Financial Conduct Authority

Research Note

Annexes for Motor Insurance Pricing and Local Area Ethnicity in England and Wales

Dobrin Barakov, Edward Barrett, Daniel Bogiatzis-Gibbons, Lawrence Charles, Annette Reid, Wenjin Li

FCA research notes in financial regulation

The FCA research notes

The FCA is committed to encouraging debate on all aspects of financial regulation and to creating rigorous evidence to support its decision-making. To facilitate this, we publish a series of Research Notes, extending across economics and other disciplines.

The main factor in accepting papers is that they should make substantial contributions to knowledge and understanding of financial regulation. If you want to contribute to this series or comment on these papers, please contact David Stallibrass (David.Stallibrass@fca.org.uk).

Disclaimer

Research notes contribute to the work of the FCA by providing rigorous research results and stimulating debate. While they may not necessarily represent the position of the FCA, they are one source of evidence that the FCA may use while discharging its functions and to inform its views. The FCA endeavours to ensure that research outputs are correct, through checks including independent referee reports, but the nature of such research and choice of research methods is a matter for the authors using their expert judgement. To the extent that research notes contain any errors or omissions, they should be attributed to the individual authors, rather than to the FCA.

Authors

All authors are staff members of the FCA.

Acknowledgements

We would like to thank Rachel Stoneman and Ian Cuddy-Gibbs for their detailed industry knowledge, Yazmin Morlet Corti and Mojdeh Khandanian for extensive methodological discussions, Mark Mitchell for methodological review and Manasi Phadnis for code review.

All our publications are available to download from www.fca.org.uk. If you would like to receive this paper in an alternative format, please call 020 7066 9644 or email publications_graphics@fca.org.uk or write to Editorial and Digital Department, Financial Conduct Authority, 12 Endeavour Square, London E20 1JN.

Contents

Annex 1: Data preparation

Motor insurance policy data

An overview of the motor insurance policy dataset is outlined in the GIPP evaluation paper.

This data set initially comprised 12,577,993 observations, of which 10,561,419 were from England and 665,540 - from Wales, representing a single insurance policy at a specific year of its tenure. Individuals can be represented in the dataset more than once, as individuals can have more than one insurance policy.

For our analysis, we use a set of variables which are common to most insurance policies. These are outlined in Table 1 alongside average and median values for the numeric ones and most common values (reported with percentage prevalence in the data) for the categorical ones.

Table 1: Specification of variables used in analysis

Variable Name	Type	Description	Average Value	Median Value	Most Common Value (as % of all data)
policy_id	Integer	Unique identifier of each policy.			
ecc_core	Float	Expected claims cost of core components of insurance policy.	£329	£246	
submission_frn	Categorical string	Identifier of company that submitted the respective policy data.			

price_core	Float	Price customer paid for core parts of the insurance policy	£416	£316	[redacted]
autorenewal	Boolean	Was the insurance policy autorenewed	[redacted]	[redacted]	1 (Yes - 72%)
insurance_type	Categorical string	What type of policy was it (as defined by the firm) – used to differentiate between pricing models of firms	[redacted]	[redacted]	[redacted]
claims_handling_costs_core	Float	Costs associated with handling claims on the policy	£7	£0	[redacted]
claims_paid_core	Float	Amount paid to policy holder on the core part of the insurance policy	£223	£0	[redacted]
yob	Integer	The year of birth of the policy holder.	[redacted]	[redacted]	[redacted]
policy_tenure	Integer	How long has the policy existed for (as of current year)	[redacted]	[redacted]	0 (36%)
num_risk	Integer	How many cars are being insured?	[redacted]	[redacted]	1 (91%)
cancellation	Boolean	Was the policy cancelled	[redacted]	[redacted]	0 (No - 84%)
policy_holder_postcode_anon	String	Policy holder postcode.	[redacted]	[redacted]	[redacted]

Source: Firm data, FCA analysis

In addition to these existing variables, we created some new variables. These were generally transformations or extensions of underlying variables which were required for our modelling process and can be found in Table 2. Policy type describes the vehicle the policy was for - car, motorcycle, or other.

