Telematic driving profile classification in car insurance pricing

Wiltrud Weidner*

Institute for Risk and Insurance, Leibniz University Hannover, Otto-Brenner-Str. 1, 30159 Hannover, Germany

Fabian W.G. Transchel

Institute for Theoretical Physics, Leibniz University Hannover, Appelstr. 2, 30167 Hannover, Germany

Robert Weidner

Laboratory for Manufacturing Technology, Helmut Schmidt University, Holstenhofweg 85, 22043 Hamburg, Germany

Abstract

This paper presents pricing innovations to German car insurance. The purpose is to provide an effective approach to adapting actuarial pricing decision to incorporate telematic data, which differs substantially from established tariff criteria in complexity and volume. A vehicle mobility model and a real-world sample of driving profiles form the input into the analysis. We propose an allocation of the driving profiles based on velocity and acceleration parameters to specific driving styles for evaluating the driving behaviour to subsequently enable discounts or surcharges on the premiums to obtain usage-based insurance premiums. The result is highly relevant for actuaries, who calculate the tariffs, but also for managers, as they have to make a pricing decision.

1. Introduction

The car insurance market is globally the most important line of business in the property and casualty insurance market due to high premium volume. In Europe, the motor vehicle business represents with €130 billion in 2013 a share of about 29% of the total premium volume (see Insurance Europe, 2014). In view of significant technological changes such as assistance systems, connected or driverless cars, it is very difficult to forecast the future of car insurance. The so-called telematic box is an intermediate step in terms of technology, which can change the rules of the pricing structure. Telematic data enable innovative, usage-based insurance premium models known as pay-how-you-drive (PHYD) insurance, but in this regard they represent a special challenge for actuaries because of their complexity and volume.

Existing studies provide different insights into telematic insurance solutions. However, empirical studies on the implementation into actuarial pricing methods are still very limited. Jun *et al.* (2007, 2011) work

*Correspondence to: Wiltrud Weidner, Institute for Risk and Insurance, Leibniz University Hannover, Otto-Brenner-Str. 1, 30159 Hannover, Germany. Tel: +4951147396044; E-mail: ww@ivbl.uni-hannover.de

out the relationship between mileage as well as velocity parameters and accident risk, whereas Kremslehner & Muermann (2013) examine the effect of the number of car rides, the relative distance driven on weekend, respectively, at night and the average speeding on risk. Toledo *et al.* (2008) determine risk indices and statistics identifying manoeuvre types with methods, which are not specified in detail. A classification analysis approach for aggregated pay-as-you-drive (PAYD) rate factors to actuarial decision making with vehicle sensor data are provided in Paefgen *et al.* (2013). Currently, either in science or in practice there is no standardised method to achieve a clear "score" of driving behaviour; telematic approaches from abroad do not extend to the very precise German actuarial pricing due to the limited scope of their scoring methods (see Karapiperis *et al.*, 2015). In general, a priori and a posteriori risk profile building should be included in a comprehensive model. What is missing therefore is an effective approach for the solid evaluation of driving profiles on the basis of the limited data available, and in particular without recourse to the claims situation which is not available.

This paper wants to give a realistic and useful evaluation method for driving profiles, which in turn can be used for innovative, optimised car insurance pricing in the private client segment due to reduction of information asymmetry with regard to the user group. Our analysis uses a vehicle mobility model for generating potential driving profiles, which grants our study a sound, broad data basis. We link the model results to real-world data of a telematic portfolio of a large German car insurer. We determine driving styles of generated driving profiles creating the basis for a general risk categorisation and validate the results by comparing them with empirical data. The combination of model results and the empirical data allows us to directly test whether a differentiation according to predefined driving styles is possible and how well the data format fits to make a valid statement on the driving styles. Furthermore, we can define a first minimum data request.

Our results not only contribute to academic research, but are also of high relevance and usefulness to insurance practice. A fundamental finding is that specific driving styles can be derived from driving profiles based on validated velocity and acceleration parameters even under consideration of severe privacy protection issues. These styles can form the basis for a risk categorisation and thus for subsequent discounts or surcharges on the premiums dependent of the driving behaviour.

The remainder of this paper is divided into four sections. First of all, the key actuarial pricing aspects in German car insurance pricing are introduced, including innovative pricing strategies. In section 3, the development of driving styles based on data using a mobility model and the process of empirical data collecting are presented. Section 4 discusses the results from the empirical study. The paper closes with conclusions and limitations of our work, and outlines suggestions for further research.

2. Overview of German Car Insurance Pricing

2.1. Current pricing strategy

The pricing of German motor vehicle insurance is based on numerous relevant factors – more than any other segment in the German property and casualty insurance. Various subjective and objective tariff criteria, as well as numerous criteria for discounts influence the insurance premium (see e.g. Stadler, 2008; Laas *et al.*, 2014). This can be attributed to the increased importance of risk-adjusted premiums after the deregulation of the European Union insurance markets in 1994. A comparison of the average annual premiums and losses for each risk shows that the price is aligned to the risk over time (see Figure 1). The most important prerequisite for adequately projecting the actual risk is the selection of different, suitable and tariff-related risk characteristics.
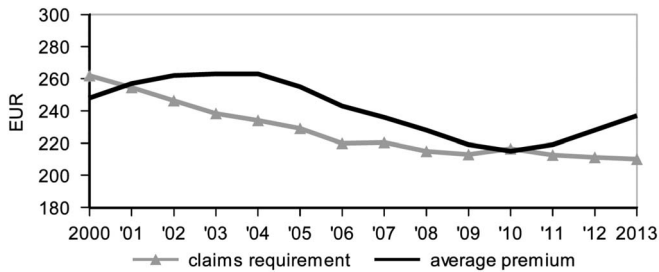
214

Figure 1. Average annual premiums and losses in the German motor vehicle third-party liability insurance from 2000 until 2013 (see Gesamtverband der Deutschen Versicherungswirtschaft e.V., 2012).

The German insurance industry distinguishes between objective and subjective criteria. Objective criteria concern the vehicle and a priori are independent of the policyholder's individual risk behaviour. These include, e.g., the type of vehicle, type classification and age of vehicle at the time of purchase. In contrast, subjective criteria relate to the person to be insured. Such criteria include the bonus-malus level, regional classification, occupation and age of the policyholder and users or the annual mileage performance (see e.g. Heep-Altiner & Klemmstein, 2001). Discounts such as home ownership, a garage or a workshop affiliation in the event of damage represent other factors (see e.g. Stadler, 2008; Laas *et al*., 2014).

The set of tariff criteria separates policies into groups, so-called tariff classes. From the historical claims data an expected average annual loss is determined for each tariff class, which corresponds to the size of the net premium (see e.g. DAV-Arbeitsgruppe Tarifierungsmethodik, 2015). Therefore, a large quantity of pricing variables induces a large dimensionality of an actuarial table. This fact in turn leads to an increasingly covert and complex market, of the one part, and high calculation capacity for the actuaries, of the other. Indeed, it can be seen that, in recent years, the insurers continuously introduce more and more subjective criteria (see Domeyer, 2005; Laas *et al*., 2014; Hartmann & Nützenadel, 2015). There are several reasons for this: the increasing knowledge of the actuarial risks due to the rising personal data availability in the course of electronic data processing, the high competitive pressure and not least the unisex judgement. This is done to calculate the tariffs more risk adjusted and more individualised.

2.2. Innovative pricing strategies

Each year, the German Insurance Association (GDV) submits a tariff recommendation for their member companies. Thus, a harmonised method of calculating tariffs has been established in Germany. Nowadays, the premium amount depends highly on the classification into a bonus-malus system. In the third-party liability and fully comprehensive insurance, a no-claims discount is set individually for each policy and is determined by the respective insurer on the basis of the bonus-malus level; the no-claims discount does not exist in the partially comprehensive insurance. Furthermore, any insurance company records statistical data, such as the insurance regional and type classification, and takes soft factors, such as age of users, group of users and occupation of the policyholder into account when setting premiums. However, the tariffs differ between the various insurance companies – the insurance companies deviate more or less from the recommendation of the GDV in some criteria, e.g., in order to meet the legal requirements of the principle of equal treatment or to regulate migration movements and selection effects (see DAV-Arbeitsgruppe Tarifierung, 2007; Laas *et al*., 2014). This includes the age group of the policyholders due to the lack of social acceptance because the age is not an acceptable reason
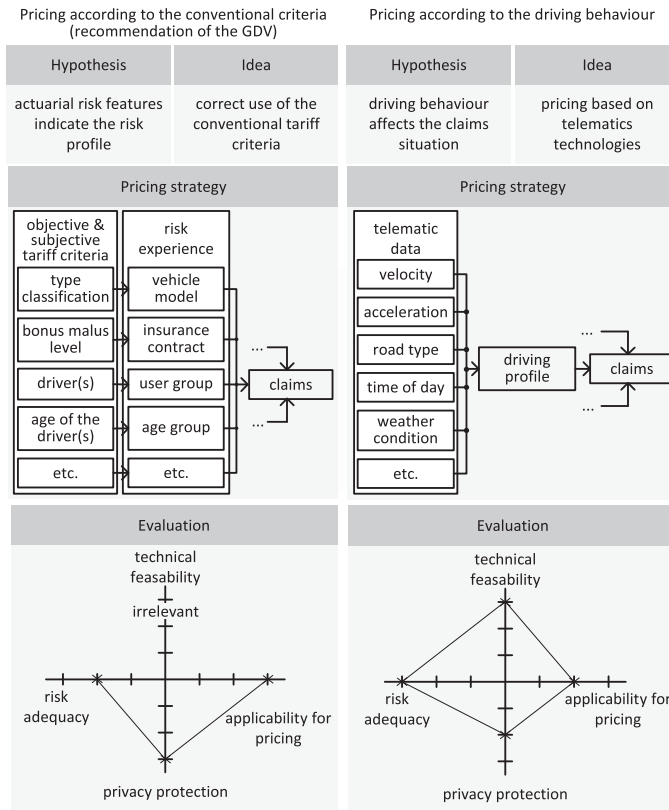
Figure 2. Conventional and telematic pricing idea comparison.

for discrimination. Therefore, the first and simplest idea to determine the premiums more risk adequately is to use the tariff criteria of the GDV correctly.

