Financial Conduct Authority

Research Note

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

Glossary

Term	Acronym	Definition
Actual Claims Costs	ACC	The amount of money paid out in claims on a policy following a claim being made and accepted as relating to the core part of the premium i.e. not including add-ons. As reported by insurers or brokers as part of our data request.
Expected Claims Cost	ECC	A modelled output produced by insurers which seeks to estimate the expected cost in claims of a policy based on underlying risk factors. As reported by insurers as part of our data request.
General Insurance Pricing Practices	GIPP	A review undertaken by the FCA to understand whether current pricing practices in home and motor insurance support effective competition and lead to good consumer outcomes.
Index of Multiple Deprivation	IMD	The Index of Multiple Deprivation ranks areas in England based on factors like income, employment, health, education, crime, housing, and environment to measure relative deprivation.
Local Super Output Area	LSOA	A geographic area used in England and Wales for the reporting of small area statistics, particularly from the census.
Office for National Statistics	ONS	A non-ministerial department responsible for the collection and publication of statistics related to the economy, population and society of the United Kingdom.
Premium		The amount of money collected by insurers for the core premium of a motor insurance policy, not including premiums related to policy add-ons, multicar policies & multi-driver policies. As reported by insurers as part of our data request.

Overview

Purpose

A well-functioning retail insurance market helps consumers navigate their financial lives, provides peace of mind, and supports growth through the effective management of risk. Motor insurance is an essential financial product, with around 34.7 million product holders in the UK as of 2024 (FCA, 2024).

The consumer group Citizens Advice has previously raised concerns that people living in areas with higher numbers of residents from minority ethnic backgrounds are paying more for motor insurance. In response to the Motor Insurance Taskforce, the FCA outlined that it would analyse the impact of insurance prices on different customer groups, including those from minority ethnic backgrounds. In this Research Note, we used a large, detailed dataset which contained information on policies, premiums and actual and expected claims costs to provide the most comprehensive analysis possible on what drove differences in motor insurance premiums between areas with differing proportions of individuals from minority ethnic backgrounds.

Key findings

Overall: There are large differences in the overall costs of motor insurance between areas of high and low numbers of residents from minority ethnic backgrounds. However, this difference is largely associated with differences in expected and actual claims costs (indicators of risk). In turn, that may be associated with risk factors correlated with higher proportions of residents from minority ethnic backgrounds, such as living in urban areas.

Regression Analysis: Using regression modelling (a way of adjusting for how local areas and individuals vary across multiple characteristics), we compare similar local areas and individuals based on the core premium a person pays for motor insurance after accounting for expected claims costs, rural-urban geography, individual age, and policy-related features. When using this technique, we find that the adjusted difference between areas with more or fewer residents from minority ethnic backgrounds is relatively small.

Size of Risk-Adjusted Price Differences: After accounting for these factors, we are left with a statistically significant but much smaller residual difference. For example, in Luton which had one of the largest residual differences and where people from minority ethnic backgrounds make up 43% of the population, the residual difference is only £28 on a total premium of £627: substantially less than the almost £300 difference when not accounting for those factors.

Explanations for this residual: We cannot be certain because we did not have access to the full algorithms or set of risk factors used by insurers. However, this smaller residual could be explained by other operating costs, uncertainty premia or other local-area or individual-level risk factors not accounted for in expected claims costs modelling.

1 Background

Regulatory context

The Consumer Duty sets high and clear standards of protection for retail customers across financial services, and came into force on 31 July 2023, and on 31 July 2024 for closed products and services. The Duty sets out that firms must regularly assess, test, understand and evidence the outcomes their customers are receiving. The Duty also supports firms' compliance with legal requirements under the Equality Act 2010 and equivalent legislation. Firms are required to monitor if any group of customers is getting different outcomes than other customers and take appropriate action if they do.

In addition to the Consumer Duty, the PROD 4 rules also require firms to monitor price and value outcomes for customer groups for insurance products. Where a firm is subject to PROD 4, it may use any monitoring or reviews it carries out under PROD 4 to comply with its monitoring obligations under the Consumer Duty.

The FCA had previously asked major motor insurers to explain how they were satisfied that their pricing did not result in unlawful discrimination. They explained that they do not collect data on race or ethnicity, nor do they use such data in pricing decisions. They also described the steps they took to ensure their pricing decisions did not unlawfully discriminate. The information provided did not indicate to us that illegal price discrimination was taking place. However, to better understand outcomes customers were receiving, we wanted to examine firms' pricing data to ascertain the extent of any difference in premiums for local areas with higher numbers of residents from minority ethnic backgrounds.

Literature review

Existing publicly available empirical research into how motor insurance pricing affects different ethnic groups in the UK is limited. Citizens Advice (CitA) produced some initial analysis in 2022. The latest version of their analysis published in 2024 took data on 20,495 individuals who received debt advice and provided information on their car insurance premiums and ethnicity. For this specific group of consumers, they conducted a regression analysis to understand the difference in price for those from minority ethnic backgrounds after controlling for gender, age, income and disability. They did not include data on claims or other potential risk factors such as location. Their analysis suggested a £307 difference in the average premium for people from minority ethnic backgrounds who sought debt advice after controlling for a limited set of demographic features. Importantly, they could not control for actual or expected claims costs.

Whilst the FCA has not undertaken analysis of differing pricing outcomes on the basis of ethnicity, the FCA has produced broader research on differences in pricing or prediction outcomes between different groups of consumers (see e.g. Lukacs et al. 2016, Starks et al. 2018, Bogiatzis-Gibbons et al. 2024, and Bogiatzis-Gibbons et al. 2025).