Table 2: Custom-made variables for use in analysis

Variable Name	Calculation	Justification	Average Value	Median Value	Most Common Value (as % of all data)
price_core_ipt	price_core * 1.12	The price after including insurance premium tax (IPT). This is the price a customer would actually see, as opposed to raw cost. This is intended to make results easier to interpret for consumers.	£436	£337	[redacted]
firm_policy_type	Concat(firm, policy_type)	Make sure different firms and vehicle policy type are treated differently. Identify pricing strategy differences across both firms and policies.	[redacted]	[redacted]	[redacted]
cost	claims_paid_core + claims_handling_costs_core	Total insurer paid core claims on the core of the policy.	£210	£0	[redacted]
age_group	Bucket(age)	Grouping ages to match buckets in the UK population and improve sample sizes for distinct groups.	[redacted]	[redacted]	31-65 years old (60%)
tenure_group	Bucket(tenure)	Bucket tenures to improve sample sizes.	[redacted]	[redacted]	0 years (32%)

ecc_core_cost_diff	ecc_core – (claims_paid_core + claims_handling_costs_core)	Used to understand if there were systemic differences in the modelled ECC in relation to actual costs.	£124	£255	[redacted]

Source: Firm data, FCA analysis

There were some smaller limitations in the core insurance pricing data we had available to us, in particular:

- We have data about expected claims costs and actual costs, although we have no data on individual driving behaviour.
- The claims process cannot be evaluated as we do not have this data, this could have an impact on consumer insurance experiences.
- The expected claims costs data used in our analysis was modelled by firms to inform insurance pricing and risk management decisions.
- Policy add-ons, multicar policies & multi-driver policies are not included in the analysis.
- Discounts applied are available in the dataset but have not been evaluated.
- We can observe actual claim amounts and compare to expected claims costs as a proxy for the accuracy of the core pricing component of insurers' models. However, we cannot evaluate actual profit margins of insurers as we cannot account for operating costs or other policy related revenue.

LSOA data on ethnicity

After joining our motor insurance pricing dataset to the census, we observed a slight under-sampling for individuals from minority ethnic backgrounds. To calculate the total proportions in each ethnicity implied from our dataset, we take the ethnicity breakdown of each LSOA and match into the dataset via LSOA. This allows us to calculate the expected ethnicity distribution across the dataset which is outlined in Table 3.

Whilst this is not fully reflective of the general UK population it could reflect several factors, including:

- differential car ownership and motor insurance product holdings between people from different ethnic backgrounds
- the sampling approach for the motor insurance data missing firms who may be more likely to have policyholders who are from minority ethnic backgrounds, although we have no evidence to suggest this has occurred
- reduced survey response rates for certain groups

To test this potential issue, we matched the GIPP motor insurance pricing data with the FCA's Financial Lives 2022 survey dataset. The findings outlined in Table 3 were broadly similar for the matched dataset. As such we found that the FLS and GIPP datasets both appear to undersample groups from minority ethnic backgrounds.[1] When comparing the combined FLS & GIPP distribution with the unmatched FLS distribution, as well as the raw population percentages, we can show that the expected GIPP & Census distribution is in line with our calculated distribution. This is as the FLS sample rate is known, whilst the GIPP motor insurance data is unknown.

Overall, we were comfortable that the undersampling relative to the UK population was not significant enough to negatively impact our core findings.

Table 3: Distribution of ethnicity based on share of LSOA for matched motor insurance pricing data relative to census

Ethnicity	Full Data	Cleaned	Data%	Census%
White	10,997,600	5,340,257	87.6%	81.7%
From minority ethnic backgrounds	1,524,560	754,543	12.4%	18.3%
• **Asian**	774,107	379,895	6.2%	9.3%
• **Black**	293,436	146,189	2.4%	4.0%
• **Mixed Race**	284,298	143,295	2.4%	2.9%
• **Other**	172,719	85,164	1.4%	2.1%
Total	**12,577,993**	**6,094,802**	**100%**	**100%**

Source: ONS data, FCA analysis

Note: Due to not possessing data about the ethnicity of every policyholder and using ethnicity proportions on LSOA level instead, there could be some rounding errors and, as a result, the individual ethnicity numbers might not add up to the total number of observations exactly.

LSOA data on other characteristics

We identified data sources for other area characteristics which may be relevant risk factors which we were able to link to the core motor insurance pricing dataset and use for our core or supplementary analysis. These additional characteristics are outlined in Table 4.

[1] The Financial Lives survey sample has fewer adults from minority ethnic backgrounds than would be expected based on actual population data. This is primarily because adults from minority ethnic backgrounds tend to return lower survey response rates than non-minority ethnic groups. Weighting is used to counteract this issue, but would not be appropriate to use in this instance.

Table 4: List of risk factors used in analysis

Variable Name	Justification
Rural, Urban	Boolean to indicate recorded geography of an LSOA (Urban areas often have higher premiums).
Total Crimes	Crime data for LSOA.
Low Income Percentage	Deprivation data for LSOA.
Number of Casualties	Collisions data for LSOA.
Accidents with Severity 1-3	Collisions data for LSOA.
Number of Vehicles	Vehicles data for LSOA.

