This item was submitted to Loughborough's Research Repository by the author.
Items in Figshare are protected by copyright, with all rights reserved, unless otherwise indicated.

A time-series analysis of motorway collisions in England considering road infrastructure, socio-demographics, traffic and weather characteristics

PLEASE CITE THE PUBLISHED VERSION

http://dx.doi.org/10.1016/j.jth.2015.10.005

PUBLISHER

© The Authors. Published by Elsevier Ltd.

VERSION

VoR (Version of Record)

PUBLISHER STATEMENT

This work is made available according to the conditions of the Creative Commons Attribution 4.0 International (CC BY 4.0) licence. Full details of this licence are available at: http://creativecommons.org/licenses/ by/4.0/

LICENCE

CC BY 4.0

REPOSITORY RECORD

Michalaki, Paraskevi, Mohammed Quddus, D.E. Pitfield, and Andrew Huetson. 2019. “A Time-series Analysis of Motorway Collisions in England Considering Road Infrastructure, Socio-demographics, Traffic and Weather Characteristics”. figshare. https://hdl.handle.net/2134/20631.

ARTICLE IN PRESS

Journal of Transport & Health ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Journal of Transport & Health

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

A time-series analysis of motorway collisions in England considering road infrastructure, socio-demographics, traffic and weather characteristics

Paraskevi Michalaki [a,*], Mohammed Quddus [a], David Pitfield [a], Andrew Huetson [b]

[a] School of Civil and Building Engineering, Loughborough University, Loughborough, Leicestershire LE11 3TU, United Kingdom
[b] Connect Roads – Balfour Beatty, Bramham Maintenance Compound, Spen Common Lane, Tadcaster LS24 9NS, United Kingdom

ARTICLE INFO

Article history:
Received 16 June 2015
Received in revised form
19 September 2015
Accepted 22 October 2015

Keywords:
Public health
Motorway
Hard–shoulder
Vector autoregressive

ABSTRACT

Traffic injuries on motorways are a public health problem worldwide. Collisions on motorways represent a high injury rate in comparison to the entire national network. Furthermore, collisions that occur on the hard–shoulder are even more severe than those that happen on the main carriageway. The purpose of this paper is to explore motorway safety through the identification of patterns in the sequence of monthly hard–shoulder and main carriageway collisions separately over a long period of time (1993–2011) by using reported collision data from British motorways. In order to examine the trends of hard–shoulder and motorway collisions over the same period, a Vector Autoregressive (VAR) model is developed; this allows the inclusion of two time-series in the same model and the examination of the effect of one series on the other and vice-versa. Exogenous variables are also added in order to explore the long-term factors that might affect the occurrence of collisions. The factors considered are related to the infrastructure (e.g. length of motorways), socio-demographics (e.g. percentage of young drivers), traffic (e.g. percentage of vehicle-miles travelled by Heavy Goods Vehicles) and weather (e.g. precipitation). The results suggest different patterns in the sequences in terms of the lingering effects of preceding observations for the two time-series. In terms of the significance of exogenous variables, it is suggested that main carriageway collision frequency is affected by weather conditions and the presence of Heavy Goods Vehicles, while hard–shoulder collisions are decreased by the presence of Motorway Service Areas, which allow a safe exit off the motorway to stop and rest in case of fatigue.

© 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Background and scope of the study

Road traffic injuries have been an important global public health problem, being the eighth leading cause of death and the number one cause among those aged 15–29 years according to the World Health Organisation (2013). Indicatively, from 2007 to 2010, there were approximately 1.3 million road traffic deaths worldwide each year. Specifically in Great Britain (GB), there were 183,670 casualties in 2013, of which 23,370 were killed or seriously injured casualties (Department for Transport, 2014).

Motorway collisions appear to have a much greater severity level than the average of the national road network collisions in GB. This is reflected by higher fatal injury rates on the motorway network. Motorway collisions can also be distinguished according to whether they occur on the hard–shoulder or not. The hard–shoulder is the lane on the verge (edge) of the road which is generally only permitted to be used for emergency purposes by the road users, for example in the case of a breakdown. Systematically the lane is used by road maintenance vehicles and emergency vehicles (e.g. ambulances and police vehicles) (Highway Code, 2007). Sometimes those vehicles are obliged to remain on the hard–shoulder for a long period of time while the personnel are performing their tasks. Working on the hard–shoulder and roadside verges on motorways is therefore considered to be unsafe and risky. Data from GB suggest that a hard–shoulder (HS) collision is five times more possible to be fatal in comparison to collisions occurring on the main traffic lanes, also called the main carriageway (MC).

Exploring the potential long-term factors affecting the frequency of collisions on the motorway HS and MC can assist in formulating safety measures to reduce them. This way, it should also be possible to identify any differences between the two types of collisions that

* Corresponding author. Tel.: +44 7964379764.
E-mail addresses: P.Michalaki@lboro.ac.uk (P. Michalaki), M.A.Quddus@lboro.ac.uk (M. Quddus), D.E.Pitfield@lboro.ac.uk (D. Pitfield), Andrew.Huetson@balfourbeatty.com (A. Huetson).

http://dx.doi.org/10.1016/j.jth.2015.10.005
2214-1405/© 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Please cite this article as: Michalaki, P., et al., A time-series analysis of motorway collisions in England considering road infrastructure, socio-demographics, traffic and weather characteristics. Journal of Transport & Health (2016), http://dx.doi.org/10.1016/j.jth.2015.10.005

would lead to a different approach by safety professionals to prevent them. The goals of this analysis are to examine the motorway collisions over time and to investigate various control and exposure variables affecting such collisions that can potentially explain the evolution of the two different series.

The rest of the paper is organised as follows. Firstly, a review of literature on factors affecting motorway collisions and the statistical methods employed to investigate time-series data is presented. This is followed by the data collected for the HS and MC collisions. This section also looks at various descriptive statistics so as to draw facts about the sample data. Subsequently, the statistical method chosen to fulfil the objectives of this paper is presented, followed by discussion of the results from the statistical modelling. Finally, a summary of the research and some conclusions are drawn.

2. Literature review

The purpose of this section is to identify some exposure and control variables that may affect motorway collisions and to explore the literature on the statistical methods employed in analysing time-series data, especially collision data.

2.1. Factors affecting motorway collision occurrences

Safety literature on motorway collisions spans several decades and is extremely rich. The factors that have appeared to affect road collision likelihood are generally divided into two categories: *engineering* and *human*.

Engineering factors are related to the *road infrastructure characteristics* and *traffic conditions*, as well as the *weather conditions*. The first category refers to attributes such as road curvature and lane width (*e.g.* Haynes et al., 2007; Kononov et al., 2008). Traffic related conditions that primarily affect the collision occurrences are vehicle speed – which is further split in various forms such as average speed, speed variance and speed limit (*e.g.* Elvik et al., 2004; Aarts and van Schagen, 2006) – as well as traffic flow and the composition of traffic, such as the percentage of Heavy Goods Vehicles (HGVs). Weather conditions also appear to be a quite significant factor in many aspects: precipitation (rain or snow), wind and fog (*e.g.* Edwards, 1994; Brijs et al., 2008).

Human factors play a very important role in road collisions. The risk of being involved in a collision varies in different driving groups. For example, HGVs drivers have been proved to be a high risk group (*e.g.* Charbotel et al., 2002). Driver age too appears to be important as younger drivers are more commonly involved in collisions whether from inexperience, in-vehicle distraction or other reasons (*e.g.* NHTSA, 2008; Hassan and Abdel-Aty, 2013). Gender has also been found to be important as male drivers generally are more often involved in collisions than females. It is further recognised that they are more prone to driving violations, and risky driving (*e.g.* Evans, 1991; Constantinou et al., 2011). Conversely, alcohol impairment is a common causal factor of collisions, especially the ones involving injury, amongst men and women (*e.g.* Holubowycz et al., 1994).

