Transportation Research Part A 71 (2015) 31–45

Contents lists available at ScienceDirect

Transportation Research Part A

journal homepage: www.elsevier.com/locate/tra

# Investigating the nature and impact of reporting bias in road crash data

Kibrom A. Abay

*Department of Economics, University of Copenhagen, Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K, Denmark*

ARTICLE INFO

*Article history:*
Received 10 November 2013
Received in revised form 27 October 2014
Accepted 3 November 2014

*Keywords:*
Reporting bias
Under-reporting
Injury severity modeling
Random coefficients bivariate ordered-response probit
Maximum simulated likelihood

ABSTRACT

This paper investigates the nature, and impact of the reporting bias associated with the police-reported crash data on inferences made using this data. In doing so, we merge a detailed emergency room data and police-reported crash data for a specific region in Denmark. To disentangle potentially common observable and unobservable factors that affect drivers' injury severity risk and their crash reporting behavior, we formulate a bivariate ordered-response probit model of injury severity risk and crash reporting propensity. To empirically identify the reporting bias in this joint model, we exploit an exogenous police reform that particularly affects some specific municipalities of the region under consideration. The empirical analysis reveals substantial reporting bias in the commonly used police-reported road crash data. This non-random sample selection associated with the police-reported crash data leads to biased estimates on the effect of some of the explanatory variables in injury severity analysis. For instance, estimates based on the police-reported crash data substantially underestimate the effectiveness of seat belt use in reducing drivers' injury severity risk.



## 1. Introduction

The incredible societal and economic burden of traffic accidents has induced considerable research investment in the safety and transportation literature. These studies mainly evolved around investigating the frequency of road crashes and the injury severity outcome of road users involved in these crashes. In doing so, almost all the existing studies use police-reported crash registers. However, the reporting bias associated with the police-reported crash data has received little attention. For instance, it is generally conceived that crashes involving minor injuries are less likely to be reported to the police while the rate of reporting for crashes involving fatal injuries is expected to be fairly high (see Blincoe et al., 2002; Elvik and Mysen, 1999; Hauer, 2006). This non-random sampling in the police-reported data, if not accounted properly, leads to biased estimates and erroneous inferences on the effects of different safety measures on the crash outcome of road users. Furthermore, some studies highlight that crash reporting (to the police) behavior of road users also vary depending on potentially observable characteristics of road users and crashes (see Amoros et al., 2006; Alsop and Langley, 2001).

While there is some consensus on the existence of reporting bias in the police-reported crash data, there have not been rigorous studies that investigate the empirical foundation and impact of this problem on inferences made using this data. Although the aforementioned descriptive studies hint for possible under-reporting of crashes in the commonly used police-reported crash registers, they do not indicate the nature and consequences of the problem as they fail to disentangle

*E-mail address:* Kibrom.Araya.Abay@econ.ku.dk

http://dx.doi.org/10.1016/j.tra.2014.11.002
0965-8564/

potentially common observable and unobservable factors that affect both crash outcome and crash reporting decision of road users. In a strict sense, these aggregate descriptive studies do not inform the breadth and impact of the reporting bias in the police-reported crash data on inferences made using this data. As highlighted by Savolainen et al. (2011), and Mannering and Bhat (2014), understanding the overall features of this reporting bias deserves substantial attention because fitting the standard econometric models to under-reported data could misguide public road safety policy measures.

The main challenge in understanding the nature and consequences of the reporting bias in the police-reported crash data is unavailability of alternative data sources. This paper attempts to fill this gap by using a detailed emergency room data along with the commonly used police-reported crash data for a specific region in Denmark. This detailed emergency room data, besides enabling us to analyze crash reporting behavior of road users, provides objective and professional measure of injury severity unlike the commonly used subjective injury severity measures in the police reports. This is appealing because significant discordance between hospital-based and police-based injury severity assignment is evidenced in the safety research (see Tsui et al., 2009). Such misclassification in the response outcome, if not accounted properly, leads to inconsistent variable estimates (see for example, Hausman et al., 1998).

