

## TRENDS IN ROAD ACCIDENTS ON BRAZIL'S HIGHWAYS: EVIDENCE OF PERSISTENCE

Peter Wanke, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

Luis A. Gil-Alana<sup>1</sup>, University of Navarra, Pamplona, Spain and Universidad Francisco de Vitoria, Madrid, Spain

Yong Tan\*, University of Huddersfield, UK

### Abstract

This paper deals with the analysis of trends in road accidents on major highways in Brazil. Using updated time series techniques, our results indicate that a low degree of long memory was detected in the series with shocks having transitory effects over time. We further find that the number of accidents taking place in Brazil has been reducing over time, though in the presence of negative shocks, the recovery is not going to be immediate due to the long memory nature of the data. Despite the absence of relevant investment relating to infrastructure expansion, it is worth mentioning the consolidation of a nationwide tolled road system in Brazil involving concessions to private administrators, alongside more severe traffic laws that can impose limitations on driving licences.

**JEL Classification:** C22; L91; O54

**Keywords:** Road accidents; Brazil; fractional integration; long memory

\*: Corresponding author:

Dr Yong Tan  
Department of Accounting, Finance and Economics  
Huddersfield Business School  
University of Huddersfield, Huddersfield  
UK, HD1 3DH

email: a.y.tan@hud.ac.uk

<sup>1</sup>Luis A. Gil-Alana gratefully acknowledges financial support from the Grant PID2020-113691RB-I00 funded by MCIN/AEI/ 10.13039/501100011033. An internal Project from the Universidad Francisco de Vitoria is also acknowledged.

## 1. Introduction

Road accidents are tragedies that lead to serious injuries and fatalities. Based on the report published by the World Health Organization (WHO) on the 21st June 2021, the number of people who died from road traffic crashes reaches around 1.3 million every year. Looking at the statistics more specifically among countries with different income levels, the report found that low and middle income countries in the world account for 93% of the world's fatalities caused by road accidents, although the total vehicles possessed by these countries only account for 60%. An ambiguous target has been set by the United Nations General Assembly to reduce the number of injuries and deaths caused by road accidents by half across the globe by the end of 2030.

The tremendous harm caused by road accidents is not only limited to the individual or the family level as reflected by the pain and loss derived from injuries or death, but its influence is extended further to a country level and in particular, it affects a country's economy. Based on the report from the WHO, road accidents entailed a cost 3% of gross domestic product for most countries in the world.

Research in this area has been attracting the attention of the academic researchers and studies have investigated this issue from different perspectives due to the fact that it can potentially bring benefits to individuals, families and entire countries by reducing the number of accident blackspots where injuries or death occur. Indeed, a WHO report shows that among other types of diseases, the road accident is the major cause of death of children and young adults aged 5-29 years.

Looking at the previous studies in the related topic area, researchers have focused on different issues including the modelling of the relationship between weather and road accidents (Amin et al., 2014); modelling road accident in the production process (Pal and Mitra, 2016); the connection between vehicle insurance and road accidents (Hsu et al.,

2015); the link between vehicle load capacity and road accidents (Castillo-Manzano et al., 2016); the effect of cell phones on vehicle accidents (Fowles et al., 2013); the influence of willingness-to-pay on road accidents (Haddak et al., 2016); among others.

It is not difficult to notice from the previous research studies that efforts have been made to investigate the road accident from the perspective of production economics by modelling the production process; also a great degree of concern has been focussed on investigating relevant factors which might reduce the volume of road accidents. Previously published studies contribute to both the empirical and theoretical perspectives, however, so far little attention has been made to look at the data in a more careful and thorough manner. More specifically, very few studies have looked at the pattern of road accidents by investigating persistence although this issue has been examined in other topic areas such as the persistence of urban air pollutant (Pan et al., 2017); persistence of inflation (Benigbo and Lopez-Salido, 2006); persistence of efficiency (Wang and Huang, 2007); among others. One exception is the paper by Gil-Alana et al. (2021). In that paper, the authors investigate road accidents in Spain using also a fractional integration approach. Their results indicate that the series examined display very low degrees of persistence, with the orders of integration being around 0 and thus showing a short memory pattern. The persistence of road accidents does not only affect the number of injuries and deaths in the long-term, but also hinders the management and control of road accidents. We contribute to the previous literature by providing an innovative model to evaluate the persistence of road accidents using a data sample from Brazil.