Driver fatigue appears to be one of the most often reported factors in road collisions, especially with HGV drivers (*e.g.* Nordbakke and Sagberg, 2007; Zhang and Chan, 2014). On British motorways, drivers have been encouraged to take a break when feeling tired with the establishment of safety signs in the 1990's by the then Highways Agency (now Highways England). These permanent messages read 'Tiredness can kill – Take a break' and are normally placed ahead of the Motorway Service Areas (MSAs). Their aim is to remind drivers to stop when necessary in safe areas and not in inappropriate places such as the HS (Horne and Reyner, 2001). The MSAs are government-approved facilities in the United Kingdom (UK) designed to provide a safe exit and rest areas for drivers using the motorway network (Motorway Services Online, 2015).

2.2. Statistical methods for modelling time-series collision data

Initial studies have focused on the development of a safety performance function based on traffic flow and density as predictors (*e.g.* Cedar and Livneh, 1982; Golob et al., 2003, 2004). Several studies have, however, concentrated on the application of advanced statistical techniques, including controlling for spatial dependency among adjacent segments and taking into account issues associated with nested collision data by employing multilevel modelling. These are to reliably investigate the contributory factors (e.g. vehicles miles travelled, rainfall, precipitation, road geometry and other vehicle related factors) in explaining the variation in road collisions (*e.g.* Jones et al., 1991; Miaou, 1994; Shankar et al., 1996; Caliendo et al., 2007; Lord and Mannering, 2010).

Most existing studies have focused on the total number of collisions that occurred on the motorways (highways/freeways) rather than distinguishing HS collisions from the MC collisions (*e.g.* Shankar et al., 1996; Lord et al., 2005; Davis and Swenson, 2006; Golob et al., 2008; Wang et al., 2009). Although the width of the HS has sometimes been employed as a predictor of motorway collision frequency (*e.g.* Noland and Oh, 2004), there is a dearth of literature on HS collisions and consequently, there is little known about the scale of safety occurrences on the HS.

A well-established choice for detecting patterns, trends and seasonality in a continuous time-series data set is the use of the Autoregressive Integrated Moving Average (ARIMA) model proposed by Box and Jenkins (1976), which has been applied to model time-series count data in many applications over the last few decades including road collisions (*e.g.* Zimring, 1975; Houston and Richardson, 2002; Goh, 2005; Noland et al., 2008). Since road collisions are non-negative, integer and random event counts, Karlis (2006) argued that a classical ARIMA model may not be suitable in modelling time-series count data. However, Quddus (2008) pointed out that an ARIMA model performs well in dealing with aggregated time-series count data, especially when the mean of the count is relatively large.

Although modelling data sets for MC and HS collisions independently would reveal patterns, trends and seasonality of individual time-series, a way of combining and examining them together would appear more valuable. In terms of revealing the relationship between two time-series data sets, the cross-correlation function has been used in time-series modelling in various disciplines (*e.g.* Haugh, 1976; Lopez-Lozano et al., 2000; Pitfield, 2005; Zebende and Filho, 2009; Brockwell, 2010; Zebende et al., 2011). For example, Haugh (1976) suggested the residual cross-correlation function which provides a means for testing the hypothesis that two stationary time series are independent (Koch and Yang, 1986; Duchesne et al., 2011).

An alternative method, which has had a long tradition as a tool for multiple time-series analysis is the Vector Autoregressive (VAR) models. (*e.g.* Quenouille, 1957). This method involves the simultaneous modelling of more than one time-series, while, it allows for exogenous variables to be included in the model (e.g. Lütkepohl, 2005). These variables can be related to and possibly explain the trends that the time-series follows. Several past studies have incorporated explanatory variables to investigate the correlation between the

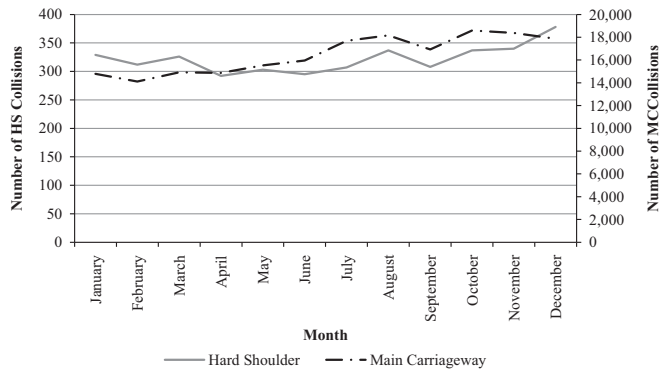

Fig. 1. Monthly distribution during the year of the number of motorway collisions on hard-shoulder and main carriageway in GB (1985–2011).

Table 1
Severity of collisions on motorways in Great Britain.

Severity of collision	***Hard–shoulder collisions* (1985–2011)**		***Main carriageway collisions* (1985–2011)**		***GB all road collisions* (2011)**
	Frequency	Relative frequency (%)	Frequency	Relative frequency (%)	Relative frequency (%)
Fatal	391	10.7	3647	2.3	1.2
Serious	912	25.0	20,392	13.0	13.8
Slight	2352	64.3	133,595	84.7	85.0
Total	3655	100	157,634	100	100

frequency of road collisions and characteristics of the road, driver and weather (*e.g.* Beenstock and Gafni, 2000; Bergel-Hayat et al., 2013; Gomes, 2013; Theofilatos and Yannis, 2014).

In this study, the time-series of HS collisions is examined separately from the MC in order to reveal any differences in the way they have evolved throughout the years. The VAR model would allow for the simultaneous modelling of the two series to investigate the possible effect on each other, while lags could be included. The addition of exogenous variables is necessary to include driver exposure to risk and to control for other factors that might be associated with the collisions, such as traffic or road related conditions. As risk of having a collision on the HS might depend on the presence of HGVs travelling in the nearside lane, the study focuses on this group of drivers, along with the other exogenous variables. In addition, as the HS is often abused by drivers stopping for rest or other non-emergency reasons, the possible impact of the presence of the MSAs is investigated.

3. Collision data collection and descriptive statistics

National collision data for GB were obtained from the police, who have since 1985 stored details using the STATS19 data collection system. Only the collisions in which one or more persons are killed or injured, and involved at least one motor vehicle are recorded. The motorway collisions, including A-roads that have been upgraded to motorways, known as A(M) roads, are extracted and two sets of monthly time-series data are created according to where the first impact happened, thus collisions that occurred on the MC and collisions that happened on the HS of the motorways in GB. The time period of data availability for this study is from 1985 to 2011. Over these 27 years, a total of 199,388 injury collisions (in which 2.3% are fatal, 13.0% involve a serious injury and 84.7% involve a slight injury) occurred on the GB motorways of which 2% occurred on the HS.

All tables and figures in this section were produced for this study using the STATS19 data, unless stated otherwise. Fig. 1 shows the monthly distribution of the reported road collisions for each of the two groups, MC and HS. They generally follow similar increasing/decreasing patterns from month to month. The lowest values for HS collisions are observed in April/May/June, while the highest are in November/December. MC collisions are mostly increased in October/November, while their lowest frequencies occur in January/February. The relative difference between the two extremes is approximately 25% for HS and 35% for MC collisions. Traffic and weather conditions are also fluctuating throughout the year; thus it needs to be examined whether they affect the collision monthly frequencies. The daily distribution during the week follows the same trend for both (peaks on Monday and Friday) and within a day both series exhibit similar patterns of hourly traffic flow with two defined peaks. In terms of types of vehicles involved in motorway collisions, a significant difference is observed in the rate of cars and HGVs between the two groups. More specifically, the percentages for HGVs involved in collisions are found to be 17.6% for MC and 34.7% for the HS collisions.