Apart from this, pricing according to driving behaviour, as shown in Figure 2, is another possibility to explore in order to achieve the aim of more individualised premiums (see Weidner & Weidner, 2014). Here, the emphasis has been placed on GPS technology (telematic) with such aspects as velocity, acceleration, road type, time of day and weather condition. Various sensors like GPS are required for measuring these parameters. Indeed, modern car technology is quite advanced and some sensors are even part of the standard equipment of each vehicle. On the basis of so-called variables of driving behaviour and driving situation (see Gerpott & Berg, 2012), the driving style and performance can be evaluated. Finally, the insurance premium can be aligned with the risk tolerance of the driver. Customers who drive safely and defensively should pay lower premiums in a telematic tariff, whereas customers who drive risky and aggressively should pay higher premiums. Such PHYD systems are already well established in the United States and in Great Britain; in the United States, every seventh new contract is based on telematics (see Stiftung Warentest, 2014). Meanwhile in Germany, first insurers now offer telematics products, which grant discounts on the premiums calculated with traditional tariff criteria.

The evaluation of the two pricing approaches (see Figure 2 to the right) shows that the inclusion of telematic data in the premium calculation is associated with great benefits in particular for differentiated tariffs. This is the result of two factors: first of all, the concrete verification of driving profiles
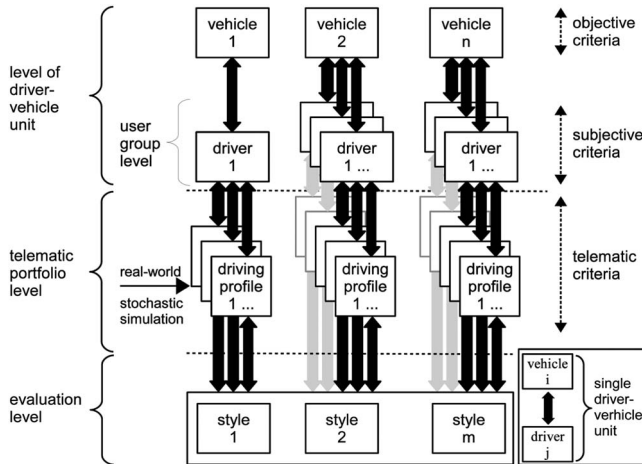
216

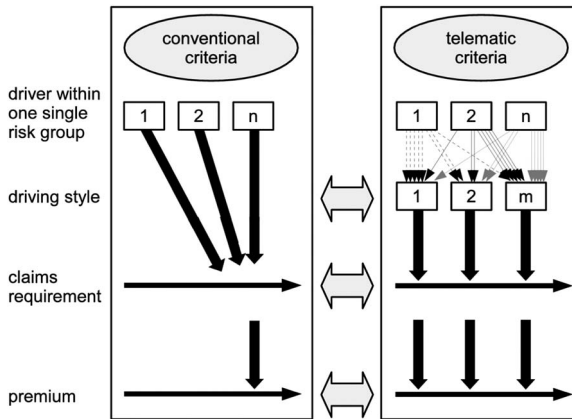**Figure 3.** Schematic definitions of abstraction levels.

Figure 4. Valuation perspective of pricing according to conventional and telematic criteria using the example of a user group consisting of single drivers.

of specific driver-vehicle units at the telematic portfolio level, second, the weighting of differing driving behaviour at the level of a single driver as well as within a user group (see Figure 3). With conventional tariff criteria, only vehicle and user group can be evaluated – not the actual driving behaviour. This has concrete implications to conventionally derived risk groups, where average premiums are calculated based upon the average claims requirement from the respective risk categories (see Figure 4 to the left). For example, novice drivers as a collective user group feature, on average, have a much higher claims requirement than more advanced drivers. Using only the conventional tariff criteria "novice driver", the calculated claims category for every driver from this user group is identical, thus implying same risk exposure and same premium. Telematic variables on the other hand allow for a differentiated perspective on individual driving styles, such that deviations from the average risk exposure (see Figure 4 to the right) can be identified and incorporated into the risk category classification. Ultimately, this will enable insurers to partly remedy information asymmetry as far as risks depending on actual driving behaviour are concerned. Excess value is created for recently closed contracts, where the bonus-malus level as major tariff criterion is not yet
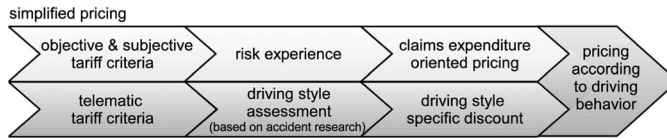
Figure 5. Incorporation of telematic data in the actuarial pricing process.

adapted to the actual risk due to missing perennial claims experience. On the other hand, there also exists information asymmetry concerning different users of a vehicle. Although the policyholder is required to name additional vehicle drivers, their usage can only be understood in terms of averages entering into the analysis of claims requirement. For an individualised risk assessment of a contract of a 45-year-old father including his 18-year-old son, the precise attribution of both usage time and risk associated with driving behaviour is critical for understanding the correct claims requirement – especially if both vehicle users exhibit vastly different usage patterns. These usage patterns are one of the key elements telematic analysis of driving behaviour can reveal.

A particular challenge regarding the successful implementation of a telematic product regards data protection issues. Complex motion profiles can be generated collecting telematic variables concerning the driving behaviour and the use of vehicle. Therefore, strict data protection criteria are applied to all processes used for the telematic tariff so that the insurer is able to adhere to all of the respective national privacy rules at all time. Concerning the German insurer's perspective, the prime principles of data minimisation and data avoidance in accordance with section 3, German Data Protection Act (BDSG) as well as the national data protection norms of the German insurance industry themselves, the Code of Conduct, processing or using as little personal data as possible, must be considered. Moreover, it must be ensured that every collected data point actually affects the tariff system complying with the purposeful limitation principle according to the BDSG under which personal data should only be collected and processed for specified and explicit purpose.

As regards the formal feasibility, the technical conditions are already in place, although it is essential to determine the measuring and transmission logic. The application of telematic technology is in particular among forerunners in the external data to optimise the pricing. Applicability of other approaches is possible but not as far developed as telematics (see Weidner & Weidner, 2014) – they will certainly be discussed, once their respective technologies reach a level of mass market sophistication, like telematics do now.

Note that there are various outstanding issues as far as the applicability for pricing of the telematic approach is concerned. Beside the question of a suitable choice of telematic tariff criteria, particularly the handling of the missing risk experience is also of major relevance to be able to put the approach into perspective with conventional pricing. This means that the premiums according to the driving behaviour cannot be directly determined based on the analysis of the claims requirements as the pricing is normally made. For this reason, the present approach investigates the pricing innovation that various premium discounts will be granted to the policyholders depending on their style of driving which requires a certain classification of driving behaviour. The discount on the premium per contract will be the weighted evaluation of all the driving styles belonging to the insured vehicle, because our analysis allows differentiating between different driving styles both for individual drivers as well as a mixed group of vehicle users. However, it is obvious that scoring of such differentiated findings requires further research in order to allow for risk adequacy premiums. Finally, the resulting simplified actuarial pricing process is shown in Figure 5.

218

3. Telematics Technologies: Evaluation of Driving Behaviour

3.1. Methodology

The procedure for the statistical analysis of driving profiles requires high-quality data and a broad database to achieve sufficient pricing precision. As representativity of the early stages of the empirical telematic portfolio can neither be trusted nor easily verified with the existing lack of claims experience, the most efficient way to build a reliably precise telematic database is extensive stochastic modelling. Therefore, a vehicle mobility model is used for stochastic simulation of driving profiles. The modelling is based on the random waypoint (RWP) principle (see Johnson & Malz, 1996). Weidner *et al.* (2015) provide a detailed introduction to the method, whereas the modified underlying assumptions are listed in the Appendix. The main point about choosing kinematic observables for this investigation is the realisation that contextual quantities describing the driving situation by likes of road type, maximal acceleration strength or time of day, which are currently in use in telematic assessments (see Karapiperis *et al.*, 2015), can be treated like classical risk factors, whereas for the driving behaviour, given by the actual movement of the vehicle, this is not the case.

