Economic indicators for automobile claim frequencies*

Indicadores económicos para las frecuencias de siniestralidad de automóviles

Eva Boj**
Anna Castañer**
M. Mercè Claramunt**
Teresa Costa**
Oriol Roch**

Abstract

This article examines the relationship between observed claim frequencies in the automobile insurance line and the evolution of selected economic magnitudes. From a variety of economic variables, we aim to identify the main factors affecting claim frequencies, while controlling for other legislative and demographic factors. Through a dynamic regression model, the analysis is conducted for three different categories of vehicles and for a variety of coverages. A comprehensive dataset from the main Spanish insurance companies is used to calibrate the model. The evidence might assist companies to improve ratemaking.

Key words: *Categories of vehicles, dynamic regression, external predictors, motor insurance, time series.*

JEL Classification: *C10, C53, E32.*

Resumen

Este artículo examina la relación entre las frecuencias de siniestralidad observadas en la línea de seguros de automóviles y la evolución de magnitudes económicas seleccionadas. A partir de una variedad de variables económicas,

nuestro objetivo es identificar los principales factores que afectan a las frecuencias de siniestralidad, controlando al mismo tiempo otros factores legislativos y demográficos. A través de un modelo de regresión dinámica, el análisis se realiza para tres categorías diferentes de vehículos y para distintas coberturas. Se utiliza una base de datos de las principales compañías de seguros españolas para calibrar el modelo. La evidencia podría ayudar a las empresas a mejorar su tarificación.

Palabras clave: *Categorías de vehículos, regresión dinámica, predictores externos, seguros de automóviles, series temporales.*

Clasificación JEL: *C10, C53, E32.*

1. Introduction

An adequate understanding of the impact of the economic cycle in the financial results of the company is one of the big challenges of the insurance sector. The inherent variability of the economic situation affects the fundamental factors that determine the equilibrium between profits and losses. So, an in-depth analysis of the performance of the portfolio claims with respect to the evolution of economic variables is an important step to understand and ultimately be able to modify the underwriting cycle of the insurance sector.

A number of studies have examined the underwriting cycle, e.g., the succession of high and low profitability periods in the insurance sector. Most of the research focuses attention on insurance pricing as the key element of discussion. The two factors involved in the determination of the pure premium of the insurance, the claim frequency and the cost per claim, evolve along the economic situation. When pricing insurance policies, the technical department of the insurance company must take into account all the relevant information regarding expected losses of the claim. The divergence between the realized factors and those implicit in the pricing determines the profit or loss of the policy. For example, Cummins and Outreville (1987) describe a counter-cyclical component to the profit cycle. The lag in compiling and incorporating the necessary information into new rates would produce a counter-cyclical effect on the results. This is explained by the fact that, in expansive periods, premiums would be set with costs related to recession periods. So, profits would decrease. As data becomes available and costs are adjusted, premiums increase with profits.

A detailed analysis of the impact of the economic cycle into the business model can be traced back to Tarbell (1932), where the influence of the economic condition in the insurance business is studied for different lines. In the case of automobile lines, and according to the author, in recession periods we observe a decrease in income that leads to a reduction of traveled distance. However, for the same reason, there exists an incentive for drivers of reduced income to profit from trivial accidents, also discouraging to keep motor vehicles in proper

conditions and repair. Also, higher unemployment rates would encourage burglaries, thefts and robberies, so an increase of claim frequencies should be expected.

The direct quantitative relationship between the underwriting cycle and the economic cycle has been less studied in the literature. Early studies, such as Fairley (1979), find negative correlations between loss costs in the insurance sector and stock returns, as pointed out by Feldblum (2001). These results would suggest a link between underwriting and economic cycle. However, the relationship is not conclusive, as shown by D'Arcy and Garven (1990) or Kozik (1994).

Another research line closer to our objectives is based on the use of dynamic macro models to study the relationship between the number of road accidents and potential explanatory factors (Hakim *et al.*, 1991). Applying this methodology, it is possible to assess the impact of economic, demographic or legislative factors in the frequency of road accidents from the point of view of road safety. For example, fuel price has been identified as a major influencing factor in the number of traffic crashes, as studied in Grabowsky and Morrisey (2006), Chi *et al.* (2010) or Chi *et al.* (2015). The reason is that, in response to gasoline price increases, we observe a less frequent, more conservative and more fuel-efficient driving behavior, therefore improving traffic safety. Also, higher employment rates and poverty levels have been associated with road casualties (Noland and Quddus, 2004) and crashes (Aguero-Valverde and Jovanis, 2006).

More recently, some authors have studied the long-term effects of economic growth in road safety. For example, the results in Elvik (2010) and Nghiem *et al.* (2013) suggest that traffic accidents initially increase with economic development before declining after reaching certain thresholds of per capita income. According to the results, it is expected that technological innovation in vehicle manufacturing and road construction leads to a reduction of claim frequency rates. These factors would outweigh other countervailing factors related to economic growth such as the increase in travel frequency.

In our study we aim to extend the literature identifying, from a wide list of economic factors, those with potential to affect vehicle claim frequency within an insurance context. From the insurance company point of view, the possibility to model the relationship between the economic cycle and the claim frequency would become a key element to adjust the subscription cycle to the economic cycle. A precise rate making system would allow insurers to compute adequate provisions for contingencies.

The analysis is performed for three different categories of vehicles, according to their common characteristics. The first category corresponds to passenger cars and vans. In the second category, we include bigger commercial vehicles such as trucks, buses, agricultural machinery, trailers and semi-trailers. Finally, we consider a third category of vehicles that comprises motorcycles and scooters.

As a novelty, for each of the three categories of vehicles we present data segregated by different coverages: general liability, which includes liabilities for corporal damage and material damage, own damage (without cost for the insured party per accident), theft and legal defense. Although the claim frequencies of

the different coverages are expected to follow similar trends, we aim to find particular behaviors that have not been identified in previous literature.

To identify the main factors affecting claim frequency data we employ a methodology based on dynamic regression models with ARIMA errors, taking into account calendar effects and outlier detection. The methodology combines elements from regression models with elements from time-series analysis.

Our study suggests that it is possible to identify several economic factors related to the claim frequency for different categories of automobiles and for different types of coverages. This information may be useful in order to determine the pricing policy in insurance companies taking into account elements of future claim frequencies, and therefore, serving as a basis to produce more accurate pricing models.

The rest of the paper proceeds as follows. In Section 2 we discuss the data. Section 3 presents the regression model used to identify the main factors explaining the claim frequency for different types of categories and coverages. Results from the estimation are reported in Section 4. Finally, Section 5 concludes.

## 2. DATA AND REGULATION

2.1. Claim frequency

In this paper we consider claim frequency data as defined by the ratio of reported claims to total exposure. By reported claims we mean reported claims during the corresponding period, independent from the moment of its occurrence, that are qualified for coverage and imply compensation from the insurance company. With respect to the total exposure, policy insurances are accounted considering the fraction of the corresponding period that provide coverage.

The number of reported claims and the risk exposures for this study have been provided by an association of the major Spanish insurance companies. From this data, the claim frequency series have been constructed. At the origin of the sample period, the collection comprised data from 55 companies, which represent the vast majority of the insurance sector in Spain. During the period under consideration, a significant number of mergers and consolidations have been taken place, so that the final number of companies has been reduced to 47.

We consider quarterly data for the period 2005-2015. The first datum corresponds to the second quarter of 2005, whereas the last datum corresponds to the fourth quarter of 2015. In total, we use 43 data points in the series. No missing observations arose during the compilation of the data. The considered period comprises the last years of the economic expansion cycle, a subsequent severe long phase of economic crisis and the first years of recovery.