Data cleaning

We cleaned the data using a rules-based methodology to be able to track the impact it would have on the various aspects of the results we would need to interrogate. The cleaning rules and their respective impact on data sample size are described in Table 5.

We initially produced eight rules which were related to the variables we intended to use, but distinct from any cleaning required directly by the model. These included:

- ensuring that the policy was not cancelled,
- that there was a single car on the policy, and
- excluding negative values from expected insurance prices.

We also excluded some extreme outliers, as we found no reasonable justification for them to remain in the dataset.

After these steps we applied data quality rules explicitly related to the modelling process (e.g. we needed data on the proportion of residents from minority ethnic backgrounds in an LSOA, we needed an ECC). This resulted in removal of data from outside England and Wales.

Additionally, as some insurance firms acted only as an intermediary, they did not have actual claims costs included in the dataset.

Whilst the majority of tests were passed at an above 90% rate indicating high data quality, we removed approximately half of the data due to the combined effects of removing data for certain countries and lack of claims paid data.

While not included, we again analysed the cleaned data for representativeness, which remained good.

Table 5: Table of data-cleaning rules and their impact on sample size

Condition	Justification	Cumulative Data Removals	Condition Fails	Pass Rate %
All Data		**12,577,993**		
Core price is not null	Allowing only sensible values for core price, which should be always positive. Used in modelling and visualisations upon transformation.	0	0	100
Core price is non-negative		156	156	>99.9
Core price is not zero		200	44	>99.9
Policy is not cancelled	Ensuring working with active policies only.	768486	768316	93.9
Number of cars on the policy is less than 2	Allowing for comparability between different policies.	1855543	1177264	90.6
Policyholder's postcode is available	Needed for modelling when grouping by LSOA level and for visualisations.	1855543	0	100
Policyholder does not reside in the Channel Islands or the Isle of Man	Excluded as different taxation rules apply.	1862114	8750	99.9
Autorenewal status is provided	Needed as a modelling variable and in visualisations.	2408066	574068	95.4
Expected Claims Costs (ECC) are provided	Allowing only sensible values for ECC, which should be always positive. Used in modelling upon transformation and in visualisations.	2490290	87293	99.3
Expected Claims Costs (ECC) are non-negative		5352402	3724224	70.4
Expected Claims Costs (ECC) are not zero		5352924	629	>99.9
Removal of duplicates	Ensuring no sampling bias is introduced in modelling.	5368712	547885	95.6
Removal of outlier prices situated below the 0.25^{th} or above the 99.75^{th} percentile	Ensuring that outlier prices do not exert significant influence on regression coefficients' estimates.	5404770	36058	99.7
Proportion of residents from minority ethnic backgrounds available	Needed for modelling and in visualisations.	6166647	1156099	90.8
Final data used for visualisations		**6166647**		**51.0**
LSOA code is available	Needed for modelling when grouping by LSOA level.	6166647	47129	99.6
Rural-urban geography indicator is provided	Needed for modelling.	6483186	1078416	91.4

Customer’s age group is available	Needed for modelling.	6483187	1	>99.9
Policy tenure group is available	Needed for modelling.	6483191	4	>99.9
Final data used for modelling		**6483191**		**48.5**

Source: Firm data, ONS data, FCA analysis

Annex 2: Methodology

Logarithmic transformation

We chose to apply a logarithmic transformation to key continuous variables, specifically, policy price and expected claims cost, for the following reasons:

- **Improved linearity and model fit:** The relationship between expected claims cost and price is multiplicative rather than additive. Taking logs stabilises the variance and helps linearise this relationship, leading to better model performance.
- **Interpretability of coefficients:** In the log-log specification, coefficients can be interpreted as elasticities — e.g. a 1% increase in ECC is associated with an approximate β% increase in price. For ethnicity, because it is a dummy variable and the functional form is log-linear, a 1 unit increase in the proportion of individuals from minority ethnic backgrounds (that is moving from 0% to 100% of a local area population being from a minority ethnic background) is associated with an approximate (100*β)% increase in price.
- **Reducing skew:** Both price and ECC are right-skewed, with long tails. Logging mitigates the influence of extreme values and makes model residuals more homoscedastic.

Our decision aligns with best practice in applied regression modelling, particularly in settings where outcomes are strictly positive and right skewed.