Table 1 shows the severity of collisions for each of the two groups. In the case of HS, the percentage of fatal injury collisions is almost five times higher than that of MC. The proportion of serious injury collisions is also relatively high in the case of HS collisions. It can therefore be stated that collisions on the HS are relatively more severe than those on the MC, which stresses the need for their investigation. A previous study (Michalaki et al., 2015) has shown that HGV involvement and fatigue tend to increase collision severity on the HS.

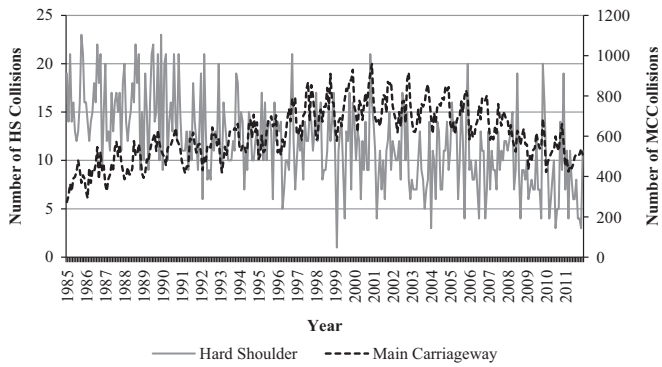

Fig. 2. Evolution of the number of motorway collisions on hard–shoulders and main carriageway per month in GB (1985–2011).

Table 2
Summary statistics for the exogenous variables in the UK.

Variable		*Period*	*Source*	*Obs.*	*Mean*	*Std. Dev.*	*Min*	*Max*
HS collisions (month)		1993–2011	STATS19	228	10.61	3.78	1	21
MC collisions (month)		1993–2011	STATS19	228	664.18	115.57	420	959
VMT motorways (millions)		1993–2011	DfT	228	55.68	6.37	42.4	62.5
% VMT by HGVs on motorways		1993–2011	DfT	228	12.67	0.58	11.4	13.3
% cars exceeding speed limit		2002–2011	DfT	120	52.80	2.90	49	57
% cars exceeding speed limit by more than 10 mph		2002–2011	DfT	120	16.90	2.22	13	20
Motorway Service Areas (# per 100 miles of motorways)		1993–2011	Motorway Services Online	228	4.98	0.42	3.86	5.50
Total length of motorways (miles)		1993–2011	DfT	228	2144.32	71.10	1995	2218
Road surface condition (%)		2003–2011	DfT	108	5.12	0.78	3.65	5.88
Public expenditure on transport (£m)		1993–2011	HM Treasury	228	17.59	4.51	10.60	24.50
Number of vehicles		1994–2011	DfT	216	30458	3141	25231	34228
Average age of cars (year)		1994–2011	DfT	216	6.92	0.25	6.61	7.54
% Driving licence holders aged	from 17 to 20	1993–2011	DfT, National Travel Survey (NTS)	228	36.74	5.56	27	47
	from 21 to 29			228	69.05	4.40	63	75
	from 30 to 39			228	82.32	1.63	78	85
	from 40 to 49			228	83.11	1.07	81	85
	from 50 to 59			228	79.95	2.77	75	83
	from 60 to 69			228	71.26	5.94	61	80
	over 70			228	45.95	7.07	35	59
% Trips by young drivers (under 29)		1995–2011	DfT, NTS	204	0.14	0.02	0.11	0.17
% VMT by young drivers (under 29)		2002–2011	DfT, NTS	120	0.19	0.01	0.17	0.20
Precipitation (mm)		1993–2011	Met Office	228	84.55	36.96	11.94	198.89
Temperature (°C)		1993–2011	Met Office	228	9.10	4.32	−0.90	17.80
Sunshine (hours per month)		1993–2011	Met Office	228	116.80	54.95	20.50	253.30

In order to see whether there are trends and seasonality in the motorway collision data, two time-series plots are produced (Fig. 2). They show strong seasonality for the MC collisions while there are both upward and downward trends. HS collisions steadily decrease over the study period.

In addition to the endogenous (dependent) variables, which are the number of monthly collisions, data for exogenous variables that could possibly explain the variation of the collision frequency throughout the years are collected. These variables are related to the use of motorways, the infrastructure, vehicle and driver characteristics, as well as the weather conditions. However, the method of collection of historical traffic data on British motorways was changed in 1993. For consistency across variables, data are collected since 1993, or otherwise (see Table 2) from when these have become available. Their time interval is either a year or a month. The sources of these data were all available online: Department of Transport (DfT) in the UK, the UK Met Office, HM Treasury and Motorway Services Online. Table 2 provides the summary statistics for the exogenous variables and Fig. 3 the trends of some for the years available.

3.1. Motorway traffic

- *Vehicle miles travelled (VMT)* on motorways in GB in millions by year
  Vehicle miles travelled refer to the actual use of motorways by any type of vehicle and are used as a measurement of the exposure of drivers to the risk of having a collision. A steady increase is observed from 1993 before becoming stable after 2006.

