Notes on Longitudinal Data Concerning Domestic Violence

Bruce Chapman
Australian National University
May 2024

Background

What follows makes some basic and broad observations concerning the suitability or otherwise of different types of data sets prospectively available to examine the causes and consequences of domestic violence (DV) in Australia and promotes for consideration the generic value of a longitudinal (panel) individually based survey motivated to examine DV. The exercise begins with the recognition that there are many ways in which to classify empirical data in the social sciences, with quite disparate benefits and limitations. Data can be, *inter alia*, administrative or based on surveys, and/or aggregative or reflecting the circumstances of individuals.

There are two parts. The first involves brief commentary on several different data classification approaches which highlights their lack of suitability with respect to understanding the causes and consequences of DV. This discussion leads to the promotion of the use of longitudinal data as the most useful empirical basis for DV research, and several key reasons for this endorsement are provided and explained in a second section.

1 Data Classifications and DV

1 (i) Administrative data (AD)

Administrative data generally relate to the measurement of people's experience with respect to government activities and programs. Some examples of AD include: the receipt and amount of social security assistance, such as for *JobSeeker* and aged pensions; electricity bill payments; and income taxes. In general AD have the advantage of measurement accuracy but have critical limitations through very little information being available concerning key variables, such as education, health, and the number and age of children. As well, there are no AD sets available which identify the experience of DV meaning that if used by themselves these data are uninformative for our project.

1 (ii) Aggregative Survey Data (ASD)

There are many different types of survey data with one being aggregative, meaning not based on the experience of individuals. Some examples of ASD include: cross-sectional regional experience of social and economic experience, such as local Census District (or State/Territory) jurisdictional averages concerning unemployment rates, age distributions by gender, or reported criminal activity; and measurement of incidence in a time series of various phenomena, such as the ethnic and immigrant status of populations. Such information is important in trying to understand location differences in demographic situations, and broad changes over time in social experience. However, for our purposes ASD suffer from the same confounding limitation as AD because no information is available concerning the most important as aspects of the experience of DV for individuals.

1 (iii) Individual Unit-record Survey Data: Cross-sectional (IUSDCS)

The basis of IUSDCS has important advantages for properly informed research concerning key aspects of the experience of DV. The critical positive points are that with all unit record data there is potential for: identification of the incidence, and the extent of, different types of DV; and, a substantial amount of information concerning demographic, economic and other characteristics that differ significantly between individuals and will be associated in various ways with DV. However, by its very nature cross-sectional data are only poorly suited to the issues of both the causes and the consequences of DV simply because the information usually only relates to the experience of people at any one point in time.

The absence of a time dimension with IUSDCS means that researchers are unable to assess the importance of factors leading up to and likely causing DV, nor is it possible to determine what the on-going consequences of DV are. These shortcomings are critical for DV policy designed to address ways in which governments might influence issues such as the incidence and duration of the effects of DV on key welfare measures including safety, income, employment, and physical and emotional well-being. For this we need panel data.

2 The Right Data to Analyse DV: Longitudinal Panels

2 (i) Introduction

The discussion above leads inexorably to one conclusion: because of the limitations of all other alternatives, individual unit-record survey data involving longitudinal panels (henceforth referred to as "longitudinal panels" (LP)) are the preferred statistical basis with which to analyse DV. LP is taken to mean a repeated cross-section of a significant number of the same people (at least 1000, and often as many as 20,000 are often available) are interviewed, usually annually. LP surveys typically involve the collection of a plethora of information related to individual experience concerning such things, *inter alia,* as: household circumstances; education; income levels and sources; employment status; health; location; ethnicity; age; religion; labour market experience; and, ideally for our project, the experience of DV.

What now follows highlights and explains the advantages of LP. While the issues are raised are generic and apply broadly to all LP, illustrations are made with respect to analyses of the causes and consequences of DV. There are other less important benefits and potential of LP not considered, nor is there any consideration of some of the key factors limiting the availability and utility of LP, such as costs and the difficulties and problems associated with attrition.

