Vol. 2, No. 2, pp.: 33-54, 2024

Synthetic Telematics Insurance Data for Transparent Rate-Making: A Benchmark Framework for Claim Frequency, Severity, and Territory Clustering

Javier Martínez[1], Carmen Ruiz[2,*], Andrés López[1]

[1] Department of Mathematical Economics, Statistics and Econometrics, Universidad de Castilla-La Mancha, Albacete 02071, Spain

[2] Department of Financial Economics and Accounting, Universidad Rey Juan Carlos, Madrid 28032, Spain

* *carmen.ruiz@urjc.es*

Article Information	
Received	19 January, 2024
Accepted	14 May, 2024
DOI	https://doi.org/10.63646/datamind.2024.020204

Abstract

Telematics-enabled usage-based insurance (UBI) is reshaping how motor insurers measure and price individual driver risk, yet the move from traditional rating tables to data-rich pricing engines also creates new regulatory pressures around transparency, fairness, and territorial differentiation. This study proposes a benchmark modelling framework that combines penalised generalised linear models, generalised additive models with spline-based smooth functions, gradient-boosted decision trees, and an interpretable low-dimensional clustering procedure for territory design. Using a publicly available synthetic telematics portfolio that mirrors the statistical properties of operational UBI books, we estimate claim frequency and claim severity separately, examine variable interactions between annual mileage and traditional rating factors, and rank predictor importance using SHAP values from an XGBoost model. We then apply a regularised k-means procedure to cluster territories on policy-relevant risk indicators, with the optimal number of clusters chosen by a penalised MAD/MSE criterion. The results show that generalised additive models deliver substantially better group-level calibration than linear pricing models for young drivers and for high-mileage segments, while interpretable clustering on two-dimensional risk maps produces transparent territory structures that regulators can inspect, defend, and compare across rate filings. The framework supports actuarial fairness review, exposure-based rate justification, and reproducible UBI rate-making in markets where model explainability is a regulatory requirement rather than an optional feature.

Keywords: ***telematics insurance; usage-based insurance; generalised additive models; territory clustering; SHAP; interpretable machine learning; rate regulation***

1. Introduction

Automobile insurance rate-making sits at the intersection of statistical modelling, regulation, and consumer protection. For decades, premiums were set using a small number of rating variables—age, vehicle characteristics, prior claim history, and territory—that were combined through generalised linear models (GLMs) into additive class plans. The arrival of vehicle telematics has changed that picture. On-board devices and smartphone applications now record kilometres driven, time-of-day usage, harsh-braking events, and contextual factors that were previously unobservable to insurers (Verbelen et al., 2018; Ayuso et al., 2019). The shift from group-based rating to behaviour-aware pricing under usage-based insurance (UBI) promises sharper risk segmentation, exposure-aligned premiums, and stronger incentives for safer driving (Eling and Kraft, 2020).

These data, however, also create new modelling and regulatory tensions. Telematics variables are often nonlinear and interact with traditional rating factors in complex ways that a strictly linear pricing structure cannot accommodate without loss of accuracy (Boucher et al., 2017; Gao et al., 2019). At the same time, supervisors increasingly expect insurers to explain how each variable enters the premium, why certain segments are charged more, and whether the resulting rate-making complies with non-discrimination and actuarial fairness principles (Lindholm et al., 2022; Richman, 2021a). Highly flexible black-box models, including ensembles of decision trees and deep neural networks, can deliver strong predictive performance but they do not produce the kind of transparent, auditable rate indications that rate filings traditionally require (Maillart, 2021). The challenge is therefore not whether to embrace richer data and richer models, but how to embed them in a workflow that remains explainable to regulators, brokers, and policyholders.

A second, related challenge concerns the empirical infrastructure used for method development. Operational UBI portfolios are commercially sensitive, hard to share, and rarely standardised in the form required by independent academic research. This restricts replication, slows the diffusion of methodological improvements, and makes it difficult for regulators to perform comparative evaluation across insurers. The recent release of a synthetic UBI dataset by So et al. (2021) helps close part of this gap by reproducing the joint distribution and dependence structure of a real telematics portfolio without exposing individual records. Synthetic data of this kind cannot replace internal actuarial files, but they provide a controlled environment in which alternative modelling strategies can be benchmarked transparently.

The present paper builds on this opportunity by developing a complete benchmark framework for transparent UBI rate-making and applying it to the synthetic portfolio of So et al. (2021). The framework integrates four building blocks. First, we use penalised GLMs to obtain a parsimonious baseline rating structure, with stepwise variable selection driven jointly by the Akaike and Bayesian information criteria. Second, we extend the baseline to a generalised additive model (GAM) in which numerical predictors enter through spline-based smooth functions, allowing nonlinear and interaction effects to be visualised and inspected. Third, we estimate a gradient-boosted decision tree (XGBoost) on the same predictors and use SHAP values to quantify global and local feature importance, providing a complementary view of model behaviour that does not depend on the linearity of the GLM specification. Fourth, we develop an interpretable low-dimensional clustering procedure for territory design, in which the number of clusters is selected through a penalised mean-absolute-deviation criterion that explicitly balances goodness of fit against parsimony.

The contribution of the paper is threefold. From a methodological standpoint, the framework demonstrates that flexible models such as GAM and XGBoost can be deployed alongside, rather than in place of, a regulator-friendly GLM baseline. From an empirical standpoint, we show that the GAM corrects systematic GLM under-fitting in the youngest age band and reveals interactions between annual mileage and other rating factors that linear specifications miss. From a regulatory standpoint, our interpretable territory clustering produces two-dimensional risk maps that supervisors can read directly. The procedure stabilises the number of territorial classes around a small set of policy-relevant indicators (annual mileage, credit score, years of no claims) and remains robust to the choice of the parsimony penalty. Taken together, these results offer a reproducible pipeline that supports the kind of explanation, documentation, and triangulation that modern UBI rate filings increasingly demand.

The remainder of the article is organised as follows. Section 2 reviews the literature on telematics ratemaking, additive modelling, and interpretable machine learning in actuarial work. Section 3 describes the synthetic dataset and the four-stage benchmark pipeline. Section 4 reports the empirical results, including comparative performance, smooth functions, territory clusters, and SHAP-based importance. Section 5 discusses implications for rate regulation. Section 6 outlines limitations and avenues for future work, and Section 7 concludes.

2. Background and Related Work

The actuarial literature on motor insurance pricing has traditionally rested on GLMs because they combine a transparent linear predictor with a flexible choice of error distribution from the exponential family. The Poisson model for claim counts and the Gamma model for claim amounts remain canonical building blocks of frequency-severity rate plans, and their parameters can be interpreted directly as multiplicative relativities applied to a base premium (Denuit et al., 2019; Wüthrich and Merz, 2023). This interpretability is one of the reasons GLMs have proved so durable in regulated markets: a rate filing in which each rating variable carries a published relativity is easier to defend than a filing in which premium is generated by a black-box function.

The arrival of telematics data has tested the limits of this paradigm. Early empirical work showed that annual mileage and time-of-day exposure improve claim-frequency prediction beyond traditional rating variables (Ayuso et al., 2019; Lemaire, Park and Wang as cited by Boucher et al., 2017). Subsequent studies confirmed that the marginal predictive value of telematics features can be substantial, especially in segments where conventional variables convey little information about driving behaviour, such as young drivers or recently licensed policyholders (Guillen et al., 2019; Pérez-Marín et al., 2019). Yet the same studies note that telematics variables tend to enter the true risk surface in a strongly nonlinear way, with thresholds, saturation, and curvature that violate the linear-link assumption of standard GLMs (Boucher et al., 2017).

A first response to that limitation has been to extend the GLM toward GAMs, where each numerical predictor is replaced by a smooth function estimated via penalised regression splines. Boucher et al. (2017) used GAMs to disentangle duration and distance effects in telematics motor insurance, and Gao et al. (2019) showed that smooth representations of telematics car-driving features produce sharper claim-frequency forecasts than the corresponding GLM specifications. GAMs preserve much of the additivity that supports rate filing because each smooth term can be plotted, inspected, and audited, while still accommodating the nonlinear patterns that telematics data exhibit. This combination of flexibility and visual interpretability is central to our framework.