2 Data & research design

Data on insurance policies

The FCA collected real-world transaction data on motor insurance policies as part of the General Insurance Pricing Practices (GIPP) evaluation. The data was drawn from insurers representing 57% of the motor insurance market by Gross Written Premium (GWP) in 2022. A detailed outline of that data is provided in the evaluation paper.

In total the dataset contained 12.6 million observations for policies providing coverage between 2019 and 2024 Q1, of which 10.6 million were from England and 0.7 million from Wales. The dataset contained data on a range of features regarding policy premium, costs and characteristics as well as the policyholder's age and location (see Annex 1: Data preparation for a full list of variables).

In line with best practice, we implemented data quality rules to ensure the data was suitable for our analysis. We provide a detailed overview of all 18 data quality rules in Table 1 of Annex 1: Data preparation. The rules aimed to ensure policies would be useable (e.g. no missing ECC), and that policies would be relevant (e.g. policy was not cancelled). After these data quality rules were applied, we were left with 6.1 million policies.

Our analysis and review of the data revealed two main potential limitations which we could not otherwise control for:

- There was some sample bias in the data set with underrepresentation of policies in areas with a higher proportion of residents from minority ethnic backgrounds (see Annex 1: Data preparation).
- Insurers may have interpreted guidance differently for certain variables leading to differences across the dataset. We judged this most likely to occur for actual claims costs paid, where we had a much higher rate of missing observations.

In addition, there were some smaller data limitations which we do not believe would have substantial impacts on our findings but for transparency we outline in Annex 1: Data preparation.

Overall, even after accounting for the potential impact of the limitations outlined, the high level of data quality after applying our cleaning rules means we are confident in our findings.

Data on local area ethnicity and choice of grouping

The data on individuals' insurance policies does not contain information on ethnicity or local-area risk factors that might determine insurance premiums. To proxy individuals' ethnicity, we used data from the 2021 UK census on the ethnic composition of their local area. This data provided us with the proportion of individuals from each of the five main ONS (Office for National Statistics) defined ethnic group categories (Asian, Black, Mixed, Other, White) for Local Super Output Areas (LSOAs). We matched to data on motor insurance policies using postcode (see Annex 1: Data preparation).

LSOAs are a local area defined by the ONS for the purpose of data collection, management, and analysis. Each LSOA represents around 1,000 to 3,000 residents, comprising 400 to 1,200 households. We chose to use LSOA as a consistent measure of size for areas as it was the lowest level of geography for which consistent data on ethnicity could be sourced. Whilst it is in an imperfect proxy for individual data, the relatively granular size gives us confidence in our measure of ethnicity. However, due to differences in measurement across the UK the analysis presented in this paper only looks at England and Wales as:

- Scotland has its own geographical area definitions which do not align with the LSOA area splits used in England and Wales.
- Northern Ireland uses different ethnicities in its census making it difficult to compare to England and Wales, it also lacked data on some local area factors we explored.

For the purposes of our analysis, we grouped the information on ethnicity into a binary classification: white and those from minority ethnic backgrounds (consisting of Asian, Black, Mixed and Other). While this combines people from different ethnic backgrounds, very low sample sizes at an LSOA level for some ethnic groups made it infeasible to reliably estimate differences in their premiums. Using this grouping allows us to analyse overall average differences across individuals from white and minority ethnic backgrounds. To ensure robustness, we also tested findings through exploratory analysis using ungrouped data. These results did not alter the conclusions of our analysis.

To measure local-area characteristics that might determine risk, we sourced additional data from the ONS and other Government sources at an LSOA level on urbanity, crime, and deprivation. Similarly, these were linked to insurance policies via postcode.

As this LSOA level data was primarily collected from government sources and the ONS the data quality was high with limited processing required.

Analytical approach

Our analysis was undertaken in two strands:

1. Descriptive analysis and visualisation to understand general patterns and trends in insurance premiums and costs between different local areas.
2. A regression analysis using a linear mixed-effects approach to understand how prices between local areas differed after controlling for expected claim costs, a limited set of local area factors for which we have data, and other unobservable characteristics of the local area (via random effects modelling). In lay terms, this means that we adjust for both observed differences in individuals, like how risky they were predicted to be when buying a policy as measured by ECC, and both observed and unobserved differences in local areas. We do this to try and make like for like comparisons between policies.

Table 1 summarises the model and which features were used. A detailed overview of our regression model and our reasoning is outlined in Annex 2: Methodology.

Table 1: Summary of the response and explanatory variables used in the main regression model

	Fixed effects	Random effects
Log-transformed Premium	• log-transformed Expected Claims Costs (ECC) • LSOA-level proportion of residents from minority ethnic backgrounds • LSOA-level rural-urban geography • customer's age group • policy tenure • policy autorenewal status • firm-policy type (a joint variable specifying a policy type within each firm)	• LSOA (random intercept only)

Source: FCA analysis

Overall, we are confident in the robustness of this approach given the data available as described above. However, as with any empirical analysis there are some limitations.

Ideally, our model would include data on all relevant individual and area risk factors. However, this is not reasonably practicable as not all data is available to us. Instead, we use firms' own ECC as a proxy for these factors supplemented by additional variables on some other local area risk factors. These additional variables include:

- whether a local area is rural or urban,
- a customer's age group, and
- policy details such as tenure, autorenewal status and policy type.

We also sought to use data on crime, accidents and deprivation as measured by IMD in our main model. However, the addition of such data had a limited effect partly due to data limitations and sparseness. Additional analysis examining the relationship between these factors and other key variables in our analysis is outlined in Annex 3: Additional exploratory analysis. To account for unobserved differences between local areas (those for which data is not available to us) we introduce random effects into our modelling.

