Feedback Form

School of Geography
FACULTY OF ENVIRONMENT

UNIVERSITY OF LEEDS

Students to complete all of the following (or work will not be accepted)

Name (optional):

Richard Thomas

Student ID:	Assessment Name (e.g. Essay 2, Group Project etc):
200818341	Assignment #1: Fertility analysis
Degree Programme & Level (e.g. BA1 Geog):	Assessment Marker:
MSc GIS / MA	Paul Norman
Module Title & Code:	Word Count: (2,000 words max)
GEOG5520: Quantitative & Spatial Methods	1830

Content, research and reading

Structure and argument

Writing, presentation and referencing

Areas for improvement to prioritise

Marker's signature:	Date:	Provisional mark:

Any mark given here is provisional and subject to moderation by the School's Board of Examiners. This is to ensure comparable marking standards for all students. In a minority of individual cases moderation can lead to either the raising or lowering of the provisional marks.

Is education more influential on area variations in fertility than ethnicity?

INTRODUCTION

Fertility varies widely across local authorities (LA) in England. This report investigates the distribution of that variation, focusing on the relative influences of education and ethnicity. The indicator used is the 2001 English census measure of Total Fertility Rate (TFR) which is the average number of children a woman would have if she continued to have children (at the same rate as at the time of the census) for the rest of her reproductive life. After a review of previous literature, this report describes the methodology used to build an appropriate statistical model for TFR, an analysis of the characteristics of this model, and discusses what these findings (in conjunction with the literature) tell us about the key influences on fertility.

BACKGROUND

In a literature review of the effect of education on fertility, Diamond et al (1999) indicated that overall it has a small negative effect but the relationship is complex and dependent on both the level of education and the demographic and socio-economic structure of the population considered. In addition to the *direct* effect of education on the individual concerned, there is the *indirect* effect of influencing others (either informally through enforcing society norms, or more formally in influencing policy and its application). In a “low fertility society” such as in England, education can have as strong an influence on timing (delaying child-bearing) as on final family size (Diamond et al, 1999). In Europe, this postponement leads to some unwanted childlessness (techniques such as IVF only having a minor corrective effect), but is not a major factor affecting TFR (Te Velde et al, 2012). The accepted reason for higher education leading to first birth postponement is the difficulty in concurrently managing being a student and a mother (Rindfuss et al, 1996). Consequently, age-specific fertility rate (AFSR) can sometimes give a clearer picture than TFR of the influence of education on fertility.

Ethnicity in England cannot be linked directly to TFR using the census. However, the "Own-Child" method (using data on co-resident minor children to estimate births and link children to their supposed mothers) is widely used to *indirectly* give reliable estimates, confirmed by comparison with *directly*-linked mother-child relationships such as from the labour force survey (LFS) (Wohlan et al, 2010, p42). In general, fertility drops the longer an ethnic group has been established in England due to ageing (initial immigrants tend to be younger) and successive generations acting closer to English norms; however, the Bangladeshi ethnic group was unusual in still having in 2001 a young age-structure which continued to grow rapidly (Rees and Butt, 2003). Using the LFS, Coleman and Dubuc (2010) noted that during 2001-2005 overall UK TFR was 1.71. The only UK ethnic groups with well above average TFR were Bangladeshi (2.97±0.33) and Pakistani (2.79±0.2) - the lowest TFR being Chinese (1.24±0.23). Bangladeshi and Pakistani groups also exhibited the highest teenage fertility, implying a possible link between education level and fertility. Based on 2001 census figures, Large and Ghosh (2006) gave similar but slightly less extreme TFR figures: overall TFR of 1.63, the highest being Pakistani (2.12), then Bangladeshi (1.94), the lowest Chinese (1.29).

Norman (2011, p105) also reported similar extremes of TFR for these 3 ethnic groups in the UK, but also emphasized lower fertility for students and higher fertility for those in deprivation. Townsend (1987) noted deprivation involved a wide range of factors which can be difficult to encompass with available statistics. He suggests that one method involves using the following statistics available in

the census: unemployment, overcrowding, lacking a car (reflecting lack of resources) and not owning a home (reflecting residential insecurity).

In this analysis of census results for influence on fertility, the effect of the Bangladeshi, Pakistani and Chinese ethnicity is thus expected to dominate. Education effects might be split: those currently students having a strong negative effect on TFR, but females with graduate level education having a smaller effect. Although not the key focus of this study, a consideration of deprivation might help improve the accuracy of a model.

