# “How does income inequality across different ethnicities in the UK affect health outcomes?”

# Abstract

This research paper aims to investigate the impact of ethnic minorities experiencing income inequality and how it affects their overall health. Our findings reveal that as income increases, the impact on health outcomes worsen for most ethnic groups. Our model included other variables such as age, sex and benefit recipients in addition to income and ethnicity. Whilst our results may not be expected, they were significant enough for us to draw conclusions and present possible reasons for such findings, therefore our research is relevant in suggesting policy changes and economic interventions.


## I. Introduction

There are many factors that contribute to lower levels of income for ethnic minorities, such as racial discrimination, language barriers and differing cultural norms. According to a study conducted by Ciphr, "Over two thirds (69%) of survey respondents from ethnic minorities say they have experienced some form of discrimination in the workplace (61%) or when applying for a new job (65%) or both." (Ciphr, 2025). Throughout this paper we will try to compare the links between income inequality and the health levels, for example, we have assumed that since Black people and Pakistanis on average have the lowest income (Peter Matejic, 2024), their health outcomes will be the poorest amongst all ethnic groups. In one paper we found it states that "people from minoritized ethnic groups tend to have much poorer health outcomes over the life course than the white majority group" (Sarah Stopforth, 2021) .Therefore, it could be argued that even after economic improvements, ethnic minorities will have poor health in general due to other factors, such racism in medical institutions.

In this empirical research we will use econometrics methods such as OLS to conclude whether income inequality does determine differing levels of health outcomes. By using various studies and research papers, we aim to deepen our understanding and suggest possible solutions.

## II. Literature Review

Many studies provide evidence between the correlation of ethnic minority individuals and low-income brackets. For example, in the field of economics students from an ethnic minority are "less likely to study at Russell Group universities" where specifically 20% less of Black students and 25% less of Pakistani and Bangladeshi students' study in those universities (Arun Advani, 2022).

An additional study examines health disparities by race and ethnicity. Findings include that amongst Hispanic and Black men and women; obesity and diabetes are more common. Evidence suggests that “38.8% of white women find it difficult to “climb several flights of stairs” whilst the same applies for Hispanic women (51.5%) and Black women~~.~~ (53.5%). Although, the research suggests that the medical diagnosis of numerous conditions is generally less frequent (lnstitute for Fiscal Studies, 2024). This has many economic impacts as higher rates of health disparities can result in lower labour participation as well as higher medical costs. Furthermore, those who cannot afford it may remain undiagnosed leading to a cycle of poverty. This has many economic impacts as higher rates of health disparities can result in lower labour participation as well as higher medical costs. Furthermore, those who cannot afford it may remain undiagnosed leading to a cycle of poverty.

Next, this study highlights the existing pay gaps within the UK in the public and private sector, particularly focusing on the public sector. This paper finds that pay is bad in the public sector where UK-born white employees are paid 25% more than Pakistanis/Bangladeshis and 15% more than Black groups (Nigel O’Leary, 2023). Alternatively in the public sector, it has been conducted that the average pay of Bangladeshis and Pakistanis is 8% better than UK-born white workers, explained by their higher endowment of productivity that enhances their characteristics. To summarize, Black people see a pay disadvantage by approximately 3%. Whilst the paper doesn’t delve into the interconnection of income for ethnic minorities and poorer health, it does briefly mention that those from an ethnic background are more likely to face worsening health, indirectly affecting employment opportunities and income.

Following on, another paper explores potential different experiences for ethnic minorities. It outlines the potential reasons of inequality amongst ethnic groups like cultural differences. This report uses the UK Household Longitudinal Study as a reference, revealing factors such as health behaviours, identities and preferences vary for different ethnicities. These factors may also discover the effects on academic achievements and earnings for adults and the younger generation. Here, (Renee Luthra, 2020) finds that people who have a “greater ethnic attachment” (Lucinda Platt, 2020) are part of healthier engagements by smoking or drinking less. Hence, it can be suggested that cultural dynamics deliver protection in health for individuals, allowing this paper to highlight the importance of considering such cultural aspects in policy interventions to address issues such as “employment penalties” (Yaojun Li, 2020), to successfully reduce gaps in income and health for ethnicities.

Moving onto a different study, this paper uses different methodologies such as narrative synthesis and data extraction strategies on multiple datasets (including MEDLINE and ASSIA) in order to find evidence that minoritized ethnic groups face a greater chance in getting multiple long-term conditions (MLTCs) (Brenda Hayanga, 2023). Specifically, the paper shows certain ethnic groups, especially South Asians and Black Caribbean and Africans have a higher likeability in facing higher MLTCs rates compared to White people. Furthermore, it also discovers that those individuals develop MLTCs at a younger age. These inequalities are mainly due to people experiencing intense racism and discrimination, creating problems for them as they may potentially have limited access to important resources (e.g. good jobs) to remain in good health conditions. Also, the paper distinguishes few limitations in their research, predominantly "the lack of consensus in how MLTCs are captured" (Brenda Hayanga, 2023). By identifying such limitations, the paper offers advice for conducting future studies. one such recommendation is to take account of the post-pandemic effect on health inequalities, whilst another one is to combine the effect of variables like gender to understand influences on health.