In addition, we evaluate how prices and differences in prices between groups are impacted by differences in claims cost and other factors listed above, but not the underlying drivers or causes of differences in risk factors between groups with different characteristics. Such analysis would require much more detailed data and an appropriate quasi-experimental design to allow us to make causal estimates. As such our analysis is descriptive and not causal.

3 Main Results

Without adjusting for risk, there are large differences in premiums by proportion of individuals from minority ethnic backgrounds

We began our analysis by examining how average policy premiums and claims costs differ across LSOAs based on the proportion of individuals from minority ethnic backgrounds within them. In all cases sample sizes were lower for LSOAs with higher proportions of individuals from minority ethnic backgrounds as few of these areas exist. For example, 69.4% of LSOAs have an average or less share of residents from minority ethnic backgrounds and 90% have a less than 48.5% share of residents from minority ethnic backgrounds.

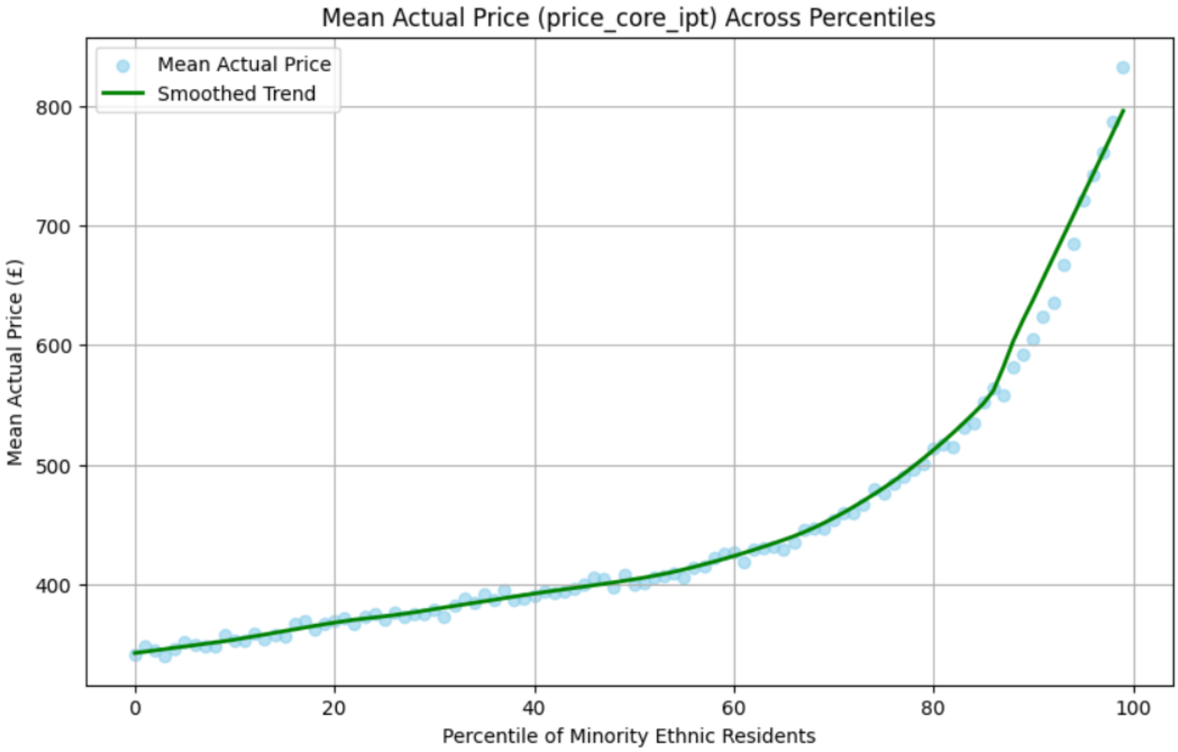
Note that for charts presented in this section we are using percentile of individuals from minority ethnic backgrounds as opposed to percentage. These graphs go from approximately 0% of individuals from minority ethnic backgrounds to around 80% of individuals from minority ethnic backgrounds (see Annex 3: Additional exploratory analysis for more explanation). This was chosen to make certain that the values provided have a real-world basis behind them, as opposed to extrapolating to areas with proportions of individuals from minority ethnic backgrounds which don't exist. Figure 1 shows that average premiums increase as the proportion of individuals from minority ethnic backgrounds in an LSOA increases.

However, there are also large differences in risk (as measured by expected claims costs)

A key determinant of the premium paid by the consumer is the risk cost to the insurer. Firms calculate this cost through modelling the ECC for each policy. This model will include a range of relevant risk factors based on individual, vehicle, and area characteristics. Figure 2 shows that like average premiums, the average ECC of a policy is higher in areas with a larger proportion of individuals from minority ethnic backgrounds.