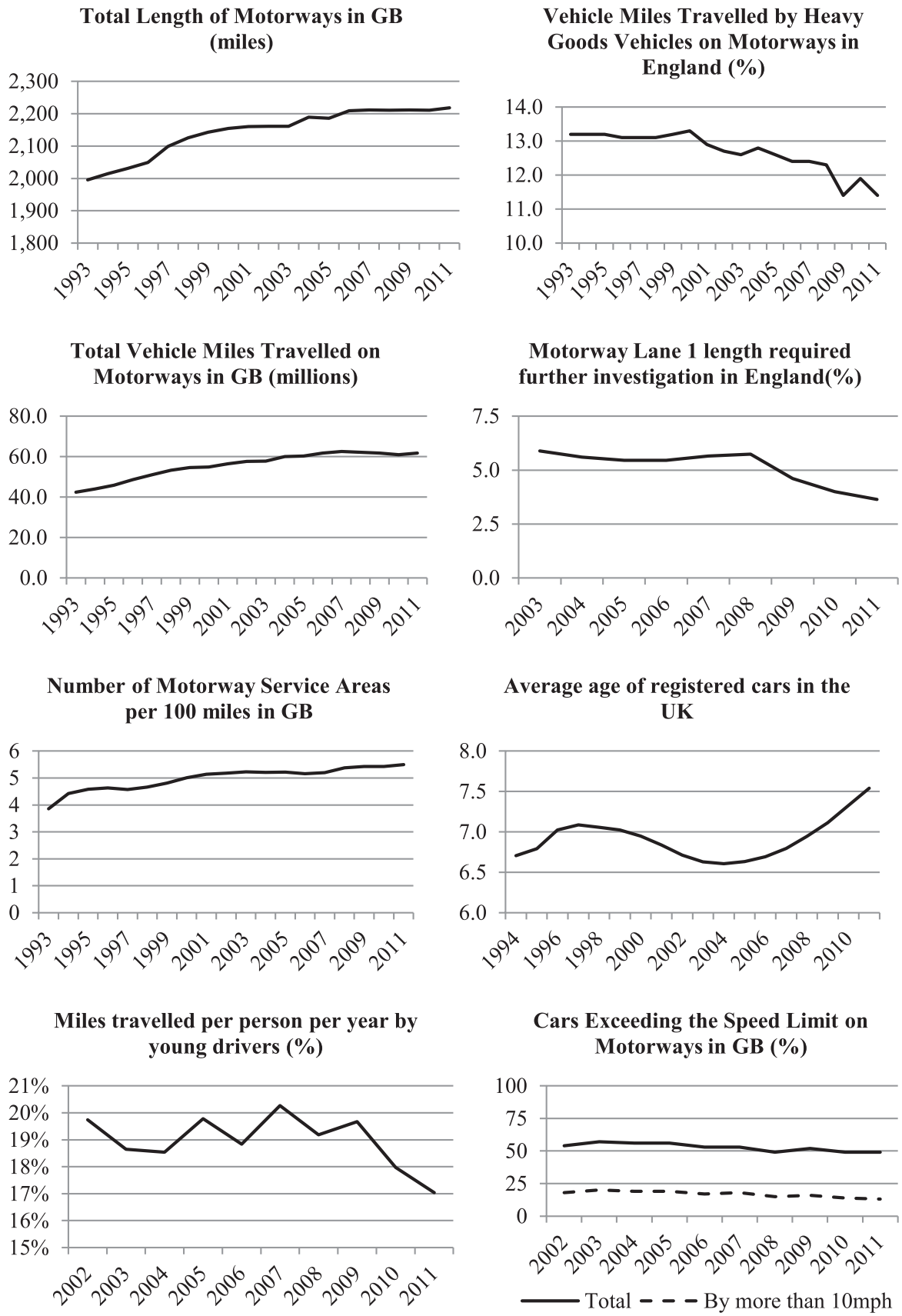

Fig. 3. Trends of exogenous variables for the years available.

Please cite this article as: Michalaki, P., et al., A time-series analysis of motorway collisions in England considering road infrastructure, socio-demographics, traffic and weather characteristics. Journal of Transport & Health (2016), http://dx.doi.org/10.1016/j.jth.2015.10.005

- *Percentage of miles travelled by HGVs* on motorways in England by year

It is of interest to investigate whether the presence of HGVs on motorways affects the likelihood of a collision, as they appear to be involved in a high percentage of motorway collisions, especially in the case of HS. This percentage has been decreasing since 1993 while a small peak is observed in 2010. It is noted that the absolute value of VMT by HGVs has been increasing. However, the reduction of the percentage is perhaps due to a greater increase of VMT by other vehicles.

- *Percentage of cars exceeding the speed limit* on motorways in the UK (overall or by more than 10 mph) by year

Contradictory results have been suggested regarding the relationship between speed of vehicles and collision likelihood. The speed limit in the UK is normally 70 mph (113 km/h). Cars exceeding the speed limit by more than 10 mph (16 km/h) are selected as an indicator of the driver's behaviour. Both percentages have been slowly decreasing since 2002 showing a general improvement in drivers' compliance.

3.2. Infrastructure characteristics

- *Motorway Service Areas (MSAs)* in the UK by year
  These are the rest areas where drivers can leave the motorway. Since the first MSAs opened in 1959, new ones have been installed across the country. This variable is defined as the number of MSAs per 100 miles of motorways. There has been a 43% increase between 1993 and 2011.
- *Total length of motorways* in GB in miles by year
  It is noted that the length of the HS is the same as that of MC, as all motorways, by definition, have an emergency lane. The width of HS is not included as a variable, as it is generally following the British Standards and varies only when a Departure from Standards has been granted for physical or other reasons. There has been a 11% increase in the total length of British motorways between 1993 and 2011.
- *Road surface condition* in England by year
  It is defined as the percentage of lane 1 length (next to the HS) surveyed requiring further investigation. This is stable from 2003 to 2008, when a significant decrease is noticed.
- *Public expenditure on transport* in the UK by year

The values were adjusted to 2012–13 price levels using Gross Domestic Product (GDP) deflators from the Office for National Statistics in the UK. There has been a fluctuation throughout the years of data collection.

3.3. Vehicle characteristics

- *Total number of vehicles* registered in the UK by year
  The number of vehicles has been constantly increasing since the data has been collected (1994).
- *Average age of cars* registered in the UK by year

It represents the progression of technology in the car industry, which might be related to the reduction of motorway collisions. The average age increased from 1994 to 1997 and then again after 2004.

3.4. Drivers' characteristics

- Percentage of population per age group that hold a driving licence in the UK by year
- Percentage of trips in the UK by young drivers (aged from 17 to 29) by year
- Percentage of miles travelled per age group as a car/van driver by year

These variables are included to control for driver experience. The most representative variable is the one referring to the miles travelled; however this information is only available since 2002. In the study, focus is on the miles travelled by young drivers. A drawback of this data is that the trips/miles do not only refer to motorways; however, it is assumed is that they still capture the general level of experience of a driver.

3.5. Weather conditions

- Total precipitation in the GB in mm by month.
- Average temperature in the UK in Celsius degrees by month.
- Total hours of sunshine in the UK by month.

Generally, according to previous studies (*e.g.* Hermans et al., 2006; Caliendo et al., 2007), a correlation between the increased collision frequency and either total precipitation, temperature or sunshine hours is expected. However, opposite results for the effect of precipitation have, to a smaller extent, also been suggested by others (*e.g.* Karlaftis and Yannis, 2010; Bergel-Hayat et al., 2013).

In the case of precipitation data, these were available for England+Wales (EW) and Scotland (SC) separately. Weighted average values were used according to the length of motorways, as in Scotland their lengths are limited in comparison to the rest of GB (England and Wales contain 87% of the total British motorway network).

4. Time-series statistical modelling using the vector autoregression analysis

The objectives of the study are to examine the relationship between the frequency of HS and MC collisions over time and to relate their frequencies to exogenous factors. Therefore, a statistical model that addresses both objectives must be selected. As discussed in Section 2.2, the Vector Autoregression (VAR) model can simultaneously analyse the relationship between two time-series datasets, while its variation, termed VAR(X), allows the inclusion of exogenous variables. In this section, attributes of this model and its suitability for this study are discussed.

The stationarity of a time-series is an assumption commonly required in statistical models. When a stochastic process is stationary, the statistical properties of the process are not a function of time (Box and Jenkins, 1976). Besides reducing the mathematical complexity of a stochastic model, the stationarity assumption may reflect reality. In certain situations, the statistical characteristics of a process are a function of time – known as a non-stationary series. To model an observed time series that possesses non-stationarity, a common procedure is to first remove the non-stationarity by invoking a suitable transformation – such as using the differences (regular, seasonal or both) of the original series – and then to fit a stationary stochastic model to the transformed sequence (Box and Jenkins, 1976). An advantage of developing a VAR model in opposition to other time-series modelling methods is that the stationarity assumption of time series data set is not mandatory (Canova, 2007).

Cointegration of the dependent variables – in this study the number of collisions on the HS and MC – needs to be tested. This would indicate any linear relationship among them and thus that the VAR specification would not be the most suitable representation. If there is at least one cointegration equation, a different version of the model should be applied, known as the Vector Error Correction (VEC) model, which contains the cointegration relations. For this check, Johansen's test may be used (Johansen, 1991).

The VAR model fits a multivariate time-series regression of each dependent variable on lags of itself and on lags of all the other dependent variables. The lags are treated as explanatory variables. A variant of the VAR model can also be used to allow the inclusion of exogenous variables in the regression to possibly explain the evolution of the dependent ones, known as VAR(X). These two characteristics of the VAR model suit the objectives of this study as the time-series collisions on the HS and MC can be included in the same model along with all the engineering and human factors that possibly affect their occurrence. A VAR model is easy to use and interpret (Watson, 1994). The analysis typically proceeds by specifying and estimating a model and then checking for its adequacy. Model revisions are performed until a satisfactory model has been found.