As stated in the introduction, we regroup the data in three different categories of vehicles. Category I corresponds to passenger cars and vans. Commercial vehicles such as trucks, buses, agricultural machinery, trailers and semi-trailers are included in Category II. Finally, Category III comprises motorcycles and scooters.
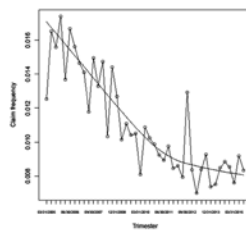
FIGURE 1

CLAIM FREQUENCY SERIES FOR DIFFERENT CATEGORIES AND COVERAGES
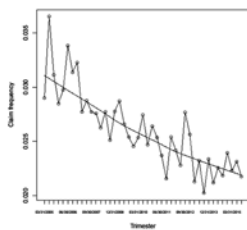
(a) Gen. liability. Category I.
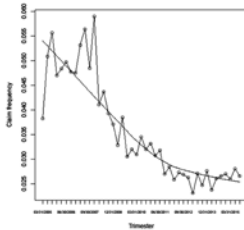
(b) Gen. liability. Category II.

(c) Gen. liability. Category III.
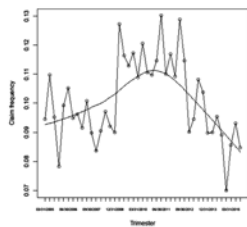
(d) Own damage. Category I.
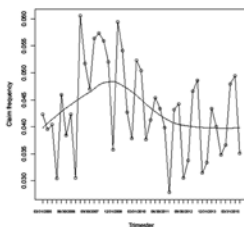
(e) Own damage. Category II.
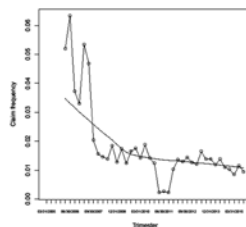
(f) Own damage. Category III.
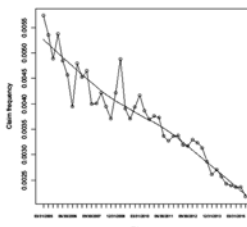
(g) Theft. Category I.
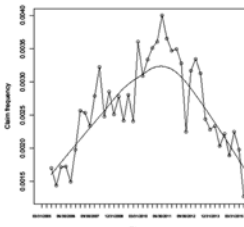
(h) Theft. Category II.
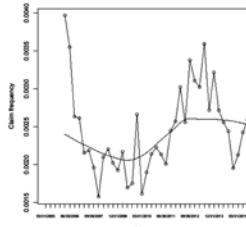
(i) Theft. Category III.
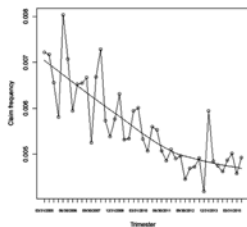
(j) Legal Defense. Category I.
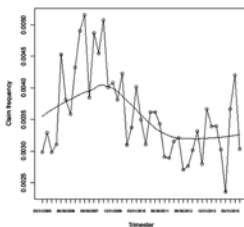
(k) Legal Defense. Category II.
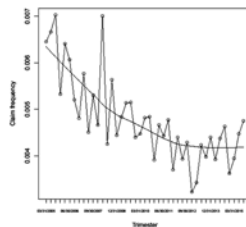
(l) Legal Defense. Category III.

For each of the three categories, we consider different coverages. The first is general liability, which includes liabilities for body injury and property damage to third parties and their property, but not to the driver or the driver's property. The second one is own damage, which offers protection against damage to the insured vehicle incurred during and accident caused by the policyholder.

Damages caused by theft of the vehicle are excluded. Theft is the third coverage included in the analysis. Through this coverage, the insurer will indemnify the insured for the total or partial removal of the vehicle. Also, damages caused during the time that, as a consequence of the theft, the automobile was in the hands of outsiders, as well as the damages caused by committing the theft or attempt of it, are included in the policy. Legal defense is the last coverage. It covers the legal assistance and expenses generated both in processes initiated to claim compensations or damages and in those that have the purpose of defending the policyholder, always within the limits established in the policy. It should be noted that expenses derived from the suspension of the driving license are excluded from coverage.

After the data cleaning process described in Section 3, Figure 1 depicts the evolution of claim frequencies for the considered period. We observe a clear decreasing trend on the general liability coverage since the beginning of the analyzed period. Although there is some variability associated, the claim frequencies quickly decrease, especially during the first years. In the later years, however, although we still observe a clear decreasing trend, the reduction is less pronounced. For the rest of coverages, we find different patterns depending on the category of vehicles. Own damage coverage shows a bump in claim frequency ratios around the central years of the sample period in Categories I and II. Category III, instead, shows a decreasing trend similar to the general liability case.

The theft coverage also performs differently across categories. The claim frequency in Category I halves during the sample period. Categories II and III, however, experience increases along the years, although in the last period the claim frequencies return to the level of initial registers. Legal defense series also differ across coverages, with a pronounced decrease in Categories I and III and an oscillation in Category II.

Table 1 provides some statistics of the claim frequencies. It can be clearly observed that the coverage with the smallest claim frequencies is theft. Not only claims related to stolen vehicles are less common: its variability is also smaller than in other coverages, and therefore, smaller differences are generally detected across quarters. Also legal defense presents small claim frequencies and volatilities, at least during the sample period. In the opposite situation we find own damage, particularly Category I, that presents a much higher claim frequency rate and volatility. It suggests a more difficult task of forecasting future values of this variable for insurers. In the medium term we find general liability, with more stable claim frequencies over time compared to own damage.

2.2. Explanatory variables

This study makes use of data relative to the global activity of the Spanish economy drawn from a variety of public sources: National Statistics Office, Ministry of Economy and Finance, Ministry of Public Works and Bank of Spain. The criteria for variable selection is to choose those variables, published

TABLE 1
SUMMARY OF STATISTICS OF THE CLAIM FREQUENCY SERIES

		Mean	Standard deviation	Minimum	Maximum
General Liability	Category I	0.02550	0.00326	0.02022	0.03649
	Category II	0.03402	0.00973	0.02305	0.05888
	Category III	0.01062	0.00276	0.00699	0.01738
Own Damage	Category I	0.10314	0.01422	0.06994	0.13012
	Category II	0.04276	0.00836	0.02779	0.06054
	Category III	0.01038	0.00990	0.00224	0.06323
Theft	Category I	0.00352	0.00082	0.00217	0.00573
	Category II	0.00265	0.00068	0.00127	0.00399
	Category III	0.00245	0.00051	0.00158	0.00396
Legal Defense	Category I	0.00548	0.00085	0.00418	0.00803
	Category II	0.00356	0.00067	0.00235	0.00515
	Category III	0.00469	0.00085	0.00322	0.00702

on a monthly or quarterly basis, that may well represent the state of the economy.

In some cases, the variables may be leading indicators of the economy. That is usually the case of the stock index, which tends to anticipate the state of the economy. In our case, since we employ Spanish data for the claim frequency, we select the corresponding stock index, the IBEX-35 index (IBEX35). Also, we include a broader leading indicator, the composite leading indicator (CLI) from the OECD. The CLI is constructed from a series of short-term economic components that tend to precede changes in the total economy, such as the rate of capacity utilization in the manufacturing sector or the production of total construction. Although the OECD computes the CLI for a wide range of countries and economies, in our analysis we just use the CLI for Spain.

General economy activity is represented through the variables real GDP per capita in euros (GDP), Industrial Production Index (IPI), unemployment rate (UPR), volume of total retail trade sales (RET) and Consumer Price Index (CPI), whereas the structure of the economy is taken into account through the variables number of worked hours in construction sector relative to the total worked hours (CONS) and the number of worked hours in service sector relative to the total worked hours (SERV). The last two variables are also stated in index terms.

In order to cope with the financial factors affecting the economy, we use the one-year Euribor (EURIBOR) and the return on invested capital minus cost on capital (CAP) excluding financial sector.