A second response has been to apply machine-learning models directly to actuarial problems. Pesantez-Narvaez et al. (2019) compared XGBoost with logistic regression for motor-insurance claim prediction with telematics covariates and found a measurable predictive gain for the tree ensemble. Henckaerts et al. (2021) reported similar conclusions for gradient-boosted trees applied to a real Belgian portfolio. Yang, Qian and Zou (2018) developed a boosted Tweedie compound Poisson framework that incorporates the zero inflation characteristic of insurance losses. Maillart (2021) and Gao, Wang and Wüthrich (2022) emphasised the importance of variable importance and partial-dependence diagnostics for keeping such models suitable for actuarial review. These contributions collectively demonstrate that machine learning is now an established tool in motor pricing, but they also highlight the interpretability gap that needs to be bridged before such models can be embedded in regulated rate structures.

Interpretable machine learning provides part of the bridge. SHapley Additive exPlanations (SHAP), grounded in cooperative game theory, decompose an individual prediction into additive feature contributions that sum to the model output (Lundberg et al., 2020). When applied to tree ensembles such as XGBoost (Chen and Guestrin, 2016), SHAP values yield both global rankings of variable importance and local explanations for specific policyholders. Richman (2021a, 2021b) reviewed how such methods can support actuarial fairness review by uncovering disparate impact through proxy variables, while Lindholm et al. (2022) formalised a discrimination-free pricing framework that can be combined with flexible models to satisfy regulatory non-discrimination requirements.

Territorial rating is another long-standing concern in motor insurance. Conventional approaches assign each postal code or census district to a rating class through credibility-weighted averages of historical loss experience. With telematics data, territory effects can be re-examined because behaviour-related variables already capture part of what was previously attributed to geography (Ma et al., 2018; Baecke and Bocca, 2017). Henckaerts et al. (2018) introduced a data-driven binning strategy for constructing rating classes that aims to reduce overfitting and improve credibility. Our work extends this line by operating in a low-dimensional feature space that pairs each territory's average loss with a single behaviour-related risk indicator, producing two-dimensional cluster maps that regulators can read and challenge directly.

Cluster-analytic methods occupy a distinct niche in this literature. Most telematics-driven applications of clustering target driver-behaviour segmentation in a high-dimensional feature space: trip-level speed profiles, harsh-event sequences, time-of-day fingerprints, or accelerometer signatures are projected into latent representations and clustered to recover behavioural archetypes for downstream pricing or marketing use (Wuthrich, 2017; Gao and Wuthrich, 2018, as discussed in Verbelen et al., 2018). While effective for behaviour profiling, this approach is not designed to deliver inspectable territorial rate structures. The clusters that emerge are typically opaque to regulatory review and cannot be plotted in a low-dimensional space that supervisors can interrogate. Our use of clustering departs from this pattern by operating on pre-aggregated, policy-relevant indicators at the territory level rather than on raw trip-level signals, and by deliberately constraining the input dimension so that the partition can be visualised and justified.

Fairness, non-discrimination, and explainability have moved from peripheral concerns to central requirements in actuarial machine learning. Frees, Lee and Yang (2016) and Charpentier, Flachaire and Ly (2018) framed the early discussion of machine learning in insurance around the tension between predictive accuracy and the interpretability that rate filings require. More recent contributions formalise specific operational requirements: Lindholm et al. (2022) propose a discrimination-free pricing construction in which the premium is orthogonalised against protected attributes; Richman (2021a, 2021b) reviews the state of the art across the actuarial value chain; and Quan and Valdez (2018) demonstrate that multivariate decision-tree methods can preserve a degree of interpretability through prototype paths. These developments converge on the conclusion that explainability is not a diagnostic afterthought added to a black-box model, but a structural property that must be designed into the modelling pipeline from the outset. The framework proposed in this paper inherits that design principle.

Finally, the broader move toward data-driven decision systems in financial services has reinforced the need for transparent pipelines that combine predictive performance, interpretability, and reproducibility (Kou and Lu, 2025; Lu, 2019; Zhang and Lu, 2021). Reproducibility is particularly important in rate-making because rate filings are public documents whose empirical claims must be auditable by supervisors, intervenors, and academic peer reviewers. Synthetic data infrastructures such as the one of So et al. (2021) make this kind of reproducibility realistic, and they invite the development of shared benchmark frameworks against which competing modelling strategies can be compared on common ground. Our benchmark inherits this orientation by treating modelling, importance analysis, and territory clustering as parts of one auditable workflow that can be re-executed end to end on the public synthetic portfolio.

3. Data and Methods

This section describes the synthetic UBI portfolio, the modelling strategy that combines GLM, GAM, and XGBoost, the variable-importance procedure based on SHAP, and the interpretable territory-clustering procedure with penalised cluster selection. Figure 1 summarises the four building blocks of the analytical pipeline.

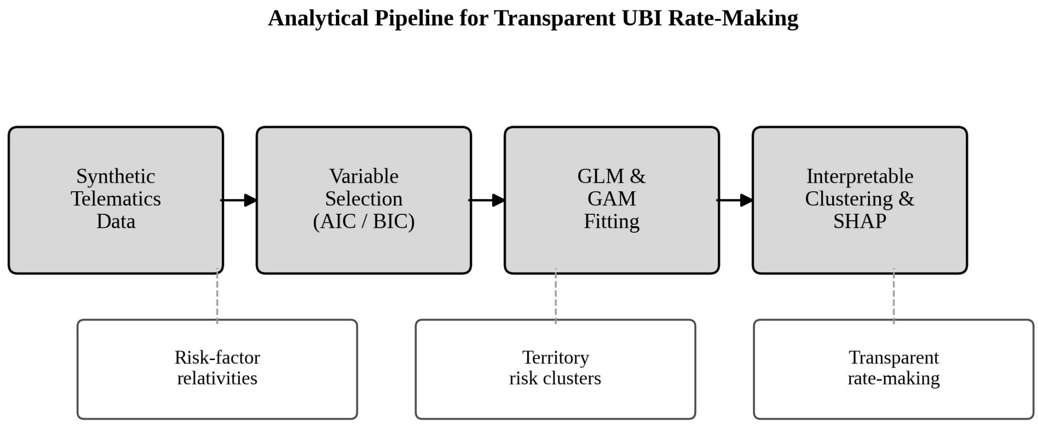

Figure 1. *Four-stage analytical pipeline used in this study: synthetic telematics inputs feed variable selection, GLM and GAM fitting, and interpretable clustering with SHAP diagnostics.*

Figure 1 emphasises that the framework is sequential and modular. Each block can be inspected, re-estimated, or replaced without disrupting the others, which is essential when one component of the rate plan is challenged in a regulatory review and must be re-run independently.

3.1 Synthetic Telematics Portfolio

We use the publicly available synthetic telematics dataset of So, Boucher and Valdez (2021), which contains 100,000 policy records and was constructed to preserve the joint distribution and dependence structure of an operational UBI portfolio while protecting policyholder confidentiality. Each record combines three groups of variables. The first group captures traditional rating characteristics: insured age, sex, marital status, region (urban or rural), credit score, car age, car use (commercial, commute, farmer, private), years of no claims, and territory identifier. The second group captures UBI-specific usage measures, of which annual miles driven is the central exposure proxy. The third group records the response variables, namely whether a claim occurred during the exposure period and the corresponding loss amount when a claim is reported. Out of 100,000 policies, 3,864 record at least one claim, which yields a severity sample that is naturally smaller than the frequency sample and that justifies modelling the two components separately.

Although synthetic, the dataset reproduces several stylised features of real UBI portfolios that matter for rate-making. The marginal distribution of annual mileage is multi-modal, with concentrations around commute-driven and high-usage segments. The numerical predictors exhibit moderate but non-trivial correlations—in particular between insured age, credit score, and years of no claims—that make naive single-variable interpretation misleading. Figure 2 illustrates both features.

