The Geography of Ethnic Homeownership

Introduction

Background

According to the 2021 Census, 63 per cent of households in England and Wales own their homes. When these statistics are broken down by ethnicity, however, substantial gaps emerge between White British households and many 'minority' (here defined as non-White British) groups. 68 per cent of White British households are in homeownership, versus just over 40 per cent of Bangladeshi or Caribbean households, and a quarter or less of African or Arab households. And while some minority groups – namely the Chinese, Pakistani, Irish and Indian communities – boast rates of ownership above the national average, this can in fact represent a decline in housing fortunes over time.

71 per cent of Indian households are in homeownership, the highest proportion of any ethnic group and the only rate to exceed that of White British households. This, however, represents a decline of 11 percentage points since the time of the 1991 Census, when 82 per cent of Indian households were homeowners. Pakistani households have likewise seen their rate of ownership fall by 12 percentage points since 1991, from 77 to 65 per cent. The White British rate, by contrast, has remained stable at 68 per cent. The picture, therefore, is one of considerable variation in homeownership attainment by ethnicity, with the minority groups who previously experienced a very high degree of homeownership success seemingly struggling to sustain this over time.

Ethnic gaps in homeownership attainment have emerged as socially and academically significant alongside Britain's emergence as a 'homeowner society'. Britain experienced a tenurial "revolution" over the course of the twentieth century, with the number of renter households falling from an estimated 90 per cent at the dawn of the First World War to less than 50 per cent in 1971 (Saunders, 1990, p13-14). The country had thus reached a homeowning majority almost a decade before Margaret Thatcher's election as prime minister, but it was during the Thatcher years (1979 to 1990) that homeownership assumed a more central and purposive role within social organisation. With the rolling back of the state's welfare function, its commitment to look after its citizens from cradle to grave, more emphasis was placed on the owned home as a site of individual responsibility and a means of meeting current and future welfare needs (Ronald, 2008; King, 1996). In this context, homeownership signifies and allows one to express conformity with the prevailing moral order, structured by the neo-liberal virtues of independence and self-reliance, while promising an escape from a world marked by uncertainty and depletion.

In today's landscape of residualised social housing and deregulated private rented housing, homeownership is unique in promising both tenancy and ontological security – protection against eviction and the ability to express oneself and exert control over one's environment. Homeownership likewise promises the ability to build up a capital asset while reducing housing costs over time, meaning financial shocks and the burdens of old age can be managed in the absence of state support. And finally, with homeownership becoming not only the modal but the normal tenure, invested with neo-liberal morality and heralded as one of life's "proper paths" (Richards, 1990), renting becomes a powerful (albeit, given the challenges that first-time buyers are acknowledged to face, not uncomplicated) marker of non-belonging, one which perhaps intersects with and reinforces processes of racial othering. In other words, with myriad forms of security (tenancy, ontological, financial, social) now premised on homeownership, any group-level variation in homeownership attainment requires urgent scrutiny, and perhaps ethnic homeownership gaps most of all.

However, the gaps reported above are unadjusted, in the sense that there may be systematic differences between White British and minority households which explain their varying propensity to own their homes. It has been suggested in grey literature (Gulliver, 2016), for example, that gaps may reflect underlying differences in the age structure and geographical distribution of ethnic groups, with areas of historic minority settlement being most affected by homeownership decline. While attempts to explain ethnic housing disadvantage should be careful not to explain away – by and large, the consequences of being locked out of ownership will be the same regardless of the drivers, and underlying group-based differences in income or education are themselves a cause for concern – the suggestion that ethnic homeownership gaps may reflect the relative simplicity of demographic or geographic difference warrants further investigation.

Research questions, data and methodology

This research note will pursue one of these lines of enquiry, that of the relationship between ethnicity, geography and homeownership. It will do so by answering the following questions:

1) **Are ethnic homeownership gaps a function of differences in geographical distribution?**

   Do gaps get smaller when we control for geography, indicating that variation in homeownership attainment is in part a function of where groups are living? Or do gaps get larger, suggesting, perhaps, that levels of minority homeownership would be even lower were it not for their concentration in more deprived areas with lower house price to income ratios (Longhi, 2019)?

2) **Are ethnic homeownership gaps themselves geographically variable?**

   Do homeownership gaps vary in size or are they relatively stable across areas? How does this relate to any geographic variation in White British homeownership, for example do gaps tend to be larger (or smaller) in areas where more White British households rent?

To answer the first set of questions, a single-level logistic regression model was fit to data from the Annual Population Survey (APS), a national sample survey combining a subset of responses from the quarterly Labour Force Survey (LFS) and a series of local sample boosts. Its larger sample size (approximately 320,000 respondents in 122,000 households) allows for more robust analysis of both smaller geographical areas and smaller population groups, although here the decision was taken to use a dataset containing additional variables at the household level, which limited the level of geographic disaggregation which could be achieved. This decision was taken as the household dataset allowed for easier identification of the Household Reference Person (generally speaking, the person in whose name the accommodation is owned or rented, and whose ethnicity is then imputed to the rest of the household), as well as easier calculation of the number of children in a household.

In addition, the decision was taken to use the January to December 2019 APS dataset rather than more recent (2020, 2021) releases. This is because the 2019 dataset is the last available release to have been entirely unaffected by the COVID-19 pandemic, the onset of which caused a significant drop in the achieved sample size and necessitated the suspension of face-to-face interviewing. While interviewing proceeded on a telephone basis, there was a notable shift in the non-response bias of the survey, with a lower proportion of renter households and a higher proportion of outright owners being included (Office for National Statistics, 2020). While the Office for National Statistics (ONS) has sought to mitigate this by taking the significant step of introducing housing tenure into the LFS weighting methodology, it is not clear from the available online materials when face-to-face interviewing was resumed (if at all) and whether there have been improvements in this new source of non-response bias over time.

Furthermore, the ONS acknowledge that tenure is “a strongly correlated symptom of the bias, [but] not necessarily the actual cause”, which may instead “come down to factors such as the length of time that someone has lived at an address, or whether or not they have a landline” (ibid). In other words, it may have been a certain type of renter (and potentially also a certain type of owner) who was liable to be missed by the new contact methodology, so the bias introduced may run deeper than a simple overrepresentation of outright owners and underrepresentation of tenants. In addition, and notwithstanding the fact that the pandemic may have had long-term implications for tenure and its correlates, the focus of this research note is not on the pandemic as a relatively bounded phenomenon, but on the more general relationship between homeownership, ethnicity and geography. It therefore makes sense to use the annual dataset covering the period immediately preceding the pandemic (January to December 2019); not only does this avoid any potential sample bias related to the change of contact methodology and uncorrected by the tenure weighting, it enables a set of pre-pandemic baselines to be estimated, paving the way for future research into how the pandemic has affected ethnic homeownership gaps and whether such effects appear more or less entrenched.