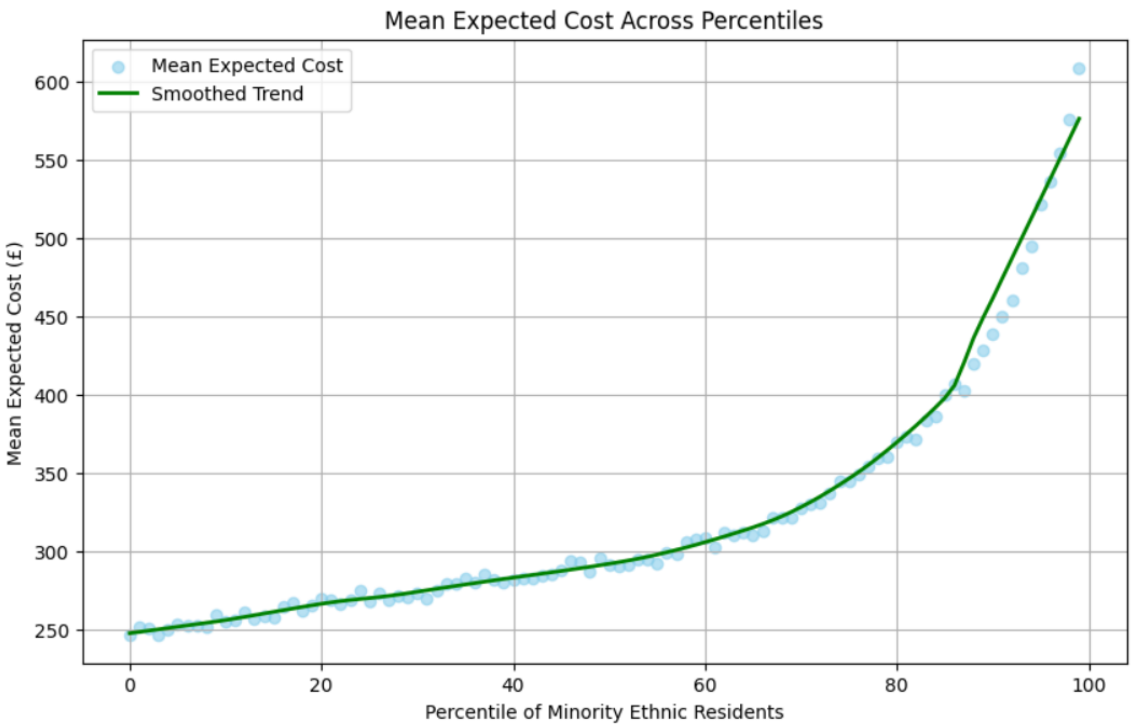
In Figure 3, we show that the same trend is generally true of actual claims costs (ACC). We explore the robustness of our results to expected claims costs as a proxy for actual claims costs in the section on Further Results.

Figure 1: Average premiums across LSOA by percentile of proportion of individuals from minority ethnic backgrounds



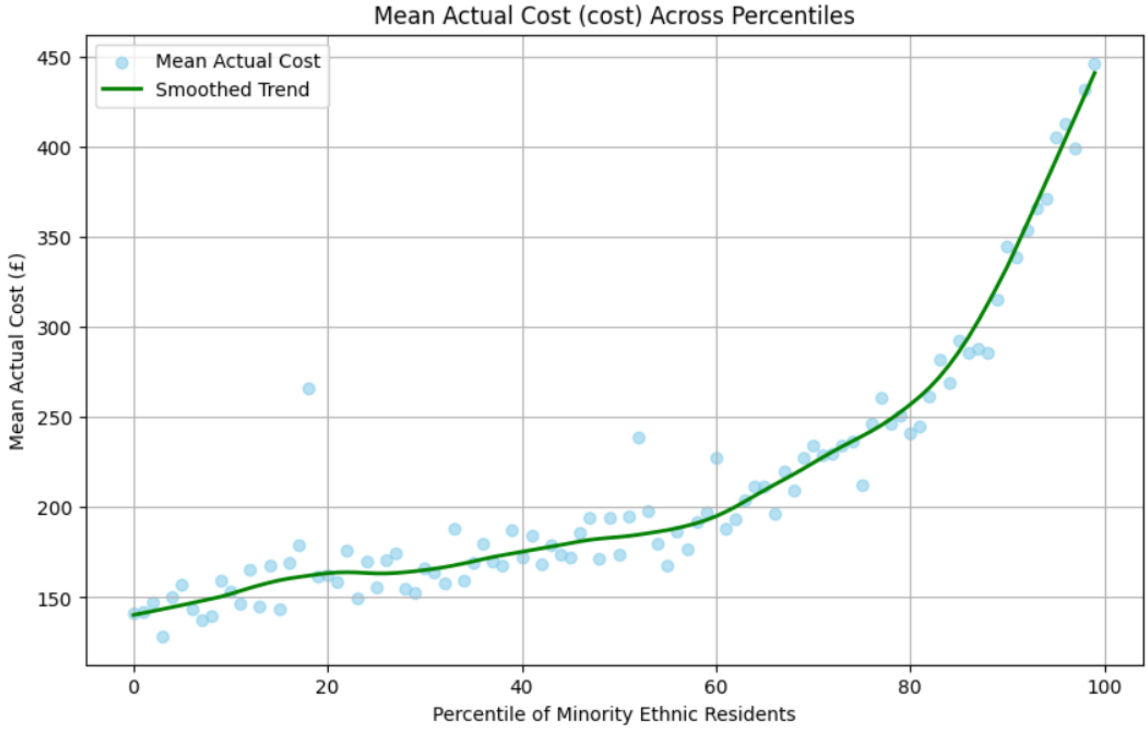
Source: Firm data, ONS data, FCA analysis

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



Source: Firm data, ONS data, FCA analysis

Figure 3: Average ACC across LSOA by percentile of proportion of individuals with minority ethnic backgrounds



Source: Firm data, ONS data, FCA analysis

When accounting for individual risk, and other local area characteristics using a random effects model, the differences are much smaller

To understand how pricing is associated with ethnicity of a local area we ran a regression analysis comparing premiums across policies. This allowed us to compare premiums while accounting for a range of factors for which we had data, including expected claims costs and some individual, firm and area level characteristics.

Whilst we were unable to control for all relevant risk factors given data availability, our model was well specified and had a relatively high measure of fit with an R- squared of c.83%. This means that our model explains about 83% of the variation in the log-transformed premium variable we use as our outcome.