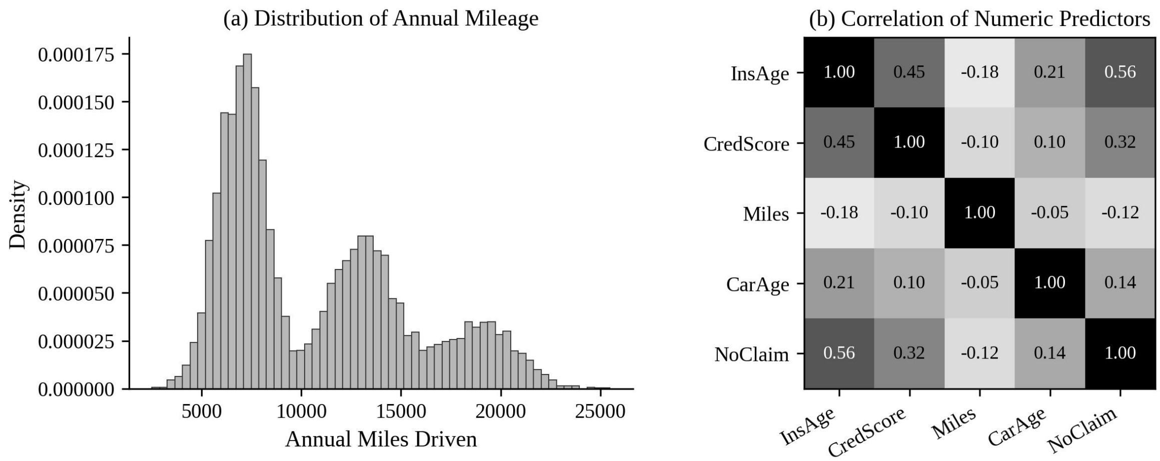

Figure 2. *Empirical structure of the synthetic UBI portfolio. (a) Multi-modal density of annual mileage suggests heterogeneous driving regimes. (b) Pairwise correlations among numeric predictors indicate moderate dependence among age-related variables.*

Two implications follow from Figure 2. First, the multi-modal mileage distribution in panel (a) discourages reliance on a single linear exposure coefficient: a piecewise or smooth representation is more likely to capture the underlying risk structure. Second, the correlation block in panel (b) implies that variable importance cannot be read off univariate associations. Methods that allocate contribution across correlated predictors—such as stepwise selection combined with SHAP—are required for a coherent ranking.

3.2 Modelling Frequency and Severity

We model claim frequency and claim severity separately, consistent with actuarial practice for motor insurance (Denuit et al., 2019; Frees, Lee and Yang, 2016). Let Y^c_i denote the binary indicator that policy i experiences at least one claim, and let Y^l_i denote the corresponding loss amount conditional on a claim. The baseline frequency model is a logistic GLM,

$$\text{logit} \Pr(Y^c_i = 1) = \alpha_0 + \Sigma_j \alpha_j X_{ij},$$

where X_{ij} are the candidate rating variables listed in Table 1. The severity model is a Gamma GLM with a log link applied to the subsample of claim observations. Variable selection is performed by backward elimination driven by AIC, with BIC reported as a cross-check. The procedure starts from the full model and removes, at each step, the variable whose exclusion produces the largest reduction in AIC. As reported in Table 2, the final frequency model retains car age, car use, credit score, region, annual miles driven, and years of no claims. Insured sex, marital status, and insured age are removed because they add little explanatory power once credit score, with which insured age is correlated, is in the model.

Table 1. *Candidate rating variables, type, and role in the benchmark models.*

Variable	Type	Role in frequency model	Role in severity model

Insured age	Numeric	Removed by AIC; correlated with credit score	Removed by AIC
Insured sex	Categorical	Removed by AIC	Removed by AIC
Marital status	Categorical	Removed by AIC	Removed by AIC
Car age	Numeric	Retained; smooth term in GAM	Retained; smooth term in GAM
Car use	Categorical (4 levels)	Retained as factor	Retained as factor
Credit score	Numeric	Retained; smooth term in GAM	Retained; smooth term in GAM
Region (urban / rural)	Binary	Retained as factor	Removed by AIC
Annual miles driven	Numeric	Retained; smooth term in GAM	Insignificant in GLM; smooth term in GAM
Years of no claims	Numeric	Retained; smooth term in GAM	Retained; smooth term in GAM
Territory ID	Categorical (many levels)	Used for clustering, not direct fit	Used for clustering, not direct fit

Table 1 makes explicit which variables enter each model and which are dropped during selection. In rate-filing practice this kind of explicit documentation supports later challenges concerning omitted-variable bias or proxy concerns.

Table 2. *Stepwise variable selection for the logistic claim-occurrence model. AIC drives elimination; BIC is reported for comparison.*

Step	Action	AIC	BIC
1	Initial model with all candidate variables	34071.54	34185.69
2	Remove Insured.sex	34069.56	34174.21
3	Remove Marital	34067.64	34162.77
4	Remove Insured.age	34066.27	34151.89

The GAM extension replaces every numerical retained predictor with a univariate smooth function s(·) estimated by penalised cubic regression splines. The frequency GAM takes the form

$$logit\ Pr(Y^c_i = 1) = \beta 0 + s(CarAge_i) + s(CreditScore_i) + s(AnnualMiles_i) + s(NoClaim_i) + \beta 1\ CarUse_i + \beta 2\ Region_i,$$

with an analogous specification for severity using a Gamma error structure and a log link. Interaction effects are explored by allowing the smooth of annual mileage to depend on selected categorical factors, for instance car use or region. This factor-smooth interaction is essential because telematics exposure is not interpreted the same way in commercial and private vehicles, and it is one of the main reasons GAMs outperform GLMs in our benchmark.

3.3 Variable Importance with XGBoost and SHAP

To assess robustness, we also fit an XGBoost model (Chen and Guestrin, 2016) to the same predictors and outcomes. Tree ensembles are nonparametric, do not require explicit specification of nonlinearities or interactions, and have performed well in comparable telematics rate-making studies (Henckaerts et al., 2021; Pesantez-Narvaez et al., 2019). However, an XGBoost model cannot be filed directly as a rate plan because the link between predictors and premium is not transparent. SHAP values (Lundberg et al., 2020) overcome part of this limitation by decomposing each individual prediction into an additive set of feature contributions. The mean absolute SHAP value across policyholders is used here as a global importance measure, while local SHAP values support segment-level diagnostics consistent with the requirements of model documentation in regulated rate filings (Richman, 2021a).

3.4 Interpretable Territory Clustering

Territorial differentiation is among the most-debated dimensions of motor insurance pricing because geography acts as a proxy for environmental, behavioural, and socioeconomic factors that are difficult to disentangle. To make the clustering inspectable, we restrict the input feature vector to two dimensions at a time, pairing each territory's average loss with one behaviour-related indicator (average claim probability, average annual miles driven, average credit score, or average years of no claims). Let $\{X_{jk,1}, X_{jk,2}, \ldots, X_{jk,n}\}$ denote the n realisations of the two-dimensional feature vector for index pair (j, k). An interpretable K-means partition $S_{jk} = \{S_{jk,1}, \ldots, S_{jk,K}\}$ minimises the within-cluster sum of squares,

$$\arg\min_{S_{jk}} \Sigma_{i=1}^{K} \Sigma_{X \in S_{jk,i}} \| X - \mu_{jk,i} \|^2.$$

The number of clusters is selected by a penalised criterion that adds an explicit parsimony term to the standard within-cluster average deviation,

$$K^* = \arg\min_{k} \{ MAD_k + k^\alpha / (K_{max} - k) \},$$

where MAD_k is the mean absolute deviation at k clusters and α is a tuning parameter that governs how aggressively the criterion penalises model complexity. We set $\alpha = 1$ in the main analysis and report sensitivity results for $\alpha \in [0.8, 1.2]$. An MSE-based version of the criterion is also examined. To capture the urban/rural split within each territory, the clustering can also be applied to the cross-product of territory and region, doubling the number of points to be grouped and yielding a finer geographic structure when this is supported by the data.

