g) attempts must be made to ensure these results are maintained and improved with an increase in appropriate budgetary, institutional and social measures, with the aim of lowering even further the figures achieved to date.

References

Black MC Basile KC Breiding MJ Smith SG Walters ML Merrick MT Chen J and Stevens MR (2011) *The national intimate partner and sexual violence survey (NISVS): 2010 summary report*. Atlanta, GA: National Center for Injury Prevention and Control.

Box GEP Jenkins GM Reinsel GC and Ljung GM (2015) *Time series analysis: Forecasting and control* (5th ed.). Hoboken, NJ: Wiley.

Box-Steffensmeier JM Freeman JR Hitt MP and Pevehouse JCW (2014) *Time series analysis for the social sciences.* New York: Cambridge University Press.

Carr A and McNulty M (2016) *The Handbook of Adult Clinical Psychology: An Evidence Based Practice Approach* (2nd ed.). NY: Routledge.

Claes C Van Loon J Vandevelde S and Schalock R (2015) An integrative approach to evidence based practices. *Evaluation and Program Planning* 48: 132–136.

DEVCO (2018) Directorate General for International Cooperation and Development, European Commission. *Evaluation policy*. From https://ec.europa.eu/europeaid/evaluation-policy_en

DSS (2018) Deputy Secretary of State, Department of State, United States. *Department of State Evaluation Policy*. From https://www.state.gov/s/d/rm/rls/evaluation/index.htm

Dunn WN (2017) *Public policy analysis*. NY: Routledge.

Dunteman GH (1984) *Introduction to linear models*. Beverly Hills, CA: Sage.

Dworkin RM (1986) *Law's Empire*. Cambridge, Massachusetts: Harvard University Press.

Eller WS Gerber BJ and Robinson S.E (2018) *Public administration research methods. Tools for evaluation and evidence-based practice* (2nd ed.). NY: Routledge.

Ellsberg M Arango DJ Morton M Gennari F Kiplesund S Contreras M and Watts C (2015) Violence against women and girls 1. Prevention of violence against women and girls: what does the evidence say? *Lancet* 385: 1555–1566.

Engle RF and Granger CWJ (1987) Co-integration and error correction representation, estimation and testing. *Econometrica* 55: 251-76.

Fleming PJ Gruskin S Rojo F and Dworkin SL (2015) Men's violence against women and men are inter-related: Recommendations for simultaneous intervention. *Social Science and Medicine* 146: 249–56.

Foshee VA Reyes LM Agnew-Brune CB Simon TR Vagi KJ Lee RD and Suchindran C (2014) The effects of the evidence-based safe dates dating abuse prevention program on other youth violence outcomes. *Prevention Science* 15: 907-16.

Foshee VA Reyes LM Tharp AT Chang LY Ennett ST Simon TR Latzman NE and Suchindran C (2014) Shared longitudinal predictors of physical peer and dating violence *Journal of Adolescent Health* 56: 106-12.

Freedman DA (2009) *Statistical models and causal inference*. Cambridge: CUP.

Gujarati DN Porter DC and Gunasekar S (2013) *Basic econometrics* (5th ed.). New Delhi: Tata McGraw Hill.

Hardy MA (1993) *Regression with dummy variables.* Newbury Park, CA: Sage.

Harvey AC (1993) *Time Series Models* (2nd ed.). New York: Harvester Wheatsheaf.

Hayes AF (2013) *An introduction to mediation, moderation, and conditional process analysis: A regression-based approach.* New York: Guilford Press.

Hayton K (2015) Evaluating the evidence –a move to more realistic evaluation: A case study of regional selective assistance in Scotland. *Evaluation* 21(2): 248–62.

Hedeker D and Gibbons RD (2006) *Longitudinal data analysis*. Hoboken: John Wiley.

Huckfeldt RR Kohfeld CW and Likens TW (1982) *Dynamic modelling: An introduction*. Newbury Park, CA: Sage.

Jaccard J Turrisi R and Wan CK (2003) *Interaction effects in multiple regression* (2nd ed.). Thousand Oaks, CA: Sage.

Jakobsen H (2014) What's gendered about gender-based violence?: Grounded theoretical exploration from Tanzania. *Gender & Sociology* 28: 537-61.

Jewkes R (2002) Intimate partner violence: causes and prevention. *Lancet* 359: 1423-29.

Jose PE (2013) *Doing statistical mediation and moderation.* New York: Guilford Press.

Kantorowicz H (1958/2014) *The definition of law*. Cambridge: CUP.

Kaufmann F-X (2014) Prevention and intervention in the analytical perspective of guidance. In: Hurrelmann K Kaufmann F-X and Lösel F (eds.) *Social Intervention: Potential and constraints*. Berlin: Walter de Gruyter.