2 (ii) The key role of unobservable and missing variables

Many of the factors which influence the life circumstances of people are difficult to measure (and are typically not measured) yet can have effects which impact critically on personal outcomes. Moreover, some of these factors can be highly correlated with included variables implying a strong potential for statistical tests to be confounded by ("biased", to use econometric terminology) and thus difficult to interpret. In the area of labour economics, for

example, there is a classic case of a confounding variable with respect to the identification of the effect of education on earnings, which is now explained.

Psychological tests reveal that "highly motivated" (non-indolent) people are more likely to both attain higher levels of education and to receive high earnings. Thus, while simple human capital models estimating the effects of education on earnings typically find strong positive relationships, it is erroneous to use these results as identifying with precision the effect of education on earnings. This is because the education measure will be capturing the individuals' motivation effects that positively affect both educational attainment and earnings, disguising and overstating the true effect of education.

However, there is very good news for researchers examining these relationships with an LP, which is that because the same people are observed multiple times, there is the facility to take account of the unobservable and missing characteristics of each individual. Because there are repeat observations the (assumed to be unchanging) unobserved characteristics of people can thus be controlled for by including in the estimation a unique variable for each person in the sample, a technique known as "fixed effects". In the above example, a fixed effects estimation allows the effects of education to be measured correcting for the confounding influence of embodied individual factors.

There are bound to be similar examples of important unobservable and missing factors in the DV research space, with one possibility being as follows. It might be the case that perpetrators of DV are more likely to have personality traits, such as poor anger management, that lead both to them being unemployed *and* committing DV. If this is the case a fixed effects approach using LP will be able to estimate the pure unemployment effect associated with people committing DV.

Similarly with survey data there might be unobservable or missing traits more commonly occurring in victims of DV that are also associated with, for example, health or labour market outcomes. Again if this is case fixed effects statistical approaches that are only available with LP will allow control for these inherent traits in order to accurately isolate and measure the consequences of DV.

2 (ii) The generic importance of dynamics from LP

The most obvious benefit of LP is the facility to illustrate how individual experiences change over time. As the great applied economist Marc Nerlove has commented: "Economic behavior is inherently dynamic so that most econometrically interesting relationships are explicitly or implicitly dynamic" (Nerlove, 2002). This is a profound point because without LP researchers typically rely on cross-section data, often without recognising the inherent shortcomings of static data with respect to making inferences associated with time.

This broad issue is emphasised in Dearden (2019), which focusses on the problems of, and provides solutions to, the use of cross-section age-earnings profiles to make inferences about future projections of lifetime earnings. Dearden highlights that such an approach leads to

significant inaccuracies in lifetime earnings projections which then led to misleading assessments of the efficacy of different approaches to higher education financing. The Dearden conclusion is that there are two useful ways to address the problem: by adopting techniques to impose hypothetical earnings dynamics on the cross-sectional data; or using LP. The second is the much-preferred empirical method.

There will similarly be very important dynamic issues associated with DV which cannot be captured satisfactorily with cross-sections, regarding some basic descriptive statistics. They include data necessary to address the questions: what is the frequency of, and time lags associated with, experiencing DV?; as aging proceeds, or the duration of a relationship increases, does the probability of experiencing or perpetrating DV change, and by how much?; and what are the effects of a household's changing external environments – such as from the state of the macro-economy or the imposition of business/social interaction lockdowns from a pandemic – concerning the experience of DV?. None of these issues can be addressed properly without LP.

2 (iii) The critical role of dynamics in establishing the causes and consequences of DV

Perhaps the biggest question for policy in this area is: "what causes DV?". Causation, is of course, the most posed question in all social science, such as what are the factors that: determine poor health; lead to unemployment; contribute to educational success/failure; influence income; result in marital dissolution; and make for successful and unsuccessful government policy interventions. The obvious point is that for an event to have an effect - an outcome - it is almost always the case that the event precedes the outcome, as in a person losing a job becoming depressed, or a long period of smoking cigarettes leading to a higher probability of some forms of cancer.