Turning to the second set of research questions, these were examined by way of a multi-level logistic regression model, which in its fullest expression allowed both the intercept (levels of White British homeownership) and the slope (for each minority group, the gap between its homeownership rate and the White British rate) to vary by area. While a multi-level modelling strategy would seem well-suited to research into ethnic housing disadvantage given the more or less local nature of housing markets, it appears to have been under-utilised insofar as the specific question of ethnic homeownership gaps is concerned, with a notable exception being Leloup et al’s (2011) work in the Canadian context. This research note borrows more heavily from Longhi’s (2019) account of the impact of geographical location on ethnic wage gaps, therefore.

Variables

Across all fitted models, the dependent variable is **tenure**, a binary variable taking the value of 1 if the household owns their home (including on a shared ownership basis) and 0 if they are renting or otherwise accommodated. The aim is to estimate the log odds of homeownership for households of a given ethnicity, or in other words the odds of a household from a given ethnic group scoring 1 on the **tenure** variable, measured on the logit scale. Odds are a ratio of probabilities, here the probability of a household scoring 1 (being a homeowner) relative to the probability of them scoring 0 (not being a homeowner).

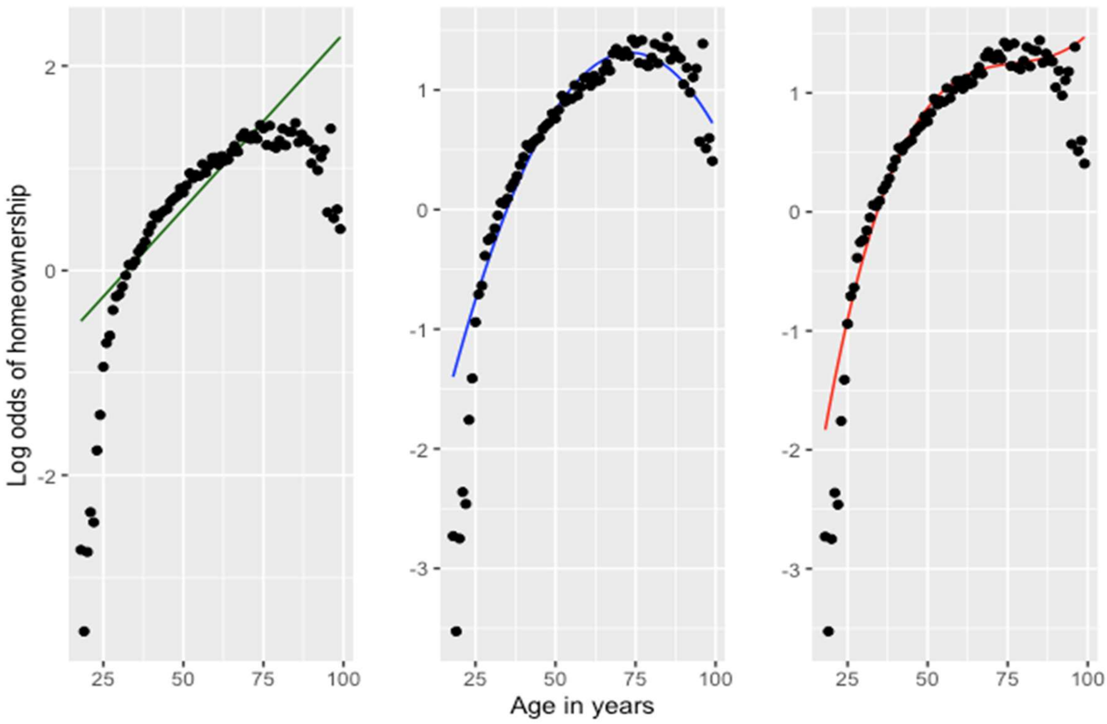
The main independent variables of interest are **ethno_fact**, a set of dummies for ethnicity with 11 categories and White British as the baseline, and **region**, a set of 13 regional dummies with the North East of England as the baseline. The APS datasets include a range of ethnicity variables at varying levels of aggregation – it was felt that the 11-category variable on which **ethno_fact** is based offered the best balance between simplicity and precision, with the proviso that ethnicity is a social construct and may inadequately reflect how people experience themselves and the world around them, even at the very highest level of disaggregation. Note that while **region** technically comprises 13 regional dummies, including one for Northern Ireland, the analysis is restricted to households in Britain as this was the level at which the aforementioned ethnicity variable was derived.

For the single-level regression, the model was ultimately expanded to include a number of additional potential predictors of homeownership. These are **age**, measured in years, **sex**, a binary variable taking the value of 1 if the Household Reference Person is female and 0 if male, **married**, a binary variable taking the value of 1 if the Household Reference Person is married, in a civil partnership, or cohabiting (and 0 otherwise), **born_overseas**, where the value of 1 denotes a Household Reference Person born outside the UK, **n_children**, a numeric variable recording the number of children in the household, and **occ_class**, a set of dummies for occupational class with eight categories and unemployment (or never having worked) as its baseline. While the APS datasets, being designed for labour market analysis, contain a wealth of information on respondent incomes including a net weekly pay, the sheer volume of missing data on this variable (87,627 non-responses from a total of 131,484 cases) and the high likelihood that those who did offer a response will be a self-selecting, potentially high-earning, group mitigated against its use. Occupational class is instead used as a proxy for both income and wealth, but such a solution is far from perfect.

Note that logistic regression assumes linearity on the logit or a linear relationship between the independent variables and the log odds of the dependent variable. It was hypothesised, however, that age may not be linearly related to the log odds of homeownership given its role in the life course, with households expected to buy at younger ages and remain in homeownership thereafter (or at least until the ageing process necessitates a different kind of housing). It was therefore hypothesised that, for younger age groups, the probability of homeownership may increase sharply with a unit increase in age, with this effect levelling out after a certain point. To test this, three regressions were carried out, the first a simple regression of **tenure** on **age**, the second adding a squared term for **age** and the third a cubed term. A series of plots were then generated, comparing the predicted values from each model to the observed log odds of homeownership at each age point recorded in the data. These plots are displayed in Figure 1 below, with the squared term (represented by the blue curve on the middle plot) seeming to best capture the association between age and the log odds of outcome. This term was therefore included in the final model.