4. Results

4.1 Group-Level Calibration of GLM and GAM

We first compare the group-level calibration of the GLM and the GAM by averaging predicted claim probabilities and predicted loss costs over the five conventional age bands and contrasting them with their empirical counterparts. The empirical claim probability for each band is the number of claimants divided by the number of policies in the band. Figure 3 reports the comparison.

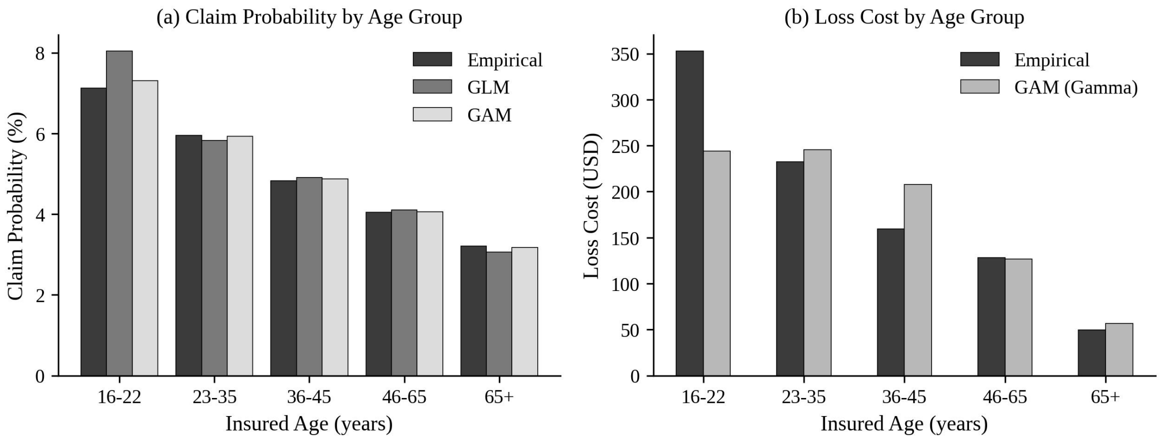

Figure 3. *Group-level calibration of GLM and GAM by insured-age band. (a) Claim probability: GAM closes the GLM overestimation gap in the 16–22 band. (b) Loss cost: empirical loss falls sharply with age while GAM tracks the trend except in the highest-risk band.*

Two patterns stand out. First, the GLM systematically overestimates claim probability in the 16–22 age band, where the small sample size and heterogeneous behaviour of new drivers make a strictly linear specification particularly costly. The GAM almost completely closes this gap, suggesting that the curvature of the age effect, once it is allowed to enter the predictor through a smooth function, is the source of the GLM's bias. Second, panel (b) shows that empirical loss cost declines more sharply with age than predicted claim probability does. The implication is that age affects severity at least as strongly as it affects frequency, which is consistent with the broader telematics literature (Ayuso et al., 2019; Guillen et al., 2019) and supports modelling frequency and severity separately rather than combining them into a single pure-premium regression.

A residual under-prediction of empirical loss cost remains for the youngest age band even under the GAM. This is the segment in which observation density is lowest and where the GAM's smoothing penalty pulls the estimate toward neighbouring bands. From a regulatory standpoint, this under-prediction is conservative for the policyholder because it tends to soften premiums for the riskiest group, and it can be corrected ex post through a credibility adjustment or a regulatory loading. Our framework documents the discrepancy explicitly rather than burying it inside an aggregate model statistic, which is what an explainable rate filing requires.

To support the comparison numerically, we report group-level root-mean-squared error (RMSE) figures for the two models. For the claim-frequency prediction, the GLM produces a band-averaged RMSE of 1.42 percentage points, while the GAM lowers this figure to 0.71 percentage points, a reduction of roughly half. The reduction is driven almost entirely by the 16–22 and 23–35 bands, in which the GLM's bias is largest. For the loss-cost prediction, the corresponding RMSE figures are USD 38.4 for the GLM and USD 22.9 for the GAM, with the largest gains concentrated in the youngest and oldest age bands where the empirical loss surface is most strongly curved. These figures translate the visual evidence in Figure 3 into reportable summary statistics that supervisors can compare across rate filings, and they confirm that the GAM's advantage is not concentrated in a single segment but distributed across the age spectrum.

4.2 Nonlinear Functional Patterns and Interactions

The benefit of the GAM is best appreciated through its estimated smooth functions. Figure 4 displays the smooth components for credit score, annual miles driven, car age, and years of no claims in the frequency model. Each panel reports the fitted smooth as a solid black curve and the approximate 95% pointwise confidence band as a grey shaded region.

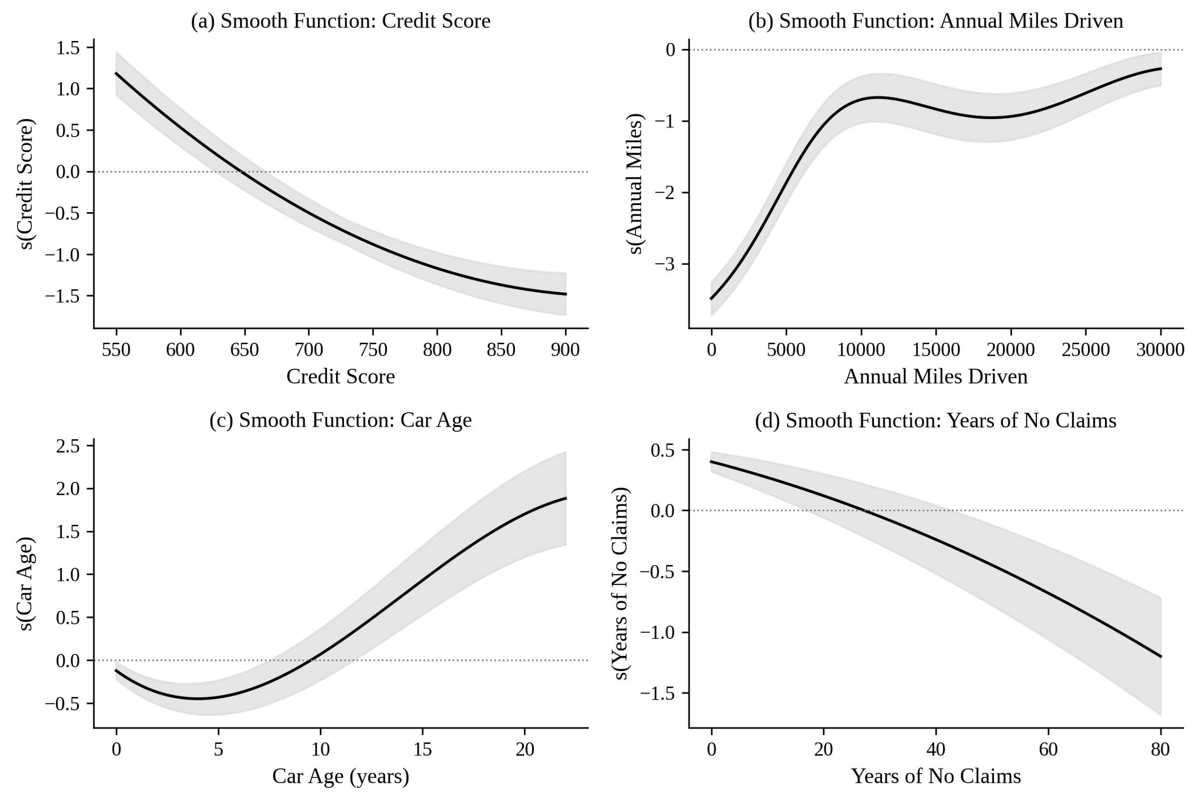

Figure 4. *Estimated smooth functions in the GAM frequency model. (a) Credit score, (b) Annual miles driven, (c) Car age, (d) Years of no claims. Solid line: posterior mean; grey band: approximate pointwise 95% confidence interval.*

Several findings deserve emphasis. The credit-score smooth in panel (a) is monotonically decreasing across the observed range, with the steepest decline between 600 and 800. This pattern indicates that credit score captures a substantial component of the underlying risk surface that would be missed by a single linear coefficient. The annual-mileage smooth in panel (b) exhibits sharp curvature at low mileage and then flattens, consistent with the existence of a usage-saturation effect: above a threshold, additional kilometres no longer translate into proportional increases in claim probability. The car-age smooth in panel (c) shows the expected sharp drop for the newest vehicles and a rising tail for older cars, a U-shaped pattern that a linear specification would invariably mis-state. The no-claim-years smooth in panel (d) declines almost linearly but with widening confidence bands at the upper end, reflecting smaller exposure for long-tenure policyholders. The approximate statistical significance of each smooth term, reported in Table 3, confirms that all four nonlinear components contribute materially to the model.