Variables related to construction and real estate market are also incorporated. We select the index of construction cost (CONSCOST), the housing price index (HOUSE), the number of mortgages relative to the total housing properties (MORT) and the total number of real estate transactions (REAL, in thousands), with this purpose.

The study also controls for other factors that may affect traffic accidents, such as number of registered vehicles in absolute value (TOTNVEH, in thousands) and per capita (NVEH), and fuel price (FUELPRICE, in index terms). Demographic factors such as population (POP, in thousands), population above 65 years (OLDPOP, in thousands) and proportion of population above 65 years (AGEPOP) are included. Also, electricity consumption in absolute value (ELEC, in ktep) and per person (PERELEC, in ktep per thousand) are considered.

Finally, we take into account four calendar effects: trading day effect (TRAD), holidays effect (HOL), Easter effect (EAST) and leap-year effect. Calendar effects will be handled through the TRAMO-SEATS application, as will be described in the next section.

Table 2 provides a summary of statistics of the main explanatory variables used in the empirical analysis. Initially, we consider all the factors described above. However, as will be explained in Section 3, for every category and coverage the number of factors will be reduced. Those factors that may be redundant or cause reverse causality will be dropped from consideration.

2.3. Regulatory factors

It is a common approach (e.g., Castillo-Manzano *et al*., 2011 or Dadashova *et al*., 2014) to include the effect of legislative actions that might affect the behavior of drivers, and therefore, reduce the number of accidents. In terms of our problem, the effect of the legislative modifications is expected to reduce the claim frequency in those coverages that are related.

For our particular set of data, we take into account major modifications of the Spanish regulation related to accidents. In the period under consideration, two major laws have been enacted to induce a reduction of the number of accidents.

Through Law 17/2005, to apply from July 2006, the Spanish government established a penalty point system with the aim to raise awareness and penalize those drivers committing severe traffic offenses. Departing from a total of twelve points, offenders are penalized with a reduction of points. The severity of the traffic infringement determines the number of lost points. The regulation also allows to credit drivers with a null record of traffic offenses with extra points, as well as the re-education of those drivers sanctioned with a partial or total loss of the assigned points.

The second reform we will take into account is the penal code reform enacted trough Law 15/2007. According to the reform, certain violations of speed limits shall be punished by imprisonment for a term of three to five months, jointly with a license disqualification from one to four years. Also, driving under the influence of alcohol, illicit drugs or related substances shall be punishable in the same terms. The Law entered into force in December 2007.

Both reforms enter the model in the form of dummy variables (DRIVLIC and CRIMCODE, respectively). In particular, the penalty point system variable takes a value of 0 until the second quarter of 2006 and a value of 1 afterwards.

TABLE 2
DESCRIPTIVE STATISTICS OF EXPLANATORY VARIABLES

Name	Units	Mean	Std Dev	Min	Max
AGEPOP	Per mille	171.89	7.47	164.36	187.31
CAP	Percentage	5.79	1.68	3.30	9.40
CLI	Index number, base=100	100.39	1.25	98.17	102.54
CONS	Number hours per 100 hours	103.05	1.72	100.75	107.04
CONSCOST	Index number, base=100	100.69	4.61	89.06	107.90
CPI	Index number, base=100	96.16	6.86	82.12	104.56
CRIMCOD	Binary				
DRIVLIC	Binary				
ELEC	Ktep (thousands of tonnes)	5214.03	303.91	4660.06	5873.23
EURIBOR	Percentage	2.04	1.52	0.06	5.38
FUELPRICE	Index number, base=100	87.70	16.11	58.99	114.25
GDP	Euros per inhabitant	5656.97	329.21	4791.13	6287.84
HOUSE	Index number, base=100	81.29	14.04	63.64	101.42
IBEX35	Index number	10441.88	2076.68	6738.10	15890.50
IPI	Index number, base=100	106.11	14.62	85.44	131.60
MORT	Thousands of mortgages	263.17	153.05	68.31	520.10
NVEH	Units of vehicles per 100 inhabitants	70.46	1.98	64.76	72.19
OLDPOP	Thousands of inhabitants	7862.80	492.48	7150.47	8699.89
PERELEC	Ktep per 1000 persons	114.13	7.10	100.31	127.06
POP	Thousands of inhabitants	45712.99	1256.52	42703.31	46792.31
REAL	Thousands of transactions	142.54	61.95	54.84	251.65
RET	Index number, base=100	98.16	12.77	71.10	120.31
TOTNVEH	Thousands of vehicles	32234.33	1730.15	27653.46	33498.50
SERV	Number hours per 100 hours	99.68	0.29	98.96	100.44
UPR	Per mille	170.45	67.81	79.32	269.35

The penal code reform takes a value of 0 until the third quarter of 2007 and a value of 1 afterwards.

Note that other important modifications concerning vehicle safety as the compulsory use of seat belts for all passengers, helmets and other safety devices, driving and rest time, are previous to the period under study, so we will not take them into account.

3. METHODOLOGY

To identify the main factors affecting claim frequency data we employ a methodology based on dynamic regression models with ARIMA errors. The methodology, that combines elements from regression models with elements from time-series, is a common tool to identify the most relevant factors affecting a given dependent variable by analyzing time dependency in the observed data. For a detailed description of this method, see Pankratz (1991). The complexity

of this procedure is due to the non-stationary nature of the relationship between explanatory variables and claim frequency. Similar approach has been followed by Wagenaar (1984), Van den Bossche *et al*. (2004) or García-ferrer *et al*. (2007) in the context of traffic safety.

In the following, we describe a unified stepwise procedure to establish an appropriate regression model relating explanatory and dependent variables, while examining the computational aspects of the process. A good starting point discussing the methodology is Andrews *et al*. (2013). In our explanation, we establish two stages. The first stage corresponds to the necessary steps to ensure non-stationarity of the time series and to reduce the potential number of explanatory variables. The second one corresponds to the regression procedure. In the later stage we use the TRAMO-SEATS methodology to choose the best model and provide the estimation and significance of the regression parameters. To choose the best final model TRAMO-SEATS uses the regression variables input by the user and generates trading day (with several specifications), Easter, intervention variables and outliers. Then, fits exact maximum likelihood estimation of the regression ARIMA model; makes diagnostics; detection and correction of outliers (Additive Outlier, AO, Transitory Change, TC, Level Shift, LS, and Innovational Outlier, IO); and optimal interpolation of missing observations and associated standard errors. Finally, computes optimal forecasts and associated standard errors.

3.1. Data preparation and processing

The first step in our approach is data cleaning, to avoid reporting errors that may arise during the compilation of the claim frequency information. Otherwise, errors may appear as outlier observations, which are sampled from the tail of the distribution and therefore have a strong influence on parameter estimation. If $q_{i,t}$ denotes the claim frequency of company i during the quarter t, we eliminate those observations exceeding certain thresholds given by

$$(1) \qquad\qquad \overline{q}_t - 3\sigma_{q_t} < q_{i,t} < \overline{q}_t + 3\sigma_{q_t},$$

where $\overline{q}_t$ denotes the weighted claim frequency mean in quarter t and σ_{q_t} denotes the weighted standard deviation in the same quarter t.

After the data cleaning process we take natural logarithm of all variables, except for the dummy variables. The advantage of this procedure is twofold. In the one hand, we mitigate potential problems derived from the skewness of the variable distributions. On the other hand, estimated coefficients can be interpreted in an easy way as elasticities.

In some cases it might be necessary to difference the series in its seasonal and non-seasonal components in order to transform the non-stationary series into a stationary one. Non-stationarity frequently arises in series with trend or with seasonality. If we estimate the model while any of the series are non-stationary, the estimated coefficients can be incorrect. In order to determine the number

of non-seasonal differences we apply augmented Dickey-Fuller, ADF, unit root test. For the seasonal component, we use the Canova-Hansen test (Canova and Hansen, 1995), testing the null hypothesis of deterministic seasonality. As suggested in Andrews *et al*. (2013), we test the dependent time series and apply the same differencing scheme to all the exogenous candidates. Nevertheless, we do not prevent the automatic estimation procedure to further difference the residuals if considered necessary.