Figure 1

Model predictions versus observed values: untransformed, squared and cubed age



Analysis

Exploratory analysis

Exploratory analysis of the independent and dependent variables was conducted prior to the model fitting. Within the parameters of a research note, only an outline of the most relevant distributions and relationships can be provided. As a first step, estimates were made of the proportion of homeowners within each ethnic group and region. The results by ethnicity are given in Table 1 and by region in Table 2.

For both variables, the mean of **tenure** (the proportion of homeowners at each level of the variable) was calculated with and without the use of survey weights and the 'Difference' column gives the difference between the two sets of estimates. For all the ethnic groups bar the White British, the inclusion of weights results in a lower estimate of the proportion in homeownership, which most likely reflects an under-sampling of younger households and the concentration of those households in the private rented sector (age being a key component of the weighting methodology). While the difference between estimates is never dramatic, for many groups it is in the region of three to four percentage points.

Table 1

Weighted and unweighted homeownership estimates by ethnicity

Ethnicity	Unweighted_n	Unweighted_mean	Weighted_mean	Difference
White British	113,584	0.703	0.705	-0.001
White Irish	900	0.677	0.640	0.037
Other White	5,651	0.381	0.365	0.016
Mixed/Multiple	903	0.473	0.469	0.004
Indian	2,306	0.720	0.690	0.030
Pakistani	1,601	0.637	0.633	0.004
Bangladeshi	542	0.441	0.399	0.042
Chinese	457	0.584	0.540	0.044
Other Asian	958	0.520	0.493	0.026
Black	3,046	0.309	0.302	0.008
Other	1,536	0.362	0.323	0.039

By and large, the APS estimates are comparable to the 2021 Census figures, although the APS returns a higher value for White British homeownership (70 per cent, versus a Census figure of 68 per cent) and a considerably lower value for Chinese homeownership (58 per cent unweighted and 54 per cent weighted, versus a Census figure of 64 per cent). The Chinese unweighted estimate is considerably closer to its Census figure than its weighted estimate, and there are other groups for whom this is the case. However, it must be noted that the Census enumerates a different population to this subset of APS data – England and Wales versus Great Britain.

Table 2 shows considerable regional variation in levels of homeownership, ranging from around 51 per cent in London to around 71 per cent in the South East, South West and Wales. However, the bulk of this variation can be attributed to London as an outlier, with homeownership in the majority of regions tracking the weighted sample mean of 66 per cent. While the data lend some support to the idea that housing markets operate on a regional level, excluding London from the analysis would result in a much narrower range of estimates.

Table 2

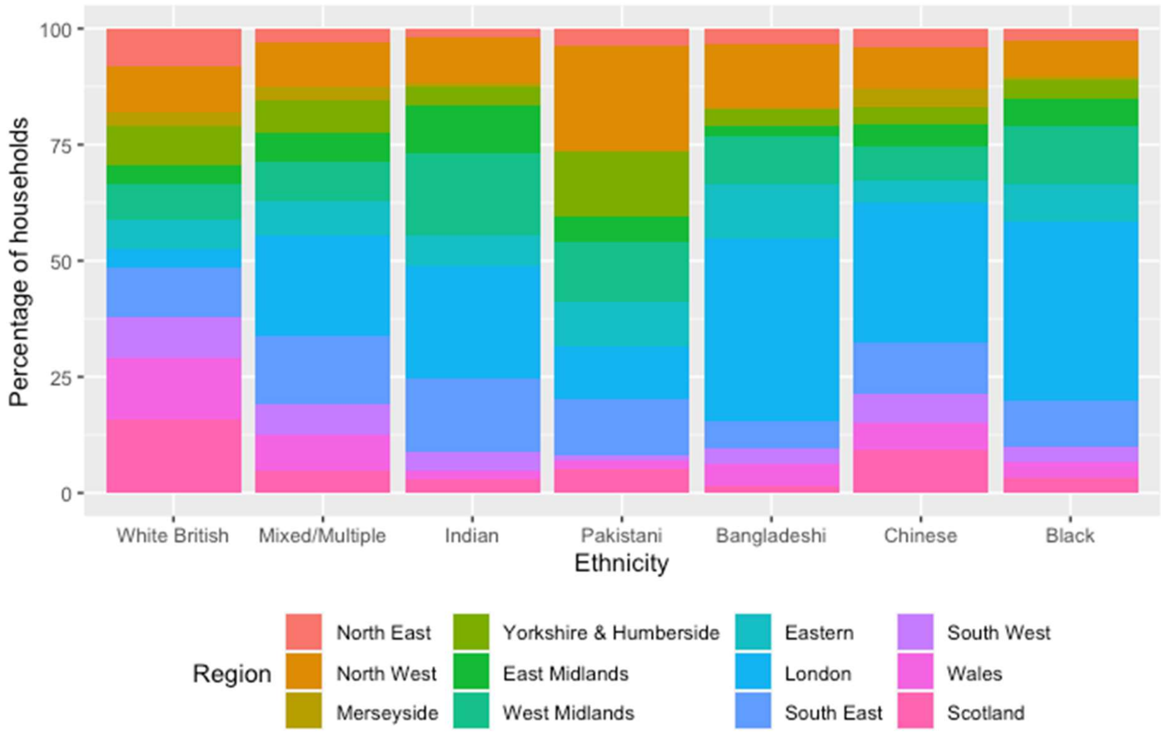
Weighted and unweighted homeownership estimates by region

Region	Unweighted_n	Unweighted_mean	Weighted_mean	Difference
North East	9,590	0.626	0.631	-0.005
North West	12,873	0.677	0.678	-0.001
Merseyside	3,621	0.689	0.689	0.000
Yorkshire & Humberside	10,655	0.673	0.663	0.011
East Midlands	5,898	0.664	0.678	-0.014
West Midlands	10,562	0.672	0.668	0.004
Eastern	8,179	0.684	0.684	0.001
London	9,794	0.519	0.511	0.008
South East	14,122	0.710	0.719	-0.009
South West	10,932	0.711	0.704	0.007
Wales	15,998	0.707	0.706	0.001
Scotland	19,260	0.679	0.653	0.026

This may suggest that geography plays less of a role in structuring ethnic homeownership gaps than has been previously assumed, although histories of migration to the UK emphasise that London drew significant numbers of South Asian and Caribbean migrants during the post-war period (Peach, 1998), with these early settlement patterns tending to influence the settlement of later cohorts through chain migration (Hatton & Wheatley Price, 2005). And indeed, Figure 2 – which shows, for a selection of ethnic groups, the percentage of households living in each region – suggests that minority households are overrepresented in London when compared to White British households. However, and with the possible exception of Bangladeshi and Black households, it is by no means the case that minority groups are overwhelmingly concentrated in London and some groups, particularly the Indian community, show a reasonable degree of concentration in the South East where homeownership is at its highest. Note that this graph was plotted on the unweighted dataset, which will be used in all subsequent analyses. This is mainly due to the complexity of including survey weights within a multi-level modelling framework and the desire to maintain a consistent approach from model to model.