The VAR(p) model with exogenous variables is written as (Lütkepohl, 2005)

$$y_t = AY_{t-1} + Bx_t + u_t$$

where:

p is the number of lags,
K is the number of endogenous variables
M is the number of exogenous variables
y_t is the vector of endogenous variables (the length of vector is K),
A is a $K \times K \cdot p$ matrix of coefficients,
B is a $K \times M$ matrix of coefficients
x_t is the vector of exogenous variables (the length of vector is M)
u_t is the vector of white noise innovations (the length of vector is K)
Y_{t-1} is the vector given by $Y_{t-1} = \begin{pmatrix} y_{t-1} \\ \vdots \\ y_{t-p} \end{pmatrix}$. The length of vector is $K \cdot p$ and is formed by stacking the y_t vectors for all the lags.

Intercept terms in the model are included in the x_t. The coefficients are estimated by iterated seemingly unrelated regression, a three-stage least squares (3SLS) method (see Zellner and Theil, 1962; Weesie, 1999).

Since the restriction of stationarity is not imposed before applying the model, the residuals of the models are checked for non-stationarity. Firstly, the residuals are plotted in order to identify any patterns (Canova, 2007). In addition, the Lagrange multiplier (LM) test for autocorrelation in the residuals resulting from the VAR model can be employed (Johansen, 1995) in the postestimation phase. The null hypothesis is that there is no autocorrelation between the residuals.

In order to investigate whether one series can cause the other, the Granger causality test can be employed. A variable x is said to Granger-cause a variable y if, given the past values of y, the past values of x are useful for predicting y (Granger, 1969). A common method for testing Granger causality of variable x on variable y is to test the null hypothesis that the estimated coefficients on the lagged values of x are jointly zero using Wald tests. Failure to reject the null hypothesis is equivalent to failing to reject the hypothesis that x does not Granger-cause y.

5. Estimation results of the VAR(X) model and discussion

Two time-series of aggregated data (i.e. all motorways in GB) are formed: the monthly number of collisions that occurred on the motorway HS in GB and the monthly number of collisions that occurred on the motorway MC in GB over the same period of time. Different time-series collision models are developed using the VAR(X) method, as described in Section 4, for HS and MC collisions in order to identify the 'best-fitted' collision model. Exogenous variables are included for the time period that data are available.

Table 3
Model Fit Statistics for the VAR model including the logarithms of hard–shoulder and main carriageway collisions (data from 1993 to 2011).

Lag	*FPE*	*AIC*	*HQIC*	*SBIC*
0	0.0052	0.4185	0.4312	0.4498
1	0.0021	−0.5104	−0.4725	−0.4166
2	0.0021	−0.5119	−0.4488	−0.3556*
3	0.0021	−0.4822	−0.3938	−0.2635
4	0.0021	−0.4806	−0.3669	−0.1993
5	0.0022	−0.4638	−0.3249	−0.1201
6	0.0022	−0.4423	−0.2782	−0.0360
7	0.0020	−0.5317	−0.3423	−0.0629
8	0.0019	−0.5839	−0.3692	−0.0526
9	0.0019	−0.5725	−0.3326	−0.0213
10	0.0018	−0.6616	−0.3964	−0.0053
11	0.0016	−0.7704	−0.4800	−0.0516
12	.0014*	−0.9166*	−0.6010*	−0.1353

5.1. Lag selection and cointegration

In order to apply the VAR and VAR(X) models, the number of lags included needs to be selected and cointegration between the series to be tested. In the preestimation phase, a set of criteria are estimated for models without any exogenous variables. This is applied for each of the models, including lags up to 12. In addition, the existence of the cointegration equation is tested for several forms of the dependent variables, such as the number of MC and HS collisions, their natural logarithms and their first and seasonal differences. Based on the final prediction error (FPE) (Lütkepohl, 2005), Akaike's Information Criterion (AIC) (Akaike 1973, 1974), Hannan and Quinn Information Criterion (HQIC) (Hannan and Quinn, 1979), 12 lags should be included in the model for all cases of dependent variables. Only the Bayesian Information Criterion (BIC) (Schwarz, 1978) suggests the use of 2 lags. Indicatively, Table 3 shows the results of the tests for the VAR model where two dependent variables are included; the logarithms of HS and MC collisions (years 1993–2011). When the natural logarithms of the dependent variables are taken and 12 lags are included, there are no cointegration equations; suggesting that the VAR model can be used, while the stationarity of residuals is tested at the end.

5.2. Results of VAR(X) model

Table 4 shows the results for the VAR(X) model for the MC and HS motorway collisions using data from 1993 to 2011, as allowed by the exogenous variables' availability. Since some of these variables were available for an even shorter period of time, models including variables available since 2003 were also tested; however, a more suitable model was not indicated. Due to autocorrelation between the residuals, some of the lags were excluded from the model.

The residuals of the model estimated are then tested for stationarity to check whether a way to control non-stationarity should have been taken into consideration. The residuals of the two series are plotted along with the 1-lag and 12-lag residuals (Figs. 4 and 5). Since no correlation pattern is observed, the choice of this VAR model appears to be valid.

For further confirmation of validity, the Lagrange-multiplier test is used (Table 5). The null hypothesis of no autocorrelation cannot be rejected at a 90% level, suggesting no misspecification of the model.

5.3. Discussion

The first part of the results (‘***Lags***’) shows the relationship between the number of MC and HS collisions and the previous values of the same series (monthly lags) up to one year (lag 12). In addition, it presents the relationships of MC collisions with the lags of HS collisions and vice versa. From Table 4, it is noticed that there are differences in which lags are significant across the two equations.

The coefficient value for the MC in relation to lag 1 of MC collisions (MC–MC) is +0.223, while to lag 12 is +0.467, showing a positive relationship between the values of a month with the month before and the corresponding one of the previous year. This suggests that there is monthly and annual seasonality in MC collisions. In addition, a negative relationship is expected between current values and values 6 months later, as the MC–MC coefficient at lag 6 is –0.102 ($z=-2.070$). Accordingly, seasonality in the HS series (HS–HS) is not apparent in this model, as none of the coefficients are significant.

The simultaneous modelling of the two-series provides the opportunity to relate the current values of one series to the lags of the other. As it is shown in the MC equation, several coefficients of the HS lags are statistically significant; more specifically lags 2, 5 and 6. For instance, the value of lag 6 coefficient for MC–HS is –0.025, showing a negative relationship between the current values of MC collisions and the values of HS collisions six months before. The same applies in the other equation, in the case of HS collisions as the dependent variable, where the value of lag 12 coefficient for HS–MC is +0.596. This suggests that when the values of MC collisions have increased, the values of HS collisions in a year's time will also increase, although this does not imply any cause–causality relationship.

The Granger test results (Table 6) suggest that for the MC collisions, the null hypothesis can be rejected; thus, the coefficients of the lags of HS collisions are not jointly zero and would be useful for MC collision prediction. On the contrary, MC collisions do not seem to be useful for the HS collisions' equation, as the null hypothesis cannot be rejected. This also supports that the evolution of HS collisions does not have a strong relationship with MC collisions, showing that the two series have to be examined separately.

Out of all the exogenous variables investigated in this model (‘***Exogenous variables***’), some appear to be significant in the equation of MC collisions and one of them for the HS collisions. The proportion of vehicle miles travelled by HGVs on motorways appears to have a positive relationship with the evolution of collisions confirmed by the positive coefficient and statistically significant value (+0.057). This

Table 4
Results of VAR(X) model for motorway collisions (1993–2011).