Kitagawa G (2010) *Introduction to time series modeling*. Boca Raton, FL: Chapman & Hall.

Kirchgässner G (2013) *Introduction to modern time series analysis* (2nd ed.). Berlin: Springer.

Kmenta J (1997) *Elements of econometrics* (2nd ed.). Ann Arbor, MI: University of Michigan Press.

Langbein L (2015) *Public program evaluation: A statistical guide* (2nd ed.). New York: Routledge.

Leeuw FL (2010) Can legal research benefit from evaluation studies? *Utrecht Law Review* 7(1): 52-65.

Leeuw FL and Schmeets H (2016) *Empirical legal research: A guidance book for lawyers, legislators and regulators*. Cheltenham, UK: Edward Elgar.

Lösel F (2018) Evidence comes by replication, but needs differentiation: the reproducibility issue in science and its relevance for criminology. *Journal of Experimental Criminology* 14(3): 257–278.

McCleary R McDowall D and Bartos BJ (2017) *Design and analysis of time series experiments*. NY: Oxford University Press.

McCollister KE French MT and Fang H (2010) The cost of crime to society: New crime-specific estimates for policy and program evaluation. *Drug and Alcohol Dependence* 108: 98-109.

McDowall D McCleary R Meidinger E and Hay RA (1980) *Interrupted time series analysis*. London: Sage.

Ministry of Equality, Government of Spain (2009) *Organic Law Assessment 1/2004 of 28th of December, 2004. Protection Measures against Gender Violence*. Madrid. [Ministerio de Igualdad, Gobierno de España. 2009. *Evaluación de la Ley Orgánica 1/2004 de 28 de diciembre. Medidas de protección contra la violencia de género*. Madrid]. From http://www.msssi.gob.es/fr/ssi/violenciaGenero/Documentacion/seguimientoEvaluacion/DOC/InformeTresLey.pdf

Ministry of Health, Social Services and Equality, Government of Spain (2018) [Ministerio de Sanidad, Servicios Sociales e Igualdad, Gobierno de España. 2018]. From http://estadisticasviolenciagenero.igualdad.mpr.gob.es/

Monsen KA (2018) *Intervention effectiveness research: Quality improvement and program evaluation*. Berlin: Springer.

Morrison K (2012) Searching for causality in the wrong places. *International Journal of Social Research Methodology* 15: 15–30.

Newton PM and Miah M (2017) Evidence-based higher education – Is the learning styles 'myth' important? *Frontiers in Psychology* 8: Art. 444. https://doi.org/10.3389/fpsyg.2017.00444

Organic Law 1/2004 (2004) Comprehensive Protection Measures Law against Gender Violence. *Spain's Official State Gazette, 313 (29th of December, 2004),* 42166-42197 [Ley Orgánica 1/2004. 2004. Ley de Medidas de Protección Integral contra la Violencia de Género. *Boletín Oficial del Estado, 313 (29 de Diciembre de 2004),* 42166-42197].

Parker R (2018) Comparative social policy and the politics of comparison. In: Gandy J Robertson A Sinclair S and Spencer JC (eds.) *Improving social intervention*. London: Routledge.

Partington M (2012) Empirical legal research and policy-making. In: Cane P and Kritzer H (eds.) *Oxford handbook on empirical legal research* (pp. 1002-1024). NY: Oxford UP.

Rouanet H and Lépine D (1970). Comparison between treatments in a repeated-measurement design: ANOVA and multivariate methods. *British Journal of Mathematical and Statistical Psychology* 23: 147–163

Shadish WR Cook TD and Campbell DT (2002) *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: H. Mifflin.

Samarasekera U and Horton R (2015) Prevention of violence against women and girls: a new chapter. *The Lancet* 385: 1480–82.

Scriven M (2008) The concept of a transdicipline; and Evaluation as a transdicipline. *Journal of Multidisciplinary Evaluation* 10: 65-66.

United Nations (2015) *The millennium development goals report.* New York: United Nations. From http://www.un.org/millenniumgoals/2015_MDG_Report/pdf/MDG%202015%20rev%20%28July%201%29.pdf

Vasileiadou E and Vliegenthart R (2014) Studying dynamic social processes with ARIMA modelling. *International Journal of Social Research Methodology* 17(6): 693-708.