Figure 2

Percentage of households from each ethnic group in each region – a subset of ethnicities



Model specification and discussion

The first model to be fit was a single-level regression of tenure on ethnicity:

$$\boldsymbol{log\left(\frac{p}{1-p}\right) = \alpha + \beta(ethno\_fact) + \epsilon}$$

Where $\boldsymbol{p}$ is the probability of a household being in homeownership. The coefficients for ethnicity represent unadjusted ethnic homeownership gaps – the difference between the log odds of homeownership for White British households and for households of a given minority ethnicity. Log odds can be easily converted into odds and from there into probabilities.

The model was then expanded to include the set of predictors for region:

$$\boldsymbol{log\left(\frac{p}{1-p}\right) = \alpha + \beta(ethno\_fact) + \beta(region) + \epsilon}$$

And, finally, a full version of the model was run incorporating all potential predictors set out in the 'Variables' section, plus a squared term for age. The aim of including regional predictors is examine how raw homeownership gaps change once region is accounted for. The aim of including the broader set of potential predictors is to establish which factors have a significant marginal association with the log odds of homeownership, and in particular to ascertain whether any significant association between region and homeownership retains its significance once other predictors are added – Hoffman & Walters (2022) argue that a random intercept, creating a new level in the model, is required for each sampling dimension in which mean differences remain after accounting for the fixed slopes of predictors.

Table 3 gives two sets of odds ratios for ethnicity and the difference between them. The adjusted odds ratios can be interpreted as the change in the odds of homeownership with a shift from White British to a given minority ethnicity, after controlling for region. The unadjusted odds ratios are the change in the odds with a shift away from White British ethnicity, without controlling for region. An odds ratio of less than 1 indicates lower odds of homeownership for the minority group, and a positive value for 'Difference' indicates an improvement in the odds ratio once region is controlled for. In other words, a positive value for 'Difference' represents a narrowing of the ethnic homeownership gap, except in the case of Indian households, who are more likely to own than the White British, and even more so after adjusting for region.

Chinese households, for example, are 0.59 times as likely to own as White British households before adjusting for region, but 0.63 times as likely after. Under the simple model, the predicted probability of homeownership for a Chinese household is 58 per cent, versus 70 per cent for White British households (note that these estimates are identical to the unweighted group means given in Table 1). Under the region-adjusted model, the predicted probability of White British homeownership stays within rounding distance of the unadjusted figure, but the predicted probability of Chinese homeownership rises to 60 per cent[1]. Hence the homeownership gap narrows, but only by around two percentage points. Of all the ethnic homeownership gaps, it is in fact the Chinese gap which shows the biggest improvement with the addition of regional controls, and hence we can conclude that the distribution of ethnic groups across regions plays only a small role in explaining their differential access to homeownership.

Table 3

Weighted and unweighted homeownership estimates by region

[1] Note that here **region** has been set to the East Midlands, where homeownership levels are closest to their overall sample mean.

Ethnicity	Adjusted_OR	Unadjusted_OR	Difference
White Irish	0.906	0.883	0.023
Other White	0.266	0.260	0.006
Mixed/Multiple	0.386	0.378	0.008
Indian	1.122	1.084	0.038
Pakistani	0.752	0.741	0.011
Bangladeshi	0.362	0.333	0.029
Chinese	0.631	0.593	0.038
Other Asian	0.475	0.457	0.018
Black	0.203	0.189	0.014
Other	0.256	0.239	0.017

Table 4 gives the regression output (coefficient estimates, standard errors, confidence levels and model fit metrics) for the three single-level models: simple, region adjusted and full. Under the full model, the vast majority of predictors are highly statistically significant, meaning we can be confident that each predictor exerts a unique (although not necessarily causal) influence on the odds of homeownership. Moving from the baseline of occupational class (unemployment) to its highest level (higher managerial and professional) is associated, for instance, with a change in the odds of homeownership of 6.54 – people in the latter group are, on average, 6.54 times as likely to be homeowners. Being married, in a civil partnership or cohabiting is likewise associated with a 3.20 change in the odds – people in this group, on average, are 3.20 times as likely to be homeowners as single people. By contrast, being born overseas is associated with considerably lower odds of homeownership than being born in the UK – the former group are 0.60 times as likely to be homeowners than the latter, again on average. Ethnicity remains a significant predictor of homeownership in most cases, although Pakistani, Chinese and Irish households now have higher odds of homeownership than White British households (the latter not to a statistically significant extent), indicating that their homeownership gaps would be reversed were it not for observed differences on the independent variables. The coefficients on the regional dummies are all highly significant. This suggests that mean differences remain – at least between the baseline and each additional category – after all fixed effects have been included, and hence there are grounds to proceed with a multi-level specification with region as a Level 2 predictor, as per Hoffman & Walters (2022).

Table 4

Single-level regression output – simple, region-adjusted and full models

	Dependent variable:
	tenure

	(1)	(2)	(3)
White Irish	-0.124	-0.099	0.056
	(0.072)	(0.072)	(0.083)
Other White	-1.348***	-1.322***	-0.602***
	(0.028)	(0.029)	(0.045)
Mixed/Multiple	-0.972***	-0.952***	-0.383***
	(0.067)	(0.067)	(0.078)
Indian	0.081	0.115*	0.399***
	(0.047)	(0.047)	(0.057)
Pakistani	-0.300***	-0.285***	0.372***
	(0.052)	(0.053)	(0.064)
Bangladeshi	-1.100***	-1.016***	-0.427***
	(0.087)	(0.087)	(0.100)
Chinese	-0.523***	-0.461***	0.488***
	(0.095)	(0.096)	(0.118)
Other Asian	-0.784***	-0.745***	-0.187*
	(0.065)	(0.066)	(0.080)
Black	-1.667***	-1.595***	-0.938***
	(0.040)	(0.041)	(0.051)
Other	-1.430***	-1.364***	-0.744***
	(0.053)	(0.054)	(0.068)
North West		0.285***	0.252***
		(0.029)	(0.032)
Merseyside		0.304***	0.267***
		(0.042)	(0.047)
Yorkshire and Humber		0.256***	0.197***
		(0.030)	(0.034)
East Midlands		0.282***	0.189***
		(0.035)	(0.040)
West Midlands		0.308***	0.194***
		(0.030)	(0.034)
Eastern		0.410***	0.211***
		(0.033)	(0.037)
London		0.029	-0.191***
		(0.031)	(0.035)
South East		0.500***	0.220***
		(0.029)	(0.032)
South West		0.445***	0.252***