The RWP model, slightly modified for this approach, describes the movement behaviour of a driver-vehicle unit in a given system area. Each movement occurs in a manner, such that a driver-vehicle unit randomly selects a destination (waypoint) in the area and a respective speed limit. The driver-vehicle unit moreover moves to this destination with its specific vehicle and driver characteristics, i.e. desired velocity depending on speed limit, specific acceleration rate and brake performance, and pauses for a certain time before selecting a new destination and speed. In this way, a stochastic simulation run generates simultaneously several thousand routes and types of drivers, from which specific driving profiles can be derived (see Figure 6). The choice of model parameters was made such that a wide range of individual characteristics of the driver-vehicle units and specific local conditions are depicted focussing on the representation of average city and rush-hour traffic. The present analysis of driving profiles is based on 10,000 driver-vehicle units, each with a sequence of 20 waypoints. Sampling frequency was set to 1 Hz in order to obtain detailed driving profiles without any loss of information[1].

To identify different driving styles, the generated driving profiles are classified using a cluster analysis[2]; such that the main objective of reducing the vast range of individual driving styles to a single driving style of each cluster is achieved[3]. The different clusters constitute the specific driving styles which will be characterised in the section below. At the beginning of the clustering process, the

---

[1] Weidner & Weidner (2014) present an analysis of the systematic error that is linked with the sampling frequency.

[2] In this simulation study, the clusters are identified using Ward's minimum variance method (Ward, 1963), implemented recursively using the Lance–William algorithm (Lance & Williams, 1967). Applying Ward's method appears to be plausible because it is directly related to the Euclidean distance. Driving profiles with, e.g., similar acceleration values have an Euclidean distance close to 0. They are pooled into one group, and the within-group sum of squares is small. In practice, this method generates highly homogeneous, compact groups and constitutes a stable procedure (Mangiameli *et al.*, 1996).

[3] The necessity of classification of the driving profiles is based upon the fact that the tariff calculation must be carried out on robust data (DAV-Arbeitsgruppe Tarifierungsmethodik, 2015). High robustness is achieved through a reduction of the number of parameters in the calculation, and through various significant premium differences between the strongly diverse driving styles. Examples where this practice has been successfully studied for many years include the aggregation of districts (Zulassungsbezirk) to regional classes (Regionalklasse) or vehicle models to type classes (Typklasse), respectively, performed over the associated claims requirements.
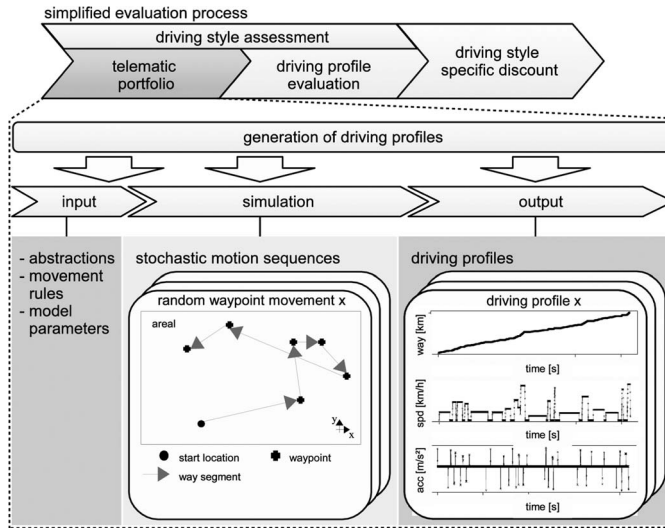
Figure 6. Model structure for generating driving profiles in the context of driving behaviour analysis.

following variables are selected for clustering because they have been identified as being most representative of the driving behaviour of the specific driver-vehicle unit: vehicle velocity and respective acceleration in the longitudinal direction (further referred to as acceleration and deceleration)[4,5]. The clustering process itself is based on medians (as robust measure of central tendencies) and interquartile ranges (as dispersion measure) which describe the distributions of each driver-vehicle unit considering the aspect of information aggregation[6].

3.2. Simulation results

The actual simulation shows that six driving styles, shown in Figure 7, can be distinguished at a significant level taking into account all three variables collectively. There are two groups of driving styles depending on the velocity process: those with lower velocity values and those with higher velocity values with the same or similar acceleration and deceleration value distribution.

[4] The chosen variables of motion accurately and completely describe the physical movement of the vehicle up to the measurement frequency's granularity. In addition, any other variable of motion, like position, momentum, jerk or meaningful aggregated quantities like averages or extreme value considerations of the aforementioned quantities can be derived from the used format of velocity and acceleration/deceleration, given a sufficiently fine discretisation. In the first step of investigating driving profiles in this way, only longitudinal acceleration was probed for more clarity. W.l.o.g. method can be extended to the measurement of lateral acceleration, representing the steering behaviour.

[5] As the actual correlation with the risk the insurer aims to measure and, subsequently, price into the tariff are unknown with respect to any telematic observable quantities at this stage of actuarial investigation of telematics technologies, the aforementioned set of variables is chosen to ensure that any correlations of risk regarding the vehicle movement can be resolved. Once reliable information about claims experience is available, insurers might be able to further refine the measurement variables both semantically and quantitatively.

[6] This approach delivers robust results even with originally incomparable extreme values, e.g., different peak values of the acceleration as a result of, e.g., a different quality of vehicle damping.

220
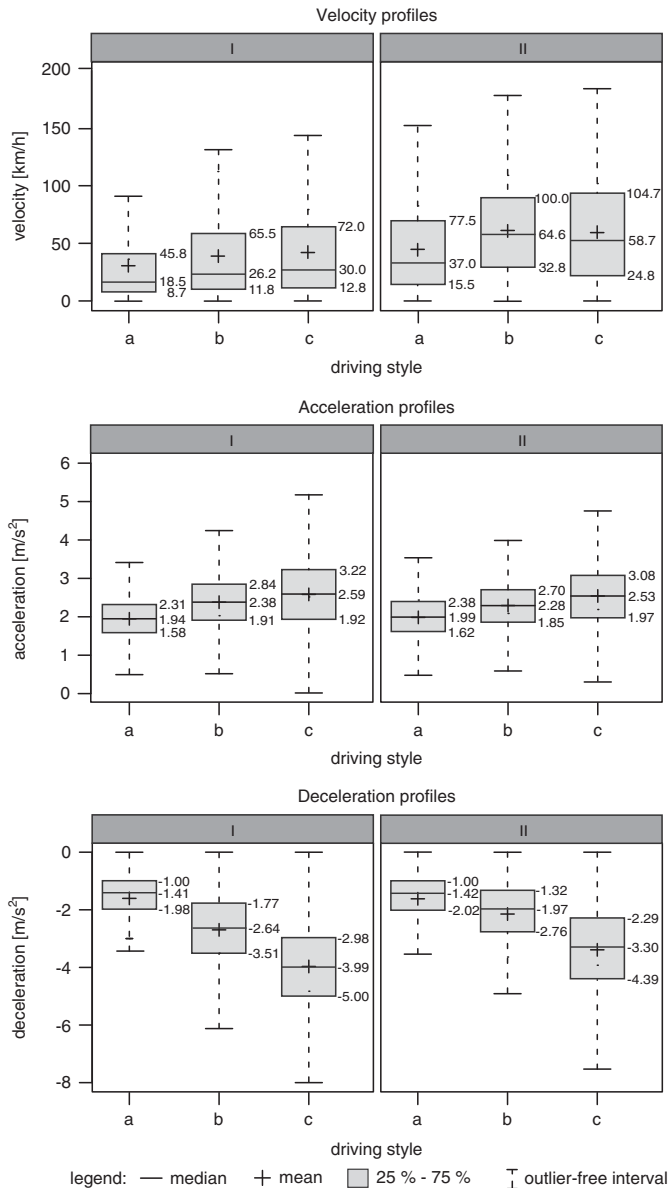
Figure 7. Box plots of data from stochastic simulation of acceleration, deceleration and velocity values.

Concerning the acceleration values, the values vary by up to $5.2\,\text{m/s}^2$, not considering outliers, and the medians of the six styles lie between 1.9 and $2.6\,\text{m/s}^2$. These values match to Lange (2006), who states a normal acceleration interval of 1.5 up to $2.5\,\text{m/s}^2$ and a maximum acceleration interval of 3.0 up to $6.0\,\text{m/s}^2$ for vehicles driving straight ahead. Extreme values, which deviate upwards and downwards from the limits of the first and third quartile by more than 1.5 times, are not shown in the box plots. However, they appear in each of the styles and amount to around 2.3% of the sample. Furthermore, the box plots indicate an increase in the interquartile range of the data with growing

221

median, which in turn indicate the degree of dispersion. The interquartile ranges of different styles vary between 0.8 and 1.3 m/s^2.