One of the main methodological problems that may arise within this context is to mistake the causal relationship between the variables. In order to make sure that the independent variable does not receive feedback from the dependent variables, we employ the Granger causality test. According to this method, and for the bi-variate case, the first variable is said to cause the second variable in the Granger sense if the forecast for the second variable improves when lagged variables for the first variable are taken into account (Granger, 1969). In our procedure, a causality analysis of the mutual relationship between the claim frequency and each of the explanatory variables is conducted. If the test detects reverse causality, the exogenous variable is eliminated. We employ a causality test at 5% level departing from the log transformed variables using the lmtest package for R (Zeileis and Hothorn, 2002).

To easy the complexity of the estimation process and to reduce the number of explanatory factors, two extra steps are taken for each category and coverage. Firstly, as explanatory factors are only included those predictors with a significant correlation with the dependent variable. For this aim, function cor.test of the R package stats is used in computing (Pearson) correlation tests at a 95% confidence level. Factors that do not show significant correlation with the corresponding claim frequency are discarded. Secondly, in order to avoid multicollinearity issues (see, e.g., Allen, 1997 and Belsley, 1991), an ad-hoc selection process is carried out by taking into account the correlation matrix among risk factors jointly with the correlation between risk factors and claim frequency, and through the use of the VIF (Variance Inflation Factor), dropping those variables with a VIF higher than 5 (Menard, 1995). Figure 2 summarizes the process.

3.2. Estimation procedure

Once the first stage is finished and the data is prepared, the second stage proceeds to construct the regression model and to estimate its parameters. Let $y_t = (y_1, \ldots, y_T)$ be the vector of observed values of the dependent variable. If we denote by x_{it} the observation of the regression variable i ($i = 1, \ldots, n$) in period t, and β_i its corresponding regression coefficient, we assume that the endogenous variable y_t is generated by the model

$$(2) \qquad y_t = \sum_{i=1}^{n} \beta_i x_{it} + z_t,$$

where the errors z_t follow an ARIMA $(p, d, q)(P, D, Q)_s$ process.

FIGURE 2
DATA PREPARATION FLOW CHART
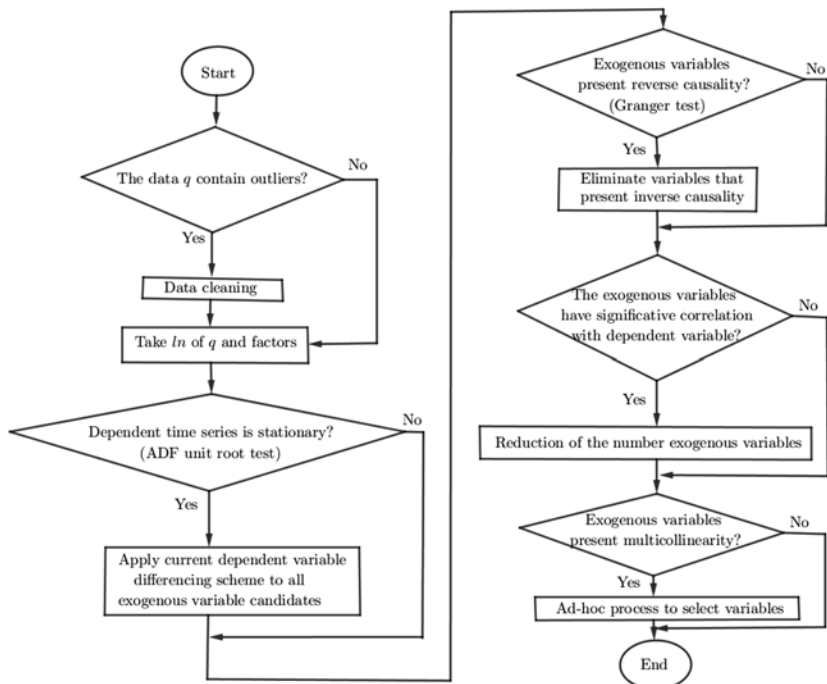
In terms of the backshift operator (or lag operator) B, such that $Bz_t = z_{t-1}$, and more generally, $B^n z_t = z_{t-n}$, the ARIMA process can be rewritten as

(3) $$\phi(B)\Phi(B^s)\delta(B)z_t = \theta(B)\Theta(B^s)a_t.$$

Under the backshift notation, on the left hand side of the equality, $\phi(B)$ denotes the autoregression operator $\phi(B) = (1+\phi_1 B + \cdots + \phi_p B^p)$, $\Phi(B^s)$ is the seasonal autoregressive operator $\Phi(B^s) = (1+\Phi_s B^s + \cdots + \Phi_{P_s} B^{P_s})$ and $\delta(B)$ reflects the seasonal and non-seasonal difference $\delta(B) = (1-B)^d (1-B^s)^D$. On the right hand side, we have $\theta(B)$, the moving average operator $\theta(B) = (1+\theta_1 B + \cdots + \theta_q B^q)$, and $\Theta(B^s)$, the seasonal autoregressive operator $\Theta(B^s) = (1+\Theta_s B^s + \cdots + \Theta_{Q_s} B^{Q_s})$. The a_t's are independent normal variates with mean 0 and variance σ_a^2. Being the data of quarterly nature, we set $s = 4$.

The estimation is carried out using TRAMO-SEATS tool, which allows specifying ARIMA errors taking into account calendar effects and outlier detection. For our purpose, however, only the TRAMO (Time series Regression with ARIMA noise, Missing observations and Outliers) part is necessary. Some of the

regression variables can be inputted by the user (we introduce several economic and legal variables) while other ones are generated by the program (trading day, Easter effect and intervention variables of several types). The program estimates both the parameters of the regression model and the ARIMA model for the errors by exact maximum likelihood (by default) although the unconditional and conditional least squares methods are available as options. It also detects and corrects for several types of outliers, computes optimal forecasts for the series, yields optimal interpolators of the missing observations and performs an automatic model identification of the ARIMA models. The basic methodology used is described in Gómez and Maravall (1994) but for complete details we refer to Gómez and Maravall (1996) and Maravall *et al*. (2014). Although some authors highlight the pitfalls of automatic variable selection methods (Harrell, 2001 or Burnham and Anderson, 2002), we are confident in following this approach for the problem considered in this paper.

4. ESTIMATION RESULTS

In this section we present estimation results of the model described in Section 3. As explained above, the procedure can be roughly summarized in two stages. The first stage in the process comprises data collection and cleaning. Once the data has been treated, we take logarithms and test for reverse causality and non-stationarity of the series. After selecting the better candidates as explanatory variables and properly differencing, in a second stage regressions are run and results are examined.

Along the section, we report the regression results grouped by coverages (general liability, own damage, theft and legal defense), so that each table collects results for the three automobile categories under the same coverage. Regression coefficients are provided jointly with the usual t-statistics (in parenthesis), and asterisk symbols are facilitated in order to indicate the significance for different levels. We also report variables that correspond to important outliers (OUT) and calendar effects (TRAD/HOL/EAST).

Since the methodology applied allows the residuals to follow an unknown ARIMA structure, its parameters must be estimated for its seasonal and non-seasonal parts. Therefore, parameter estimation results are also reported for each case. The optimal values of the parameters are automatically computed trough the TRAMO-SEATS adjustment procedure.

Under the hypothesis of the model, residuals are assumed to be normally distributed and serially independent. So, a complete analysis must be performed to check that the desired statistical properties are met. TRAMO-SEATS provides a battery of tests to ensure the quality of the adjustment (see, pp. 23-24 of Maravall *et al*., 2014). The residual diagnostics include checks on zero mean, lack of residual autocorrelation, normality, skewness, kurtosis, random residual signs, seasonality detection, and others.