		(0.030)	(0.034)
Wales		0.379***	0.278***
		(0.028)	(0.031)
Scotland		0.256***	0.152***
		(0.026)	(0.030)
Female			-0.092***
			(0.015)
Age			0.087***
			(0.003)
Squared age			-0.0003***
			(0.00002)
Married			1.163***
			(0.016)
Born overseas			-0.517***
			(0.033)
Number of children			-0.093***
			(0.008)
Routine occupations			0.122***
			(0.030)
Semi-routine			0.332***
			(0.028)
Lower supervisory and technical			0.857***
			(0.033)
Small employers and own account workers			0.980***
			(0.030)
Intermediate occupations			1.208***
			(0.030)
Lower managerial and professional			1.569***
			(0.025)
Higher managerial and professional			1.878***
			(0.029)
Constant	0.863***	0.564***	-4.249***
	(0.006)	(0.021)	(0.077)
Observations	131,484	131,484	131,484
Log Likelihood	-80,663.330	-80,383.740	-66,251.740
Akaike Inf. Crit.	161,348.700	160,811.500	132,573.500
Note:			*p<0.05; **p<0.01; ***p<0.001

A multi-level model was therefore fit to the data, first varying only the intercept by region, then both the intercept and slope of ethnicity. The full specification of the model was as follows:

- Level 1

$$\boldsymbol{Y_{ij} = a_j + b_j(ethno\_fact)_{ij} + \epsilon_{ij}}$$

- Level 2

$$\boldsymbol{a_j = \alpha_0 + u_j}$$

$$\boldsymbol{b_j = \beta_0 + v_j}$$

Here, Y denotes the log odds of homeownership, i denotes a household and j a region. We thus have 131,484 households nested in 12 regions, and the final model assumes that both levels of White British homeownership (α_0) and the size of ethnic homeownership gaps (β_0) vary by region.

To begin, a simple varying intercepts model was fit, without the second Level 2 equation above, meaning the slope of ethnicity or the size of each ethnic group's homeownership gap is assumed to be constant across regions. This model is equivalent to the single-level model with a set of fixed effects for region, the second column in Table 4, and allows us to extract more information on regional variation in White British homeownership. Table 5 shows the fixed effect of the model intercept – with a slope for ethnicity included, White British homeownership – which can be interpreted as the average level of White British homeownership across all regions. Also given are the model random effects, or the region-by-region deviation of White British homeownership from its national average. The estimate is the log odds of White British homeownership in a given region, with 95 per cent Confidence Intervals, and the final column converts these log odds to probabilities. Under this simple model, the predicted probability of White British homeownership ranges from 64 per cent in the North East to 74 per cent in the South East. In London, the predicted probably of White British homeownership is 65 per cent, considerably higher than the sample estimate for the region of around 51 per cent.

Table 5

Variation in White British homeownership by region – varying intercepts model

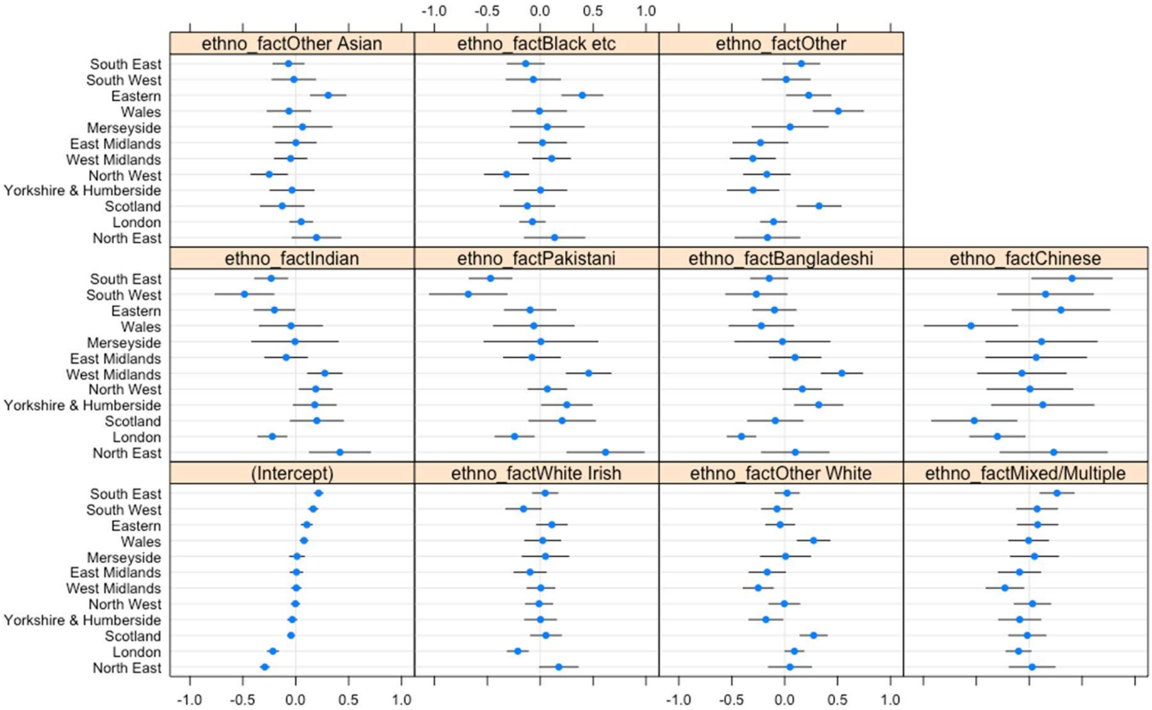
Region	Random	Fixed	Estimate	Lower	Upper	Probability
North East	-0.282	0.852	0.570	0.529	0.612	0.639
North West	-0.003	0.852	0.850	0.812	0.887	0.700
Merseyside	0.015	0.852	0.867	0.799	0.936	0.704
Yorkshire & Humberside	-0.031	0.852	0.821	0.780	0.861	0.694
East Midlands	-0.005	0.852	0.847	0.793	0.901	0.700
West Midlands	0.020	0.852	0.872	0.831	0.913	0.705
Eastern	0.118	0.852	0.970	0.923	1.018	0.725
London	-0.252	0.852	0.600	0.558	0.641	0.646
South East	0.208	0.852	1.060	1.023	1.097	0.743
South West	0.153	0.852	1.005	0.964	1.047	0.732
Wales	0.090	0.852	0.942	0.908	0.976	0.719
Scotland	-0.031	0.852	0.821	0.791	0.851	0.694