Overall, our regression model showed that:

- Controlling for ECC and a limited set of local area factors for which we have data explained most of the difference in average price between local areas with different proportions of individuals from minority ethnic backgrounds
- There remained a small positive, statistically significant effect of the percentile of individuals from minority ethnic backgrounds in a local area on average motor insurance premiums.

To display the results of our analysis more clearly, we have grouped local areas to show how controlling for ECC and some additional risk factors explains most of the difference in premium. Figure 4 displays the same chart as Figure 1 but with highlighted estimates for selected cities. For these cities Table 2 outlines the:

- share of residents from minority ethnic backgrounds
- average premium for policyholders in our dataset
- estimated average premium implied by our regression after controlling for ECC, the rural-urban geography of the area, age, policy tenure, autorenewal status, firm, policy type, and other unobservable characteristics of the local area (via random effects modelling) compared to a hypothetical area which was entirely white

In all cases, the difference in prices between areas is broadly explained by ECC, which is closely correlated with the actual claims cost, and the additional factors we can control for, with a small residual left.

We cannot be certain what causes this residual. There exist multiple other factors which firms will consider as part of their pricing but which we have been unable to include in our analysis. This could explain the difference where such factors are correlated with the proportion of residents from minority ethnic backgrounds in a local area. Such factors could include:

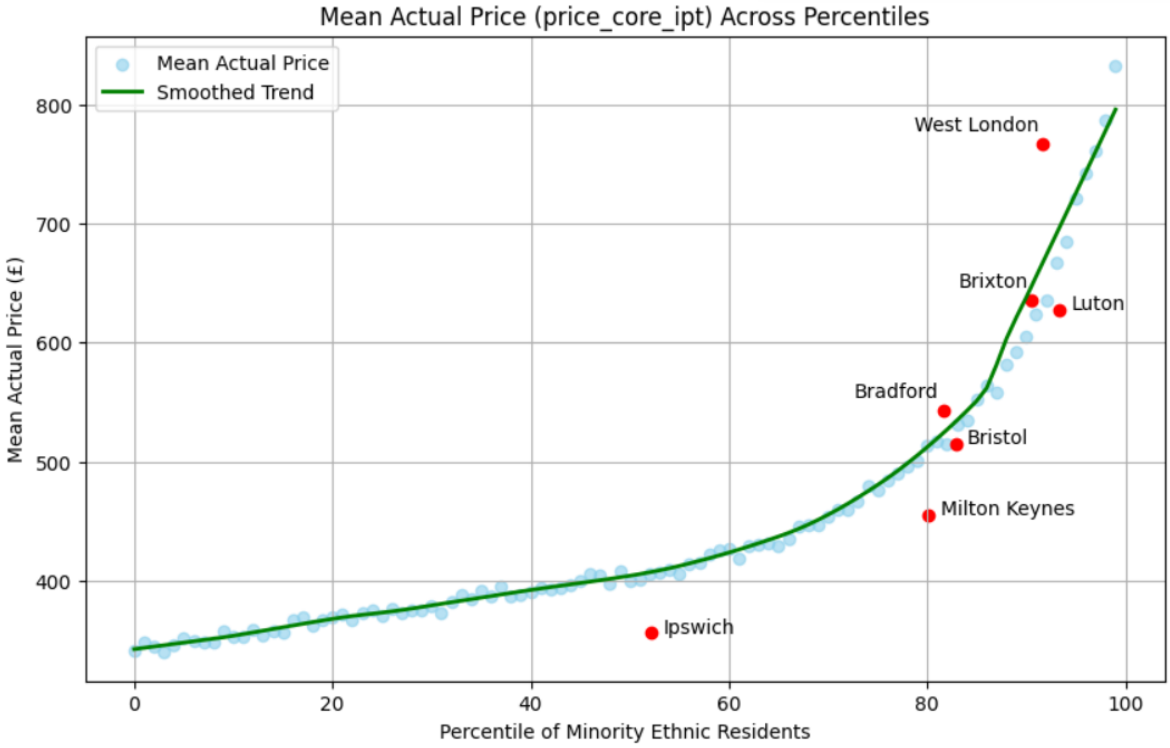
- Other costs relevant to a policy but which we were unable to include in our analysis due to lack of data, such as operational expenses or the allocation of fixed overhead costs
- Differences in risk estimation due to scarcity of data or concentrated policy holdings resulting in an uncertainty premium being applied
- Other risk-related factors not captured in expected claims costs.

As an example of how such factors can result in higher premiums when correlated with the proportion of individuals from minority ethnic backgrounds consider the effect of living in an urban area, which we controlled for in our analysis. People from minority ethnic backgrounds are more likely to live in urban areas. Urban areas are more likely to experience higher crime and therefore present higher risk cost to insurers. This would translate through to higher expected claims costs and therefore higher average prices for such consumers.

Further detail on the indicative example above is outlined in Annex 3: Additional exploratory analysis, alongside details on:

- local area factors which we were unable to include in our regression modelling due to sparsity of data available but for which we undertook exploratory analysis of the relationship with both the proportion of individuals from a minority ethnic background in a local area and ECC
- the associations between other factors included in our regression modelling and ECC

Figure 4: Average premium across LSOA and selected cities by proportion of individuals from minority ethnic backgrounds



Source: Firm data, ONS data, FCA analysis

Table 2: Average and implied premium across selected cities after controlling for ECC and some additional risk factors

City	Share of individuals from minority ethnic backgrounds	Average premium	Implied premium from regression analysis	Unexplained Difference
Ipswich	6%	£356	£352	£4
Milton Keynes	19%	£455	£443	£12
Bradford	21%	£543	£530	£13
Bristol	22%	£515	£501	£14
Brixton	35%	£635	£613	£22
West London	38%	£767	£743	£24
Luton	43%	£627	£599	£28

Source: Firm data, ONS data, FCA analysis

4 Further Results

In this section, we detail further analysis we conducted, first on the robustness checks we did to ensure that our results are valid, and second, more limited analytical work relating to age of the policyholder.