The current paper is structured as below: Section 2 presents the historical context, followed by Section 3 which outlines the innovative methodology adopted by the current study. Section 4 provides the data and the descriptive statistics. Section 5 presents and discusses the results. Finally, the concluding remarks are provided in Section 6.

## 2. Historical context

The World Health Organization (WHO) provides a number of indicators related to road crashes in most countries in the world, including Brazil. In particular, some of the data is related to the number of deaths in road traffic crashes provoked by alcohol-related factors. The data for Brazil shows that in 2016, 36.7% of male deaths in road accidents were attributed to alcohol related road traffic crashes, ranking 66<sup>th</sup> in the world, while for females, 23% of deaths in road traffic crashes were connected to alcohol related factors, ranking 74<sup>th</sup> in the world.

Not only does the WHO provide data regarding alcohol related deaths from road traffic crashes but relevant classifications have also been drawn up showing the distribution of road traffic deaths by type of road user. Based on the statistics for the year 2016, 23.2% of deaths involved drivers/passengers of 4-wheeled vehicles, 31.4% were from drivers/passengers of motorized 2- or 3- wheelers, deaths of pedestrians accounted for 18.1% with cyclists accounting for the lowest percentage, which was 3.4%. Based on the data, it is recommended that priority should be given by the government and regulatory authorities to instigate policies to enhance safety for motorized 2-4 wheeled vehicles. According to a report from the World Bank global road safety facility, the population of Brazil was 207,652,864 in 2016, while the number of reported road crash fatalities in that year was 38,651 and the number of estimated serious injuries was 615,105. Logically, this resulted in substantial costs for the country. In fact, the reported cost reached \$118,799 million, accounting for 6.6% of the country's GDP in 2016.

In the previous section, based on the report from the WHO, road accidents are the number one cause of death in the world for people aged 5-29 years old, while looking at the overall data for the whole population in Brazil, the following Table shows the top 10

causes of death in Brazil and their world ranking. As we can see 20.18% of deaths are related to road traffic accidents in Brazil ranking 77<sup>th</sup> in the world, following violence, lung disease and prostate cancer, the road accident related death is the fourth most serious social issue in Brazil.

**Table 1: Top 10 causes of death in Brazil, their rate and ranking**

	Causes	Rate	world ranking
1	Coronary Heart Disease	79.33	138
2	Stroke	51.38	122
3	Influenza and Pneumonia	42.68	76
4	Diabetes Mellitus	30.34	89
5	Violence	29.95	<b>10</b>
6	Lung Disease	28.67	61
7	Prostate Cancer	20.52	63
8	Road Traffic Accidents	20.18	77
9	Breast Cancer	15.78	97
10	Lung Cancers	13.93	87

**Table 2. Deaths per age group**

Age Group	Deaths by traffic accident		Deaths by other causes	
	Number	%	Number	%
Less than 1	95	0.31	61846	6.44
1 to 9	1450	4.75	16096	1.68
10 to 19	3602	11.81	25691	2.67
20 to 29	7332	24.04	52104	5.42
30 to 39	6002	19.68	61901	6.44
40 to 49	4619	15.14	86987	9.06
50 to 59	3071	10.07	113311	11.80
60 to 69	2110	6.92	154776	16.11
70 and over	1934	6.34	383608	39.93
Ignored	286	0.94	4294	0.45
Total	30501	100	960614	100

Source: Ministry of Health

Finally, we can look at the death from traffic accidents by age group. The Ministry of Health of Brazil provided the statistics, which are shown in Table 2. As we can see from the table, the number of deaths from road accidents is mainly focused on three age groups, 20-29, 30-39 and 40-49 with the largest number of deaths corresponding to the first group. In comparison, the next three age groups are not particularly influenced by traffic accidents, these are 70 and over, 60-69 as well as 50-59, and they have the largest number of death derived from non-traffic accident causes. Based on the data, we suggest that extra attention and effort should be made by the government and regulatory authorities to bring in relevant policies aimed at the 20-50 age group .