Wooldridge JM (2010) *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

Table 1

Month-by-month DVM data (from January 2000 to June 2016)

	Month													
Year *(t)*	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Mean (by year)	SD (by year)
2000 *(1-12)*	6	5	5	2	6	8	6	4	7	7	2	5	5.25	1.86
2001 *(13-24)*	5	2	4	5	9	3	1	4	5	4	3	5	4.17	1.99
2002 *(25-36)*	5	3	2	9	3	3	6	7	3	5	4	4	4.50	2.02
2003 *(37-48)*	8	4	6	5	6	8	8	8	4	2	9	3	5.92	2.31
2004 *(49-60)*	2	5	6	6	8	6	7	6	7	7	4	8	6.00	1.71
2005 *(61-72)*	6	5	3	5	4	6	4	6	4	5	6	3	4.75	1.14
2006 *(73-84)*	9	4	9	5	5	2	8	9	4	6	2	6	5.75	2.56
2007 *(85-96)*	5	9	4	4	7	10	8	7	4	5	7	1	5.92	2.54
2008 *(97-108)*	7	8	4	5	3	5	5	8	7	7	6	11	6.33	2.15
2009 *(109-120)*	0	5	5	2	9	5	6	6	6	6	1	5	4.67	2.50
2010 *(121-132)*	3	4	7	8	6	5	10	2	8	6	7	7	6.08	2.27
2011 *(133-144)*	8	5	3	3	8	2	7	4	6	6	3	6	5.08	2.07
2012 *(145-156)*	8	2	2	4	3	10	3	2	2	7	6	3	4.33	2.74
2013 *(157-168)*	4	4	8	6	6	2	2	2	8	5	5	2	4.50	2.24
2014 *(169-180)*	7	5	8	4	2	3	3	8	2	2	5	5	4.50	2.24
2015 *(181-192)*	3	2	4	1	4	5	9	6	2	7	7	10	5.00	2.86
2016 *(193-198)*	8	4	2	4	4	2	6	4	2	4	5	4	4.08	1.73
2017 *(199-216)*	6	10	3	4	6	3	2	4	2	5	3	3	4.25	2.26
2018 *(217-226)*	2	2	3	4	1	5	6	6	10	4	-	-	4.30	2.63
Mean (by month)	5.37	4.63	4.63	4.53	5.26	4.89	5.63	5.42	4.89	5.26	4.72	5.06		
SD (by month)	2.50	2.27	2.17	1.93	2.33	2.60	2.54	2.17	2.47	1.56	2.14	2.65		

Table 2

Results of the statistical regression analysis

ANOVA[a]

Model	Sum of Squares	df	Mean Square	F	p
Regression	101.120	5	20.224	4.109	.001
Residual	1023.763	208	4.922		
Total	1124.883	213			

[a] Dependent Variable: DVM.

Coefficients[a]

Model	Unstandardised Coefficients		Standardised Coefficients	t	p
	B	Std. Error	Beta		
Constant	6.959	.708		9.823	>.001
Y_{t-1}	-.006	.069	-.006	-.092	.926
Y_{t-12}	-.152	.069	-.149	-2.188	.030
Law Effect	-4.081	.970	-.869	-4.205	>.001
1/t	-35.053	15.925	-.199	2.201	.029
(1/t)·Law Effect	443.173	124.678	.667	3.555	>.001