By merging a detailed emergency room data with police records, this paper investigates the nature and implication of the reporting bias associated with the police-reported crash data. At least two important issues are worth of noting in such an investigation: (i) reporting to the police could be driven by potentially *observable characteristics* of road users and crashes. In this case, as far as we can control for these observable characteristics in the specification, using the under-reported police-reported data in injury severity analysis has no severe consequence on the effect of the explanatory variables considered. (ii) Reporting to the police can also be driven by *common unobservable* (to the analyst) factors which affect both road users' injury severity risk and their crash reporting propensity. This type of reporting bias misleads the variable effects that can be estimated in a standard injury severity model using the police-reported crash data.

To investigate the association between injury severity risk and crash reporting behavior of road users, we formulate a bivariate ordered-response probit model of injury severity risk and reporting propensity (to the police). This joint model assumes that road users' injury severity risk and their crash reporting propensity (to the police) are determined by some common underlying observed and unobserved (to the analyst) factors. Thus, simultaneous modeling of injury severity outcomes and crash reporting decision of road users helps to disentangle common observable and unobservable factors which affect both outcomes. To empirically identify the reporting bias in this joint model, we exploit an exogenous police reform that is expected to affect drivers' crash reporting propensity but not their injury severity risk.

The bivariate ordered-response probit model we set up in this paper also accommodates unobserved heterogeneity effects associated with the explanatory variables of the system of equations. This applies for the explanatory variables that enter the injury severity risk as well as crash reporting propensity equation. Accommodating unobserved heterogeneity effects is not just a simple econometric exercise to improve the fit of the model, but it is critically important for the reason that ignoring such effects can lead to erroneous and inconsistent estimates (Chamberlain, 1980; Bhat, 2001; McFadden and Train, 2000).[1] The choice of ordered probit over unordered response models is primarily motivated by the fact that ordered response models are extensively used in the safety and transportation literature despite their restrictive nature (see Savolainen et al., 2011; Mannering and Bhat, 2014). Furthermore, we understand the restrictive nature of the standard ordered response models compared to the newly introduced generalizations (see Eluru et al., 2008), but for parsimonious and empirical identification purposes we will stick to the standard bivariate ordered response probit model.

To sum up, this study uses a detailed emergency room data and adopts a joint model to capture the dependence among unobserved factors that affect road users' injury severity risk and crash reporting propensity to the police, while also accommodating unobserved heterogeneity in the effects of the explanatory variables. This joint model along with its dependence structure informs the size, nature and impact of the reporting bias in the police-reported crash data. Our slightly structural approach provides more detailed insight on the road use and reporting behavior of individuals. Besides to the main purpose of identifying reporting bias, our structured framework opens for further investigation on what is driving the potential reporting bias in the police records. Unfolding the behavioral and physical mechanisms driving the reporting bias in the police records provides interesting input to design convivial public road safety policies. For instance, drivers holding better insurance coverage may drive aggressively, perhaps due to the notion of moral hazard, and the same type of drivers might be more likely to report to the police to secure their insurance claims. In the absence of insurance data, this type of behavioral pattern creates common unobservable trend that can drive part of the reporting bias potentially expected in the police-reported crash data.[2]

The remaining sections of this paper are organized as follows: the next section reviews previous evidence on under-reporting in road crash data and its implication on injury severity modeling. Section 3 presents the data and sample description. Section 4 describes the model specification. Section 5 presents the empirical results and Section 6 summarizes the key findings.

[1] Some recent studies that allowed for random coefficients specification in injury severity models underscore the importance of accounting for unobserved heterogeneity effects in the explanatory variables (see, for example, Abay, 2013; Abay et al., 2013; Anastasopoulos and Mannering, 2011; Christoforou et al., 2010; Eluru and Bhat, 2007; Paleti et al., 2010; Milton et al., 2008).

[2] We thank a referee for suggesting this type of behavioral pattern.