The values of the driving styles concerning the deceleration vary by up to –8 m/s^2, the range of the medians is between –1.4 and –4.0 m/s^2. This means that the values for the deceleration are within the required wide range marked by not only individual factors such as tyres, brakes and specific breaking behaviour, but also by road conditions such as the road surface, gradient and weather condition. A deceleration of –7.0 up to –8.0 m/s^2 (and up to –10.0 m/s^2 in isolated cases) can be achieved under ideal conditions. Due to less grip on wet or smooth surface, the maximum deceleration decrease to –6 up to –7 m/s^2, respectively, at most to –1 m/s^2 (see Becke, 2005). In contrast to the medians and means of the acceleration, which are on top to each other, as far as the deceleration is concerned, there are three driving styles showing means clearly below the medians. This is due to the fact that an emergency braking constitutes an exception when driving anticipatorily, but extreme values are necessarily part of even the most defensive driving profiles, although much rarer. Outliers here amount to only 1.5% of the sample; there are so many values close to 0 that they are not even classified as outliers. This is due to the fact that in contrast to acceleration behaviour, where the driver aims to reach the recommended speed as soon as possible, whereas breaking behaviour is much more anticipatory with regard to junctions or (heralded) obstacles. Also, one can see that the interquartile range tends to increase with growing absolute value of the median. The range lies between 1.0 and 2.1 m/s^2.

As far as the velocity is concerned, the box plots show medians between 18.5 and 64.6 km/h. It is evident that the values scatter more widely above the median than below. This means that the distribution of the values is (slightly) skewed to the right, and thus the distribution in the lower speed range is much smaller than in the upper range. This is the result of three factors: first of all, starting and braking due to obstacles, such as traffic lights. Second, the speed limit. Finally, it matters to what extent the driver's desired velocity and speed limit could be attained (see Schüller, 2010). Moreover, there are only upward outliers (2.3% of the sample), because significant deviations downwards cannot occur due to the minimum speed of 0 km/h, exemplary a standstill at a traffic light. Again, the interquartile range of the velocity values tends to grow with increasing medians and varies between 37.0 and 79.8 km/h.

Interesting is the symmetry comparing driving styles regarding the acceleration with those regarding the deceleration. That means in particular, that corresponding driving styles with strong acceleration also show strong deceleration and vice versa. This indicates that the driving style is the result both of acceleration and deceleration and cannot be inferred of either alone. But not only the acceleration and the driver's braking behaviour, but also the velocity processes depend on the technical features of the vehicle and the attitude of the driver. Thus, the velocity is also linked with acceleration and braking behaviour. However only slightly[7], as there are some factors countervailing this natural tendency. Drivers with identical acceleration and deceleration processes can choose between several road types, i.e. national roads, urban roads and highways, with different recommended speeds which is underlined by the two groups of driving styles with regard to velocity range. Furthermore, a defensive driving leads, as a result of uniform traffic flow avoiding special acceleration and deceleration processes, to a higher average velocity. Again, only incorporating velocity as third parameter fully differentiates the picture of driving styles.

---

[7] This is due not least to the link between the velocity and acceleration: the derivative of the velocity function with respect to time represents the instant of the acceleration.

222

4. Application to an Empirical Telematic Portfolio

4.1. Methodology

For this study, we use data obtained from a database of a large German car insurer that covers several hundred vehicles from the private client segment for a period of 12 months. Figure 8 shows the telematic portfolio structure by age of the policyholders and vehicles in comparison with the total German portfolio structure. It should be noted that only specific customer segments demand for telematic policies. The actual demand for PHYD contracts is higher for both younger and older than average policyholders with new and therefore expensive vehicles.

The data set contains detailed information about vehicle activities, i.e. vehicle position data and several vehicle movement parameters; information about the concrete driver are not directly at hand due to various vehicle users, but special driving behaviour of each vehicle trip might allow conclusion to be drawn to individual person[8]. A data point is recorded regularly by a telematic device which is installed in the insured car. The respective device sends a record after a change of direction or after 2 km driven, respectively, after 2 minutes provided that the measured velocity is smaller than a certain threshold. After the initial editing of the large-scale data material, including data cleansing, there remain 2,466,257 data points and a distance driven of 1,131,573 km considering 91,716 vehicle trips from 110 vehicles[9]. Contractual data, especially tariff criteria that are unnecessary for evaluating and pricing the driving behaviour as well as claims histories are not available, because of privacy concerns[10].

In the approach taken in this paper, the evaluation of driving behaviour included velocity, acceleration and deceleration data analyses to match the stochastic model already mentioned. Every reported data point contains the vehicle velocity measured at the time of record and, in addition, the maximum velocity (average value over 5 seconds) since the last record. As the acceleration sensor measures data at high frequency of 100 Hz, enabling precise data of fast-moving objects, in practice, raw and discrete data are not available in the data set due to the large volume. Instead, the acceleration and deceleration behaviour refer to aggregate data because of the necessary information reduction. This means that the sensor provides grouped frequency distributions with

---

[8] To make things easier to understand, in this section, we use the term "vehicle trip" synonymously to the term "driving profile".

[9] We restrict the used data set to vehicle trips where a clear start and end time can be determined without resorting to movement profiles. Moreover, we correct vehicle trips with unbalanced distributions of acceleration parameters (e.g. with drift in x (left-right) and y (front-back) direction) turning the system of coordinates by the angle determined by the relation of the number of total movement trackings in the respective axis. Such drifts are likely to be caused by either a tilting of the acceleration sensor and/or incorrect calibration of the zero-point hysteresis. In addition, we eliminate vehicle trips with unrealistically high values of velocity or distance between two data points. These trips reflect errors that may be attributed to connection failure with the GPS satellite and depend on the technical realisation rather than the presented method.

[10] It is important to note that, although contractual data is not available for this examination, this is not necessary to evaluate the results. At the current stage of development, the important question this publication aims to answer is whether telematics technologies can be used to obtain differentiable insights into the driving behaviour of a vehicle driver. Indeed, we find that there is differentiation potential in the observed driving profiles, such that an insurer with available contractual data can match their results with future claims requirements concerning the evaluation of the actual risk.
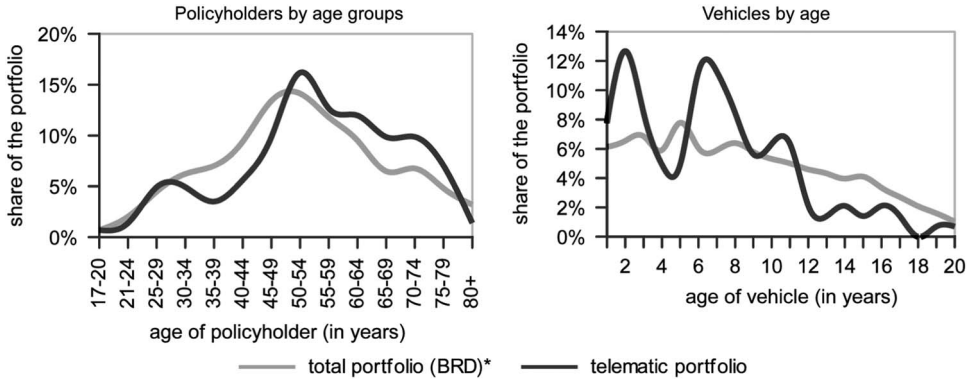
Figure 8. Structure of the telematic data. Note: *Vehicle registrations (passenger cars) in Germany (Kraftfahrt-Bundesamt, 2014).
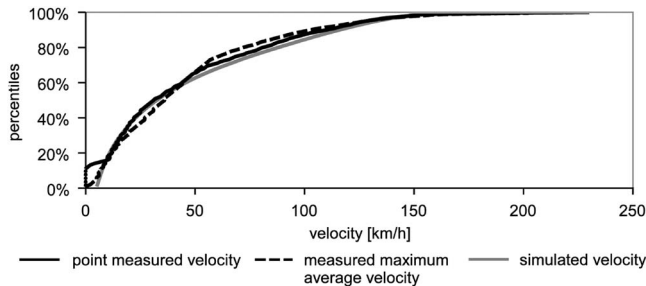

Figure 9. Distribution of the velocity.

various class widths when a data point is recorded; whereby the class width is to be based on pricing criteria[11,12].

We examine two characteristics of individual driving behaviour from the telematic data set, i.e. the velocity and acceleration performance. Figure 9 shows an overview over the distribution of the velocity data. The distribution functions are based on either the point measured velocity (at the time of record) or the measured maximum average velocity (highest average value over 5 seconds between two data points) and differ about equal to a level shift[13]. The maximum average velocity values dominate the point measured velocity for small amplitudes, because the dynamic of vehicles is different for the respective velocity regimes; apart from super sport vehicles, accelerating from standstill to around 50 km/h (an vice versa) goes much faster than from 50 km/h onwards.

---

[11] For example, if stronger braking is considered as risk prevention and therefore should not be punished, then a single interval for all high deceleration values is enough.

[12] The exact intervals for the acceleration values are $(0,0.98]$, $(0.98,2.45]$, $(2.45,4.41]$, $(4.41,6.86]$, $(6.86,12.00)$ and those for the deceleration are $(0,-0.98]$, $(-0.98,-1.96]$, $(-1.96,-2.94]$, $(-2.94,-3.92]$, $(-3.92,-12.00)$.