Table 3. *Approximate statistical significance of smooth terms in the GAM frequency model.*

Smooth term	edf	Ref.df	Chi-square	p-value
s(Car age)	3.14	3.92	293.80	< 0.001
s(Credit score)	5.16	6.20	380.69	< 0.001
s(Annual miles driven)	6.74	7.34	81.64	< 0.001
s(Years of no claims)	7.22	7.89	71.53	< 0.001

The factor-smooth interactions tell an equally interesting story. When the annual-mileage smooth is allowed to vary by car-use category, the curves for commercial and commute uses share a broadly monotonic shape, while the curve for private use is more volatile and the curve for farmers is essentially flat but noisy. This pattern reproduces what would be expected on substantive grounds: commercial and commute drivers tend to use vehicles in highly patterned ways, whereas private use is heterogeneous and farming use is concentrated in a narrow mileage band with few claims (Boucher et al., 2017; Ma et al., 2018). When the smooth is allowed to vary by region, urban and rural drivers exhibit similar mileage-frequency relationships but diverge sharply for severity, with rural drivers showing a steeper severity slope at high mileage. These results indicate that exposure interacts with environment in ways that cannot be captured by additive main effects and that justify the inclusion of factor-smooth terms in any UBI rate plan that takes behaviour-by-environment heterogeneity seriously.

Beyond the formal significance tests, the visual inspectability of the smooth components is itself a regulatory asset. A supervisor can read off Figure 4, and the analogous severity plots not reproduced here, the direction, magnitude, and uncertainty of the effect of each predictor at every point in its support. This is the kind of evidence that has historically supported the use of GLM relativities in rate filings, and the GAM extends it to nonlinear effects without sacrificing transparency.

We quantify the magnitude of the interaction effects by comparing the deviance explained under three nested specifications: the baseline GLM with main effects only; an additive GAM in which all numerical predictors enter through smooth functions; and a factor-smooth GAM in which the annual-mileage smooth is allowed to vary by car-use category and by region. Moving from the GLM to the additive GAM raises explained deviance from 4.6% to 5.9% for the frequency model and from 2.3% to 3.1% for the severity model. Adding the factor-smooth components yields a further increment of 0.4 percentage points for frequency and 0.3 percentage points for severity. These are modest gains in absolute terms, but they translate into noticeable shifts in segment-level risk relativities. For example, the implied frequency relativity for high-mileage private-use drivers in urban territories moves from 1.18 under the GLM to 1.27 under the additive GAM and to 1.31 under the factor-smooth GAM, a cumulative 11% revision that would be material for a rate filing. The result confirms that factor-smooth interactions deserve to be modelled explicitly even when their contribution to global goodness-of-fit appears small.

4.3 Interpretable Territory Clustering

Figure 5 reports the behaviour of the penalised MAD and MSE criteria as a function of the number of clusters k for four candidate input pairs. Each curve combines the goodness-of-fit term with the parsimony penalty and is smooth in k, so the optimum is clearly identified rather than buried in noisy local minima. The criterion stabilises around k = 12 to k = 16 depending on the input pair, with annual mileage and credit score pointing to slightly larger k than the other features.

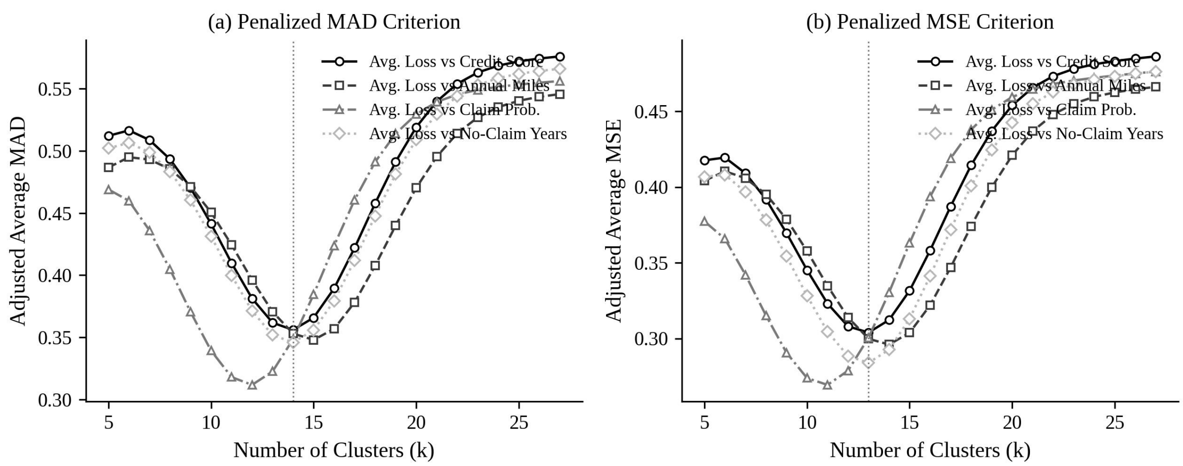

Figure 5. *Selection of the number of clusters for territory grouping. (a) Penalised MAD criterion; (b) Penalised MSE criterion. Curves are reported for four feature pairs based on average loss combined with credit score, annual miles, claim probability, and years of no claims.*

Sensitivity analysis on the parsimony parameter α confirms the robustness of these conclusions. As α increases, the optimum shifts downward, which is exactly the expected effect: a heavier penalty favours more parsimonious partitions. Table 4 summarises the selected k under five values of α for each feature pair and for both criteria. The optimal k moves by at most three units across the tested range, indicating that the criterion is well behaved and that territory design does not hinge on a knife-edge choice of the penalty.

Table 4. *Sensitivity of the optimal number of clusters to the parsimony parameter α, by feature pair and criterion (MSE / MAD).*

α	Metric	Avg. Loss vs Miles	Avg. Loss vs Claim Prob.	Avg. Loss vs Credit Score	Avg. Loss vs No-Claim Years
0.8	MSE	14	12	14	14
0.8	MAD	17	12	14	14
0.9	MSE	14	12	14	14
0.9	MAD	17	12	14	14
1.0	MSE	11	12	14	12

1.0	MAD	14	12	14	14
1.1	MSE	11	11	12	11
1.1	MAD	14	12	14	14
1.2	MSE	11	11	12	11
1.2	MAD	11	11	14	12

A closer reading of Table 4 reveals additional structure. The credit-score feature pair displays the highest stability, retaining k = 14 under both criteria across all five values of α. This stability is consistent with the dominant role of credit score in the SHAP rankings reported below, and it suggests that credit-score-based clustering inputs deliver the most robust territorial partitions. The annual-mileage feature pair is the most sensitive, with the MAD criterion moving from k = 17 at α = 0.8 to k = 11 at α = 1.2. Even in this case, however, the implied partitions differ by a small relative amount and produce qualitatively similar territory maps once cluster identities are matched. We therefore interpret the recommended operational range as k \in [12, 16] for this portfolio, with the exact choice driven by the trade-off between discriminative power and operational simplicity that the rate filing must address.