In the DV causality research space there will be both external and person-specific factors that can change over time and thus influence the probability of the experience. For example, in the first category would include a recession or lockdowns due to a pandemic, and in the second might be individual traumas associated with a death in the family, or personal illness. Neither category can be captured very usefully without recognising time dimensions associated with events.

Perhaps the second biggest question for policy in this area is: "what are the consequences of DV?"; importantly all the above-noted methodological issues just as readily apply to this question. That is, there is an event – a person experiencing DV – which has consequences, and these are very likely to be multi-faceted. The following issues of interest are the effects of DV on: the economic situation of both victims and perpetrators; the health, mental and physical, of victims; and, the circumstances of children, with respect to emotional well-being, health, financial welfare, and educational aspirations and outcomes Without identification of the precise timing of both DV and its effects these issues can't be addressed in any convincing fashion; LP is required.

2 (iv) Leads, lags and DV frequency in the context of LP

Health, social and economic circumstances, and their associations with respect to causes and consequences, are rarely instantaneous and can be quite disparate in terms of the time involved. For example, with respect to health it can take 20-40 years from ingesting asbestos to the development of mesothelioma, but less than a week before a severe bacterial infection leads to gangrene. With respect to the experience of DV there are some critical issues concerning leads and lags, with important lessons for policy, and some illustrations are as follows.

On the causal side, there might be situations contributing to the likely of DV being perpetrated which depend on the duration of a negative event, for example the experience of unemployment. Most unemployment is of very short duration, but for some the experience can be extended and this is known to be associated with increasing probabilities of negative outcomes, such as the commitment of a crime, including an act of violence. In this context researchers and governments need to know about these time dimensions for policy interventions to be most effective and desirable.

Similarly, there will be fundamental issues associated with time lags with respect to the probabilities of various health, social and labour market consequences of DV for victims. As examples, the following issues seem to be of great importance: the length of time after experiencing DV that the effects last; the circumstances that affect the length of time involved in the continuation and intensity of these negative experiences; and, how the issues differ depending on the nature of DV (for example, physical compared to financial).

2 (v) Are there repeat occurrences of DV?

One of the critical issues associated with the measurement of all human experience relates to what could be referred to as “bunching”. To illustrate the potential significance of this, imagine the following scenario related to a hypothetical community of 100 people which is observed over a given period to have had 10 separate incidences of DV. With respect to perpetration, at one extreme this could mean that in the period one person committed all 10 acts of DV or, at the other extreme, that 10 people committed one DV act each. Similarly for this community, at one extreme the data could mean that one person was a victim of DV 10 times or, at the other extreme, that 10 people were victims of DV just once.

For these illustrations of the extremities is not difficult to see how profoundly different our interpretations of the bases for understanding DV and what the concomitant efficacies of alternative policies would be. For example, if there was only one perpetrator in the community, the right response would be to find this person and put a stop to their behaviour. But if the perpetuation of DV is common-place and widespread, policy responses would need then to be more broadly-based. The same sort of conclusions can be drawn with respect to the bunching or otherwise in the experience of DV for victims.

3 Final Comment

The fundamental point with respect to all the above is that without a LP it is difficult to understand the nature, incidence, and the usefulness of policy responses to DV unless the data involve following the same people over time. As well, the sample sizes need to be sufficiently large to allow accurate measurement of the statistical associations, and the time periods sufficiently long to maximise the prospects of the data capturing meaningfully the complexities of DV and what might be done to improve the current abhorrent situation.

May 20, 2024

References

Dearden, Lorraine (2019) 'Evaluating and designing student loan systems: an overview of empirical approaches'. *Economics of Education Review*, 71, 49-64.

Hsiao, C. (2003). *Analysis of Panel Data*, Vol. 34 of Econometric Society monographs. Cambridge University Press, Cambridge, 2nd ed.

Nerlove, M. (2002). *Essays in Panel Data Econometrics*. Cambridge University Press, Cambridge.