### 3. Methodology

Dealing with persistent time series data, the most standard approach is to consider an autoregressive of order 1, i.e., AR(1) process of form

$$(1 - \alpha L) x_t = u_t, \quad t = 1, 2, \dots, \quad (1)$$

where  $L$  is the lag-operator ( $L^k = x_{t-k}$ ) and where  $u_t$  is an integrated of order 0 (I(0) or short memory) process. By this we mean a covariance stationary process where the infinite sum of the autocovariances is finite. In this context, the parameter  $\alpha$  measures the degree of persistence, and if  $\alpha = 1$ ,  $x_t$  is said to contain a unit root and be nonstationary, with shocks having permanent effects on the evolution of the series. The model can be generalized to an AR(p) process of the form:

$$\phi_p(L) x_t = u_t, \quad t = 1, 2, \dots, \quad (2)$$

where  $\phi_p(L) = (1 - \phi_1 L - \dots - \phi_p L^p)$ , and persistence is measured then as the sum of the AR coefficients, i.e.,  $\phi = (1 + \phi_1 + \dots + \phi_p)$ .

A problem with these autoregressive approaches is that they produce an abrupt change in the behaviour of the series as  $\alpha$  (or  $\phi$ ) approaches unity, moving from the stationary to the nonstationary case. This is solved by using alternatives based on fractional integration, which use the following equation,

$$(1 - L)^d x_t = u_t, \quad t = 1, 2, \dots, \quad (3)$$

and where  $d$  can be any real value. Note, that for any real value of  $d$ , the polynomial in  $L$  in the right hand side of equation (3) can be expanded as:

$$(1 - L)^d = \sum_{j=0}^{\infty} \binom{d}{j} (-1)^j L^j = 1 - dL + \frac{d(d-1)}{2} L^2 \quad (4)$$

and thus,  $x_t$  in (3) depends on all its past history if  $d$  is a fractional value. Moreover, and based on the above equation, the higher the value of  $d$  is, the higher the dependence between the observations is, and thus,  $d$  can be considered as an indicator of the degree of persistence in the data. Note that in this context, the unit root nonstationary model corresponds to  $d = 1$  in (3)

In addition, and based on standard approaches in the unit root literature (Barghava, 1986; Schmidt and Phillips, 1992; etc.) we permit  $x_t$  to be the errors in a regression model that incorporates a linear time trend of the form:

$$y_t = \alpha + \beta t + x_t, \quad t = 1, 2, \dots, \quad (5)$$

where  $y_t$  are the original time series data.

We estimate the order of integration of the series,  $d$ , using a frequency domain version of the Whittle function (see, Dahlhaus, 1989) by using a testing procedure developed in Robinson (1994) that is particularly appropriate compared with other methods. Thus, for example, it behaves well in finite samples (Gil-Alana and Robinson, 1997) which is important in the present work given the limited number of observations

used in this application. Moreover, the method remains valid even in nonstationary contexts (i.e.,  $d \geq 0.5$ ) unlike most estimating/testing procedures where stationarity is a pre-requisite; thus, using the method of Robinson (1994) we do need the requirement of preliminary differencing in the case of nonstationary series. Finally, it has a standard null limit distribution (unlike most unit roots and fractional methods), and it is the most efficient method in the Pitman sense against local departures from the null.