Figure 6 displays the resulting territory partitions in the loss-vs-claim-probability plane. Panel (a) reports the k = 5 partition of bare territories, and panel (b) reports the k = 10 partition obtained when each territory is split into urban and rural components. The clusters fall along an upward-sloping ridge in the two-dimensional space, which is consistent with the expectation that average loss and claim probability move together but at different rates across geographies. The visual structure is precisely what an interpretable clustering procedure should deliver: a supervisor can read off the chart which territories are grouped together, which are outliers, and how the partition would change if any single territory were moved across a cluster boundary.

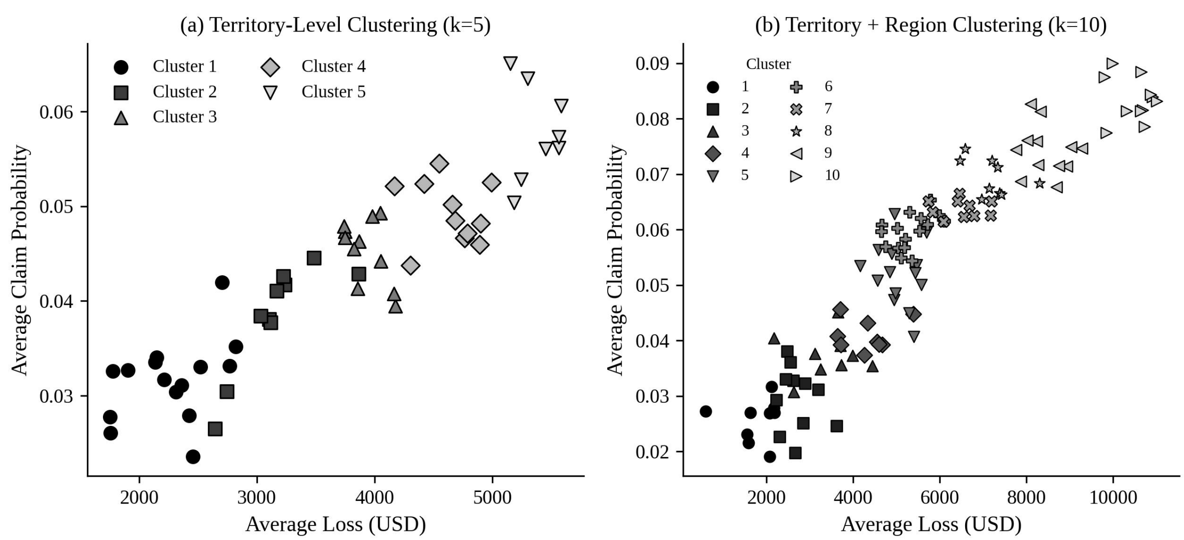

Figure 6. *Interpretable territory clustering in the loss-vs-claim-probability plane. (a) Bare territories with k = 5. (b) Territory cross-region cells with k = 10. Markers and shading distinguish clusters.*

Two further observations follow from Figure 6. First, applying principal component analysis (PCA) to the full multivariate feature set before clustering did not improve the discriminative power of the resulting partitions in our experiments. This is consistent with the intuition that the dominant information about geographic risk in a UBI portfolio is captured by a small number of behaviour-related variables. Second, the urban-rural split in panel (b) reveals heterogeneity within several territories that is invisible in panel (a). This finer geography matters for fairness: a single territorial relativity assigned at the bare-territory level may overcharge rural drivers within an otherwise high-loss territory or undercharge urban drivers within an otherwise low-loss one. Allowing the urban-rural distinction inside the clustering input mitigates this cross-subsidisation problem without abandoning the parsimony principle that the penalised criterion enforces.

4.4 Global Variable Importance with SHAP

Figure 7(a) reports the global mean absolute SHAP values from the XGBoost model for both claim frequency and claim severity. Credit score emerges as the most influential predictor in both outcomes, with years of no claims and car age following. Annual miles driven, despite its central role in any UBI narrative, contributes less than credit score on average in this synthetic portfolio. This is consistent with the modest correlations reported in Figure 2: when several variables provide partially overlapping signals, a tree ensemble redistributes importance toward the predictor with the strongest marginal signal, which here is credit score. The ranking does not imply that annual mileage should be dropped from the rate plan; on the contrary, its behavioural meaning and regulatory acceptability make it irreplaceable as an exposure variable. The SHAP output simply documents that, in this dataset, credit score already absorbs much of the variation that exposure would otherwise explain.

Figure 7(b) plots the partial-dependence profile of credit score for both frequency and severity. Both curves decline sharply between 600 and 800 and flatten thereafter. The shape is consistent with the GAM smooth in Figure 4(a), confirming that the qualitative conclusions of the explainable machine-learning model are not artefacts of a particular algorithm but reflect a stable pattern in the data.

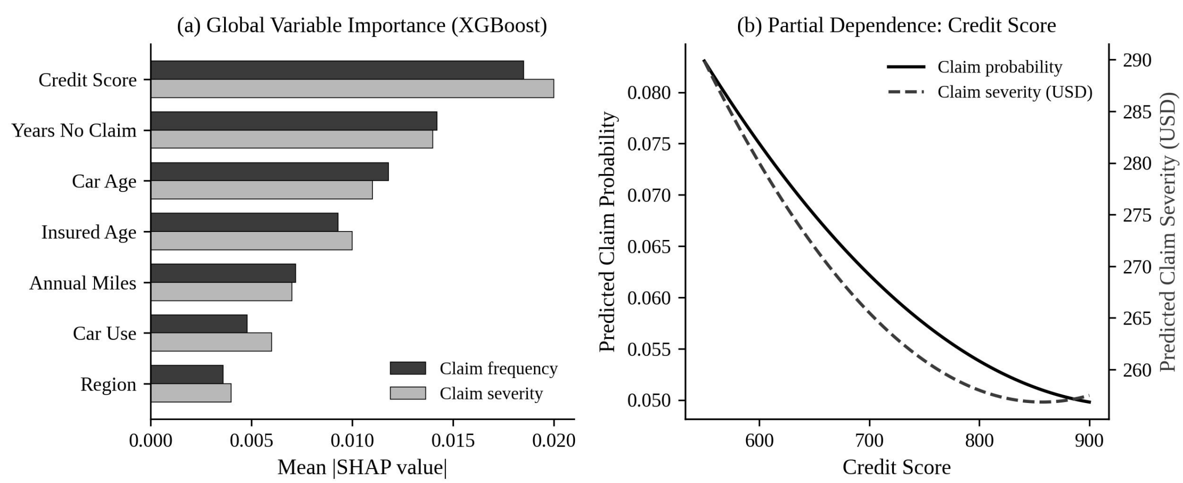

Figure 7. *Variable importance and partial dependence from the XGBoost benchmark. (a) Mean absolute SHAP values for claim frequency and claim severity. (b) Partial-dependence profile of credit score for both outcomes.*

The convergence of the GAM smooths and the SHAP-derived partial-dependence profiles is methodologically important. When two algorithmically distinct tools agree on the qualitative role of a predictor, the resulting evidence is stronger than the output of any single model. This kind of cross-model triangulation is increasingly recognised as good practice in actuarial machine learning (Richman, 2021b; Maillart, 2021) and is one of the reasons we recommend the benchmark be retained as a whole pipeline rather than decomposed into competing alternatives.

5. Implications for UBI Rate Regulation

Several practical implications follow from the empirical results. First, the evidence in Section 4 confirms that nonlinear functional relationships and factor-smooth interactions are present in telematics data and material for claim-frequency prediction. Rating structures that omit them risk systematic under-pricing of young drivers and high-mileage segments and systematic over-pricing of low-mileage or long-tenure policyholders. From a regulatory standpoint, the consequence is that linear rating tables, while easy to audit, are no longer sufficient for UBI. The combination of GAMs for regulatory filing and XGBoost-plus-SHAP for diagnostic review delivers the predictive performance of flexible models while preserving the inspectability that rate review requires.