[13] Note that although this global average difference seems to be of little significance, the two observed measures may differ greatly when comparing individual samples. This can however not be presented due to the data security regulations of the tariff itself.
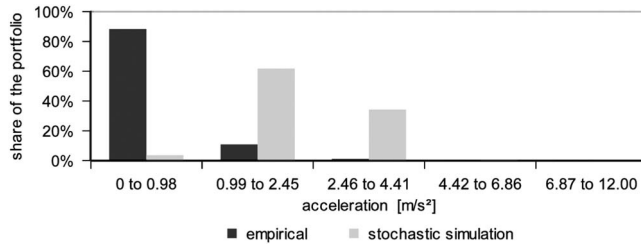
224

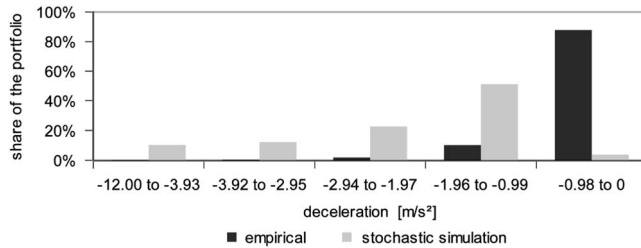**Figure 10.** Frequency distribution of the acceleration.

Figure 11. Frequency distribution of the deceleration.

A 5-second maximum average naturally overestimates velocity compared with random, noisy point measurements. For larger velocities, however, the maximal average is much closer to the real values, because changes in speed behave more inert. In addition, absorbing the fact that the respective distributions are weighted for comparison means that the point measured velocity dominates in the second part of the figure. As the point measured velocity fit better the simulated velocity, the analysis in the following part is limited to the point measured velocity instead of the maximum average velocity[14].

The number of measured acceleration and deceleration values at intervals of unequal length is shown in Figures 10 and 11 in terms of percentages with interval limits. It is obvious that the empirical measurement of the telematic portfolio and the stochastic simulation yield different results. First of all, it should be mentioned that high demands are placed on the acceleration sensor whose technical realisation is still poorly conceived, but is expected to improve with increasing demand for telematic tariffs, respectively, as for the insurer more precise measurements are directly connected to more precise tariffs and will thus spark innovation of technical realisations. The required high sensitivity of the sensor to permit coverage of low acceleration range in the same manner as a high acceleration range induces inaccuracies and therefore many measured values close to 0[15]. On the other hand, the model is designed in such a way that it focusses on real velocity adjustments, which is of real interest for describing acceleration profiles, i.e. the value range without so-called noise. Empirical data is subject to the limits of the realisation and as such permeated by skews

---

[14] However, we carried out the classification for the maximum average velocity, as accident research shows that speed above a prescribed speed limit or too high for the prevailing condition is one of the main causes of accidents (Statistisches Bundesamt, 2015). The results are comparable apart from a shift of the velocity variable and therefore do not require additional elaboration.

[15] These slightly different values close to 0, i.e. deviations in either direction from 0, are referred to as random errors. Random errors like these are considered as noise floor of the acceleration sensor that can be evaluated through a large number of observations, i.e. the ensemble of driver-vehicle units (Bevington, 1969).

induced by real-world fluctuations such as sensor failures or misaligned equipment. Furthermore, the model excludes high-risk ranges such as full braking or acceleration behaviour of sport cars to obtain a genuine result, such that the value range in the model is smaller than that of the real data. It is also noted that the empirical acceleration and deceleration distribution – excluding noise – is quiet unbalanced across intervals. Acceleration and deceleration values in the upper two intervals are rather rare, despite its wide range of up to $-12 \, \text{m/s}^2$. This indicates that the number of intervals of the technical realisation is not sufficiently detailed in order to exploit the full potential of differentiation.

The classifications of the driving profiles are deduced from the characteristics of driving behaviour of the telematic data set allocating them to the driving styles determined in section 3.2 based on the simulation results of the implemented vehicle mobility model. For this purpose, general statistical measures, which describe the distributions of the variables of the driving behaviour for each individual vehicle trip, need to be identified and carefully studied from the data set. The medians and interquartile ranges of the velocity, acceleration and deceleration values form the basis for classification analogous to the clustering process. We derive the velocity parameters from individual point measured velocities weighting by the duration of the measurement between the two relevant data points[16]. In contrast to velocity, where discrete values are available, the medians and interquartile ranges have to be estimated with respect to acceleration and deceleration. Therefore, we make an assumption of a normal distribution for the acceleration behaviour and a gamma distribution for the deceleration, respectively, based on the interval and its next neighbours. In this context, we filter the noise, i.e. replace the empirical frequencies of both intervals adjacent to the origin with those of the simulation results. In addition, we do not consider high-risk ranges, such that we only investigate the range of -8 to $+8 \, \text{m/s}^2$.

The classification is consequently made via distances using the following classification principle:

$$d(x_j, y_i) = \min_i \{d(x_j, y_i) \mid 1 \leq i \leq 6\} \Rightarrow \text{driving profile } j \in \text{driving style } i$$

where $x_j = (x_1, \ldots, x_6)'$ is the measured, respectively, gathered feature vector (i.e. standardised medians and interquartile ranges of the velocity, acceleration and deceleration) of the examined driving profile, $y_i = (y_1, \ldots, y_6)'$ the feature vector of driving style i and $d(x_j, y_i) = \|x_j - y_i\|_2$ the Euclidean distance. Thus, the underlying classification principle states that a driving profile is attributed to the driving style to which the feature vector has smallest distance. To achieve this, the medians and interquartile ranges are standardised so that they can be directly compared with one another[17].

4.2. Results

In addition to the categorisation of the vehicle trips into six driving styles in Figure 12, the corresponding summary statistics are also shown. Approximately 99% of the trips can be attributed

---

[16] Lemaire *et al.* (2015) pointed out that annual mileage is by far the most important factor determining the accident frequency and consequently is a common tariff criterion. In order to avoid interrelations with other conventional criteria, annual mileage does not appear to be an ideal weighting factor, thus we decided to use the measurement duration instead.

[17] The standardisation is performed using the following standard procedure: $z = \frac{x - m_{est}}{s_{est}}$, where m_{est} is the median, as an estimate of the mean, and s_{est} the interquartile range, as an estimate of the standard deviation.

driving style Ia: 83.21%
trips: 76,321 # vehicles: 110 distance: 499,497.81 km data points: 1,676,297

driving style IIa: 15.99%
trips: 14,668 # vehicles: 110 distance: 623,422.98 km data points: 771,621

driving style Ib: 0.04%
trips: 35 # vehicles: 17 distance: 175.66 km data points: 668

driving style IIb: 0.74%
trips: 679 # vehicles: 68 distance: 7,986.94 km data points: 17.156

driving style Ic: 0.01%
trips: 6 # vehicles: 3 distance: 143.80 km data points: 202

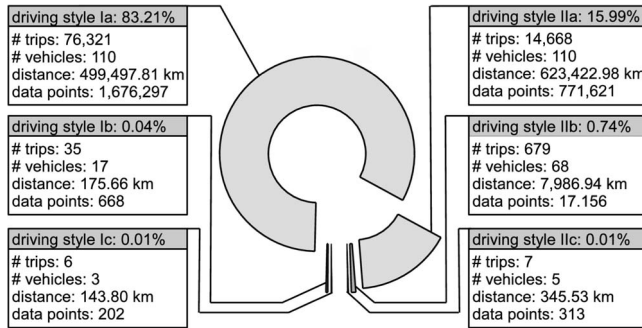| driving style IIc: 0.01% |
| --- |
| # trips: 7
vehicles: 5
distance: 345.53 km
data points: 313 |

Figure 12. Distributions of the vehicle trips and corresponding summary statistics.

to the driving styles characterised by relatively moderate acceleration and deceleration values (see driving styles Ia and IIa). However, these driving styles differ to high degree as far as velocity is concerned, depending on dominant road type. In all, 16% of all trips exhibit higher speed, whereas the velocity is average or moderate in 83% of the trips; it should be noted though that this perspective is skewed due to the fact that total distance is even larger for the driving styles with higher speed. Also, the ratio of data points compared with the distance is much smaller with higher speed. This can be attributed to the technical fact that the present measurement apparatus takes less observations at high speed without much change parameters. Driving style IIb, in which both the acceleration and deceleration behaviour as well as the velocity processes are more speedy, seldom occurs with a meagre share of 1% of the trips. Altogether, it is to be stated that sporty driving styles, with considerable high absolute values across at least the acceleration and deceleration parameters, do only occur occasionally. It is important to bear in mind that the clustered driving styles based on simulation data provide some substantial indications of the range of potential driving behaviour, however, without a statement about the probability of occurrence. This is due to the fact that the primary aim of the simulation is to model all possible driving styles, albeit not normalised over regular portfolio behaviour, as this is not known a priori, i.e. without gauging from empirical data[18]. Only the analysis of real driving profiles evidently demonstrates that adverse driving behaviour with (assumed) higher risk and consequently higher premium is rightly rare. However, the limits of the approach should be pointed out: the vehicle trips must be divided more into several sections to select exactly the rare, adverse driving styles.