In order to reinforce the validity of the results, we provide an extra analysis of the regression residuals. The normality assumption is first evaluated by visual inspection of the histogram of the residuals versus the normal probability plot, examining the resemblance of both representations. However, since graphical methods are known to be deceptive, and incorrect conclusions may be taken on the basis of a visual inspection, we use additional normality tests to obtain more conclusive assessments.

4.1. General liability

We start by analyzing the results on general liability coverage, which is the most important automobile coverage in terms of volume of premiums. It covers physical injury as well as damage to property involuntarily caused to a third party. In Spain, the compulsory insurance of motor vehicles guarantees coverage of civil liability. The amounts of the compulsory insurance coverage are 70 million Euros per claim in damages to persons, whatever the number of victims and 15 million Euros per claim in damage to property (Law 21/2007). The above amounts are updated based on the European consumer price index. If the amount of the compensation is greater than the amount of the compulsory insurance coverage, this maximum amount will be paid, and the remainder up to the total amount of the compensations will be paid by the voluntary insurance or the person responsible for the loss.

Table 3 reports the results of the relationship between claim frequency related to general liability coverage and several predictors. Before the regression, Categories I, II and III have been differenced with respect to its non-seasonal component. We only include in the table those predictors that were not discarded because of reverse causality or because they did not show significant correlation with the dependent variable and therefore have been incorporated into the dynamic regression. The table reports estimated coefficients and t-values of the regression. The estimated ARIMA $(p, d, q)(P, D, Q)$ structure for the residuals is reported at the bottom of the table.

According to the results, we find that regulatory measures are the most significant among the considered variables affecting Category I. In particular, it is crucial the introduction of the penalty point system. As expected, the tightening of traffic regulation impacts negatively the claim frequency rate of passenger cars: the threat of severe sanctions, including the possibility of loss of the driving license, represses an aggressive driving attitude.

In Category II we find a negative relationship, although feeble, with unemployment rate. This finding is consistent with previous literature: a higher unemployment rate implies less driving and a safer driving attitude, which results in lower claim frequency rates. Also, the Easter effect has been found to be significant in this category. Since vehicles included in Category II are fundamentally related to working effects, it is reasonable that the Easter holidays imply a reduction of activity within this class of vehicles, therefore incurring in lower claim frequencies.

TABLE 3
SUMMARY OF THE MODEL FOR GENERAL LIABILITY COVERAGE

Predictors	Category I	Category II	Category III
CAP	0.052238 (0.07817)	0.097122 (0.21923)	
CPI	2.466700 (2.01853)		
CRIMCODE	-0.025915 (0.04305)	-0.040669 (0.03452)	0.017256 (0.02086)
DRIVLIC	-0.184630** (0.06556)	0.040049 (0.02458)	-0.032412· (0.01932)
ELEC		-2.291100 (6.03885)	
GDP	-0.414710 (0.57861)	-0.760050 (0.89412)	
PERELEC	-0.012036 (0.30106)	3.078800 (6.12696)	
REAL	0.090039 (0.06311)	0.025352 (0.08336)	
RET	-0.331980 (0.25828)		
UPR		-0.429470· (0.22867)	
EAST1		-0.125570* (0.04887)	
OUT1 (2)	-0.367240** (0.10463)		
ARIMA model	$(0,1,1)(0,0,0)_4$	$(0,0,1)(0,1,1)_4$	$(3,0,0)(0,0,0)_4$
AIC	-79.6311	-96.2673	-47.6946
BIC	-4.4186	-4.8662	-3.8337

Notes: Standard error in parenthesis. ***, **, * and · indicate significance at 0.1, 1, 5 and 10% levels (t-statistic). The outlier observation number is indicated in parenthesis.

We found Category III, in the case of general liability, to be the most difficult series under study to model. From all the candidate explanatory variables, only two of them we considered to have a potential influence, and only the penalty point system variable is found to be weakly significant. So, the effect of economic variables on motorbike and scooter claims, for this particular coverage, keeps pending of further strong evidence.

According to the diagnostics obtained while running TRAMO-SEATS, residuals present a good behavior in all the examined components and therefore the regression is qualified as "good", in the TRAMO-SEATS terminology. Also, with the aim to reassess the quality model, we graphically check the distribution of the residuals in Figure 3, plotting its histograms for each of the three categories.

Histograms (Figure 3) show a normal pattern in the three cases except possibly in the far right tail. In order to reinforce the acceptance of the normality hypothesis, we perform three complementary normality tests on the residuals:

FIGURE 3

DISTRIBUTIONS OF RESIDUALS. GENERAL LIABLITY COVERAGE

(a) Category I. Distribution.   (b) Category II. Distribution.   (c) Category III. Distribution.
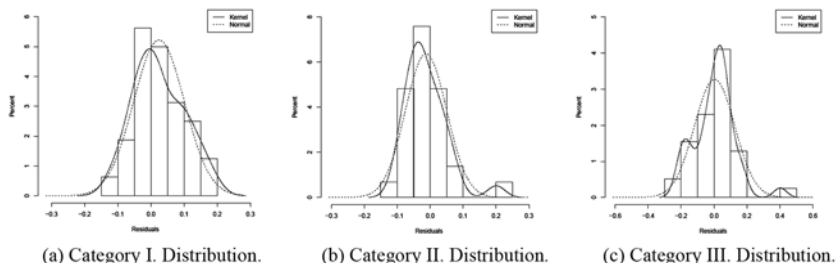
Shapiro-Wilk, Kolmogorov-Smirnov and Geary. They all test the null hypothesis that the observations come from a normal population. So, higher p-values support the normality of the residuals.

Results on tests are reported in Table 4. At 5% level, only the Shapiro-Wilk test would reject the null hypothesis of normality for Categories II and III. The rest of the tests support the normality of the residuals.

With respect to the autocorrelation of the residuals, Table 5 collects the results on the performed tests (Ljung-Box and Friedman), which show no sign of autocorrelation problems.

TABLE 4

NORMALITY TESTS OF RESIDUALS. GENERAL LIABILITY

	Shapiro-Wilk		Kolmogorov-Smirnov		Geary	
	Statistic	p-value	Statistic	p-value	Statistic	p-value
Category I	0.97450	0.6316	0.11582	0.3369	0.82240	0.3155
Category II	0.90737	0.0148	0.13624	0.1833	0.75641	0.8955
Category III	0.94083	0.0405	0.13265	0.0813	0.74795	0.9485

TABLE 5

AUTOCORRELATION AND SEASONALITY TESTS OF RESIDUALS. GENERAL LIABILITY

	Ljung-Box		Friedman	
	Statistic	p-value	Statistic	p-value
Category I	11.84	0.3754	1.65	0.6481
Category II	8.55	0.5752	0.26	0.9679
Category III	12.95	0.1647	3.27	0.3523

4.2. Own damage

We next analyze the impact of economic and related variables to claim frequencies for own damage coverage. Own damage covers damages to the policyholder's vehicle in transit and possible related incidences as fire or external explosion. It does not include damages to passengers nor mechanical failures which are not produced by accidents. Own damage is usually part of a fully comprehensive insurance policy, so it is usually found in insurance policies covering new vehicles. It should be noted that in Spain, some extraordinary risks, such as damages caused by terrorism or atypical severe natural phenomena are compensated by a special Consortium, and therefore, are not included in the own damage coverage.

In this case, Category III series has been differenced in its non-seasonal component, whereas Category II has been differenced with respect to its seasonal component, to ensure stationarity. Table 6 collects the estimation results.

Compared to the general liability case, results provide more insight on the impact of explanatory variables on claim frequencies. Category I shows a negative effect from the Industrial Production Index. An increase of industrial activity means a decrease in claim frequency rates. Also, electricity consumption has a negative impact on claim frequency rates. A possible explanation is that these variables reflect an improvement of economic activity and therefore levels of wealthiness, which are usually related to lower levels of fraud. As the economy improves, individuals are more reluctant to claim own damages in cases where that would pose to commit an act of fraud, and vice versa.