Vector autoregression				
Sample	1993–2011			
Equation	**MC collisions**		**HS collisions**	
Variable	**Coef.**	***z-statistic***	**Coef.**	***z-statistic***
Lags				
Log (MC collisions)				
Lag 1	0.223	***4.120***	0.173	0.550
Lag 2	0.139	***2.750***	0.178	0.610
Lag 4	0.108	***2.010***	−0.581	−***1.880***
Lag 5	0.043	0.810	0.254	0.820
Lag 6	−0.102	−***2.070***	0.457	1.610
Lag 9	−0.131	−***2.490***	−0.175	−0.570
Lag 11	0.147	***2.950***	−0.440	−1.520
Lag 12	0.467	***8.720***	0.596	***1.920***
Log (HS collisions)				
Lag 1	0.009	0.770	0.039	0.570
Lag 2	−0.280	−***2.360***	−0.088	−1.260
Lag 4	−0.016	−1.300	0.108	1.530
Lag 5	−0.042	−***3.470***	−0.004	−0.050
Lag 6	−0.025	−***1.970***	0.021	0.290
Lag 9	−0.006	−0.510	0.009	0.120
Lag 11	−0.017	−1.380	0.001	0.020
Lag 12	−0.001	−0.050	−0.046	−0.660
Exogenous variables				
% VMT by HGV on Motorways	0.063	***2.540***	−0.002	−0.020
Precipitation	0.001	***5.760***	0.001	0.820
Sunshine	0.001	***5.400***	0.000	−0.450
Motorway Services	−0.047	−1.240	−0.369	−***1.660***
Constant	0.233	0.800	1.060	0.630
Equation	RMSE	*R*-sq	chi^2	$P > chi^2$
Log (MC collisions)	0.06913	0.8596	1322.375	0.0000
Log (HS collisions)	0.40017	0.1977	53.238	0.0001
Parameters	21			
Model Fit Statistics				
Log likelihood	188.371			
FPE	0.00089			
AIC	−1.35528			
HQIC	−1.09014			
SBIC	−0.69898			
No of observations	216			

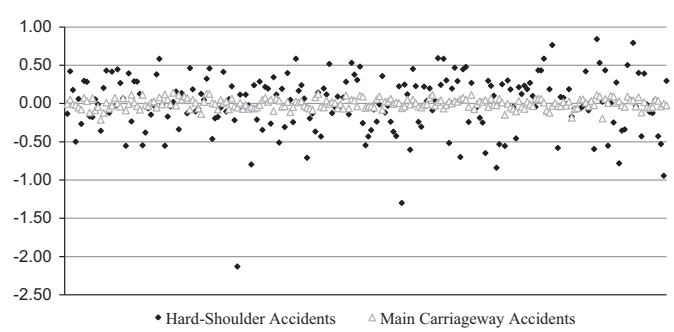

Fig. 4. Residuals of the VAR model for the hard–shoulder and main carriageway collisions equations.

finding is in line with other existing studies (*e.g.* Charbotel et al., 2002). Due to multi-collinearity, the proportion of HGVs could not be included in the model along with the total vehicle miles travelled.

Regarding the weather conditions, it is argued that both precipitation and hours of sunshine per month increase the number of MC collisions. As in other studies (*e.g.* Antoniou et al., 2013), it is also supported here that rainfall and snowfall are associated with a higher number of collisions. In addition, the hours of sunshine per month appears to have a positive relationship with the number of MC collisions, as has been suggested in other studies (Hermans et al., 2006). Under rain conditions, lower visibility is expected as well as a higher risk of skidding/losing control of the vehicle. On the other hand, sunny weather can be linked with a higher likelihood of glaring, especially on the motorway. It could also be suggested that, from a behavioural point of view, drivers become less careful and concentrate less when the conditions on the road appear good.

Motorway Service Areas appear to be a significant factor for HS collisions, having a negative relationship at the 90% confidence level. This indicates the importance for the drivers to be able to exit the motorway safely and take any preventative action required when tired.

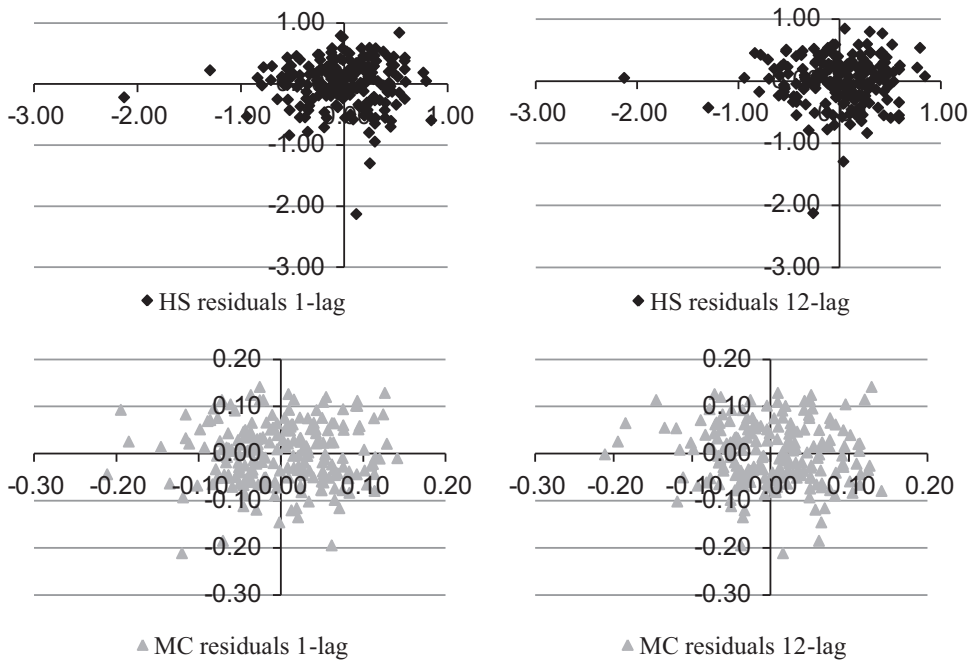

Fig. 5. Relationship of residuals of the VAR model for the hard–shoulder and main carriageway collisions equations with lagged residuals.

Table 5
Lagrange-multiplier test results for VAR(X) model.

Lag	Chi^2	df	Prob > chi^2
1	2.8678	4	0.58019
2	1.3234	4	0.85740
3	1.6214	4	0.80495
4	5.1672	4	0.27057
5	4.2478	4	0.37351
6	1.9376	4	0.74724
7	4.4969	4	0.34292
8	5.7976	4	0.21478
9	2.7492	4	0.60063
10	2.0303	4	0.73018
11	8.1139	4	0.08749
12	3.5578	4	0.46915

Table 6
Granger causality Wald tests for VAR(X) model.

Equation	Excluded	Chi^2	df	Prob > chi^2
Log (MC collisions)	Log (HS collisions)	25.335	8	0.001
Log (MC collisions)	ALL	25.335	8	0.001
Log (HS collisions)	Log (MC collisions)	12.213	8	0.142
Log (HS collisions)	ALL	12.213	8	0.142

In addition, it has been observed that drivers do stop on the HS for non-emergency reasons as to check directions or make a phone call. The presence of MSAs on the motorway, including the frequent signage that indicates the distances to the next ones, discourages the drivers from stopping on the HS – unless there is an unexpected emergency – reducing the exposure of stopped vehicles to live traffic. It would be of interest to examine how the use of MSAs by the road users has evolved throughout the years, but these data were not available. In addition, as fatigue is an important factor of motorway collisions (Michalaki et al., 2015), drivers are encouraged to have a break at the MSAs when feeling too tired to continue their journey. This is supporting Horne and Reyner's (2001) work to establish signs on motorways to encourage the use of rest areas when required.