From the number of vehicles per driving style in Figure 12, which is in the sum overall styles considerably greater than the total number of vehicles in the whole telematic portfolio, it is clearly evident that trips of the same vehicle are not necessarily solely attributed to one driving style. This means that the driving behaviour of several different trips with the same vehicle may result in differently attributed driving styles. Figure 13 shows that the trips of each vehicle belong to at least two driving styles, the trips of 62.7% of all vehicle even covers between three and six different driving styles. This particularly illustrates that indeed driving style differentiation can be achieved

---

[18] To complete this assessment, it seems opportune to address the question of self-sampling arising from an empirical telematic portfolio: it could very well be that driving styles of customers applying for telematic products in this introductory stage of the technology might be significantly different from a standard portfolio because of a selection effect.
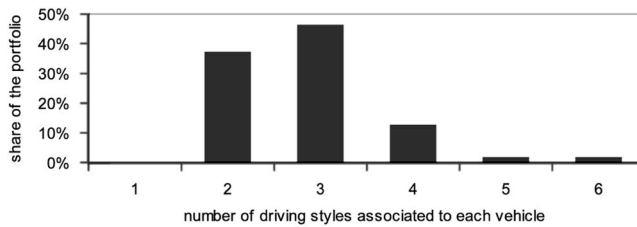
227

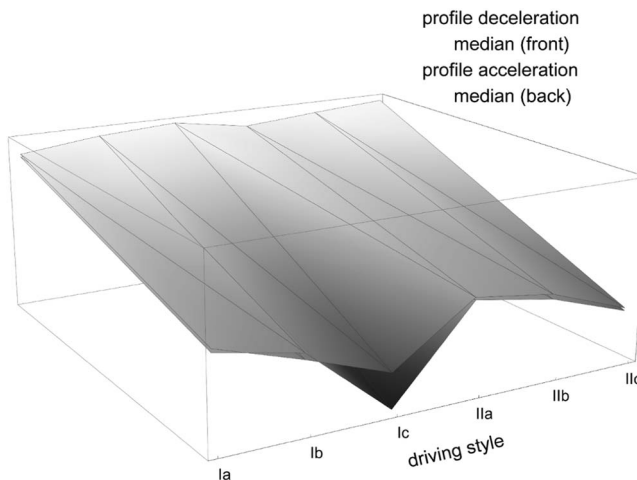**Figure 13.** Survey of the number of driving style matches per vehicle.

Figure 14. Polygon match of acceleration and deceleration medians (upper layer: stochastic simulation; lower layer: empirical).

under empirical real-world conditions. Moreover, the fact that more than half of all vehicles show trips in at least one of the supposedly less defensive driving styles. The reasons for this are switching drivers and routes with different traffic load as well as local conditions and inconsistent driving behaviour in itself. Even though the classification is performed with statistical foundation and still without risk assessment, one initial, general observation for the pricing approach can be derived from these findings, namely that a discount on the premium must be the result of weighted evaluation of the driving styles.

Overall, it turns out that the driving styles based on empirical data correspond well with the driving styles deduced from simulated driving profiles. We clearly observe that driving style acceleration medians as well as deceleration medians of simulation and empirical portfolio match closely for all driving styles (see Figure 14, Appendix B for a more quantitative listing)[19]. The good consistency of the results of the two data sets indicates that the used measures of central tendencies and dispersion, the median and interquartile range, sufficiently describe the distributions of driving behaviour for classification within the present technical realisation. Thus, the classification can be carried out on aggregate data which is important in two ways. First of all, we can reduce the total data volume

---

[19] We refrain from correlating Figure 14 with the velocity medians, as the empirical telematic portfolio covers a special customer segment, thus only realising the simulated behaviour spectrum partially.

228

before transfer and consequently minimise the costs associated with transfer and storage management. Also, this way we consider the national data protection norms of the German insurance industry, with regard to data minimisation and data avoidance.

5. Conclusion

It is not very likely that the high complexity of the actuarial tariff table will allow incorporating variables of driving behaviour into the table so long as an insurer has not yet established a risk experience in respect of telematic data. In spite of that, the question whether sensor data-derived tariff factors are substitutes for conventional tariff factors does not arise. However, subsequent discounts on the premiums dependent on the driving behaviour are a temporary opportunity to incorporate initial insights to telematic data. Premium reduction, which have to be determined on the basis of official accident statistics so long as no risk experience exist, acts as a stimulus to consumers concluding a telematic policy in order to establish a risk experience so that over the long term, variables of driving behaviour can substitute for conventional tariff criteria. In the actual literature, several questions concerning the differentiation and evaluation of driving behaviour are not answered. Thus, this paper provides a basis for optimised short- to medium-term pricing including telematic data, presenting a classification of driving profiles. Selected driving styles, which arise when driving profiles are modelled using a stochastic simulation, are considered. This approach accounts for possible parameter ranges as well as interdependencies and exemplary focusses on velocity, acceleration and deceleration behaviour.

The analysis leads to three main findings for insurance companies, each concerning the automobile product development, generating innovations, the actuarial division, calculating tariffs and management, undertaking pricing decision. First, a systematic method is introduced for generating specific driving styles using a vehicle mobility model. By connecting the model results to empirical insurance telematic data, we submit verification that driving profiles could be attributed to the identified driving styles. These styles are plausible and comprehensive because, as shown, the simulation model reflects the wide range of individual characteristics of driving behaviour.

Second, the analysis of the empirical telematic portfolio illustrates that the measurement and evaluation of driving behaviour should focus on essential information and furthermore exploits the full potential of differentiation. In spite of the fact that present telematic data is still subject to unfinished detection procedures, it is clearly established how partially inconclusive and possibly inaccurate telematic data can still be used to gain insights as long as the relevant features, such as the distribution's main parameters, can be inferred.

Third, the introduced method for analysing specific driving behaviour also shows how the classification of driving styles can be carried out on aggregate data, i.e. measures of central tendencies as well as dispersion, using only a few variables including their interdependencies. This specific result appears of particular importance in the context of national privacy protection, in particular, with regard to the prime principles of data minimisation and data avoidance be met by any telematic tariff as far as personal data is concerned.

With this in mind, some methodological and conceptual limitations should be taken into account in the interpretation of our results. The driving styles shown here should only serve as an example, encouraging the extensive use of statistical methods as a basis for analysis of driving

229

behaviour. Although we selected detailed model data in our analysis, other empirical portfolio-specific parameter sets should be applied to complete the picture. The verification of the model results should also be replicated with real-world telematic data sets of different composition of portfolio and data granularity in order to validate our findings. Moreover, it should be useful to extend the analysis by further data, in particular, variables of driving situation such as driving time or road conditions, which are far easier to analyse and can be approximated as conventional tariff criteria.

The results furthermore emphasise that the telematic risk assessment approach could have the potential for creative destruction in the matter of pricing innovation. The objective data on the individual usage of a vehicle provide new information in actuarial decision making. Thus, this approach could prospectively create a new insurance pricing structure, more differentiated and individualised than before. Insurers who use it can more adequately price the actual risk and gain competitive advantage.

References

Becke, M. (2005). Unfallrekonstruktion – Geschwindigkeitsermittlung vor Bremsbeginn, Verkehrs-rechtsReport, 1/2005, 20–24.

Bettstetter, C., Hartenstein, H. & Pérez-Costa, X. (2004). Stochastic properties of the random waypoint mobility model. *Wireless Networks*, **10**(5), 555–567.

Bevington, P. (1969). *Data Reduction and Error Analysis for the Physical Sciences*. McGraw-Hill, New York.

DAV-Arbeitsgruppe Tarifierung (2007). *Berufspflichten des Aktuars bei der Tarifgestaltung in der Schadenversicherung*. Deutsche Aktuarvereinigung e.V., Köln.

DAV-Arbeitsgruppe Tarifierungsmethodik (2015). *Aktuarielle Methoden der Tarifgestaltung in der Schaden-/Unfallversicherung*. Verlag Versicherungswirtschaft GmbH, Karlsruhe.

Domeyer, A. (2005). *Spezielle Versicherungen des privaten, des gewerblichen und des Industriegeschäfts*. Verlag Versicherungswirtschaft GmbH, Karlsruhe.

Insurance Europe (2014). Statistics No. 50 – European insurance in figures. Available online at the address http://www.insuranceeurope.eu/uploads/Modules/Publications/statisticsno50european insuranceinfigures.pdf [accessed Jul-2015].

Gerpott, T.J. & Berg, S. (2012). Pay-As-You-Drive Angebote von Erstversicherern für Privatkunden: Eine betriebswirtschaftliche Analyse. *ZVersWiss*, **101**, 3–29.

Gesamtverband der Deutschen Versicherungswirtschaft e.V. (2012). *Jahrbuch 2012 – Die deutsche Versicherungswirtschaft*. GDV, Berlin.

Hartmann, M. & Nützenadel, C. (2015). Trends in der Technologie sowie Erkenntnisse des Behavioural Pricings vereinbaren. *I·VW Management Information – St. Galler Trendmonitor für Risiko- und Finanzmärkte*, **37**(1), 3–9.

Heep-Altiner, M. & Klemmstein, M. (2001). *Versicherungsmathematische Anwendungen in der Praxis: Mit Schwerpunkt Kraftfahrt und Allgemeine Haftpflicht*. Verlag Versicherungswirtschaft GmbH, Karlsruhe.