Robustness of results to using expected claims costs as a proxy for actual claims costs

In Section 3 we showed the relationship between premiums, ECC and ACC. For our regression analysis we used ECC as it was more complete in our dataset. We wanted to check whether it is valid to use ECC as a proxy for individual risk.

The ECC attempts to forecast the potential cost of claims faced by insurers. Subject to how a firm models ECC and defines the costs of claims, we would broadly expect ECC to track with ACC i.e. the amount of money paid out by insurers on claims, where the expected risk is in line with actualised risk. Comparing Figure 2 and Figure 3, we see ACC broadly tracks ECC, albeit with more variation as they are an actual and not modelled outcome – a pattern that typically exists across modelled/predicted and actual outcomes.

Figure 5 shows the average difference between ECC and ACC. There was a positive but broadly constant difference across LSOAs regardless of the proportion of individuals from minority ethnic backgrounds up to the 90th percentile. There was a slightly larger difference after this, predominantly driven by some outlier values at the end of the distribution. This could be caused by several potential factors, and the actual claims costs not being fully realised. If the impact of this is not evenly distributed across claims, it could lead to differences between ACC and ECC. We are aware this could be an issue but are unable to say by how much it may affect our dataset.

As outlined earlier, data quality for ACC was poorer than other variables due to missing data. Claims costs paid data was missing or 0 for a considerable number of cases. As such we were not always able to identify whether a claim was made or not. This may skew our data on ACC where missing data was in fact a claim which was not made or paid out. This would be more impactful for LSOAs with smaller sample sizes where individual claims have more impact than LSOAs for which sample sizes are larger.

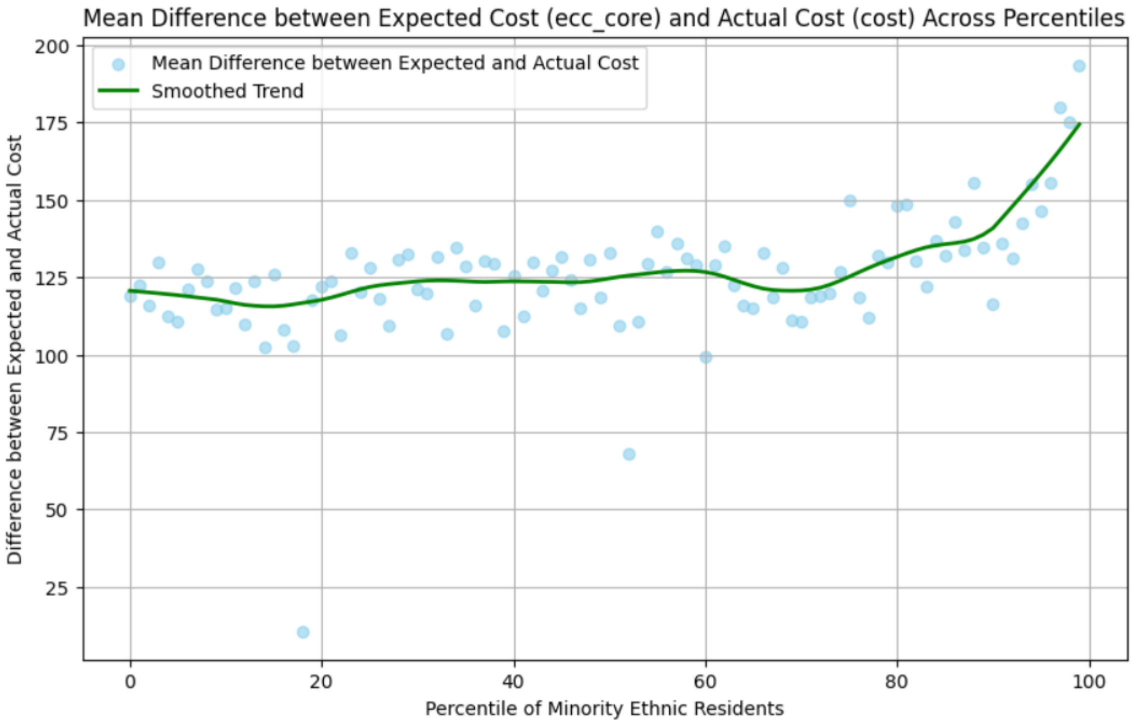
To further test this we conducted analysis on claims paid after removing 0 values. This is shown in Figure 6. When doing this we found that the pattern in Figure 5 was inversed, with ACC increasing faster than ECC as the share of individuals from minority ethnic backgrounds increased. This indicated a higher average claim paid value when a claim is paid out for such areas, which would be in line with higher premiums. Without additional information, however, we cannot distinguish between the different reasons that a 0 may be included in the dataset (including time delays, lack of reporting, and 0 value claims). This means we cannot be certain 0 values are appropriate in this instance and represent a "true" 0.

There may be additional factors which are incorporated into the ECC but do not fall into the definition of ACC as part of our data request. For instance, in some cases additional variable costs or allocation of fixed costs may be incorporated as an expense in ECC but may not be included in the ACC, although a cost will have been incurred. We are unable to account for these as we do not have additional data.

Whilst there is some uncertainty given the factors outlined above about the relationship between ACC and ECC, we believe that the broader general trend between ECC and ACC substantiates our approach to using ECC in our analysis.