Although insurance fraud has been object of multiple studies (see Artís et al., 1999, Tennyson and Salsas-Forn, 2002 or Ngai et al., 2011), the connection between fraud and economic indicators has not been much analyzed in the academic literature. Apart from Lawal et al. (2017), we could not find recent quantitative studies relating the two aspects. However, this relationship has been suggested in the professional practice. For example, Nick Starling, former director of general insurance and health of the Association of British Insurers, emphasized that "Fraud thrives in a recession, so insurers are intensifying their crackdown on insurance cheats"[1].

Categories II and III are mainly affected by legal reforms. Both the introduction of the penalty point system and the criminal code reform are found to affect claim frequency rates on both categories. However, the impact of the reforms is mixed, simultaneously affecting in a positive and negative manner. We have no explanation for this mixed effect, since the reform of the penal code or the introduction of the penalty point system should not apparently have a positive impact on claim frequencies.

Models for the three categories have been found to be good enough according to TRAMO-SEATS set of diagnostics. As in the previous subsection, we

---

[1]   Insurance Day, Issue 2902, p. 6.

TABLE 6
SUMMARY OF THE MODEL FOR OWN DAMAGE COVERAGE

Predictors	Category I	Category II	Category III
CAP	-0.118480·	-0.034923	
	(0.06511)	(0.20628)	
CRIMCODE	-0.001428	-0.083315**	0.230880·
	(0.00794)	(0.02999)	(0.12235)
DRIVLIC	-0.014022	0.076368**	-0.235690*
	(0.01424)	(0.02805)	(0.11535)
GDP		1.088000	
		(1.28372)	
IPI	-0.105300**		
	(0.02976)		
MORT		0.130260	
		(0.09679)	
OLDPOP	1.466300		
	(2.57143)		
PERELEC	-0.930730***		
	(0.21387)		
REAL			0.480220·
			(0.25762)
RET	-0.002586		
	(0.18024)		
SERV	-7.855000*	12.20700	
	(3.60476)	(9.64079)	
OUT1(16)	0.223970***		
	(0.04875)		
OUT2 (31)	-0.198000***		
	(0.04386)		
ARIMA model	$(0, 0, 1)(0, 0, 0)_4$	$(2, 0, 0)(1, 1, 0)_4$	$(0, 0, 3)(0, 0, 0)_4$
AIC	-108.6503	-41.8600	-41.5195
BIC	-5.0248	-3.6818	-1.5438

Notes: Standard error in parenthesis. ***, **, * and · indicate significance at 0.1, 1, 5 and 10% levels (t-statistic). The outlier observation number is indicated in parenthesis.

FIGURE 4
DISTRIBUTIONS OF RESIDUALS. OWN DAMAGE COVERAGE

(a) Category I. Distribution.   (b) Category II. Distribution.   (c) Category III. Distribution.
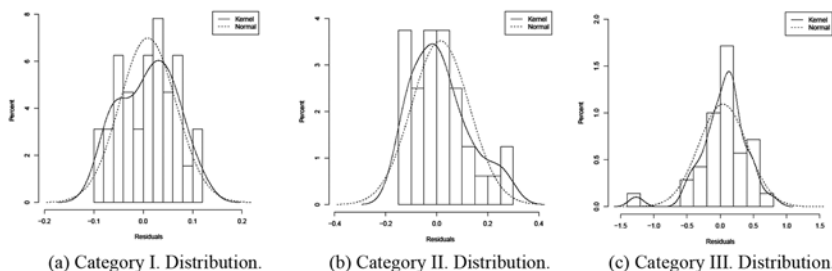
also analyze the residuals in order to check the normality and autocorrelation of the residuals (Figure 4).

With respect to the normality tests (Table 7), none of them reject the null hypothesis at 5% level except the Shapiro-Wilk applied to Category III. Also similarly, the Geary test provides a high confidence on the normality assumption for the same Category III, in opposition to what the Shapiro-Wilk test suggests.

Finally, the autocorrelation of the residuals seems to be under control, according to the results in Table 8.

TABLE 7
NORMALITY TESTS OF RESIDUALS. OWN DAMAGE

	Shapiro-Wilk		Kolmogorov-Smirnov		Geary	
	Statistic	p-value	Statistic	p-value	Statistic	p-value
Category I	0.97198	0.5556	0.11631	0.3305	0.84406	0.1295
Category II	0.93665	0.0602	0.13037	0.1804	0.79726	0.6010
Category III	0.92007	0.0143	0.11837	0.2453	0.71983	0.9845

TABLE 8
AUTOCORRELATION AND SEASONALITY TESTS OF RESIDUALS. OWN DAMAGE

	Ljung-Box		Friedman	
	Statistic	p-value	Statistic	p-value
Category I	11.28	0.4198	5.55	0.1357
Category II	9.42	0.3993	1.20	0.7530
Category III	5.00	0.8347	2.25	0.5222

4.3. Theft

Another coverage interesting of study is theft. The intuition suggests that, as the economy improves, the number of theft cases should decrease, and vice versa. Results should thus support this intuition.

Table 9 shows the impact of explanatory variables in theft coverage. In this case, it has only been necessary to difference the non-seasonal component in Categories I and II. In Category I we find a very strong positive influence between unemployment rate and theft. In periods where unemployment rates rise, so does theft. Thus, unemployment acts as a natural trigger of criminal activity as a consequence of the economic needs of the population. The effect is reinforced observing the negative relationship between construction costs and claim frequencies. In periods of higher construction activity (and therefore higher construction costs), levels of theft decrease. Along the same line of reasoning,

TABLE 9

SUMMARY OF THE MODEL FOR THEFT COVERAGE

Predictors	Category I	Category II	Category III
AGEPOP			-3.54110*
			(1.51143)
CAP		-0.247440·	
		(0.13408)	
CONS	-1.284600·		
	(0.65848)		
CONSCOST	-1.622600**		
	(0.56952)		
CRIMCODE	-0.025829·	-0.117610*	-0.281160*
	(0.01287)	(0.05323)	(0.10774)
DRIVLIC	-0.000225	0.095570·	-0.443960**
	(0.01281)	(0.04950)	(0.14442)
FUELPRICE	0.122560		0.512890·
	(0.12667)		(0.28198)
GDP			-0.413230
			(0.74177)
HOUSE			-0.000003*
			(0.00000)
MORT			-0.415940**
			(0.14764)
PERELEC			0.310790
			(0.56337)
RET	0.128070		
	(0.24236)		
SERV		-10.52600	
		(9.16695)	
TOTNVEH			3.18960
			(2.9992)
UPR	0.402380***		
	(0.07494)		
OUT1 (17)	-0.152490**		
	(0.05001)		
ARIMA model	$(0, 0, 1)(0, 0, 0)_4$	$(1, 0, 0)(0, 0, 0)_4$	$(0, 0, 0)(0, 0, 0)_4$
AIC	-125.5587	-31.3536	-38.7056
BIC	-5.5101	-3.4166	-3.4464

Notes: Standard error in parenthesis. ***, **, * and · indicate significance at 0.1, 1, 5 and 10%
        levels (t-statistic). The outlier observation number is indicated in parenthesis.

the improvement of wealth levels among the population would reduce the need
of committing illegal activities and thus deter crimes related to theft.

A similar pattern is perceived in Category III. Significant factors that imply
an improvement of the economy are found to be negatively correlated with theft.
It is the case of house pricing index and mortgage indicators. Also legal factors
are found to have a negative impact, but there is no plausible explanation on
its effects on theft.

The analysis of residuals performed by TRAMO-SEATS reflects a good fitting. A graphical representation of the density of the residuals is shown in Figure 5, where histogram plots resemble those of a normal distribution.