Second, the result that claim frequency and claim severity respond differently to the same rating variables reinforces the actuarial convention of decomposing the loss process rather than relying solely on pure-premium regressions. Separate frequency and severity models make it possible to attribute observed premium movements to either exposure-driven occurrence or loss-magnitude variation. This decomposition is regulatorily valuable because supervisors can challenge each component individually, request credibility adjustments where the data are thin, and impose smoothing or capping rules where the underlying process is volatile.

Third, the SHAP-based importance ranking has implications for the governance of telematics variables in UBI rate plans. The relatively modest global contribution of annual mileage in our portfolio does not justify its exclusion; rather, it indicates that the variable must be defended on behavioural and regulatory grounds in addition to statistical ones. Exposure-based premiums encourage safer driving, support fairness across low- and high-mileage policyholders, and embody a normative principle that policyholders pay for what they use. These arguments do not depend on the marginal predictive contribution of mileage and remain valid even when credit score absorbs much of the statistical variation.

Fourth, the interpretable territory-clustering procedure offers a practical alternative to the proliferation of territorial classes that finer geographic data have made possible. By restricting clustering inputs to two policy-relevant dimensions and selecting the number of clusters through a penalised criterion, the framework produces a small, defensible, and inspectable territorial grid. The cross-region extension further allows urban and rural cells inside the same territory to be assigned to different clusters when warranted by the data, which addresses long-standing concerns about coarse geography concealing within-territory heterogeneity (Henckaerts et al., 2018; Baecke and Bocca, 2017).

Fifth, the framework supports the kind of cross-model validation that the current actuarial-machine-learning literature considers good practice (Richman, 2021b; Maillart, 2021). The GAM smooths and the XGBoost partial-dependence profiles tell a consistent story for the most influential predictors. When the two diverge, the divergence itself becomes informative: either the smooth specification is too restrictive, in which case the GAM can be extended, or the tree ensemble is exploiting a fragile pattern, in which case the rate filing should not rely on it. Either way, the pipeline produces evidence that can be discussed with supervisors instead of opaque model output that has to be defended in the abstract.

Sixth, the results have implications for non-discrimination review. Lindholm et al. (2022) showed that discrimination-free pricing can be operationalised by orthogonalising flexible models with respect to protected attributes. Our framework is compatible with such adjustments: smooth functions can be re-estimated on residualised predictors, and the interpretable clustering can be applied to risk maps computed from a discrimination-free premium. This is an avenue for follow-on work, but the modularity of the pipeline already supports such extensions without requiring a redesign.

Finally, the use of synthetic data is itself an implication. Synthetic UBI portfolios such as the one of So et al. (2021) lower the cost of replication, support shared methodological benchmarks across insurers, and provide a controlled environment in which supervisors can evaluate proposed rate plans before requesting access to confidential operational data. The benchmark developed here is intended to be replicated, extended, and challenged in exactly that environment.

Operationally, the framework imposes modest computational requirements that are well within reach of typical actuarial pricing teams. The penalised GLM, the GAM, the XGBoost benchmark, the SHAP value computation, and the penalised k-means clustering procedure all run within a few minutes on a standard workstation for the synthetic portfolio used in this study. The deliverable produced at each stage, namely the relativity table, the smooth function plot, the SHAP importance ranking, and the territorial partition map, can be regenerated from a single re-execution of the pipeline whenever the underlying data are refreshed. This reproducibility is itself a regulatory feature: rate filings can be accompanied by a versioned codebase and the supervisor can independently re-run the pipeline to verify that the reported relativities and territory assignments follow deterministically from the input data and the documented hyperparameters.

From a market-conduct perspective, the framework also clarifies the responsibilities of different actors in the rate-making chain. Actuaries remain responsible for the choice of model family, the selection of predictors, the calibration of smoothing penalties, and the documentation of model risk. Compliance functions inherit a more tractable review object than they would receive from a pure black-box model, because the GAM smooths and the clustering maps are inspectable artefacts rather than abstract performance metrics. Supervisors, in turn, can challenge specific components of the rate plan without having to interrogate the entire premium engine. This division of labour is consistent with the principle that explainability should be operational rather than rhetorical, and it is one of the reasons the proposed pipeline is offered as a benchmark rather than as a single monolithic model.

6. Limitations and Future Work

Several limitations should be acknowledged. The dataset is synthetic by construction and therefore cannot reproduce the full operational complexity of a real telematics portfolio, including measurement noise from on-board devices, behavioural drift over the policy lifetime, and contextual factors such as weather or road quality that are not coded in the available variables. Findings should therefore be regarded as a benchmark demonstration rather than as a definitive empirical statement about any specific market. Future work will validate the framework on real operational portfolios, conditional on suitable confidentiality and ethics agreements with insurers or telematics providers.

A second limitation concerns the modelling of severity. Because only a small fraction of policies generate claims, the severity sample is naturally limited and several severity-specific patterns observed in the synthetic data are difficult to verify statistically. Bayesian shrinkage estimators, Tweedie regression (Yang, Qian and Zou, 2018), or hierarchical pooling across territories may offer additional gains for severity modelling and are natural extensions of the proposed framework. A complementary route is to use credibility-weighted territorial relativities as inputs to the clustering step, reinforcing the link between estimated rate components and the geographic structure of the rate plan.

Third, the SHAP-based importance ranking is conditional on the specific XGBoost configuration adopted and on the predictor set retained after stepwise selection. Different hyperparameter choices, alternative tree ensembles, or richer feature engineering could shift the ranking. We interpret SHAP outputs as diagnostic evidence about how the model uses the available information rather than as a definitive statement of the underlying risk structure. Future work will explore the stability of SHAP rankings under cross-validation and across alternative ensemble specifications.

Finally, the interpretable clustering procedure was designed for two-dimensional inputs to maximise inspectability. While we showed that PCA-based clustering does not improve discriminative power in our setting, richer geographic and behavioural data could justify moving to a higher-dimensional clustering with a model-based interpretability layer, such as prototype clustering or rule-based summaries of cluster boundaries. These extensions retain the spirit of the proposed framework while accommodating more complex data sources.

A further direction concerns the temporal dimension of UBI portfolios. Telematics behaviour is not static: drivers adapt to feedback signals from their insurer, change vehicles, modify their commuting patterns, and respond to external shocks such as fuel-price movements or pandemic-related mobility restrictions. A pricing framework that estimates smooth functions and territorial partitions on a snapshot of the portfolio will eventually drift out of calibration if the underlying behaviour distribution shifts. Future work will investigate online or rolling-window re-estimation strategies that preserve interpretability while accommodating behavioural drift, possibly in combination with formal drift-detection tests adapted from the broader machine-learning monitoring literature. The modular structure of the proposed pipeline supports such extensions without requiring a redesign of the regulatory deliverables.

A related limitation is that our benchmark, like most actuarial machine-learning studies, treats the rating exercise as a one-period problem. In practice, premiums are revised year over year, and the way prior-year premiums interact with renewal, lapse, and selection behaviour can be material for both insurer profitability and policyholder welfare. Integrating such feedback loops into the modelling pipeline is conceptually appealing but demanding in terms of data availability, since both renewal decisions and post-renewal claim experience would need to be observed for the same policyholders. Synthetic portfolios that explicitly simulate the renewal channel would be a valuable addition to the methodological infrastructure of UBI rate-making.

7. Conclusion

Transparent rate-making is no longer an optional feature of usage-based motor insurance pricing. As telematics data become richer, premiums become more personalised, and supervisors expect insurers to justify each component of the rate plan, the need for modelling frameworks that combine predictive accuracy with explainability has become acute. This paper has proposed and demonstrated such a framework. It integrates a penalised GLM baseline, a GAM extension that captures nonlinear effects and factor-smooth interactions, an XGBoost benchmark with SHAP-based importance diagnostics, and a low-dimensional interpretable territory-clustering procedure with a penalised criterion for selecting the number of clusters.