METHODOLOGY

A multivariate linear regression model to estimate TFR is developed using the 2001 census-derived data supplied. The key census statistics supplied can be grouped by their possible relevance to the factors outlined above:

- **Ethnicity**: Pakistani or Bangladeshi ethnicity, Chinese ethnicity
- **Education**: Females with a degree, Students
- **Deprivation**: Unemployed Persons, Overcrowded Households, Households with no access to a car, Public Renters, Persons of Low Social Class
- **Others**: Females married, Armed forces, Women working over 31 hours per week

To improve accuracy, all variables determined influential are included – not just education and ethnicity related ones. Initially, bivariate Pearson Correlations are calculated between each variable and TFR. Variables showing significance *p*-values above 0.05 (indicating below 95% confidence) are taken as not acceptably reliable and rejected at this stage. Matrix bivariate correlation of the remaining variables is used to examine collinearity: even if this is not strong enough to reject outright variables as redundant, it gives a good indication of where variable co-dependence might cause regression calculation problems. Iterations of multivariate regressions with different groups of variables are made, selecting variables based on several factors: measured statistical significance, improvement of Adjusted R^2 measure (representing overall model fit), strength of effect on TFR, and likelihood based on plausible causality and literature findings.

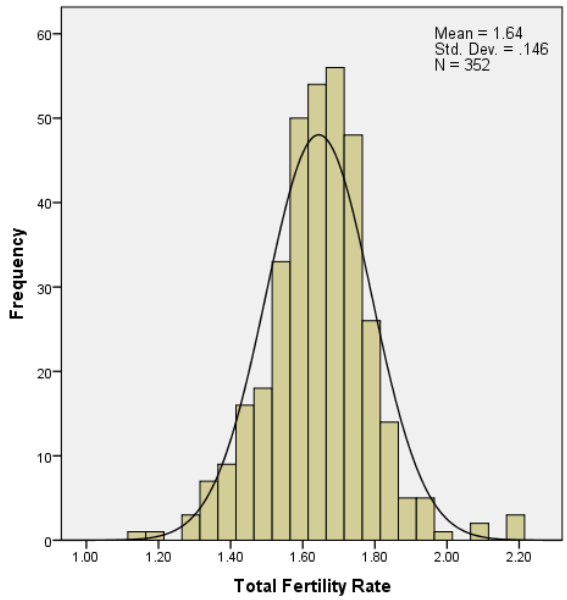

Figure 1: 2001 TFR Distribution in England (with overlay of normal distribution curve)

Several checks need to be made to confirm the model validity (Osborne and Waters, 2002). Figure 1 shows that TFR exhibits the required near-normal distribution. For each dependent variable, plots of standardized residuals from bivariate regression against predicted TFR are examined. Figure 2 illustrates this for the most influential variable: Chinese Ethnicity. Non-linearity would be indicated by visible slope or curvature in the plot. Homoscedasticity (lack of variance in errors for different dependent variable values) would be indicated by changes in residual spread (y-axis) for varying predicted values (x-axis).

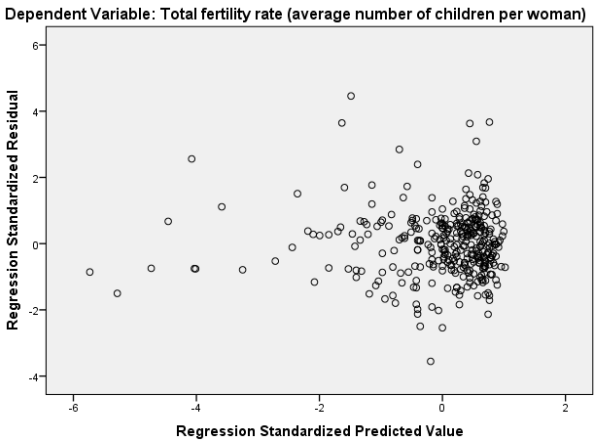

Figure 2: Linear regression-based bivariate model residual for predicted TFR dependent on Chinese Ethnicity

RESULTS AND ANALYSIS

The variables "Pakistani or Bangladeshi ethnicity", "Students" and "Chinese ethnicity" all show large (and statistically significant) apparent effects on TFR so are incorporated in the model. It is however noted that Chinese Ethnicity showed quite a high positive correlation (*p*-value <0.001) for both "Females with a Degree" (Correlation=0.661) and "Students" (Correlation=0.684), suggesting a level of collinearity - that perhaps Chinese ethnicity increases likelihood of being a student or graduate. However literature noted earlier suggests Chinese ethnicity has a very strong negative correlation with fertility, so it is kept in the model.