Normality test on the residuals (Table 10) corroborate the results. None of the tests reveal problems with respect to the normality of the residuals. Even the Shapiro-Wilk test, which has been shown to be the strictest out of the three normality tests in the previous subsections, does not reflect problems on this aspect, assigning high p-values to all categories. Also, according to the test results in Table 11, residuals do not show presence of serial autocorrelation at higher lags.

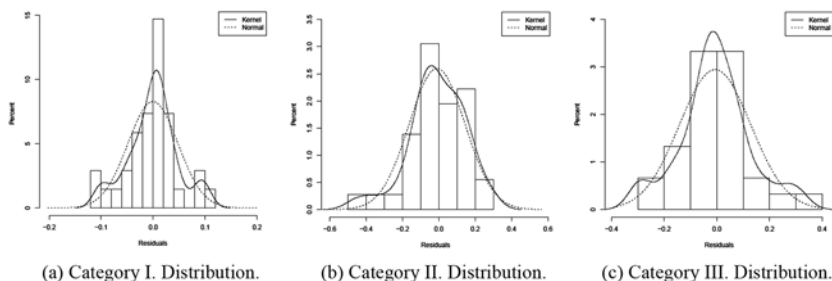FIGURE 5

DISTRIBUTIONS OF RESIDUALS. THEFT COVERAGE

(a) Category I. Distribution.   (b) Category II. Distribution.   (c) Category III. Distribution.

TABLE 10

NORMALITY TESTS OF RESIDUALS. THEFT

	Shapiro-Wilk		Kolmogorov-Smirnov		Geary	
	Statistic	p-value	Statistic	p-value	Statistic	p-value
Category I	0.96737	0.3930	0.10419	0.4591	0.74371	0.9520
Category II	0.96182	0.2444	0.10921	0.3410	0.77300	0.8145
Category III	0.96943	0.5237	0.11953	0.3355	0.73493	0.9675

TABLE 11

AUTOCORRELATION AND SEASONALITY TESTS OF RESIDUALS. THEFT

	Ljung-Box		Friedman	
	Statistic	p-value	Statistic	p-value
Category I	7.84	0.7274	9.90	0.0194
Category II	7.02	0.7977	6.07	0.1084
Category III	10.96	0.5320	1.11	0.7736

4.4. Legal defense

Finally, Table 12 collects results from the regression model concerning legal defense coverage. Categories I, II and III have been differenced in its non-seasonal components.

The increase in construction activity, usually a reflection of higher economic activity, is found to have a negative effect on Categories I and II. As the economy boosts, we observe less incentives to claim for damages. In Category II, not only construction affects the claim frequency rate; so does services. Moreover, the Industrial Production Index has a negative effect on the frequency claims, which reinforces the conclusion that economic variables might act as powerful indicators on claim frequency rates. Unfortunately, no conclusive effects are found in Category III, which again highlights the difficulty to find suitable models for motorcycles and scooters.

The analysis of the residuals does not indicate problems in Categories I and II. P-values do not reject the normality assumption even for high significance levels (Table 13), and the serial autocorrelation of the residuals behaves as expected (Table 14). Normality tests in Category III, however, point out potential problems on the fitting. The normality assumption is rejected at a 5% level according to Kolmogorov-Smirnov test, although the other tests support normality at the same level.

After analyzing the results for the twelve series, two different patterns emerge. Generally, economic variables that represent an improvement of the economic situation, such as decreasing unemployment rates, increase in construction activity or Industrial Production Index are correlated with higher rates of claim frequency. However, when considering the own damage coverage, the reverse pattern is observed. Economic variables that reflect a booming activity are found to be negatively correlated with claim frequency series. A possible explanation, that will deserve further attention, is the strong effect that fraud might cause on these particular insurance policies.

FIGURE 6

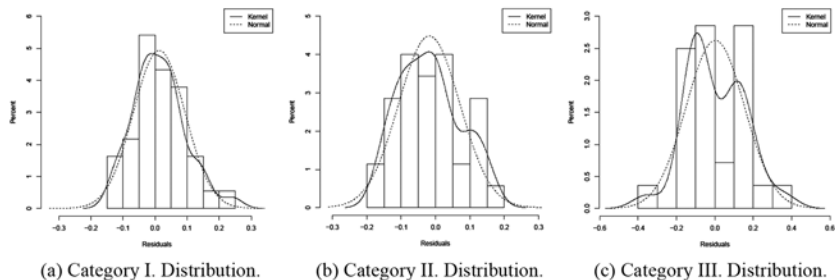DISTRIBUTIONS OF RESIDUALS. LEGAL DEFENSE COVERAGE

(a) Category I. Distribution.

(b) Category II. Distribution.

(c) Category III. Distribution.

TABLE 12

SUMMARY OF THE MODEL FOR LEGAL DEFENSE COVERAGE

Predictors	Category I	Category II	Category III
CAP			-0.672060
			(0.41738)
CONS	-2.841000**	-2.991200***	1.179600
	(0.99775)	(0.62955)	(4.21516)
CONSCOST			-1.437100
			(1.68114)
CPI		1.754300	3.647800
		(1.55633)	(3.07619)
CRIMCODE	-0.001476	-0.049416	0.082664
	(0.01575)	(0.03209)	(0.05765)
DRIVLIC	-0.005562	0.014545	0.000155
	(0.01421)	(0.03246)	(0.05116)
GDP			4.603400
			(2.85061)
IPI		-0.166040***	
		(0.03463)	
MORT			0.056296
			(0.15591)
OLDPOP		4.504500	
		(6.04279)	
REAL			-0.026852
			(0.13601)
RET	0.111200		1.486100·
	(0.32202)		(0.83474)
SERV		-25.96000***	
		(3.79472)	
TRAD1	0.027828**		
	(0.01021)		
ARIMA model	$(0, 0, 1)(0, 0, 0)_4$	$(1, 0, 0)(0, 0, 0)_4$	$(1, 0, 0)(0, 1, 0)_4$
AIC	-84.6266	-74.7665	-25.7142
BIC	-4.6257	-4.2959	-3.0558

Notes: Standard error in parenthesis. ***, **, * and · indicate significance at 0.1, 1, 5 and 10% levels (t-statistic).

TABLE 13

NORMALITY TESTS OF RESIDUALS. LEGAL DEFENSE

	Shapiro-Wilk		Kolmogorov-Smirnov		Geary	
	Statistic	p-value	Statistic	p-value	Statistic	p-value
Category I	0.96945	0.3935	0.08762	0.6716	0.78841	0.6845
Category II	0.96264	0.2740	0.10099	0.4875	0.82116	0.3240
Category III	0.94201	0.1243	0.17222	0.0328	0.85925	0.0660

TABLE 14
AUTOCORRELATION AND SEASONALITY TESTS
OF RESIDUALS. LEGAL DEFENSE

	Ljung-Box		Friedman	
	Statistic	p-value	Statistic	p-value
Category I	14.19	0.2228	1.67	0.6444
Category II	13.57	0.2576	2.25	0.5222
Category III	8.27	0.6893	0.26	0.9679

5. Conclusions

Adapting the pricing methodology to the current economic situation is a challenging goal for the insurance industry. A deep understanding of the factors involved in the determination of realized claim frequency is a fundamental element in order to counteract the natural fluctuations of losses. The actual economic situation affects the determination of insurance rates, increasing or decreasing the premiums accordingly. So, companies must anticipate future costs to set the appropriate insurance rates.

This study extends current research on insurance economics literature aiming to estimate the effect of macroeconomic and microeconomic factors in the claim frequencies of automobile lines, while controlling for major regulatory reforms. As a novelty, we use data disaggregated by coverages and automobile categories.