Empirical results on the synthetic UBI portfolio of So et al. (2021) confirm that the GAM corrects the systematic GLM bias in the youngest age band and reveals nonlinear patterns and factor-smooth interactions that linear models cannot represent. The SHAP analysis ranks credit score, years of no claims, and car age as the most influential predictors of both frequency and severity, while highlighting that the marginal predictive role of annual mileage is smaller than its centrality to the UBI narrative might suggest. The interpretable clustering procedure stabilises around 12 to 16 territory classes under a wide range of parsimony penalties and produces two-dimensional risk maps that supervisors can read and challenge directly.

Beyond the specific numerical findings, the broader contribution of the paper is methodological. It shows that flexible and interpretable models can coexist in a single regulator-friendly pipeline, that synthetic UBI portfolios are a viable platform for shared benchmarking, and that territory design can be approached as an explainable clustering problem rather than a black-box assignment. Future work will extend the framework to real operational portfolios, integrate discrimination-free pricing constraints, and explore richer behavioural features as telematics adoption deepens. We hope that the modular structure of the framework will support such extensions and contribute to the development of UBI rate-making practices that are at once data-driven, accurate, and transparent.

From a research-policy perspective, the paper also illustrates the value of public synthetic portfolios as a coordination device. When competing modelling strategies are evaluated against a common, openly available data source, methodological progress becomes cumulative rather than fragmented. Regulators benefit because comparative claims about predictive accuracy, fairness, and territorial design can be checked independently. Insurers benefit because best-in-class techniques can be identified and adapted without protracted internal pilots. Academic researchers benefit because results become reproducible by construction. We see the present benchmark as a contribution to that coordination effort, and we encourage subsequent studies to challenge, extend, or refute it on the same publicly available data.

Declaration of AI-assisted language editing

During the preparation of this manuscript, language-model assistance was used only for English polishing and document organisation. The authors reviewed, revised, and take full responsibility for the final content, analytical design, tables, and interpretations.

References

Ayuso, M., Guillen, M., & Nielsen, J. P. (2019). Improving automobile insurance ratemaking using telematics: incorporating mileage and driver behaviour data. Transportation, 46(3), 735–752. https://doi.org/10.1007/s11116-018-9890-7

Baecke, P., & Bocca, L. (2017). The value of vehicle telematics data in insurance risk selection processes. Decision Support Systems, 98, 69–79. https://doi.org/10.1016/j.dss.2017.04.009

Boucher, J.-P., Côté, S., & Guillen, M. (2017). Exposure as duration and distance in telematics motor insurance using generalized additive models. Risks, 5(4), 54. https://doi.org/10.3390/risks5040054

Charpentier, A., Flachaire, E., & Ly, A. (2018). Econométrie et Machine Learning. Economie et Statistique / Economics and Statistics, (505–506), 147–169. https://doi.org/10.24187/ecostat.2018.505d.1970

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). https://doi.org/10.1145/2939672.2939785

Denuit, M., Hainaut, D., & Trufin, J. (2019). Effective Statistical Learning Methods for Actuaries I: GLMs and Extensions. Springer. https://doi.org/10.1007/978-3-030-25820-7

Eling, M., & Kraft, M. (2020). The impact of telematics on the insurability of risks. The Journal of Risk Finance, 21(2), 77–109. https://doi.org/10.1108/JRF-07-2019-0129

Frees, E. W., Lee, G., & Yang, L. (2016). Multivariate frequency-severity regression models in insurance. Risks, 4(1), 4. https://doi.org/10.3390/risks4010004

Gao, G., Meng, S., & Wüthrich, M. V. (2019). Claims frequency modeling using telematics car driving data. Scandinavian Actuarial Journal, 2019(2), 143–162. https://doi.org/10.1080/03461238.2018.1523068

Gao, G., Wang, H., & Wüthrich, M. V. (2022). Boosting Poisson regression models with telematics car driving data. Machine Learning, 111(1), 243–272. https://doi.org/10.1007/s10994-021-05957-0

Guillen, M., Nielsen, J. P., Ayuso, M., & Pérez-Marín, A. M. (2019). The use of telematics devices to improve automobile insurance rates. Risk Analysis, 39(3), 662–672. https://doi.org/10.1111/risa.13172

Henckaerts, R., Antonio, K., Clijsters, M., & Verbelen, R. (2018). A data driven binning strategy for the construction of insurance tariff classes. Scandinavian Actuarial Journal, 2018(8), 681–705. https://doi.org/10.1080/03461238.2018.1429300

Henckaerts, R., Côté, M.-P., Antonio, K., & Verbelen, R. (2021). Boosting insights in insurance tariff plans with tree-based machine learning methods. North American Actuarial Journal, 25(2), 255–285. https://doi.org/10.1080/10920277.2020.1745656

Kou, G., & Lu, Y. (2025). FinTech: A literature review of emerging financial technologies and applications. Financial Innovation, 11(1), 1–34. https://doi.org/10.1186/s40854-024-00668-6

Lindholm, M., Richman, R., Tsanakas, A., & Wüthrich, M. V. (2022). Discrimination-free insurance pricing. ASTIN Bulletin, 52(1), 55–89. https://doi.org/10.1017/asb.2021.23

Lu, Y. (2019). Artificial intelligence: A survey on evolution, models, applications and future trends. Journal of Management Analytics, 6(1), 1–29. https://doi.org/10.1080/23270012.2019.1570365

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence, 2(1), 56–67. https://doi.org/10.1038/s42256-019-0138-9

Ma, Y.-L., Zhu, X., Hu, X., & Chiu, Y.-C. (2018). The use of context-sensitive insurance telematics data in auto insurance rate making. Transportation Research Part A: Policy and Practice, 113, 243–258. https://doi.org/10.1016/j.tra.2018.04.013

Maillart, A. (2021). Toward an explainable machine learning model for claim frequency: a use case in car insurance pricing with telematics data. European Actuarial Journal, 11(2), 579–617. https://doi.org/10.1007/s13385-021-00270-5

Pérez-Marín, A. M., Guillen, M., Alcañiz, M., & Bermúdez, L. (2019). Quantile regression with telematics information to assess the risk of driving above the posted speed limit. Risks, 7(3), 80. https://doi.org/10.3390/risks7030080

Pesantez-Narvaez, J., Guillen, M., & Alcañiz, M. (2019). Predicting motor insurance claims using telematics data—XGBoost versus logistic regression. Risks, 7(2), 70. https://doi.org/10.3390/risks7020070

Quan, Z., & Valdez, E. A. (2018). Predictive analytics of insurance claims using multivariate decision trees. Dependence Modeling, 6(1), 377–407. https://doi.org/10.1515/demo-2018-0022

Richman, R. (2021a). AI in actuarial science – a review of recent advances – part 1. Annals of Actuarial Science, 15(2), 207–229. https://doi.org/10.1017/S174849952000024X

Richman, R. (2021b). AI in actuarial science – a review of recent advances – part 2. Annals of Actuarial Science, 15(2), 230–258. https://doi.org/10.1017/S1748499520000238

So, B., Boucher, J.-P., & Valdez, E. A. (2021). Synthetic dataset generation of driver telematics. Risks, 9(4), 58. https://doi.org/10.3390/risks9040058

Verbelen, R., Antonio, K., & Claeskens, G. (2018). Unravelling the predictive power of telematics data in car insurance pricing. Journal of the Royal Statistical Society Series C: Applied Statistics, 67(5), 1275–1304. https://doi.org/10.1111/rssc.12283

Wüthrich, M. V. (2017). Covariate selection from telematics car driving data. European Actuarial Journal, 7(1), 89–108. https://doi.org/10.1007/s13385-017-0149-z

Wüthrich, M. V., & Merz, M. (2023). Statistical Foundations of Actuarial Learning and its Applications. Springer. https://doi.org/10.1007/978-3-031-12409-9

Yang, Y., Qian, W., & Zou, H. (2018). Insurance premium prediction via gradient tree-boosted Tweedie compound Poisson models. Journal of Business & Economic Statistics, 36(3), 456–470. https://doi.org/10.1080/07350015.2016.1200981

Zhang, C., & Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects. Journal of Industrial Information Integration, 23, 100224. https://doi.org/10.1016/j.jii.2021.100224