The final paper discussed is the most appropriate regarding our empirical research question. This paper applies data from over 20 years for analysis, with key findings being that ethnic health dispirited have stayed high in the UK. This notably applies to Bangladeshi and Pakistani adults over the age of 40, being reported of having higher levels of long-term illnesses, which limit their daily-life activities. Furthermore, they have self-reported their health ratings as "poor" in contrast to White British people. Whilst the paper finds that adjustments have been made on socio-economic position (like income and job types), health disparities continue to remain although decreasing (Sarah Stopforth, 2021). This helps them examine that whilst income and job opportunities are critical aspects of health gaps in the UK, other factors like structural discrimination are significant contributors. Besides, this study supports the idea of "weathering" hypothesis, denoting that people facing socioeconomic disadvantages experience early deterioration in their health. They uncover that Pakistanis and Bangladeshis are underrepresented within high-paying and professional areas, reinforcing the idea that those excluded from opportunities in the economy will face poorer health later in life.

## III. Econometric Models

We are using an OLS Model, with the health outcome being the dependent variable, specifically looking at health problems lasting or expected to last more than 1 year and ethnicity and income being the independent variables. Ethnicity is typically a categorical variable; therefore, we will be converting it into a dummy variable either 1 or 0 and White will be the reference variable which we will use to compare the results for other ethnic minorities to.

### III.I Initial Model

$$\mathrm{LNGLST}_i = \beta_0 + \beta_1\mathrm{ETHUKEUL}_i + \beta_2\mathrm{CLAIMS14}_i + \beta_3\mathrm{STAT}_i + \beta_4\mathrm{HOURPAYi}_i + \beta_5\mathrm{AGE}_i + \beta_6\mathrm{SEX}_i + \varepsilon_i$$

However, there is the possibility of endogeneity arising, for example, the issue of reverse causality where poor health is the factor that is leading to a lower income. Furthermore, other factors such as environment and different dietary habits that could have a knock-on effect on income instead as an individual may have poor health to start with. To combat this, we could use IV.

By using this model, we are also assuming that income affects all ethnic minorities in the same way, however, since the Indian ethnic group tends to have a higher income compared to others, which could mean they have better access to healthcare. Therefore, it is best to include a more complex model that would account for these differences, as income and ethnicity would be interacting.

From regressing our initial model, (refer to MODEL 1 in the appendix) the calculated coefficients are positive and there is sufficient evidence to suggest that there is a link between ethnic minorities and worse health outcomes, as the coefficient of ETHUKEUL (Ethnicity) is 0.03131, meaning that on average ethnic minorities report 3.13% more health problems. The biggest factors affecting worse health outcomes are age at 9.3 % which is reasonable as age increases health problems arise, and whether someone is a benefit recipient (CLAIMS14), a person receiving benefits is 32.1% more likely to have higher health problems. Although the effect for hourly pay (HOURPAY) is significant (p=0.004) the effect size is small, as an hourly pay increase is associated with a 0.22% increase in long term health problems, suggesting that income may impact health. Furthermore, higher employment status (STAT) is associated with an increase in 6.7% increase in health issues, which could be caused by stress related conditions.

The R-squared and adjusted R-squared values suggest that most of the variance in the data (around 83%) is explained by the model. The Prob > F (0.0000) means that the model is statistically significant overall, suggesting there are no overfitting issues.

## III.II Extended Model

$$\mathrm{LNGLST}_i = \beta_0 + \beta_1 \mathrm{HOURPAY}_i + \sum \beta_j \mathrm{ETHNICITY}_{ji} + \sum \beta_{int}(\mathrm{HOURPAY}_i \times \mathrm{ETHNICITY}_{ji})$$
$$+ \beta_2 \mathrm{CLAIMS14}_i + \beta_3 \mathrm{STAT}_i + \beta_4 \mathrm{AGE}_i + \beta_5 \mathrm{SEX}_i + \varepsilon_i$$