Regression specification

We chose to use a random effects model to account for differences between LSOAs. Our chosen model has random intercepts only. We attempted to fit the model incorporating random slopes too, however, we found that the variance of the variables representing the slopes was negligible and, hence, we decided not to include them. For a further discussion of the issues presented here, see Bell, Fairweather, and Jones (2019).

Fixed effects models could not be used at our preferred geographic measurement because the local-area ethnicity is only measured every decennial census and so is for our purposes time-invariant. Time-invariant characteristics cannot be included in fixed effects models because they are perfectly collinear with the fixed effects.

Random effects models assume that entity-specific effects are drawn randomly from a population-level distribution, which only makes sense if those entities are exchangeable, that is the modeller does not believe that some subset of the entities really would have very different behaviour from the others. They also assume that the random effects are uncorrelated with observed covariates. This assumption is testable using a Hausman test with a fixed effects model. We conducted the Hausman test using two nested models: a log-linear mixed effects model including the predictors tenure_group, age_group, autorenewal, firm_policy_type, log(ecc_core), urban and minority_ethnic_background with grouping by LSOA, and a fixed effects ordinary least squares (OLS) model including tenure_group, age_group, autorenewal, firm_policy_type and log(ecc_core). The variables urban and minority_ethnic_background were excluded from the OLS model as they are defined at the LSOA level and thus collinear with the fixed effects. The test was performed on the five variables common to both models. Under the null hypothesis, the random effects model is appropriate, assuming no correlation between the regressors and the unobserved LSOA-specific effects. The alternative hypothesis posits that the regressors are endogenous, favouring the fixed effects specification. The resulting p-value from the test was high, indicating insufficient evidence to reject the null. We therefore concluded that the random effects model is appropriate in this case, as the assumption of exogeneity appears reasonable.

The geographic random effects attempt to account for time-invariant differences between areas that are not captured by (and are independent of) ethnicity or other measured local area characteristics included in robustness checks. Exchangeability seems reasonable here because the LSOAs are quite small, whereas it might have been unreasonable if it were measured at the city level.

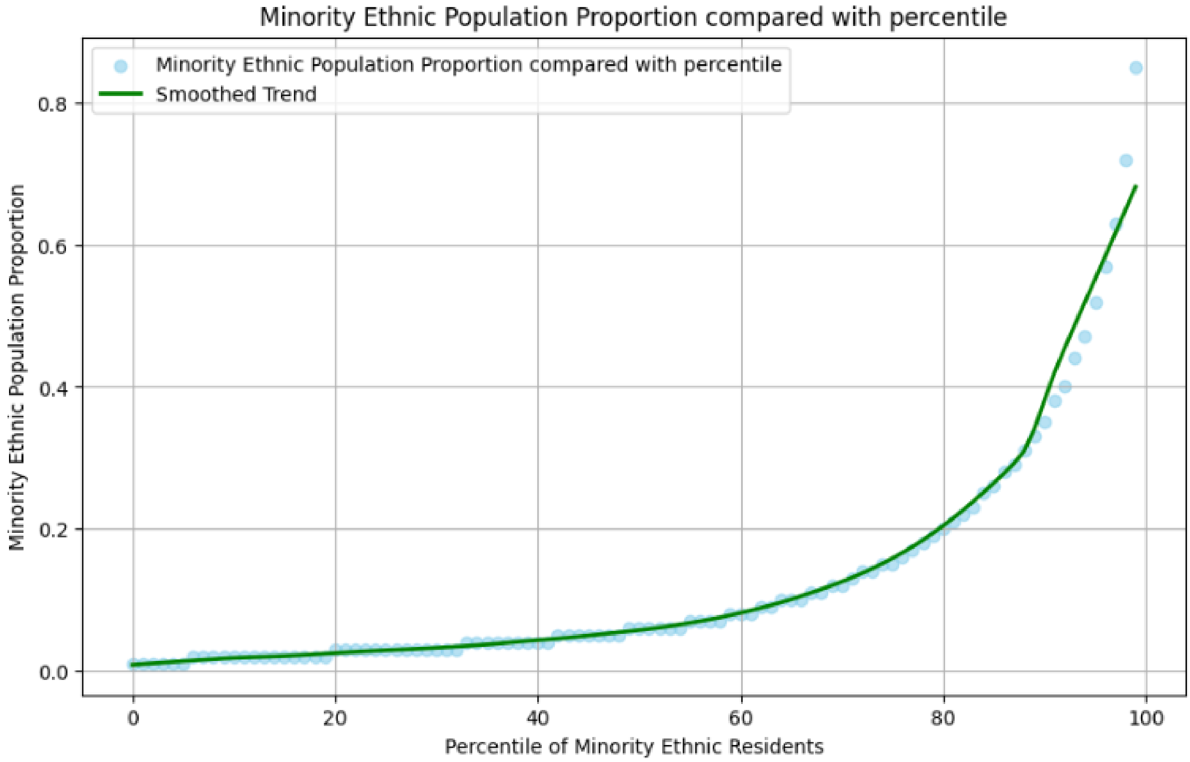
We use normally distributed random effects, though Schielzeth et al. (2020) find that random effects modelling is typically robust to the choice of distribution except in the case that the true distribution is bimodal.