6. Conclusions

This paper focused on a major public health problem, that of motorway collisions. A special interest was shown in the road collisions that occur on the emergency lanes of motorways – the hard–shoulder (HS). Using 27 years of monthly collision data from Great Britain, it was found that HS collisions are much more severe than main carriageway (MC) collisions (10.7% of HS collisions are fatal vs. 2.3% of MC collisions

being fatal). The time-series plots provided information in terms of trend and seasonality, showing evident differences. After having comprehended the data by employing descriptive statistics, time-series modelling was applied, and more specifically the Vector Autoregressive Analysis. The advantage of the VAR model is providing a way of examining the joint evolution of two time-series that are not expected to progress in isolation.

Firstly, the model, which included 12 monthly lags as explanatory variables, suggested that the relationship between the two series is not strong. In order to investigate the factors that could be related to the number of collisions, exogenous variables were also incorporated. However, due to limitations in data availability, smaller dataset was used (1993–2011). Exogenous variables that appeared to be significant in the VAR model for the MC collisions were related to the HGVs travelling on motorways, as well as the weather conditions – especially precipitation and hours of sunshine – while the presence of Motorway Service Areas (MSAs) was significant for HS collisions.

The reduction of the percentage of VMT travelled by HGVs – which does not imply a reduction in the absolute number of VMT travelled by HGVs as well – decreases the collision frequency, showing that the composition of traffic is important in motorway safety. Increasing awareness of HGV drivers regarding the safety concerns that they should have in the environment they are working through training could be crucial as they appear to be a high risk group. For future study, it would be interesting to also investigate how the age of HGVs affects the likelihood of collisions, as well as finding more measures, apart from exceeding the speed limit, for drivers' compliance to the driving rules. Furthermore, the results suggested that the presence of rest areas (known as Motorway Service Areas) has a positive effect in the reduction of motorway collisions. These areas have had a significantly increased presence on British motorways in the last decades. Their further spread is recommended, especially in geographical areas where they are not as common. In addition, all drivers – of private vehicles and HGVs – should be encouraged to use them whenever they feel tiredness during their journey or need to perform any other non-emergency task.

A multitude of private vehicles and operatives use the HS regularly for emergency or other reasons (e.g. road maintenance) and they are constantly in danger when stopped in these positions. Collisions on HS, even though their frequency is low, have serious impacts when they occur. This research supports that when devising motorway safety strategy and systems it is important that collisions on the HS are examined separately from the MC in order to be able to provide specific solutions for preventing these events. Such solutions could be campaigns to promote the appropriate use of the hard–shoulder HS and the use of service areas on the motorway, as well as campaigns to increase road users' awareness of the inherent risks when on the hard–shoulder. Also, the development of a risk-based management tool for the motorway Operator would aid safer roadworker deployment.

Acknowledgements

This research was undertaken as part of an Engineering Doctorate project jointly funded by the Centre of Innovative and Collaborative Construction Engineering (CICE) at Loughborough University and Balfour Beatty. The support of the Engineering and Physical Sciences Research Council is gratefully acknowledged (EPSRC Grant EP/F037272/1). The authors would also like to thank the anonymous reviewers for their valuable comments.

References

Aarts, L., van Schagen, I., 2006. Driving speed and the risk of road crashes: a review. Accid. Anal. Prev. 38 (2), 215–224.
Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petrov, B.N., Csaki, F. (Eds.), Second International Symposium on Information Theory. Akailseoniai–Kiudo, Budapest, pp. 267–281.
Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control 19, 716–723.
Antoniou, C., Yannis, G. & Katsochis, D. (2013). Impact of meteorological factors on the number of injury accidents. In: Proceedings of the 13th WCTR, July 15–18, Rio de Janeiro, Brazil.
Beenstock, M., Gafni, D., 2000. Globalization in road safety: explaining the downward trend in road accident rated in a single country (Israel). Accid. Anal. Prev. 32 (1), 71–84.
Bergel-Hayat, R., Debbarh, M., Antoniou, C., Yannis, G., 2013. Explaining the road accident risk: weather effects. Accid. Anal. Prev. 60, 456–465.
Box, G.E.P., Jenkins, G.M., 1976. Time series analysis: forecasting and control, Rev. ed. Holden-Day, San Francisco.
Brijs, T., Karlis, D., Wets, G., 2008. Studying the effect of weather conditions on daily crash counts using a discrete time-series model. Accid. Anal. Prev. 40 (3), 1180–1190.
Brockwell, P.J., 2010. Time series analysis. In: Peterson, P., Baker, E., McGaw, B. (Eds.), International Encyclopedia of Education, Elsevier Science Limited, Oxford, UK, pp. 474–481.
Caliendo, C., Guida, M., Parisi, A., 2007. A crash prediction model for multilane roads. Accid. Anal. Prev. 39, 657–670.
Canova, F., 2007. Methods for Applied Macroeconomic Research. Princeton University Press, New Jersey, USA.
Cedar, A., Livneh, L., 1982. Relationship between road accidents and hourly traffic flow. Accid. Anal. Prev. 14, 19–44.
Charbotel, B., Chiron, M., Martin, J.L., Bergeret, A., 2002. Work-related road accidents in France. Eur. J. Epidemiol. 17, 773–778.
Constantinou, E., Panayiotou, G., Konstantinou, N., Loutsiou-Ladd, A., Kapardis, A., 2011. Risky and aggressive driving in young adults: personality matters. Accid. Anal. Prev. 43, 1323–1331.
Davis, G.A., Swenson, T., 2006. Collective responsibility for freeway rear-ending accidents? An application of probabilistic causal models. Accid. Anal. Prev. 38 (4), 728–736.
Department for Transport. (2014). Reported Road Casualties in Great Britain: Main Results 2013.
DfT, 2015 [Online]. Available from 〈https://www.gov.uk/government/organisations/department-for-transport/about/statistics〉.
Duchesne, P., Ghoudi, K. & Remillard, B. (2011). On Testing for Independence between Several Time Series. Available at SSRN: 〈http://ssrn.com/abstract=1823626〉 or http://dx.doi.org/10.2139/ssrn.1823626.
Edwards, J.B., 1994. Wind-related road accidents in England and Wales 1980–1990. J. Wind Eng. Ind. Aerodyn. 52, 293–303.
Elvik, R., Christensen, P., Amundsen, A., 2004. Speed and Road Accidents; An Evaluation of the Power Model. Institute of Transport Economics TØI, Oslo.
Evans, L., 1991. Traffic Safety and the Driver. Van Nostrand Reinhold, New York.
Goh, B.H., 2005. The dynamic effects of the Asian financial crisis on construction demand and tender price levels in Singapore. Build. Environ. 40, 267–276.
Golob, T.F., Recker, W.W., Alvarez, V.M., 2003. A tool to evaluate the safety effects of changes in freeway traffic flow, Working Papers, California Partners for Advanced Transit and Highways (PATH). Institute of Transportation Studies (UCB), UC Berkeley.
Golob, T.F., Recker, W.W., Alvarez, V.M., 2004. Freeway safety as a function of traffic flow. Accid. Anal. Prev. 36, 933–946.
Golob, T.F., Recker, W., Pavlis, Y., 2008. Probabilistic models of freeway safety performance using traffic flow data as predictors. Saf. Sci. 46 (9), 1306–1333.
Gomes, S.V., 2013. The influence of the infrastructure characteristics in urban road accidents occurrence. Accid. Anal. Prev. 60, 289–297.
Granger, C.W.J., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37 (3), 424–438.
Hannan, E.J., Quinn, B.G., 1979. The Determination of the order of an autoregression. J. R. Stat. Soc. Ser. B 41, 190–195.
Hassan, H.M., Abdel-Aty, M.A., 2013. Exploring the safety implications of young drivers' behavior, attitudes and perceptions. Accid. Anal. Prev. 50, 361–370.
Haugh, L.D., 1976. Checking the independence of two covariance-stationary time series: a univariate residual cross-correlation approach. J. Am. Stat. Assoc. 71 (354), 378–385.