Adding "Females with a degree" to the model marginally improves the "Adjusted R^2" measure, but raises the "Chinese Ethnicity" *p*-value to 0.742. This is assumed to be a distortion in regression estimates due to the possible collinearity noted earlier and as "Chinese Ethnicity" has 10 times greater impact on TFR, "Females with a degree" is rejected. Previous literature suggested that this variable has only limited effect on TFR, so the effects of its omission should be small.

Adding "Overcrowded households" marginally improves model-fit. Though possibly a result of high fertility (thus being the dependent variable), it is more likely a good deprivation indicator (Norman, 2011, p105). Though having significant bivariate correlation (0.689) with "Pakistani or Bangladeshi ethnicity", it indicated a strong and statistically significant effect on TFR without unsettling measured contribution of other variables, suggesting overcrowding is an issue beyond these groups.

The variables "Households with no access to a car", "Unemployed persons", "Persons of Low Social Class" make either little improvement to model-fit or indicate unreliable *p*-values. Though indicative of possible deprivation and so relevant, these variables are not necessarily a particularly good method of estimating overall deprivation (Townsend, 1987). Thus, limiting the model to the inclusion of "Overcrowded households" as a measure of deprivation seems reasonable.

The multivariate linear regression coefficient estimates for the selected variables (as determined by the IBM SPSS Statistics program) are shown in Table 1, giving us the following model:

$$Predicted\ TFR = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4$$

Table 1: Total Fertility Rate (TFR) linear model coefficients

Index (n)	Census-derived Variable (x_n)	Coefficient (b_n)	Standard Error	Significance (*p*-value)
0	(Constant)	1.766	.016	.000
1	Pakistani or Bangladeshi ethnicity (%)	.018	.003	.000
2	Chinese ethnicity (%)	-.095	.019	.000
3	Students (%)	-.018	.002	.000
4	Overcrowded households (%)	.027	.006	.000

This regression analysis suggests that all the dependent variables included have a notable bearing on the outcome and are statistically significant. "Chinese ethnicity" reduces fertility with nearly 4 times the impact of any other factor, albeit with a much larger standard error in the prediction. Fertility increase with "Overcrowded Households" is next most notable, with "Student" and "Pakistani or Bangladeshi ethnicity" having equal but opposite impacts.

So how does the model perform? TFR figures recorded in the 2001 census are illustrated in figure 3, with figure 4 showing the errors in the model predictions from these. Out of the 354 English local authorities considered, only 5 of these (1.4%) give error magnitudes greater than 0.3 (±18% of TFR). These 5 cases are listed in Table 2 and show no obvious geographical patterns with both error extremes including densely populated inner-city urban and sparsely populated rural areas, in both the north and south of the country.