Note that by using this approach we do not explicitly account for non-linearities in relationships in the data.

Annex 3: Additional exploratory analysis

Individuals from a minority ethnic background percentage vs. percentile

Figure 1: Proportion of individuals from minority ethnic backgrounds by respective percentile



Source: ONS data, FCA analysis

The charts presented in this research note use percentile of residents from minority backgrounds as opposed to the total percentage. Figure 1 shows how the proportion of individuals from minority ethnic backgrounds maps to the respective percentile. Nearly half of all areas have less than 10% of their residents being from minority ethnic backgrounds. As the percentile increases the proportion increases rapidly, mostly in urban areas.

Additional analysis on the effect of ECC on insurance premiums and characteristics it correlates with

There are a multitude of factors that may be utilised by insurance firms to inform their pricing decisions. These factors will affect premiums most likely through being used as risk measures to model ECC, but potentially directly depending on pricing models.

Factors could include local area measures, such as crime, gathered from publicly available data, purchased from a third party or collected privately by the firm. It could also include information about the individual requesting insurance, such as:

- claims history
- motoring convictions (and type of motoring convictions)
- vehicle modifications (and what those modifications are)
- length licence held
- telemetry data from a black box in their car (which could be used to price dynamically or at renewal)

Finally, firms may use data based on their own experience of events that could lead to a claim in those areas, and more detailed understanding of the pricing of those incidents.

Where such factors that increase risk for firms are correlated with the proportion of individuals from minority ethnic backgrounds in a local area, this will result in higher average ECC and premiums for those local areas.

Whilst we did not have access to data on many of the factors outlined above, we considered how we could use publicly available local area factors, such as crime, as a control in our regression modelling. However, lack of data for many LSOAs meant this could result in significantly reducing the size of our dataset, which we wanted to avoid to maintain sufficient LSOA representativity and the robustness of our findings.

Instead, we undertook indicative graphical analysis for several factors to understand:

- The relationship between such factors and the proportion of residents from minority ethnic backgrounds in a local area
- The relationship between such factors and ECC

As an indicative example of such analysis consider crime. As highlighted earlier, the share of residents from minority ethnic backgrounds tends to be higher in urban areas. Urban areas also tend to have higher crime. We would expect higher crime to be associated with higher risk for insurers. As such, all other factors held equal, we would expect there to be a higher ECC for such areas, resulting in higher premiums for residents of that area. Figure 2 and Figure 3 shows how this relationship is displayed in the data. Such a relationship is indicative of the type of factor that could explain the premium residual.

As part of our main regression model, we considered ECC as an explanatory factor for the value of insurance premiums. We log-transformed this variable to improve linearity of the model fit. We found that a single-unit increase in the logged ECC leads to an approximately 0.83-unit increase in the logged premium. We also identified that the logged ECC explains 47% of the variation of the logged Gross Insurance Premium and as such is the factor which explains the highest variation amongst all regressors.

As part of our efforts to identify factors which correlate with ECC, we considered the influence of local-area characteristics on ECC as measured at LSOA level. These include:

- the urban/rural geography
- the total number of reported crimes
- the percentage of low-income individuals,
- the number of vehicles,
- the number of casualties and
- the number of accidents with varying severities,

The urban/rural geography of an area was part of the dataset for our main regression model and hence we can provide estimates for its influence on the ECC based on regression results. This suggests that urban/rural geography of an area explains 9% of the variation in ECC.