#### **4. Data and descriptive statistics**

The Brazilian road accident data were obtained from the ANTT Dados Abertos website<sup>1</sup>, available on a daily basis from 2010 to 2021. Since the original downloaded dataset indicates, per road, every km that had an accident in a given day, and we had to aggregate this information to obtain how many accidents occurred per day on each road, as displayed in Table 3, in major tolled Brazilian roads, namely: (i) Fernão Dias (links the cities of São Paulo and Belo Horizonte); (ii) Autopista Fluminense (links the cities of Rio de Janeiro and Campos, the major oil offshore production pole in Brazil); (iii) Litoral Sul (links major cities in the Southern parts of Brazil, such as Curitiba and Florianópolis); (iv) Planalto Sul (also links major cities in the Southern parts of Brazil, such as Curitiba and Florianópolis, but stretches down to the borders with Argentina in Rio Grande do Sul state); and (v) Regis Bittencourt, also known as the “road of death”, which links the cities of São Paulo and Curitiba.

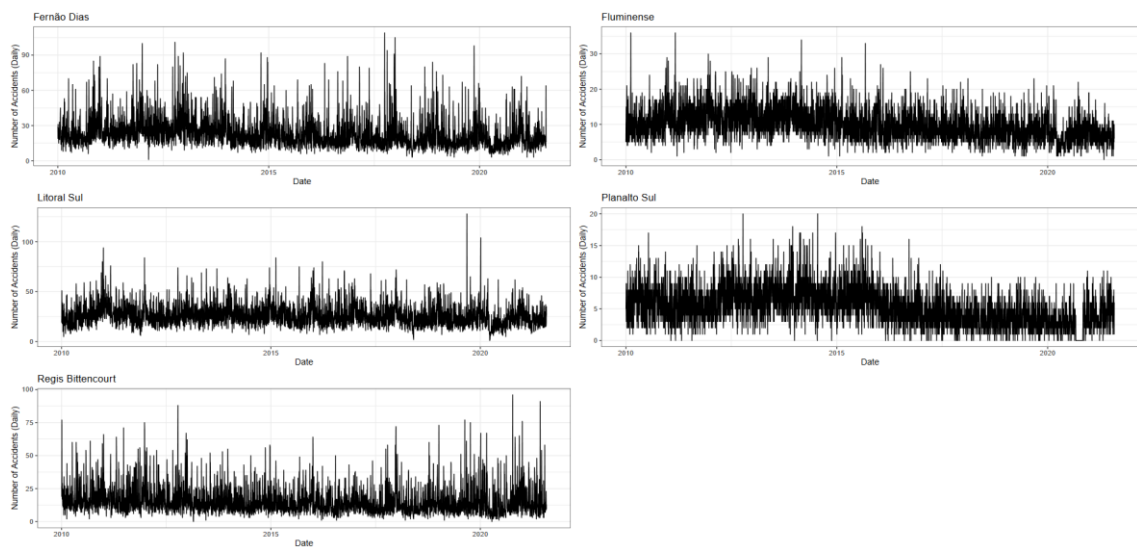
---

<sup>1</sup> <https://dados.antt.gov.br/dataset/acidentes-quilometro-rodovias>

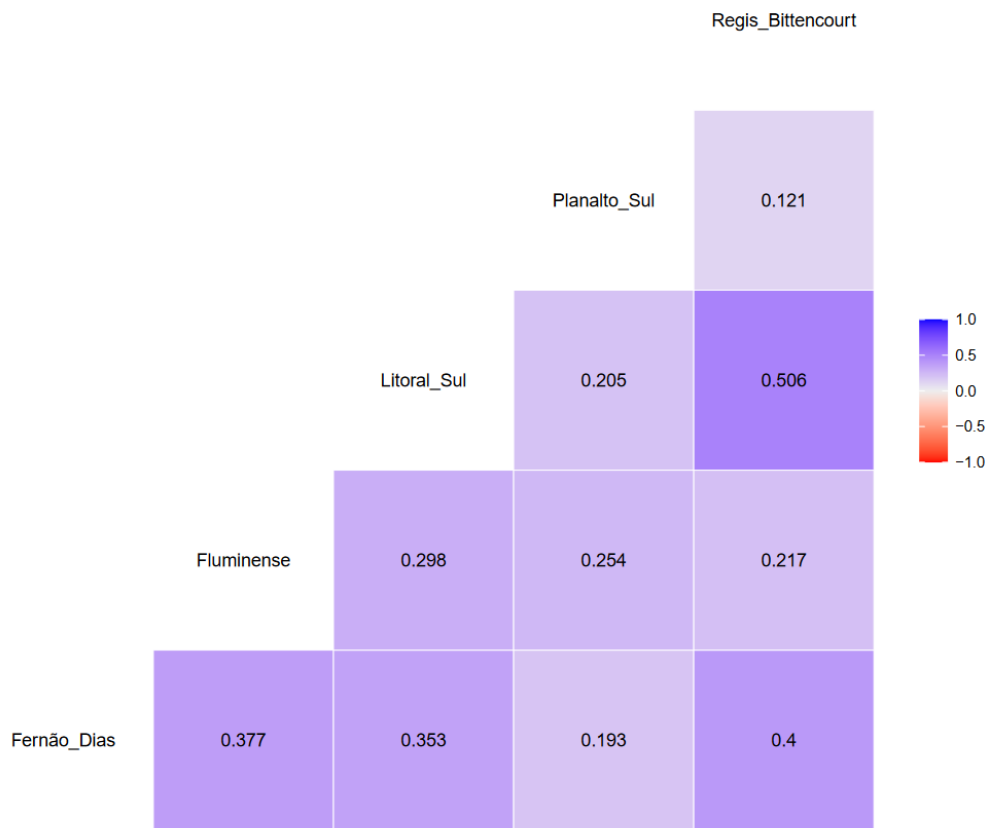


**Table 3: Descriptive Statistics for the number of accidents per day in major Brazilian Roads**

Descriptive Statistics	Fernão Dias	Fluminense	Litoral Sul	Planalto Sul	Regis Bittencourt
Min	1	0	1	0	0
Median	20	9	25	5	13
Max	109	36	128	20	96
Mean	23.38	9.97	26.32	5.22	15.63
SD	12.67	4.52	10.78	3.00	10.29
CV	0.54	0.45	0.41	0.57	0.66
Skewness	1.92	0.84	1.43	0.72	2.12
Kurtosis	5.63	1.36	4.82	0.78	7.00
Information Entropy	0.43	0.34	0.41	0.29	0.41