The full varying slopes and intercepts model was then fit to the data. Figure 3 below shows the variation of each random effect – White British homeownership and, for each minority group, the homeownership gap – around the fixed or average effect of that variable. For example, and as per Table 5 above, we see that White British homeownership is highest in the South East and lowest in the North East. Homeownership gaps do seem to vary quite considerably by area, with the Pakistani gap widening in the South and closing in the North East, for instance. Table 6 gives region-by-region information for Pakistani households, including odds ratios with 95 per cent Confidence Intervals.

As Figure 3 suggests, Pakistani households have the lowest odds of homeownership relative to White British households in the South West (0.37 times as likely to own) and the highest relative odds in the North East, where the point estimate suggests that they are more likely to own than the White British. The latter Confidence Interval contains the value of 1, however, so we cannot confidently assert that the Pakistani homeownership gap closes in the North East (although it certainly seems to narrow). This reflects a more general trend apparent in Figure 3, namely that for some ethic groups, the variation in their homeownership gap from region to region is estimated with quite a high degree of uncertainty.

Figure 3

Varying intercepts and slopes model – a dotplot of model random effects



Finally, the model output suggests considerable variation by ethnic group in terms of how the intercepts and slopes are correlated. For some ethnic groups, the correlation is positive, indicating that the group's odds of owning relative to White British households improve as levels of White British homeownership increase, or the homeownership gap closes. For households of mixed or multiple ethnicity the correlation is 0.36 and for 'Other' households it is 0.35, indicating a moderate positive association. The situation is reversed, however, for Indian and Pakistani households, whose slopes are strongly negatively correlated with the model intercepts (-0.54 and -0.60 respectively). For Indian households, who have higher average rates of ownership than White British households, this reflects a narrowing of the gap in areas where White British homeownership attainment is higher. For Pakistani households, and as Table 6 demonstrates, homeownership gaps are particularly severe in the areas where White British households enjoy the very highest rates of ownership. While this may seem intuitive given the relative nature of White British homeownership and ethnic homeownership gaps, it is not the case for all ethnic groups.

Table 6

Varying intercepts and slopes model – regional variation in the 'effect' of Pakistani ethnicity

Region	Random	Fixed	Estimate	OR	Lower	Upper
North East	0.617	-0.316	0.301	1.352	0.940	1.945
North West	0.068	-0.316	-0.248	0.780	0.650	0.936
Merseyside	0.006	-0.316	-0.310	0.734	0.429	1.256
Yorkshire & Humberside	0.252	-0.316	-0.064	0.938	0.739	1.191
East Midlands	-0.078	-0.316	-0.394	0.674	0.516	0.881
West Midlands	0.458	-0.316	0.142	1.152	0.933	1.423
Eastern	-0.096	-0.316	-0.412	0.663	0.520	0.845
London	-0.242	-0.316	-0.558	0.572	0.476	0.688
South East	-0.469	-0.316	-0.785	0.456	0.373	0.557
South West	-0.680	-0.316	-0.996	0.369	0.256	0.532
Wales	-0.060	-0.316	-0.376	0.686	0.469	1.004
Scotland	0.207	-0.316	-0.109	0.897	0.655	1.227

Bibliography

Gulliver, K. (2016). *Forty years of struggle: a window on race and housing, disadvantage and exclusion*. London: Human City Institute. Available at: https://hqnetwork.co.uk/download.cfm?doc=docm93jijm4n2932.pdf&ver=5697 (Accessed: 23 August 2021).

Hatton, T. J. & Wheatley Price, S. (2005). "Migration, Migrants and Policy in the United Kingdom", in Zimmermann, K. F. (ed.) *European Migration: What Do We Know?*. Oxford: Oxford University Press, pp. 113-172

Hoffman, L. & Walters, R. W. (2022). "Catching Up on Multilevel Modelling". *Annual Review of Psychology*, 73. pp. 659-689. doi: 10.1146/annurev-psych-020821-103525

King, P. (1996). *The Limits of Housing Policy: A Philosophical Investigation*. London: Middlesex University Press

Leloup, X., Apparicio, P. & Esfahani, F. D. (2011). "Ethnicity and Homeownership in Montréal, Toronto and Vancouver: Measuring Effects of the Spatial Distribution of Ethnic Groups Using Multilevel Modelling in 1996 and 2001". *Journal of International Migration & Integration*, 12 (4). pp. 429-451. doi: *10.1007/s12134-011-0186-4*

Longhi, S. (2020). "Does geographical location matter for ethnic wage gaps?". *Journal of Regional Science*, 60 (3). pp. 538-557. doi: https://doi.org/10.1111/jors.12469

Office for National Statistics. (2020). *Coronavirus and its impact on the Labour Force Survey*. [Online]. [Accessed: 16 July 2023]. Available at: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/articles/coronavirusanditsimpactonthelabourforcesurvey/2020-10-13

Peach, C. (1998). "South Asian and Caribbean Ethnic Minority Housing Choice in Britain". *Urban Studies*, 35 (10), pp. 1657-1680. doi: 10.1080/0042098984097

Richards, L. (1990). *Nobody's Home: Dreams and Realities in a New Suburb*. Oxford: Oxford University Press

Ronald, R. (2008). *The Ideology of Home Ownership: Homeowner Societies and the Role of Housing*. Basingstoke: Palgrave Macmillan