The rest of the local characteristics were not included in our main regression model due to either:

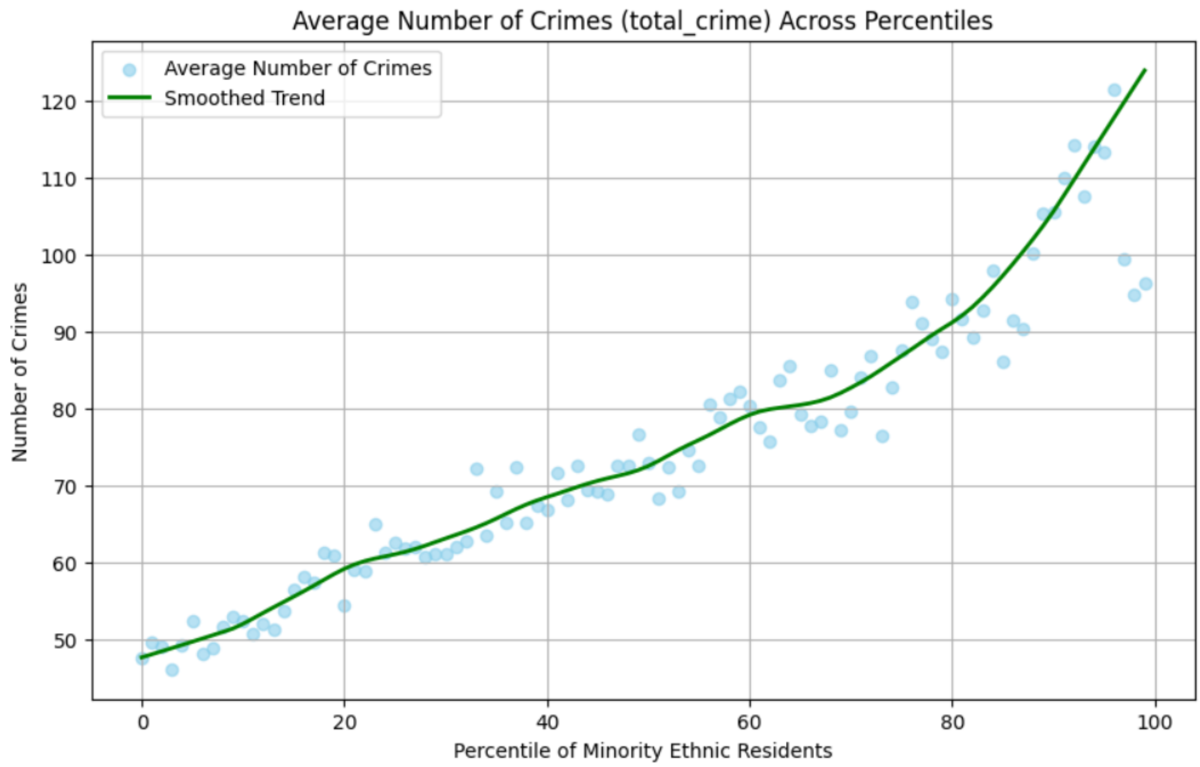
1. Sparseness of the data when measured at an LSOA level i.e. for many LSOAs there was limited information
2. Limited variation of the data across LSOAs, meaning that little additional explanatory power as measured by the Root Mean Squared Error of our model was added from including them i.e. including them did not improve our model accuracy.

As a result, we undertook graphical analysis on the relationship between these variables and ECC. Overall, we found that ECC has a:

- strong positive association with the total number of reported crimes and the percentage of low-income individuals in an area
- unclear or weaker positive associations with all the rest of the local-area characteristics examined, albeit it was unclear if this was a genuine association or driven by data quality issues given the sparseness of these variables

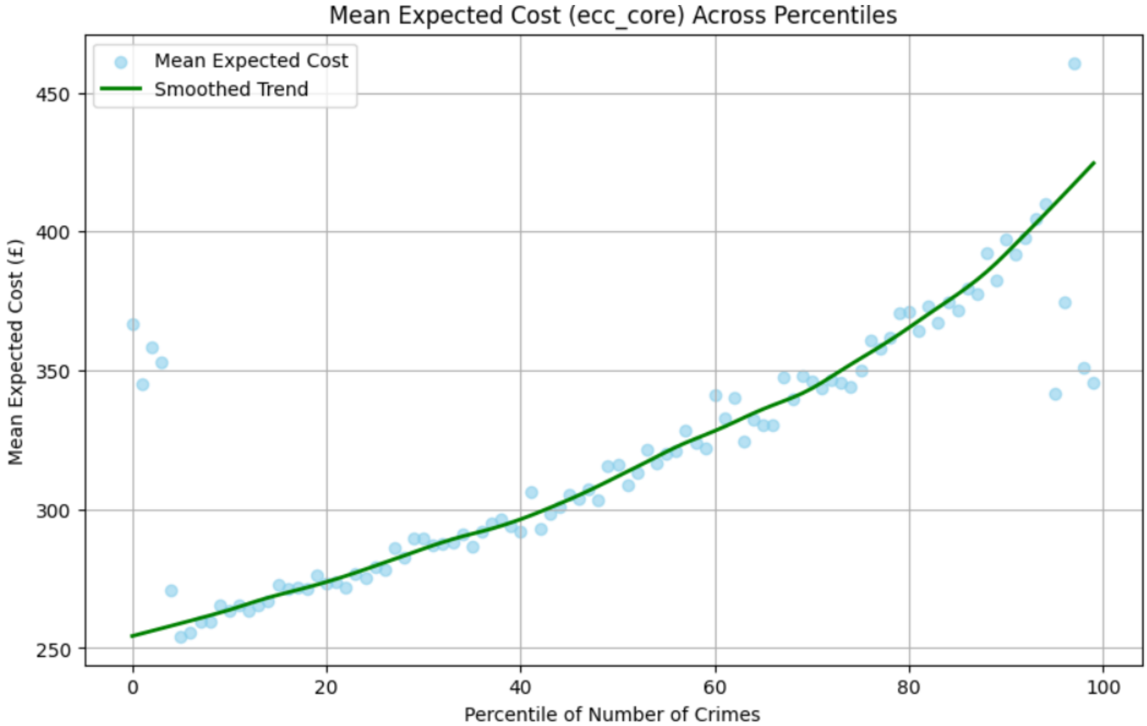
We also considered the influence of individual characteristics on ECC. These included the policy's tenure group, the age group of the policyholder, presence of autorenewal and the firm-policy type. Explanation of these variables can be found in Annex 2: Methodology. Overall, we found that they account for 20% of the variation in ECC.