**Figure 1: – Number of accidents per road (daily)**



**Figure 2: Correlogram of roads accidents**

Figure 2 displays the time series plots and suggests a strong seasonal pattern of daily accidents in major Brazilian roads, mostly driven by effects such as weekdays and holidays, no significant increasing or decreasing trend can be captured by a mere glimpse. Besides, daily accidents between these roads seem to be uncorrelated (cf. Fig. 2), confirming these roads are independent logistics systems. We can see from Table 1 that there is a large difference in terms of the maximum number of accidents that took place on the roads with very tiny difference noticed in terms of the minimum number of accidents, this also results in a large difference in the standard deviation. In summary, the statistics show that the maximum daily number of accidents took place in Litoral Sul, Fernão Dias and Regis Bittencourt, these three roads experiencing the largest standard deviation. With regard to skewness, we notice that Fernão Dias, Litoral Sul, and Regis Bittencourt were highly skewed in the distributions, with the distributions of the remaining two roads being moderately skewed. The distributions of three roads, namely Fernão Dias, Litoral Sul and Regis Bittencourt are leptokurtic, while the distributions of the other two roads were found to be platykurtic. Finally, it was found that all the roads in the sample exhibit a degree of disorder in the data, although Planalto Sul performed the best.

## 5. Empirical results

We consider the model given by the equations (3) and (5), i.e.,

$$y_t = \alpha + \beta t + x_t, \quad (1 - B)^d x_t = u_t, \quad t = 1, 2, \dots, \quad (6)$$

where  $y_t$  is the time series we observe,  $\alpha$  and  $\beta$  are unknown coefficients referring respectively to a constant and a time trend, and  $x_t$  is supposed to be  $I(d)$  where the order of integration  $d$  is an additional parameter to be estimated from the data; finally,  $u_t$  is an error term that is supposed to be uncorrelated with zero mean and constant variance.