Johnson, D.B. & Malz, D.A. (1996). Dynamic source routing in ad hoc wireless networks. In T. Imielinskis & H.F. Korth (Eds.), *Mobile Computing* (pp. 153–181). Kluwer Academic Publishers, Dordrecht.

Jun, J., Ogle, J. & Guensler, R. (2007). Relationships between crash involvement and temporal spatial driving behaviour activity patterns: use of data for vehicles with global positioning systems. *Transportation Research Record*, **2019**, 246–255.

Jun, J., Ogle, J. & Guensler, R. (2011). Differences in observed speed patterns between crash-involved and crash-not-involved drivers: application of in-vehicle monitoring technology. *Transportation Research Part C*, **19**(4), 569–578.

Karapiperis, D., Obersteadt, A., Brandenburg, A., Castagna, S., Birnbaum, B., Greenberg, A. & Harbage, R. (2015). Usage-based insurance and vehicle telematics: insurance market and regulatory implications. *CIPR Study Series*, **2015-1**, 1–79.

Kraftfahrt-Bundesamt (2014). *Fahrzeugzulassungen (FZ) – Bestand an Kraftfahrzeugen und Kraftfahrzeuganhängern*. Kraftfahrt-Bundesamt, Flensburg.

Kremslehner, D. & Muermann, A. (2013). Asymmetric information in automobile insurance: evidence from driving behavior, Working Paper, Vienna University of Economics and Business, Wien.

Laas, D., Hartmann, M., Nützenadel, C., Schmeiser, H. & Wagner, J. (2014). *Pricing-Strategien in der KFZ-Versicherung*. Solution Providers Schweiz AG, Dübendorf/Zürich.

Lance, G.N. & Williams, W.T. (1967). A general theory of classificatory sorting strategies. *Computer Journal*, **9**(4), 373–380.

Lange, F. (2006). Unfallrekonstruktion – Anfahrbeschleunigungen, VerkehrsrechtsReport, 10/2006, 377–382.

Lemaire, J., Park, S.C. & Wang, K. (2015). The impact of covariates on a bonus-malus system: an application of Taylor's model. *European Actuarial Journal*, **5**(1), 1–10.

Mangiameli, P., Chen, S.K. & West, D. (1996). A comparison of SOM neural network and hierarchical clustering methods. *European Journal of Operational Research*, **93**(2), 402–417.

Paefgen, J., Staake, T. & Thiesse, F. (2013). Evaluation and aggregation of pay-as-you-drive insurance rate factors: a classification analysis approach. *Decision Support Systems*, **56**, 192–201.

Schüller, H. (2010). *Modelle zur Beschreibung des Geschwindigkeitsverhaltens auf Stadtstraßen und dessen Auswirkungen auf die Verkehrssicherheit auf Grundlage der Straßengestaltung*. Available online at the address http://nbn-resolving.de/urn:nbn:de:bsz:14-qucosa-61493 [accessed Jan-2015].

Stadler, M. (2008). *Die Kfz-Versicherung*. Verlag Versicherungswirtschaft GmbH, Karlsruhe.

Statistisches Bundesamt (2015). *Verkehrsunfälle 2014 – Fachserie 8 Reihe 7*. Statistisches Bundesamt, Wiesbaden.

Stiftung Warentest (2014). Telematik-Tarife in der Autoversicherung: Spion fährt mit. Available online at the address https://www.test.de/Telematik-Tarife-in-der-Autoversicherung-Spion-faehrt-mit-4692642-0/ [accessed Jul-2015].

Toledo, T., Musicant, O. & Lotan, T. (2008). In-vehicle data recorders for monitoring and feedback on driver's behavior. *Transportation Research Part C*, **16**(3), 320–331.

Ward, J.H (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, **58**(301), 236–244.

Weidner, W. & Weidner, R. (2014). Identifikation neuer Ansätze zur individuellen Kfz-Tarifierung. *ZversWiss*, **103**(2), 167–193.

Weidner, W., Weidner, R. & Transchel, F.W.G. (2015). Implementierung der Pkw-Telematik in die Kfz-Tarifierung. *Zeitschrift für Verkehrswissenschaft*, **85**(2), 91–121.

Appendix A: Vehicle Mobility Model

Before we can introduce the RWP model, we need to present a set of abstract axioms and rules of movement necessary to describe the vehicle behaviour as close to realistic traffic conditions as possible.

Axioms

- Time-discrete system. A driving profile is given by a sequence of movement vectors in time-discrete succession – in general, GPS-based position measurements are also of this format.
- Microscopic model. The movement patterns of individual vehicles are based on the specific driving behaviour of the vehicle driver as well as the technical properties of the vehicle itself. The granularity of the description thus must be chosen such that distinctions can be drawn on the level of individual driver–vehicle units.
- Stochastic process. The driver–vehicle units as well as the environmental conditions (weather, traffic jams, resting times) are of stochastic nature.

Rules of movement

- Maximum velocity. The vehicle fits the individually desired velocity to the maximally allowed velocity. Limiting factors such as traffic density and road conditions and road type are included in the determination of the recommended speed, i.e. the velocity considered optimal under recognition of all external conditions. This is of particular importance in Germany, where special conditions apply: on the Autobahn highways, there is not necessarily a speed limit, but rather a speed recommendation (Richtgeschwindigkeit) of 130 km/h. This means that in principle, the driver can exceed this velocity. The model accounts for this effect by weighting the deviation from the maximum speed less strictly if it is a recommendation instead of a limit. The model thus captures driving behaviour in countries with and without strict speed limits.
- Acceleration/deceleration behaviour. Vehicles do not start and stop without transition. We decompose the route between waypoints into segments of constant acceleration, so that the vehicle will accelerate at an individual rate at the beginning of the route until it moves with the segment-specific maximum velocity and decelerates at the end, each process conditioned on the type of waypoint in question.
- Stop times. At traffic obstacles like traffic lights, zebra crossings or traffic jams, the vehicle waits for a specific, stochastic waiting time.

Simulation of movement

In essence, the RWP model describes the movement of a driver-vehicle unit in a flat, bounded rectangle $[0, a] \times [0, b] \subset \mathbb{R}^2$ (see Bettstetter *et al.*, 2004). The spatial distribution of all possible waypoints $\mathbf{P} = (P_x, P_y)$ is then uniformly distributed. For each simulated driver-vehicle unit (i.e. a so-called RWP node) j, $j = 1, 2, 3, \ldots, m$, a RWP $\mathbf{P}_0^{(j)}$ is chosen at the start of the simulation, then a sequence of waypoints $\mathbf{P}_i^{(j)}$, $i \in \mathbb{N}$, is selected from the restricted area. The movement of each node j consequently can be given as a time-discrete process:

$$\left\{ P_i^{(j)} \right\}_{i \in \mathbb{N}_0} = P_0^{(j)}, P_1^{(j)}, P_2^{(j)}, \ldots$$

The RWP node movement between waypoints is modelled by a segment-specific nominal velocity, sampled from the uniform distribution over $[v_{\min}, v_{\max}]$, where $v_{\min} > 0$ and $v_{\max} < \infty$, and

232

a component $v(r_{i,k})$ unique to the node. Thus, the speed at point $r_{i,k}$, $k = 0, 1, 2, \ldots, n_i$, is given by

$$v_i(r_{i,k}) = v_i \cdot v(r_{i,k})$$

Essentially, these nominal parameters are given by speed limits, traffic density and environmental conditions, whereas the personal part $v(r_{i,k})$ depends on the specific types of driver and vehicle. In order to make this mechanism as realistic as possible, before each step a waiting time $T_{p,i}$ is sampled from the normal distribution given by $t_{p,i} \in [0, t_{p,\max}]$ with $t_{p,\max} < \infty$. In total, the movement period of a node is given by the vector $(\mathbf{p}_{i-1}, \mathbf{p}_i, v_i(r_{i,k}), t_{p,i})$.

Driving profiles

To create specific driving profiles from the sample data we represent the stochastic process of distances between succeeding waypoints by (see Bettstetter *et al.*, 2004)

$$\{L_i^{(j)}\}_{i \in \mathbb{N}} = L_1^{(j)}, L_2^{(j)}, L_3^{(j)}, \ldots$$

with $L_i^{(j)} = \left\| P_i^{(j)} - P_{i-1}^{(j)} \right\|_2 = \left(\left(P_{x,i}^{(j)} - P_{x_{i-1}}^{(j)} \right)^2 + \left(P_{y,i}^{(j)} - P_{y,i-1}^{(j)} \right)^2 \right)^{1/2}$. Keep in mind that the velocity between waypoints is in general not constant, such that the measurement frequency crucially determines the quality of the velocity profile. In order to obtain segments $\{s_{i,k}^{(j)}, k = 0, 1, 2, \ldots, n_i\}$, we decompose every instance $l_i^{(j)}$, depending on the chosen measurement frequency. From equal-time-distance tupels $(s_{i,k}, v_i(r_{i,k}))$, $i \in \mathbb{N}_0$, $k = 0, 1, 2, \ldots, n_i$, we can then refine a velocity-position profile, where for waiting delays $t_{p,i}$ we need to add tupels of the form $(0, 0)$. In practice, any details of the movement are captured precisely if the measurement frequency is high enough, because for lower frequencies, velocity changes happening below the time resolution of the measurement cannot be resolved. However, in order to recreate an approximate velocity profile, measurements can be used as (linear) interpolation values. Consequently, for the average velocity between measurements $\mathbf{P}_{i-1}$ and $\mathbf{P}_i$ we find that

$$\bar{v}_i = \frac{\sum_{k=1}^{n_i} s_{i,k}}{\sum_{k=1}^{n_i} t_{i,k}}, \ i \in \mathbb{N}$$

where $t_{i,k}$ is the time for the segment $s_{i,k}$. Now if the measurement frequency is good enough, we can simplify the expression by taking $t_{i,k} = 1/t_A$

$$\bar{v}_i = \frac{1}{n_i} * t_A * \sum_{k=1}^{n_i} s_{i,k}, \ i \in \mathbb{N}$$