Figure 2: Average number of crimes across LSOA by percentile of proportion of individuals from minority ethnic backgrounds



Source: ONS data, FCA analysis

Figure 3: Average Expected Cost by percentile of total number of reported crimes



Source: Firm data, ONS data, FCA analysis

Annex 4: Regression analysis

Table 6: Regression table of a Linear Mixed Effects Model with log(price_core_ipt) as a response variable and tenure_group, age_group, autorenewal, firm_policy_type, urban, non_white, log(ecc_core) as regressors. Grouping is done on geography_code.

Model:	MixedLM	Dependent Variable:	log_price_core_ipt
No. Observations:	6094802	Method:	ML
No. Groups:	33647	Scale:	0.0665
Min. group size:	8	Log-Likelihood:	-409408.7165
Max. group size:	562	Converged:	Yes
Mean group size:	181.1		

	Coef.	SE	z	P>|z|	[0.025	0.975]
Intercept	1.504	0.002	714.177	0.000	1.500	1.509
C(tenure_group)[T.0]	-0.133	0.001	-156.411	0.000	-0.135	-0.132
C(tenure_group)[T.1-2]	-0.065	0.001	-76.778	0.000	-0.067	-0.064
C(tenure_group)[T.10-14]	0.098	0.001	95.649	0.000	0.096	0.100
C(tenure_group)[T.15-19]	0.128	0.001	89.879	0.000	0.125	0.130
C(tenure_group)[T.20-24]	0.171	0.002	93.263	0.000	0.167	0.175
C(tenure_group)[T.25+]	0.160	0.002	79.144	0.000	0.156	0.164
C(tenure_group)[T.3-5]	0.011	0.001	13.212	0.000	0.010	0.013
C(tenure_group)[T.6-9]	0.073	0.001	79.888	0.000	0.072	0.075

C(age_group)[T.25-30]	-0.073	0.001	-85.315	0.000	-0.074	-0.071
C(age_group)[T.31-65]	-0.139	0.001	-181.440	0.000	-0.141	-0.138
C(age_group)[T.66-70]	-0.172	0.001	-199.556	0.000	-0.173	-0.170
C(age_group)[T.71+]	-0.140	0.001	-173.570	0.000	-0.142	-0.139
C(autorenewal)[T.1.0]	0.004	0.000	12.467	0.000	0.003	0.004
C(firm_policy_type)[T.1]	0.399	0.013	29.687	0.000	0.373	0.425
C(firm_policy_type)[T.2]	0.281	0.003	87.113	0.000	0.275	0.287
C(firm_policy_type)[T.3]	0.076	0.001	66.919	0.000	0.074	0.078
C(firm_policy_type)[T.4]	0.656	0.030	21.536	0.000	0.597	0.716
C(firm_policy_type)[T.5]	-0.001	0.002	-0.454	0.650	-0.005	0.003
C(firm_policy_type)[T.6]	-0.059	0.001	-49.100	0.000	-0.062	-0.057
C(firm_policy_type)[T.7]	0.033	0.001	28.884	0.000	0.031	0.035
C(firm_policy_type)[T.8]	-0.200	0.002	-119.391	0.000	-0.203	-0.196
C(firm_policy_type)[T.9]	-0.050	0.002	-23.128	0.000	-0.054	-0.046
C(firm_policy_type)[T.10]	0.241	0.002	102.302	0.000	0.236	0.245
C(firm_policy_type)[T.11]	-0.077	0.001	-68.778	0.000	-0.079	-0.075
C(firm_policy_type)[T.12]	0.145	0.002	70.529	0.000	0.141	0.149
C(firm_policy_type)[T.13]	-0.071	0.001	-62.976	0.000	-0.074	-0.069