**Table 4: Estimates of the differencing parameter, d, and 95% confidence intervals**

Series	No terms	With a constant	With a constant and a linear time
FERNÃO.DIAS	0.17 (0.15, 0.19)	0.11 (0.09, 0.13)	<b>0.06 (0.04, 0.08)</b>
FLUMINENSE	0.21 (0.19, 0.22)	0.14 (0.13, 0.16)	<b>0.05 (0.04, 0.07)</b>
LITORAL.SUL	0.15 (0.12, 0.17)	0.07 (0.05, 0.08)	<b>0.03 (0.01, 0.05)</b>
PLANALTO.SUL	0.22 (0.20, 0.23)	0.17 (0.16, 0.19)	<b>0.12 (0.11, 0.14)</b>
REGIS.BITTENCOURT	0.11 (0.09, 0.13)	0.06 (0.04, 0.08)	<b>0.04 (0.02, 0.06)</b>

The values in parenthesis are the 95% confidence interval for the values of d. In bold, the selected model in relation with the deterministic terms.

**Table 5: Estimated coefficients in the selected models**

Series	d	Intercept (tvalue)	Time trend (tvalue)
FERNÃO.DIAS	0.06 (0.04, 0.08)	28.624 (50.28)	-0.00259 (-10.90)
FLUMINENSE	0.05 (0.04, 0.07)	12.887 (71.35)	-0.00139 (-19.11)
LITORAL.SUL	0.03 (0.01, 0.05)	29.069 (72.15)	-0.00131 (-8.04)
PLANALTO.SUL	0.12 (0.11, 0.14)	7.173 (38.83)	-0.00094 (-12.87)
REGIS.BITTENCOURT	0.04 (0.02, 0.06)	17.876 (43.26)	-0.00104 (6.27)

Table 4 reports the estimated values of d in (6) under the three classical assumptions in the unit roots literature: i) no deterministic terms, i.e., imposing that  $\alpha = \beta = 0$  a priori in (6); ii) with an intercept (only  $\beta = 0$  a priori), and iii) with an intercept and a linear time trend. If these two coefficients are significant we keep this model, otherwise we move to the model with an intercept, and if this is also insignificant, to the case with no terms. We have marked in bold in the table the significant case for each series, and we notice that the time trend is statistically significant in all five series. Focussing on the differencing parameter, we observe that for this case, the estimates of d are all significantly positive though very close to 0. Thus, though long memory is detected ( $d > 0$ ), the values are very small, with shocks disappearing extremely fast. Table 5 focuses on the estimated coefficients of these selected models, and we observe that the time trend is negative in all cases, ranging their values from -0.00094 (Planatol\_Sul) to -0.00259 (Fernaao\_Dias). These negative trends indicate that the number of accidents is

reducing over time; however, the fact that  $d$  is positive indicates the presence of long memory, so negative shocks in the series (for instance, an increasing number of accidents) disappear relatively slowly, requiring therefore some action to return to their original long term projections.

## 6. Conclusion

In this article we have examined the number of road accidents in the main roads in Brazil by using time series techniques based on fractional integration. According to our results, the five series examined display significant negative time trends, implying a reduction in the number of accidents over time; however, at the same time the series display long memory, a feature that indicates that shocks tend to be persistent and have long-lived effects.

**Table 6: Time trend coefficients under  $I(0)$  and  $I(d)$  errors**

Series	$d = 0$	$d$ estimated
FERNÃO.DIAS	-0.00265	-0.00259
FLUMINENSE	-0.00141	-0.00139
LITORAL.SUL	-0.00132	-0.00131
PLANALTO.SUL	-0.00099	-0.00094
REGIS.BITTENCOURT	-0.00105	-0.00104