ARTICLE IN PRESS

Haynes, R., Jones, A., Kennedy, V., Harvey, I., Jewell, T., 2007. District variations in road curvature in England and Wales and their association with road-traffic crashes. Environ. Plan. A 39 (5), 1222–1237.
Hermans, E., Brijs, T., Stiers, & T., Offermans, C. (2006). The impact of weather conditions on road safety investigated on an hourly basis. In: Proceedings of the 85th Annual meeting of the Transportation Research Board, January 22–26, Washington, D.C.
Highway Code (2007). Department for Transport, United Kingdom.
Holubowycz, O.T., Kloeden, C.N., McLean, A.J., 1994. Age, sex, and blood alcohol concentration of killed and injured drivers, riders, and passengers. Accid. Anal. Prev. 26 (4), 483–492.
Horne, J., Reyner, L., 2001. Sleep-related vehicle collisions: some guides for road safety policies. Transp. Res. Part F: Traffic Psychol. Behav. 4 (1), 63–74.
Houston, D.J., Richardson, L.E., 2002. Traffic safety and the switch to a primary seat belt law: the California experience. Accid. Anal. Prev. 34 (6), 743–751.
Johansen, S., 1991. Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. Econometrica 59 (6), 1551–1580.
Johansen, S., 1995. Likelihood-Based Inference in Cointegrated Vector Autoregressive Models. Oxford University Press, Oxford.
Jones, B., Janssen, L., Mannering, F., 1991. Analysis of the frequency and duration of freeway accidents in Seattle. Accid. Anal. Prev. 23 (4), 239–255.
Karlaftis, M.G. and Yannis, G. (2010). Weather effects on daily traffic accidents and fatalities: a time series count data approach. In: Proceedings of the 89th Annual Meeting of the Transportation Research Board, January 10–14. Washington, D.C.
Karlis, D. (2006). Time series model for count data. Paper presented at the Annual Conference of the Transportation Research Board, Washington, DC, January.
Koch, P.D., Yang, S.-S., 1986. a method for testing the independence of two time series that accounts for a potential pattern in the cross-correlation function. J. Am. Stat. Assoc. 81 (394), 533–544.
Kononov, J., Bailey, B., Allery, B., 2008. Development of relationships between safety and congestion for urban freeways. Transp. Res. Record: J. Transp. Res. Board 2083, 26–39.
Lopez-Lozano, J.M., Monnet, D.L., Yague, A., Burgos, A., Gonzalo, N., Campillos, P., Saez, M., 2000. Modelling and forecasting antimicrobial resistance and its dynamic relationship to antimicrobial use: a time series analysis. Int. J. Antimicrob. Agents 14, 21–31.
Lord, D., Washington, P., Ivan, J.N., 2005. Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. Accid. Anal. Prev. 37 (1), 35–46.
Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transp. Res. Part A 44, 291–305.
Lütkepohl, H., 2005. New Introduction to Multiple Time Series Analysis. Springer, Springer-Verlag, Berlin Heidelberg.
Mct Office, 2015 [Online]. Available from: 〈http://www.metoffice.gov.uk/climate〉.
Miaou, S.P., 1994. The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. Accid. Anal. Prev. 26 (4), 471–482.
Michalaki, P., Quddus, M.A., Pitfield, D.E., & Huetson, A. (2015). Exploring the factors affecting motorway accident severity in England using the generalised ordered logistic regression model. In: Proceedings of the 94th Annual meeting of the Transportation Research Board, January 11–15, 2015, Washington, D.C.
Motorway Services Online. (2015). Motorway Service Area. [Online]. Available from: 〈http://motorwayservicesonline.co.uk/Motorway_service_area〉, (accessed 15.09.15).
National Travel Survey, 2015 [Online]. 〈https://www.gov.uk/government/collections/national-travel-survey-statistics〉.
NHTSA (2008). Traffic Safety Facts: Young Drivers. National Center for Statistics and Analysis. Available at: 〈http://www-nrd.nhtsa.dot.gov/pubs/811169.pdf〉.
Noland, R., Oh, L., 2004. The effect of infrastructure and demographic change on traffic-related fatalities and crashes: a case study of Illinois county-level data. Accid. Anal. Prev. 36 (4), 525–532.
Noland, R.B., Quddus, M.A., Ochieng, W.Y., 2008. The impact of London congestion charge on traffic casualties: an intervention analysis. Transportation 35 (1), 73–91.
Nordbakke, S., Sagberg, F., 2007. Sleepy at the wheel: knowledge, symptoms and behaviour among car drivers. Transp. Res. Part F: Traffic Psychol. Behav. 10 (1), 1–10.
Pitfield, D.E., 2005. A time-series analysis of the pricing behaviour of directly competitive 'low-cost'airlines. Int. J. Transp. Econ. Vol. XXXII (No.1), 14–39.
Quddus, M.A., 2008. Time series count data models: an empirical application to traffic accidents. Accid. Anal. Prev. 40 (5), 1732–1741.
Quenouille, M.H., 1957. The Analysis of Multiple Time-Series. Griffin, London.
Schwarz, G., 1978. Estimating the dimension of a model. Ann. Stat. 6, 461–464.
Shankar, V., Mannering, F., Barfield, W., 1996. Statistical analysis of accident severity on rural freeways. Accid. Anal. Prev. 28 (3), 391–401.
Theofilatos, A., Yannis, G., 2014. A review of the effect of traffic and weather characteristics on road safety. Accid. Anal. Prev. 72, 244–256.
Wang, C., Quddus, M.A., Ison, S.G., 2009. Impact of traffic congestion on road safety: a spatial analysis of the M25 motorway in England. Accid. Anal. Prev. 41 (4), 798–808.
Watson, M.W., 1994. Vector Autoregression and Cointegration. In: Engle, R.F., McFadden, D.L. (Eds.), Handbook of Econometrics, Vol. IV. Elsevier Science B.V., Amsterdam.
Weesie, J., 1999. sg121: seemingly unrelated estimation and the cluster-adjusted sandwich estimator. Stata Technical Bull. 52, 34–47, Reprinted in Stata Technical Bulletic Reprints. 9, 231–248. College Station. TX: Stata Press..
World Health Organisation. (2013). Global Status Report on Road Safety 2013. Geneva, Switzerland.
Zebende, G.F., Filho, A.M., 2009. Cross-correlation between time series of vehicles and passengers. Physica A 388, 4863–4866.
Zebende, G.F., Da Silva, P.A., Filho, A.M., 2011. Study of cross-correlation in a self-affine time series of taxi accidents. Physica A 390, 1677–1683.
Zellner, A., Theil, H., 1962. Three-stage least squared: simultaneous estimation of simultaneous equations. Econometrica 30 (1), 54–78.
Zhang, T., Chan, A.H.S., 2014. Sleepiness and the risk of road accidents for professional drivers: a systematic review and meta-analysis of retrospective studies. Saf. Sci. 70, 180–188.
Zimring, F., 1975. Firearms and federal law: the Gun Control Act of 1968. J. Legal Stud. 4, 133–198.

Please cite this article as: Michalaki, P., et al., A time-series analysis of motorway collisions in England considering road infrastructure, socio-demographics, traffic and weather characteristics. Journal of Transport & Health (2016), http://dx.doi.org/10.1016/j.jth.2015.10.005