From a wide range of potential economic explanatory variables, a stepwise elimination procedure allows to select a small set of candidate factors that are related to the claim frequencies. The methodology employed allows to deal with the usual non-stationarity of this type of series, as well as the multicollinearity among the explanatory factors. The regression procedure is based on a dynamic regression model, which combines longitudinal and cross-sectional elements. Thus, explanatory variables can be included in the regression while residuals are allowed to follow an autoregressive integrated moving average model, both for their seasonal and non-seasonal components.

Generally, our results conform with related literature. However, the distinction between coverages reveals a different behavior of claim frequency rates depending on coverage. In particular, we find that general liability, theft and legal defense respond in opposite direction to an improvement of the general economy as does the own damage coverage. The motivation of the distinct response may be caused by the different effect that fraud causes on the coverages. In the case of own damage, fraud seems to have a strong influence on claim frequency rates, increasing the number of claims as the economy weakens according to economic indicators. However, the opposite effect is present in the rest of claims, which decrease with the economic activity. This different behavior has not been identified in previous literature and deserves further study.

We leave for further research the segmentation of companies according to its volume. In our study we use aggregate data for the whole industry, so observed effects in claim frequencies correspond to the global insurance portfolio. However, portfolios of small insurance companies could be influenced by different factors than portfolios of large companies, due to the different nature of its clients and therefore different composition of their portfolios. A detailed analysis on this aspect could reveal relevant information to companies, which could be exploited in order to provide more accurate insurance rates.

Another natural extension of the analysis is the study of the relationship between economic indicators and other lines of insurance. We find home insurance an interesting candidate to develop further research. The comparison of results between automobile and home insurance could provide more insight on the relationship between these two lines of insurance and the evolution of the economic cycle.

6. REFERENCES

Aguero-Valverde, J., and P. P. Jovanis (2006). "Spatial analysis of fatal and injury crashes in Pennsylvania". *Accident Analysis & Prevention*, 38(3): 618-625.

Allen, M. P. (1997). Understanding regression analysis. New York: Springer.

Andrews, B. H., M. D. Dean, R. Swain, and C. Cole (2013). Building ARIMA and ARIMAX models for predicting long-term disability benefit application rates in the public/private sectors. Society of Actuaries.

Artís, M., M. Ayuso, and M. Guillén (1999). "Modelling different types of automobile insurance fraud behavior in the Spanish market". *Insurance: Mathematics and Economics*, 24(1-2): 67-81.

Belsley, D. A. (1991). Conditioning diagnostics: collinearity and weak data in regression. New York: John Wiley & Sons.

Burnham, K. P., and D. R. Anderson (2002). Model Selection and Multimodel Inference, 2nd ed. New York: Springer-Verlag.

Canova, F., and B. E. Hansen (1995). "Are seasonal patterns constant over time? A test for seasonal stability". *Journal of Business & Economic Statistics*, 13(3): 237-252.

Castillo-Manzano, J. I., M. Castro-Nuño, and D. J. Pedregal (2011). "Can fear of going to jail reduce the number of road fatalities? The Spanish experience". *Journal of Safety Research*, 42(3): 223-228.

Chi, G., W. Brown, X. Zhang, and Y. Zheng (2015). "Safer roads owing to higher gasoline prices: How long it takes". *American Journal of Public Health*, 105(8): 119-125.

Chi, G., A. G. Cosby, M. A. Quddus, P. A. Gilbert, and D. Levinson (2010). "Gasoline prices and traffic safety in Mississippi". *Journal of Safety Research*, 41(6): 493-500.

Cummins, J. D., and J. F. Outreville (1987). "An international analysis of underwriting cycles in property-liability insurance". *The Journal of Risk and Insurance*, 54(2): 246-262.

Dadashova, B., B. Arenas-Ramírez, J. Mira-McWilliams, and F. Aparicio-Izquierdo (2014). "Explanatory and prediction power of two macro models. An application to van-involved accidents in Spain". *Transport Policy*, 32: 203-217.

D'Arcy, S. P., and J. R. Garven (1990). "Property-liability insurance pricing models: An empirical evaluation". *The Journal of Risk and Insurance*, 57(3): 391-430.

Elvik, R. (2010). "The stability of long-term trends in the number of traffic fatalities in a sample of highly motorized countries". *Accident Analysis & Prevention*, 42(1): 245-260.

Fairley, W. B. (1979). "Investment income and profit margins in property-liability insurance: Theory and empirical results". *The Bell Journal of Economics*, 10(1): 192-210.

Feldblum, S. (2001). "Underwriting cycles and business strategies". *Proceedings of the Casualty Actuarial Society*, LXXXVIII:175-235.

García-ferrer, A., A. De Juan, and P. Poncela (2007). "The relationship between road traffic accidents and real economic activity in Spain: common cycles and health issues". *Health Economics*, 16(6): 603-626.

Gómez, V. and A. Maravall (1994). "Estimation, prediction, and interpolation for nonstationary series with the Kalman filter". *Journal of the American Statistical Association*, 89(426): 611-624.

Gómez, V., and A. Maravall (1996). Programs tramo (time series regression with arima noise, missing observations and outliers) and seats (signal extraction in arima time series). Instructions for the user. Documento de Trabajo, 9628, Servicio de Estudios, Banco de España.

Grabowsky, D. C., and M. A. Morrisey (2006). "Do higher gasoline taxes save lives?". *Economics Letters*, 90(1): 51-55.

Granger, C. W. J. (1969). "Investigating causal relations by econometric models and cross-spectral methods". *Econometrica*, 37(3): 424-438.

Hakim, S., D. Shefer, A. S. Hakkert, and I. Hocherman (1991). "A critical review of macro models for road accidents". *Accident Analysis & Prevention*, 23(5): 379-400.

Harrell, F. E. (2001). Regression Modeling Strategies. New York: Springer-Verlag.

Kozik, T. J. (1994). "Underwriting betas-the shadows of ghosts". *Proceedings of the Casualty Actuarial Society*, LXXXI: 303-329.

Lawal, A. I., E. O. Amogu, J. O. Adeoti, and M. A. Ijaiya (2017). "Fraud and business cycle; Empirical evidence from fraudsters and fraud managers in Nigeria". *Studies in Business and Economics*, 12(1): 110-128.

Maravall, A., G. Caporello, D. Pérez, and R. López (2014). New features and modifications in tramo-seats. Banco de España.

Menard, S. (1995). Applied logistic regression analysis: Sage University Series on Quantitative Applications in the Social Sciences. Thousand Oaks: Sage.

Ngai, E. W. T., Y. Hu, Y. H. Wong, Y. Chen, and X. Sun (2011). "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature". *Decision Support Systems*, 50(3): 559-569.

Nghiem, H. S., L. B. Connelly, and S. Gargett (2013). "Are road traffic crash fatality rates converging among OECD countries?". *Accident Analysis & Prevention*, 52: 162-170.

Noland, R. B., and M. A. Quddus (2004). "A spatially disaggregate analysis of road casualities in England". *Accident Analysis & Prevention*, 36(6): 973-984.

Pankratz, A. (1991). Forecasting with dynamic regression models. New York: John Wiley & Sons.

Tarbell, T. F. (1932). "The effect of changes in values on casualty insurance". *Proceedings of the Casualty Actuarial Society*, XIX, Part 1: 1-12.

Tennyson, S., and P. Salsas-Forn (2002). "Claims auditing in automobile insurance: fraud detection and deterrence objectives". *The Journal of Risk and Insurance*, 69(3): 289-308.

Van den Bossche, F., G. Wets, and T. Brijs (2004). A regression model with ARIMA errors to investigate the frequency and severity of road traffic accidents. Proceedings of the 83th Annual Meeting of the Transportation Research Board, Washington.

Wagenaar, A. C. (1984). "Effects of macroeconomic conditions on the incidence of motor vehicle accidents". *Accident Analysis & Prevention*, 16(3): 191-205.

Zeileis, A., and T. Hothorn (2002). "Diagnostic checking in regression relationships". *R News*, 2(3): 7-10.