Saunders, P. (1990). *A Nation of Homeowners*. London: Unwin Hyman

Appendix: R Script

```
"Present your R code here"

setwd(" ")
getwd()
list.files()

library(haven)
library(tidyverse)
library(srvyr)
library(survey)
library(gridExtra)
library(jtools)
library(broom)
library(broom.mixed)
library(lme4)
library(lmerTest)
library(flextable)
library(reshape2)
library(stargazer)
library(lattice)

aps_hhs <- read_dta("apsh_jd19_eul.dta")
ls(aps_hhs)

aps_hhs_sub <- aps_hhs %>%
  filter(HRP == 1) %>%
  filter(AGE >= 18) %>%
  select(HRP, TEN1, ETHGBEUL, SEX, AGE, AAGE, MARDY6, CRY12, HDPCH4, HDC515, HDPCH18, GOVTOF,
 NSECMJ10, NETWK, PHHWTA18)

aps_hhs_sub[aps_hhs_sub < 0] <- NA
UK <- c(921, 922, 923, 924, 926)

aps_hhs_sub <- aps_hhs_sub %>%
  mutate(tenure = ifelse(is.na(TEN1), NA, ifelse(TEN1 <= 3, 1, 0))) %>%
  mutate(ethno_fact = factor(ETHGBEUL, levels = c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 ),
                         labels = c("White British", "White Irish", "Other White", "Mixed/Multiple", "Indian", "Pak
istani", "Bangladeshi", "Chinese", "Other Asian", "Black", "Other"))) %>%
  mutate(married = factor(MARDY6, levels = c(2, 1), labels = c("Non-married etc", "Married etc"))) %>%
  mutate(born_overseas = ifelse(is.na(CRY12), NA, ifelse(CRY12 %in% UK, 0, 1))) %>%
  mutate(n_children = HDPCH4 + HDC515 + HDPCH18) %>%
  filter(n_children < 8) %>%
  mutate(region = factor(GOVTOF, levels = c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13),
                      labels = c("North East", "North West", "Merseyside", "Yorkshire & Humberside", "East Mid
lands", "West Midlands", "Eastern", "London", "South East", "South West", "Wales", "Scotland", "Northe
rn Ireland"))) %>%
  mutate(occ_class = factor(NSECMJ10, levels = c(8, 7, 6, 5, 4, 3, 2, 1),
```

```
                         labels = c("Never worked, unemployed etc", "Routine occupations", "Semi-routine occup
ations", "Lower supervisory and technical", "Small employers and own account workers", "Intermediate
occupations", "Lower managerial and professional", "Higher managerial and professional"))) %>%
  drop_na(tenure, ethno_fact, SEX, AGE, AAGE, married, born_overseas, n_children, region, occ_class)

aps_hhs_sub_w <- aps_hhs_sub %>%
  srvyr::as_survey(weights = PHHWTA18)
aps_hhs_sub_w

means_full <- aggregate(tenure ~ AGE, data = aps_hhs_sub, FUN = mean)
means_full
means_full$log_odds <- log(means_full$tenure/(1-means_full$tenure))
new_data1 <- data_frame(AGE = means_full$AGE)
new_data1

age_mod4 <- glm(tenure ~ AGE, data = aps_hhs_sub, family = binomial)
new_data1$preds4 <- predict(age_mod4, new_data1, type = "link")

linear_plot1 <- new_data1 %>%
  ggplot(aes(x = AGE, y = preds4)) +
  labs(x = " ", y = "Log odds of homeownership") +
  geom_line(colour = "darkgreen") +
  geom_point(data = means_full, aes(x = AGE, y = log_odds))

age_mod5 <- glm(tenure ~ AGE + I(AGE^2), data = aps_hhs_sub, family = binomial)
new_data1$preds5 <- predict(age_mod5, new_data1, type = "link")

squared_plot1 <- new_data1 %>%
  ggplot(aes(x = AGE, y = preds5)) +
  labs(x = "Age in years", y = " ") +
  geom_line(colour = "blue") +
  geom_point(data = means_full, aes(x = AGE, y = log_odds))

age_mod6 <- glm(tenure ~ AGE + I(AGE^2) + I(AGE^3), data = aps_hhs_sub, family = binomial)
new_data1$preds6 <- predict(age_mod6, new_data1, type = "link")

cubed_plot1 <- new_data1 %>%
  ggplot(aes(x = AGE, y = preds6)) +
  labs(x = " ", y = " ") +
  geom_line(colour = "red") +
  geom_point(data = means_full, aes(x = AGE, y = log_odds))

grid.arrange(linear_plot1, squared_plot1, cubed_plot1,
             nrow = 1)

mean_weight_ethno <- aps_hhs_sub_w %>%
  group_by(ethno_fact) %>%
  summarise(weight_mean = survey_mean(tenure, na.rm = TRUE)) %>%
```

```
  as.data.frame()
mean_weight_ethno

mean_unweight_ethno <- aps_hhs_sub %>%
  drop_na(ethno_fact) %>%
  group_by(ethno_fact) %>%
  summarise(unweight_mean = mean(tenure, na.rm = TRUE)) %>%
  as.data.frame()
mean_unweight_ethno

aps_hhs_sub_w %>%
  group_by(ethno_fact) %>%
  survey_tally()

countdf <- aps_hhs_sub %>%
  group_by(ethno_fact) %>%
  drop_na(ethno_fact) %>%
  summarise(count = n()) %>%
  as.data.frame()

count <- countdf$count

weight_unweight_ethno <- data.frame(mean_weight_ethno$ethno_fact, count, mean_unweight_ethno
$unweight_mean, mean_weight_ethno$weight_mean)
weight_unweight_ethno <- weight_unweight_ethno %>%
  drop_na() %>%
  rename(Ethnicity = mean_weight_ethno.ethno_fact, Unweighted_n = count, Unweighted_mean = mea
n_unweight_ethno.unweight_mean, Weighted_mean = mean_weight_ethno.weight_mean) %>%
  mutate(Difference = Unweighted_mean - Weighted_mean) %>%
  flextable() %>%
  colformat_double(digits = 3)

mean_weight_region <- aps_hhs_sub_w %>%
  group_by(region) %>%
  summarise(weight_mean = survey_mean(tenure, na.rm = TRUE)) %>%
  as.data.frame()
mean_weight_region

mean_unweight_region <- aps_hhs_sub %>%
  group_by(region) %>%
  summarise(unweight_mean = mean(tenure, na.rm = TRUE)) %>%
  as.data.frame()
mean_unweight_region

aps_hhs_sub_w %>%
  group_by(region) %>%
  survey_tally()
```

```
countdf_region <- aps_hhs_sub %>%
  group_by(region) %>%
  drop_na(region) %>%
  summarise(count = n()) %>%
  as.data.frame()

count_region <- countdf_region$count

weight_unweight_region <- data.frame(mean_weight_region$region, count_region, mean_unweight_re
gion$unweight_mean, mean_weight_region$weight_mean)
weight_unweight_region <- weight_unweight_region %>%
  drop_na() %>%
  rename(Region = mean_weight_region.region, Unweighted_n = count_region, Unweighted_mean = me
an_unweight_region.unweight_mean, Weighted_mean = mean_weight_region.weight_mean) %>%
  mutate(Difference = Unweighted_mean - Weighted_mean) %>%