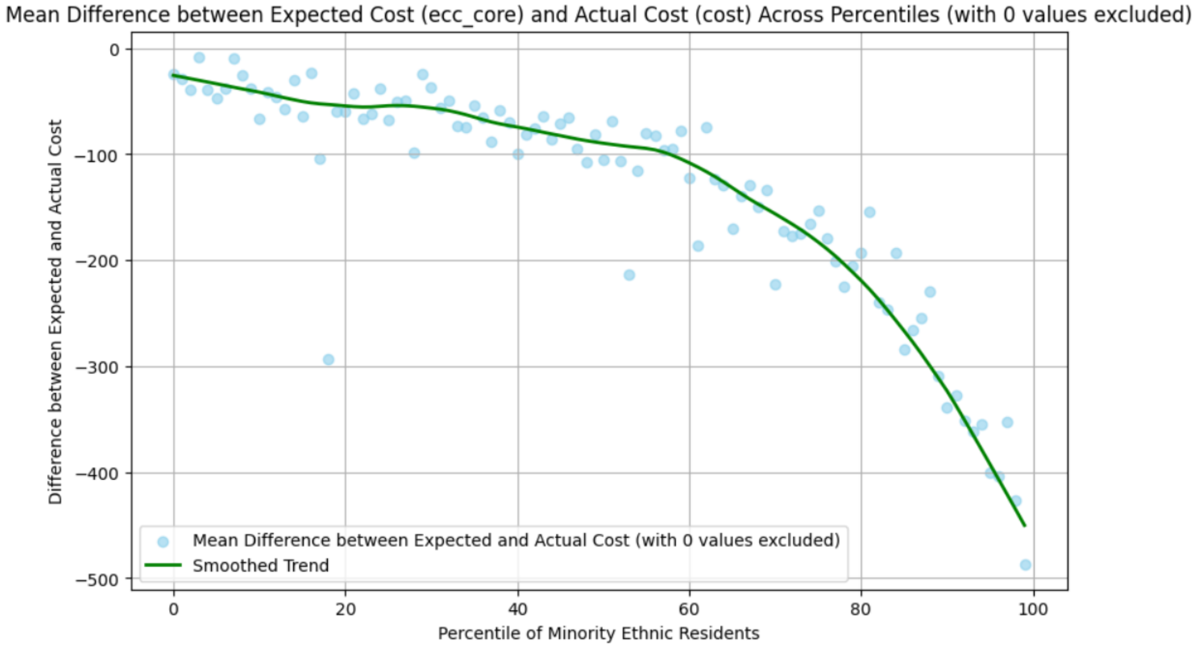
As an additional test of the relationship, Figure 7 shows the average difference between ECC and ACC as a percentage of the ECC. In this case we see that the difference between the two declines in proportional terms as the share of individuals from minority ethnic backgrounds in an area increases. This reflects that whilst the absolute difference between ECC and ACC increases, the overall increase in ECC is larger for these areas. As a result, the proportional difference between ECC and ACC declines. Whilst this does not allow us to determine drivers of absolute differences, it suggests that whatever is driving the difference is having a proportionally reduced effect as the share of individuals from minority ethnic backgrounds increases.

Figure 5: Average difference between ECC and ACC across LSOA by percentile of proportion of individuals from minority ethnic backgrounds



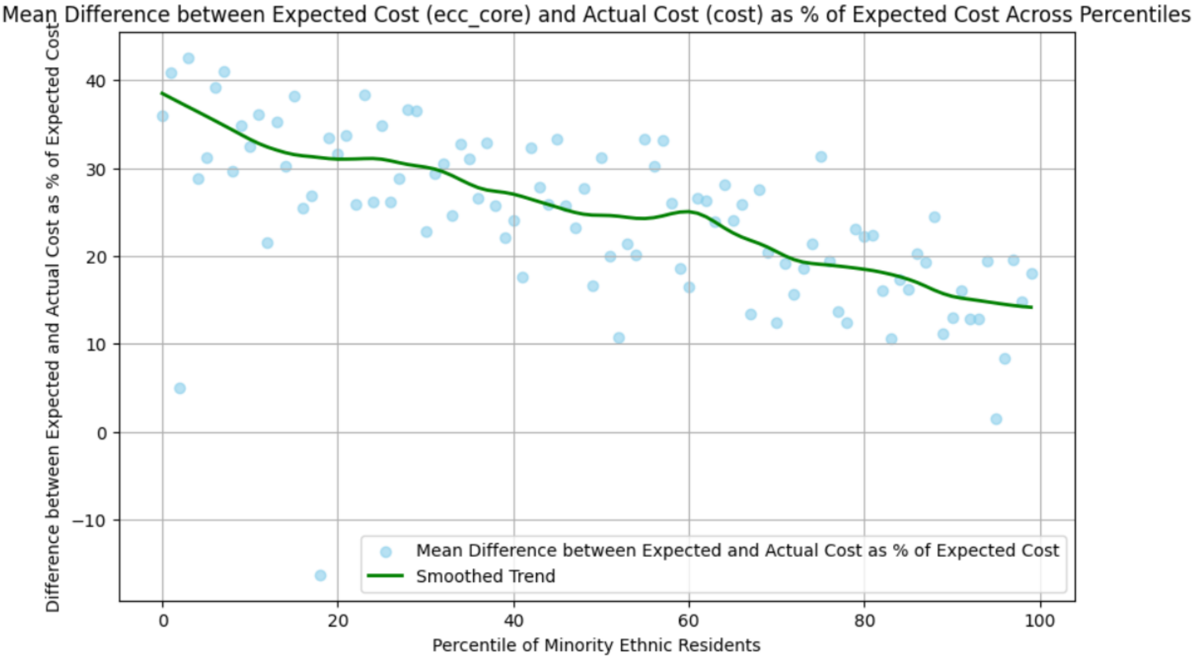
Source: Firm data, ONS data, FCA analysis

Figure 6: Mean of (ECC – Actual Claims Cost) across LSOA by percentile of proportion of individuals from minority ethnic backgrounds (Actual Claims Cost zero values excluded)



Source: Firm data, ONS data, FCA analysis

Figure 7: Average difference between ECC and ACC as a proportion of ECC across LSOA by percentile of proportion of individuals from minority ethnic backgrounds



Source: Firm data, ONS data, FCA analysis

Summary of other robustness checks

To evaluate the robustness of our analysis, the generalisability of our regression model and account for possible interactions between some of the variables we used, we conducted a series of statistical procedures outlined in Table 3.