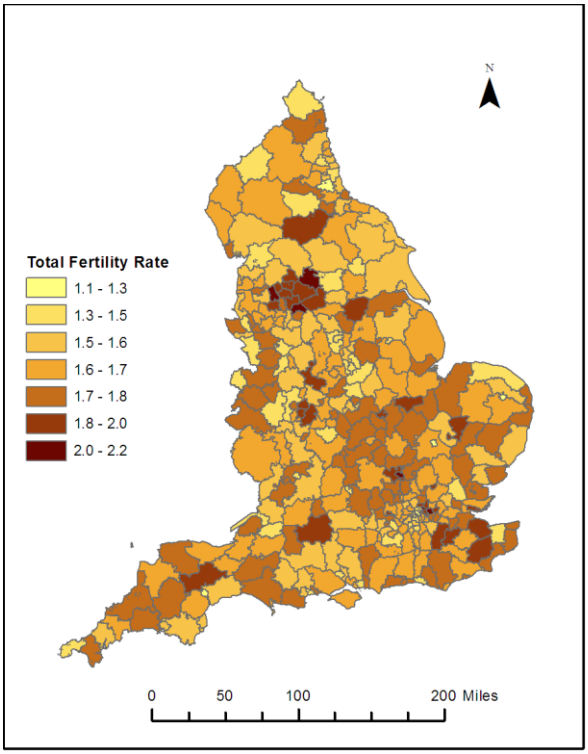

Figure 3: Total Fertility Rate (TFR) by local authority

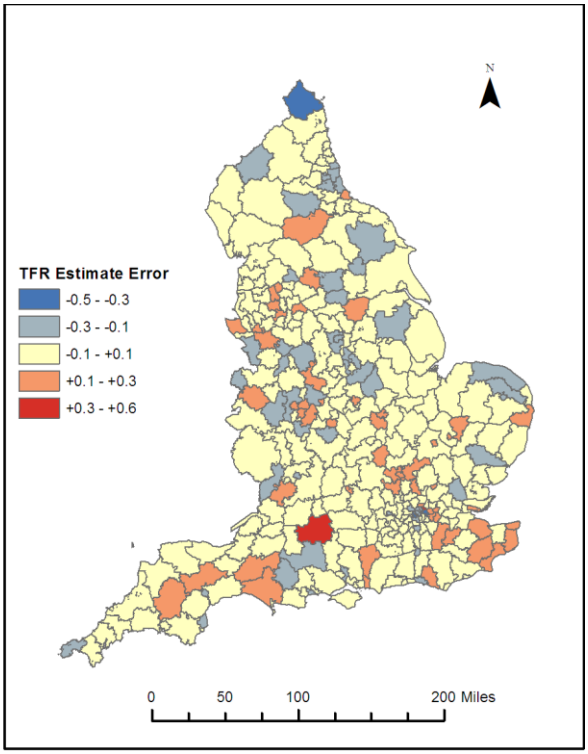

Figure 4: TFR Model Residuals by local authority

Table 2: Regression Model Errors of magnitude larger than 0.3

Local Authority	Region	TFR (Predicted)	TFR (Actual)	Error
Hackney	London	1.73	2.08	+0.35
Kennet	Wiltshire	1.65	1.96	+0.31
Berwick-upon-Tweed	Northumberland	1.68	1.38	-0.31
Hammersmith and Fulham	London	1.68	1.37	-0.31
Tower Hamlets	London	2.22	1.82	-0.40

DISCUSSION AND CONCLUSIONS

Strong positive correlation of "Pakistani and Bangladeshi ethnicities" and strong negative correlation of "Chinese ethnicity" is in agreement with general literature findings noted earlier. However, in this analysis "Chinese ethnicity" is seen as much more influential, which is the opposite of many previous quantitative findings (Coleman and Dubuc, 2010. Large and Ghosh, 2006). The relatively larger error bars in this estimation of "Chinese ethnicity" contribution to TFR may indicate a problem with using this data in the analysis, even though the correlation is reported as statistically highly significant. Although census data does not allow for direct linking of ethnicity and fertility, application of the "Own-Child" method to data from other surveys would give a more accurate estimate which could be used in this model.

The reduction in TFR for those classed as "students" agrees with general findings (Te Velde et al, 2012). However, fertility has a complex dependency on many different aspects of education (Diamond et al, 1999); the limited education-related statistics available for this analysis preclude further exploration of this. However, previous literature noted earlier indicated that overall effects of education on total fertility in England are small.

In conclusion, analysis of the 2001 census data supplied indicates that ethnicity is notably more influential on fertility than education. The map of prediction model residuals (figure 4) together with lack of obvious patterns in prediction errors suggest that for the statistics used the model is reasonably good. However, there are clearly other factors affecting fertility: previous literature noted earlier and the strong (but not obviously direct) effect of "Overcrowded Households" suggest that deprivation might be the more direct influence. Broader analysis using more deprivation-related statistics from other sources might thus improve model accuracy.

REFERENCES

Coleman, D. and S. Dubuc. 2010. The fertility of ethnic minorities in the UK, 1960s–2006. *Population Studies,* **64**(1), pp.19-41.

Diamond, I., M. Newby and S. Varle. 1999. Female education and fertility: Examining the links. *Critical perspectives on schooling and fertility in the developing world*, pp.23-48.

Large, P. and K. Ghosh. 2006. A methodology for estimating the population by ethnic group for areas within England. *Population Trends,* **123**, p21.

Norman, P. 2011. Relationships Between UK Sub-national Trends in Infant Mortality and Fertility. *In:* J. Stillwell and M. Clarke, eds. *Population Dynamics and Projection Methods.* Springer Netherlands, pp.99-114.

Osborne, J. and E. Waters. 2002. Four assumptions of multiple regression that researchers should always test. *Practical assessment, research & evaluation,* **8**(2), pp.1-9.

Rees, P. and F. Butt. 2004. Ethnic change and diversity in England, 1981–2001. *Area,* **36**(2), pp.174-186.

Rindfuss, R. R., S. P. Morgan and K. Offutt. 1996. Education and the Changing Age Pattern of American Fertility: 1963-1989. *Demography,* **33**(3), pp.277-290.

Te Velde, E., D. Habbema, H. Leridon and M. Eijkemans. 2012. The effect of postponement of first motherhood on permanent involuntary childlessness and total fertility rate in six European countries since the 1970s. *Human Reproduction,* **27**(4), pp.1179-1183.

Townsend, P. 1987. Deprivation. *Journal of social policy,* **16**(02), pp.125-146.

Wohland, P., P. Rees, P. Norman, P. Boden and M. Jasinska. 2010. *Ethnic population projections for the UK and local areas, 2001-2051.* School of Geography, University of Leeds.