  flextable() %>%
  colformat_double(digits = 3)

ethno_reduced <- c("White British", "Mixed/Multiple", "Indian", "Pakistani", "Bangladeshi", "Chinese", "
Black")

aps_hhs_sub %>%
  drop_na(ethno_fact, region) %>%
  filter(ethno_fact %in% ethno_reduced) %>%
  rename(Ethnicity = ethno_fact, Region = region) %>%
  count(Ethnicity, Region) %>%
  group_by(Ethnicity) %>%
  mutate(pct = prop.table(n) * 100) %>%
  ggplot() + aes(Ethnicity, pct, fill = Region) +
  geom_bar(stat="identity") +
  ylab("Percentage of households") +
  theme(legend.position = "bottom")

mod_sl_raw <- glm(tenure ~ ethno_fact, data = aps_hhs_sub, family = binomial)
summary(mod_sl_raw)

mod_sl_region <- glm(tenure ~ ethno_fact + region, data = aps_hhs_sub, family = binomial)
summary(mod_sl_region)

region_or <- as.vector(exp(mod_sl_region$coefficients[2:11])) %>%
  format(digits = 3)
ethno_or <- as.vector(exp(mod_sl_raw$coefficients[2:11])) %>%
  format(digits = 3)
ethno_list_10 <- c("White Irish", "Other White", "Mixed/Multiple", "Indian", "Pakistani", "Bangladeshi",
"Chinese", "Other Asian", "Black", "Other")
or_comp <- as.data.frame(cbind(ethno_list_10, region_or, ethno_or))

or_comp %>%
```

```
  rename(Ethnicity = ethno_list_10, Adjusted_OR = region_or, Unadjusted_OR = ethno_or) %>%
  mutate(Difference = as.numeric(region_or) - as.numeric(ethno_or)) %>%
  flextable()

newdata_raw_region <- data_frame(ethno_fact = c("White British", "Chinese"), region = "East Midland
s")
predict_raw <- predict(mod_sl_raw, newdata = newdata_raw_region, type = "response")
predict_region <- predict(mod_sl_region, newdata = newdata_raw_region, type = "response")

mod_sl_full <- glm(tenure ~ ethno_fact + region + SEX + AGE + I(AGE^2) + married + born_overseas + n_
children + occ_class, data = aps_hhs_sub, family = binomial)
summary(mod_sl_full)

sl_reg_table <- stargazer(list(mod_sl_raw, mod_sl_region, mod_sl_full),
        omit.stat = c("f", "rsq", "ser"),
        covariate.labels = c("White Irish",
                        "Other White",
                        "Mixed/Multiple",
                        "Indian",
                        "Pakistani",
                        "Bangladeshi",
                        "Chinese",
                        "Other Asian",
                        "Black",
                        "Other",
                        "North West",
                        "Merseyside",
                        "Yorkshire and Humber",
                        "East Midlands",
                        "West Midlands",
                        "Eastern",
                        "London",
                        "South East",
                        "South West",
                        "Wales",
                        "Scotland",
                        "Female",
                        "Age",
                        "Squared age",
                        "Married",
                        "Born overseas",
                        "Number of children",
                        "Routine occupations",
                        "Semi-routine",
                        "Lower supervisory and technical",
                        "Small employers and own account workers",
                        "Intermediate occupations",
                        "Lower managerial and professional",
```

```
                              "Higher managerial and professional"),
          type = "text",
          digits = 3,
          no.space = T,
          intercept.bottom = TRUE,
          star.cutoffs = c(0.05, 0.01, 0.001),
          out = "SL reg table.html")

exp(mod_sl_full$coefficients)

mod_ml_int_cs <- glmer(tenure ~ ethno_fact + (1|region), data = aps_hhs_sub, family = binomial)
summary(mod_ml_int_cs)

mod_ml_int_cs_fix <- fixef(mod_ml_int_cs)
mod_ml_int_cs_ran <- as_tibble(ranef(mod_ml_int_cs, condVar = TRUE))
mod_ml_int_cs_ran <- mod_ml_int_cs_ran %>%
  select(grp, condval, condsd) %>%
  rename(region = grp, ranef = condval) %>%
  mutate(fixef = mod_ml_int_cs_fix["(Intercept)"]) %>%
  mutate(est = fixef + ranef,
         lower_bound = fixef + ranef - 1.96*condsd,
         upper_bound = fixef + ranef + 1.96*condsd,
         or = exp(est),
         or_lower_bound = exp(lower_bound),
         or_upper_bound = exp(upper_bound),
         prob = exp(est)/(1+exp(est)))

mod_ml_int_cs_ran <- mod_ml_int_cs_ran %>%
  select(region, ranef, fixef, est, lower_bound, upper_bound, prob) %>%
  rename(Region = region, Random = ranef, Fixed = fixef, Estimate = est, Lower = lower_bound, Upper =
upper_bound, Probability = prob) %>%
  flextable() %>%
  colformat_double(digits = 3)

mod_ml_int_vs <- glmer(tenure ~ ethno_fact + (ethno_fact|region), data = aps_hhs_sub, family = binom
ial, control = glmerControl(optCtrl=list(maxfun=1e6)))
summary(mod_ml_int_vs)
coef(mod_ml_int_vs)
fixef(mod_ml_int_vs)
ranef(mod_ml_int_vs)

dotplot(ranef(mod_ml_int_vs, condVar = TRUE, main = "Random effects by region"))

mod_ml_int_vs_ran_Pak <- mod_ml_int_vs_ran %>%
  filter(term == "ethno_factPakistani") %>%
  select(grp, condval, condsd) %>%
  rename(region = grp, ranef = condval) %>%
  mutate(fixef =mod_ml_int_vs_fix["ethno_factPakistani"]) %>%
```

```
  mutate(est = fixef + ranef,
         lower_bound = fixef + ranef - 1.96*condsd,
         upper_bound = fixef + ranef + 1.96*condsd,
         or = exp(est),
         or_lower_bound = exp(lower_bound),
         or_upper_bound = exp(upper_bound))

mod_ml_int_vs_ran_Pak <- mod_ml_int_vs_ran_Pak %>%
  select(region, ranef, fixef, est, or, or_lower_bound, or_upper_bound) %>%
  rename(Region = region, Random = ranef, Fixed = fixef, Estimate = est, OR = or, Lower = or_lower_boun
d, Upper = or_upper_bound) %>%
  flextable() %>%
  colformat_double(digits = 3)
```