Table 3: Specification of robustness checks undertaken along with justification and interpretation of results

Procedure	Justification	Outcome
Fitting the regression model with Actual Claims Costs instead of Expected Claims Costs	Since ECC is a modelled estimate, we wanted to utilise Actual Claims Costs to better reflect the real value an insurance firm must pay out for each policy.	53% of Actual Claims Cost values were either not available (due to poor data quality) or zero (as claims costs are only generated when there is a claim, and the claims frequency is relatively low). This meant that standard regression approaches would not work as the variable lacks enough variation on which to measure.
Fitting a regression model accounting for the interaction effects between the rural-urban geography and the proportion of individuals from minority ethnic backgrounds on premium	Differences in characteristics such as traffic, road quality and population density between urban and rural areas could result in a differing approach to calculating motor insurance premiums by insurance firms. Measuring the effect of the proportion of residents from minority ethnic backgrounds in an area on price requires adjusting for urban/rural geography as rural areas typically have a higher proportion of white residents.	Upon fitting a Linear Mixed Model with an interaction effect between the urban/rural geography of an area and the proportion of individuals from minority ethnic backgrounds, we found that the average effect of the proportion of individuals from minority ethnic backgrounds on the premium is around 10 percentage points smaller in urban than in rural areas.

Calculating regression coefficients' standard errors when data is clustered on three distinct levels, namely LSOA, firm and policy	Due to the hierarchy of our data, where policies in a given year of their tenure are nested within policies, which in turn are nested within firms and LSOAs, we wanted to test estimates of standard errors for our regression coefficients. Specifically, we wanted to test the effect on premiums of the share of residents with minority ethnic backgrounds at LSOA, firm and policy level to mitigate potential bias in their initial estimation.	We found that the standard error of the coefficient specifying the effect on premiums of the proportion of individuals from minority ethnic backgrounds did not change substantially when calculated on policy and on LSOA level. There was some difference when considering a firm-level estimation, which could be attributed to firms' differing business models and pricing strategies as well as their established presence in specific areas.
Performing a 10-fold cross-validation of the regression model on our data	We wanted to assess the generalisability of our model by fitting it only on part of our full dataset and testing it on the rest being 'unseen data'. We performed this procedure ten times on different data splits to ensure an unbiased estimate of the model's performance.	We found that the RMSE (Root Mean Squared Error) as well as the regression coefficient of the percentile of individuals from minority ethnic backgrounds stayed roughly the same for all data splits, hence, we concluded that our specified regression model does not overfit and generalises well to new data.
Conducting a Hausman test (see Annex 2: Methodology for more details)	We wanted to establish whether a random-effects linear model was necessary for our use case, or a fixed-effects model was sufficient.	Upon conducting the Hausman test, we could not find evidence against the suitability of a random effects model for describing our data.

Source: Firm data, ONS data, FCA analysis

Investigating the impact of age

Alongside individuals from minority ethnic backgrounds, the Taskforce also highlighted interest in the cost of motor insurance for young and older people. Whilst this analysis has focused on how average motor insurance premiums differ by local areas depending on ethnicity, as part of our main regression model, we considered age as an explanatory factor for insurance premiums. For consistency with other FCA analysis we split age into separate groups, setting the age group '17-24 years old' as our reference one. This allows us to make some inference as to how pricing may differ between different age groups, controlling for other factors.

Using the regression coefficients, which you can refer to in Annex 4: Regression analysis, we can infer the effect of customers belonging to each of the different age groups on their insurance premiums after controlling for the other variables in our regression. Note that this comes with the same limitations as described for our other findings.

Table 4 details the interpretation for this variable from our regression. We generally observe that customers aged 17-24 years old pay the most for their insurance premiums, at an average price of £1,019.

We also find that insurance premiums decrease with age up until the 66-70 years old group and then slightly increase again for the oldest customers in the '71+ years old' age group. The variation in pricing across age groups can be at least partially explained by variations in risk in the form of ECC. It is worth noting that age is highly correlated to many of the other risk factors included in this analysis and that firms are legally entitled to charge different prices across age groups, so long as they can demonstrate that this variation reflects differences in risk.

Table 4: Effect of belonging to each of the age groups considered on insurance premiums

Age group	Effect on insurance premiums (95% confidence intervals)
17-24 years old	Reference group
25-30 years old	Customers in this age group pay 6.9-7.1% (£70-£73) less on average in insurance premiums compared to customers in the reference group.
31-65 years old	Customers in this age group pay 12.9-13.2% (£131-£134) less on average in insurance premiums compared to customers in the reference group.
66-70 years old	Customers in this age group pay 15.6-15.9% (£159-£162) less on average in insurance premiums compared to customers in the reference group.
71+ years old	Customers in this age group pay 13.0-13.2% (£132-£135) less on average in insurance premiums compared to customers in the reference group.

Source: Firm data, ONS data, FCA analysis

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