C(firm_policy_type)[T.14]	-0.535	0.009	-61.330	0.000	-0.553	-0.518
C(firm_policy_type)[T.15]	0.084	0.001	72.610	0.000	0.082	0.086
C(firm_policy_type)[T.16]	-0.365	0.001	-325.166	0.000	-0.367	-0.362
C(firm_policy_type)[T.17]	-0.272	0.001	-202.554	0.000	-0.275	-0.270
C(firm_policy_type)[T.18]	-0.012	0.001	-10.349	0.000	-0.014	-0.010
C(firm_policy_type)[T.19]	0.145	0.001	99.349	0.000	0.142	0.148
C(firm_policy_type)[T.20]	-0.279	0.001	-241.180	0.000	-0.282	-0.277
C(firm_policy_type)[T.21]	-0.184	0.001	-153.606	0.000	-0.187	-0.182
C(firm_policy_type)[T.22]	1.105	0.001	798.499	0.000	1.103	1.108
C(urban)[T.True]	0.019	0.001	34.030	0.000	0.018	0.020
non_white	0.178	0.001	148.262	0.000	0.176	0.181
log_ecc_core	0.832	0.000	4091.747	0.000	0.832	0.833
Group Var	0.001	0.000				

Annex 5: References

Bell, A., Fairbrother, M. and Jones, K. (2019). Fixed and random effects models: making an informed choice. *Quality & Quantit*y, **53**, 1051–1074. https://doi.org/10.1007/s11135-018-0802-x

Bogiatzis-Gibbons, D., Charles, L., Dewing, H., Gretschel, C., Jomy, M., Reid, A., and Slack, R. (2024). Research Note: A literature review on bias in supervised machine learning. *FCA Research Notes*, https://www.fca.org.uk/publication/financial-lives/fls-2024-general-insurance-protection.pdf.

Bogiatzis-Gibbons, D., Lawlor, L., Charles, L., Limani, B., O'Brien, E., and Staniland, S. (2025). Research Note: An empirical analysis of pricing differences by demographic characteristics in the UK mortgage market. *FCA Research Notes*, https://www.fca.org.uk/publication/financial-lives/fls-2024-general-insurance-protection.pdf.

Department for Transport (2025). Ministers bring together industry experts and consumer champions to tackle spiralling costs for drivers. Ministers bring together industry experts and consumer champions to tackle spiralling costs for drivers - GOV.UK.

FCA (2025a). EP25/2: An evaluation of our general insurance pricing practices (GIPP) remedies.

FCA (2025b). EP 25/2 Technical annex: Evaluation of our General Insurance Pricing Practices (GIPP) remedies.

FCA (2025c). Roadmap for retail insurance – empowering consumers and strengthening markets | FCA.

FCA (2025d). Financial Lives 2024 survey General insurance & protection. https://www.fca.org.uk/publication/financial-lives/fls-2024-general-insurance-protection.pdf.

Lukacs, P., Neubecker, L., and Rowan, P. (2016). Price discrimination and cross-subsidy in financial service. https://www.fca.org.uk/publication/financial-lives/fls-2024-general-insurance-protection.pdf.

Maughan, C., San, J., and Rahman, L. (2024). Popping the Bonnet: Exploring affordability issues in car insurance and the ethnicity penalty. *Citizens Advice Reports*, https://www.fca.org.uk/publication/financial-lives/fls-2024-general-insurance-protection.pdf.

Schielzeth, H., Dingemanse, N.J., Nakagawa, S, et al. (2020). Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution*, **11**: 1141–1152. https://doi.org/10.1111/2041-210X.13434.

Starks, M., Reynolds, G., Gee, C., Burnik, G., and Lachlan, V. (2018). Price discrimination in financial services: How should we deal with questions of fairness? *FCA Research Notes*, https://www.fca.org.uk/publication/research/price_discrimination_in_financial_services.pdf.

12 Endeavour Square, London E20 1JN
Telephone: +44 (0)20 7066 1000 Website: www.fca.org.uk