The acceleration profile naturally emerges from the first-order derivative of the velocity profile as $a_i(r_{i,k}) = \dot{v}_i(t, r_{i,k})$, $i \in \mathbb{N}_0$, $k = 0, 1, 2, \ldots, n_i$. The average acceleration is then given by

$$\bar{a}_i = \frac{v_{i,m_i} - v_{i,l_i}}{\sum_{k=l_i}^{m_i} t_{i,k}}, \ i \in \mathbb{N}$$

where $v_{i,m_i-1} < v_{i,m_i} \geq v_{i,m_i+1}$ and $v_{i,l_i-1} \geq v_{i,l_i} < v_{i,l_i+1}$ for $a > 0$ or $v_{i,m_i-1} > v_{i,m_i} \leq v_{i,m_i+1}$ and $v_{i,l_i-1} \leq v_{i,l_i} > v_{i,l_i+1}$ for $a < 0$, respectively.

Table A.1 Parameters of the vehicle mobility model.

Variational parameter	Symbol	Variational method	Parameterisation
Waypoint in the area $A = [0,a] \times [0,b] \subset \mathbb{R}^2$	$p = (p_x, p_y)$	$P = (P_x, P_y) \sim U(A)$, i.e. $P_x \sim U[0,a]$ $P_y \sim U[0,b]$	$a = 2\,\text{km}$ $b = 5\,\text{km}$
Pause time at waypoint	t_p	$T_p \sim N(\mu, \sigma^2)$ $t_p \in [0, t_{p,\max}]$ with $t_{p,\max} < \infty$	$\mu = 40\,\text{s}$ $\sigma = 20\,\text{s}$ $t_{p,\max} = 120\,\text{s}$
Segment-specific nominal velocity	v	$V \sim U[v_{\min}, v_{\max}]$ with $v_{\min} > 0$ and $v_{\max} < \infty$	$v_{\min} = 5\,\text{km/h}$ $v_{\max} = 200\,\text{km/h}$
Scale and limit for the velocity of a driver-vehicle unit	$sf_{v^{(j)}}$ $v^{(j)}_{\max}$	$SF_{v^{(j)}} \sim N(\mu_{\text{sf}}, \sigma^2_{\text{sf}})$ $V^{(j)}_{\max} \sim N(\mu_v, \sigma^2_v)$ $v^{(j)}_{\max} > v_{\min,\max}$, with $v_{\min,\max} < \infty$	$M_{\text{sf}} = 1.02$ $\sigma_{\text{sf}} = 0.05$ $\mu_v = 140\,\text{km/h}$ $\sigma_v = 25\,\text{km/h}$ $v_{\min,\max} = 70\,\text{km/h}$
Segment-specific acceleration of a driver-vehicle unit	a	$A^{(j)} \sim N(\mu^{(j)}, (\sigma^{(j)})^2)$ $a^{(j)} \in [a_{\min}, a_{\max}]$ with $a_{\min} > 0$ and $a_{\max} < \infty$	$M^{(j)} \sim N(2.5, 0.4^2)$ with $\mu^{(j)} \in [a_{\min}, a_{\max}]$ and $\rho(\mu^{(j)}, sf_{v^{(j)}}) = 0.90$ $\sigma^{(j)} = \min\left\{\frac{\mu^{(j)} - a_{\min}}{2.576}; \frac{a_{\max} - \mu^{(j)}}{2.576}\right\}\text{m/s}^2$ $a_{\min} = 1.0\,\text{m/s}^2$ $a_{\max} = 6.0\,\text{m/s}^2$
Segment-specific deceleration of a driver-vehicle unit	d	$D^{(j)} \sim \Gamma(n^{(j)}, 1/\lambda^{(j)})$ with $n^{(j)} = \mu^{(j)}\lambda^{(j)} \geq 1$ and $\lambda^{(j)} = (\sigma^{(j)}/n^{(j)})^2 > 0$ $d^{(j)} \in [d_{\min}, d_{\max}]$ with $d_{\min} > 0$ and $d_{\max} < \infty$	$M^{(j)} \sim N(2.5, 1.8^2)$ with $\mu^{(j)} \in [d_{\min}, d_{\max}]$ and $\rho(\mu^{(j)}, sf_{v^{(j)}}) = 0.90$ $\sigma^{(j)} = \min\left\{\frac{\mu^{(j)} - d_{\min}}{2.576}; \frac{d_{\max} - \mu^{(j)}}{2.576}\right\}\text{m/s}^2$ $d_{\min} = 1.0\,\text{m/s}^2$ $d_{\max} = 8.0\,\text{m/s}^2$

Appendix B: Quantitative Characterisation of the Driving Profiles

Table A.2 Frequency distribution of acceleration process.

Driving style	Database	Interval boundary (m/s^2)				
		0 to 0.98	0.99 to 2.45	2.46 to 4.41	4.42 to 6.86	6.87 to 8.00
Total	Simulation	3.54%	61.78%	34.15%	0.53%	0.00%
	Empirical	3.65%	88.12%	8.04%	0.17%	0.01%
I						
a	Simulation	3.83%	78.47%	17.69%	0.01%	0.00%
	Empirical	3.70%	89.15%	7.03%	0.12%	0.01%
b	Simulation	3.59%	50.91%	45.12%	0.38%	0.00%
	Empirical	4.73%	75.21%	14.39%	2.79%	2.88%
c	Simulation	3.62%	41.14%	52.89%	2.35%	0.00%
	Empirical	9.35%	85.86%	3.63%	0.34%	0.82%
II						
a	Simulation	3.53%	75.24%	21.19%	0.04%	0.00%
	Empirical	3.50%	85.58%	10.62%	0.28%	0.01%
b	Simulation	3.34%	58.14%	38.29%	0.23%	0.00%
	Empirical	4.82%	73.39%	19.70%	1.46%	0.63%
c	Simulation	3.25%	43.21%	52.03%	1.52%	0.00%
	Empirical	3.56%	87.03%	8.85%	0.56%	0.00%

Table A.3 Frequency distribution of deceleration process.

Driving style	Database	Interval boundary (m/s^2)				
		−8.00 to −3.93	−3.92 to −2.95	−2.94 to −1.97	−1.96 to −0.99	−0.98 to 0
Total	Simulation	10.24%	12.21%	22.57%	51.20%	3.78%
	Empirical	0.23%	1.93%	14.51%	79.88%	3.46%
I						
a	Simulation	0.98%	4.84%	19.46%	70.54%	4.17%
	Empirical	0.11%	1.75%	14.59%	80.15%	3.41%
b	Simulation	16.07%	24.48%	29.39%	26.30%	3.76%
	Empirical	17.32%	11.79%	23.05%	45.14%	2.70%
c	Simulation	51.76%	23.68%	13.90%	7.05%	3.60%
	Empirical	92.49%	3.31%	0.73%	1.62%	1.85%
II						
a	Simulation	0.98%	5.04%	20.63%	69.52%	3.84%
	Empirical	0.13%	1.51%	12.24%	82.51%	3.61%
b	Simulation	6.09%	14.54%	29.43%	46.42%	3.51%
	Empirical	2.74%	15.94%	45.90%	32.24%	3.19%
c	Simulation	34.63%	24.57%	22.13%	15.38%	3.30%
	Empirical	52.24%	2.29%	2.61%	39.32%	3.55%

Table A.4 Frequency distribution of velocity process.

Driving style	Database	Mean	Median	IQR	Minimum	Maximum
			Measure (km/h)			
Total	Simulation	49.49	32.48	62.26	0.00	222.73
	Empirical	45.09	31.00	55.00	0.00	271.00
I						
a	Simulation	34.41	18.49	37.02	0.00	202.17
	Empirical	29.29	22.00	32.00	0.00	221.00
b	Simulation	43.65	26.16	54.14	0.00	205.52
	Empirical	29.46	20.00	36.00	0.00	163.00
c	Simulation	46.86	29.97	59.20	0.00	222.73
	Empirical	62.39	55.00	94.00	0.00	182.00
II						
a	Simulation	49.99	36.95	62.03	0.00	201.64
	Empirical	74.87	77.00	82.00	0.00	271.00
b	Simulation	68.57	64.63	67.22	0.00	206.30
	Empirical	49.91	41.00	56.00	0.00	253.00
c	Simulation	66.40	58.65	79.83	0.00	205.82
	Empirical	86.68	102.00	81.00	0.00	174.00

Note: IQR, interquartile range.

236