Comparing the coefficients of the time trends between the  $I(0)$  specification for the error term and the  $I(d)$  approach, we notice, in Table 6 that the size of the coefficients for all the roads are smaller under the long memory approach implying that the reduction in the number of accidents is smaller than the number provided by the short memory or  $I(0)$  approach. Nevertheless, the differences are small. The differences between short and long memory approaches can be explained by the current status of infrastructure expansion investments, the granting of current road infrastructure to private concessionaires, and more rigid legal apparatus and the introduction of more severe traffic laws to punish

negligent drivers. While road infrastructure in Brazil has remained stagnant over the last four decades, conjunctural improvements emerged in terms of a nationwide toll system capable of maintaining the current road infrastructure minimally operational and safe. On the other hand, civil society pressed congressmen and authorities to approve a traffic law similar to the one ratified in EU countries, where negligent drivers can lose their driving licence after a number of incidents. Most of this pressure emerged as a consequence of the higher fatality rates in Brazilian roads, even when compared to other countries and diseases.

This article can be extended in several directions. For example, as a robustness method, the long memory methodology used in this work can be extended to other parametric or even non-parametric methods. In addition, the presence of structural breaks can also be taken into account and a comparison of these results with other smaller roads or even other countries are issues which are currently in progress.

## References

- Amin, S. R., Zareie, A. & Amador-Jimenez, L. E. (2014). Climate change modeling and the weather-related accidents in Canada. *Transportation Research Part D: Transport and Environment*, 32, 171-183.
- Benigno, P. & Lopez-Salido, J. D. (2006). Inflation persistence and optimal monetary policy in the euro área. *Journal of Money, Credit and Banking*, 38, 587-614.
- Bhargava, A. (1986). On the theory of testing for unit roots in observed time series. *Review of Economic Studies*, 53, 369-384.
- Castillo-Manzano, J. I., Castro-Nuno, M. & Fageda, X. (2016). Exploring the relationship between truck load capacity and traffic accidents in the European Union. *Transportation Research Part E: Logistics and Transportation Review*, 88, 94-109.
- Dahlhaus, R. (1989) Efficient Parameter Estimation for Self-Similar Process. *Annals of Statistics*, 17, 1749-1766.
- Fowles, R., Loeb, P. D. & Clarke, W. (2013). The cell phone effect on truck accidents: a specification error approach. *Transportation Research Part E: Logistics and Transportation Review*, 50, 18-28.
- Gil-Alana, L.A., Duarte B. & Valbuena, C. (2021), Road accidents in Spain. Are they persistent?, IATSS Research, forthcoming,
- Gil-Alana, L.A. & Robinson, P.M. (1997), Testing of unit roots and other nonstationary hypotheses in macroeconomic time series, *Journal of Econometrics*, 80, 241-268.
- Haddak, M. M., Lefevre, M. & Havet, N. (2016). Willingness-to-pay for road safety improvement. *Transportation Research Part A: policy and Practice*, 87, 1-10.
- Hsu, Y. C., Shiu, Y. M., Chou, P. L. & Chen, Y. M. J. (2015). Vehicle insurance and the risk of road traffic accidents. *Transportation Research Part A: Policy and Practice*, 74, 201-209.
- Pal, D. & Mitra, S. K. (2016). An application of the directional distance function with the number of accidents as an undesirable output to measure the technical efficiency of state road transport in India. *Transportation Research Part A: Policy and Practice*, 93, 1-12.
- Pan, W., Xue, Y., He, H. D. & Lu, W. Z. (2017). Traffic control oriented impact on the persistence of urban air pollutants: a causeway bay revelation during emergency period. *Transportation Research Part D: Transport and Environment*, 51, 304-313.
- Robinson, P.M. (1994) Efficient tests of nonstationary hypotheses, *Journal of the American Statistical Association*, 89, 1420-1437.
- Schmidt, P. & Phillips, P.C.B. (1992). LM tests for a unit root in the presence of deterministic trends, *Oxford Bulletin of Economics and Statistics*, 54, 257-287.

Wang, M. H. & Huang, T. H. (2007). A study on the persistence of Farrell's efficiency measure under a Dynamic framework. *European Journal of Operational Research*, 180, 1302-1316.