[a] Dependent Variable: DVM

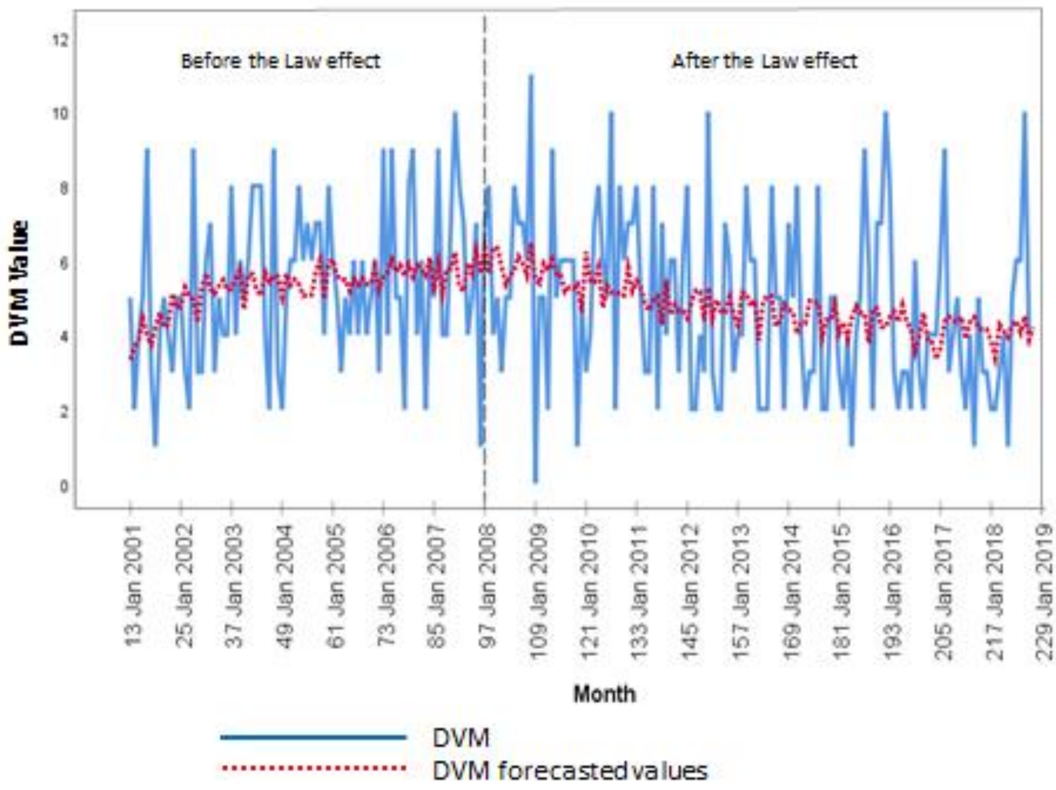

Figure 1. Real and forecasted values of DVM in accordance with Equation A4.4.

Note. t = 1, January 2000; t = 96, December 2007; t = 226, October 2018.

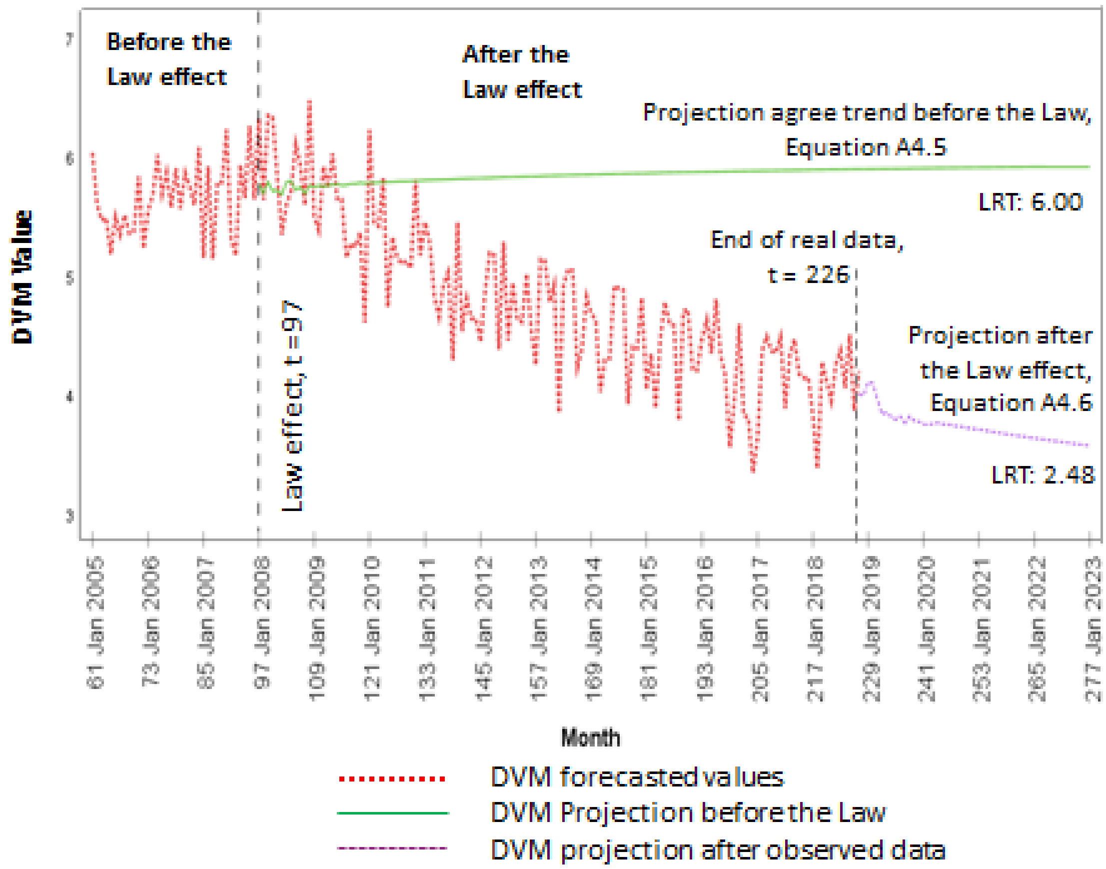

Figure 2. Forecasted values of DVM in accordance with Equations A4.4 (dashed red line), A4.5 (projection before the Law, thin green continuous line) and A4.6 (projection after the Law Effect, thick purple discontinuous line). Observe that the forecasted values are equal to those in Figure 1, but we begin Figure 2 from t = 61 (Jan 2005).