In our extended model, we include the interaction terms HOURPAY and ETHNICITY in order to delve into further analysis and investigate how income by certain ethnic groups impact health outcomes. We're particularly assessing the effect of working hours on health (expressed as LNGLST). The key variables CLAIMS14, AGE, SEX and STAT are shown highly significant in our model (refer to Model 3) with high t-values of 324.87 (AGE) and 277.23 (CLAIMS14) and all with p-values less than 0.001. This indicates that they have strong connections with health due to them being the main demographic controls in plenty of health inequality research conducted. Age is usually correlated with more health problems. For STAT and CLAIMS14, we are aware that socio-economic status directly links with worse health conditions e.g. unemployed people (who also apply for unemployment-related benefits) have limited access to healthcare and financial resources. Focusing onto the interaction terms, our results highlight that Pakistanis (p-value<0.05) and Black/African/Caribbean/Black British people (p-value<0.05) are significant, with Bangladeshis being borderline significant with a p-value of 0.071. This implies that as they work for more hours, they will face more health problems. However, we do some OLS assumption checks for our results to be reliable and valid. We conducted tests such as the Breusch-Pagan test and Ramsey RESET, where we find indications of heteroskedasticity and omitted variable bias in Model 3. Furthermore, another test called the Cameron & Trivedi IM-test is conducted (refer to test 4 in appendix). Here, we find there is presence of heteroskedasticity (with p-value = 0.000) as we found already. This suggests that the residuals' variance isn't constant, violating an assumption of the OLS regression. Further on, we find that the test for skewness and kurtosis, both with p = 0.000 indicate that the residuals don't follow the shape of a normal distribution. So, for reliable conclusions to be made, future analysis needs to be conducted as residuals aren't normally distributed.

Based on evidence, changes and improvements are made in our extended model (refer to Model 4) including combust standard errors to correct for heteroskedasticity and retained interaction terms, to tackle the potential non-linear relationship between income and ethnic groups. Notably, Indian and Bangladeshi interactions now become significant in Model 4, with p-values changing from 0.124 to 0.032 (Indian) and 0.071 to 0.004 (Bangladeshi). This emphasizes that benefits of health from income may vary for ethnic group minorities. In our extended model, we considered adding other control variables, but some weren't available in our dataset or would either conflict or create a complex model (like self-employment, marital status, and disability), which we recognize as a downside of our model and hypothesis in our conclusion. Overall, the use of robust errors in Model 4 makes our outcomes remain consistent.

## III.III Final Model

The following is a model including all the variables from our data set with ethnicity interacting with hourly pay.

$$LNGLST_i = \beta_0 + \beta_1 HOURPAY_i + \sum\beta_j ETHNICITY_{ji} + \sum\beta_{int}(HOURPAY_i \times ETHNICITY_{ji})$$
$$+ \beta_2 CLAIMS14_i + \beta_3 STAT_i + \beta_4 AGE_i + \beta_5 SEX_i + \varepsilon_i$$

From regressing our final model (refer to MODEL 2 in the appendix), we have found that increasing income for some ethnic minorities does not necessarily decrease health issues. Bangladeshis experience the largest effect, with a 3.2% increase in reported long-term health issues for each additional £1/hour earned. This is followed by individuals from the Mixed/Multiple ethnic group 2.88%, Pakistanis 1.79%, and for the Black/African/Caribbean/Black British group 1.16%. As $p < 0.05$ this means that these results are statistically significant. Although other ethnic backgrounds, Indian and other Asian, show positive coefficients, these effects are not statistically significant. However, for the Chinese ethnic group income is negatively associated with health problems, though this result is not statistically significant either.

When referring to the marginal effects (refer to MODEL 3 in the appendix) we found that for the Mixed/Multiple ethnic groups, a £1/hour increase in income is associated with a statistically significant 2.98% increase in the likelihood of reporting a long-term health condition. Being both statistically significant, Pakistanis experience an increase of 1.85% and Black/African/Caribbean individuals experience increases of 1.25%. Whilst only marginally significant, the Bangladeshi group shows the largest increase at 3.3%. The White, Indian, and Chinese groups do not show significant marginal effects, the Chinese group even suggesting a small but non-significant positive effect, this could be due to factors such as the 'healthy immigrant' effect, where only healthy individuals would be able to migrate to a different country. Overall, the analysis highlights that the marginal effect of income on health is not uniform, and that increasing wages do not mean better health, therefore reinforcing the need for equity-focused health and economic policy.

After testing for endogeneity (refer to Model 6) we have found that for first stage regression, BENTYP4 (income support) coefficient is -0.2458 with t-statistic of -8.75, meaning that BENTYP4 and HOURPAY have a statistically significant relationship, hence proving income support to be a valid instrument variable for hour pay. In the second stage, we can see that HOURPAY has a coefficient of -0.1044 with a z-statistic of -3.74 and a p-value of 0.000. indicating statistical significance. Hence, after endogeneity is taken into consideration, we can conclude that HOURPAY does have a negative effect on LNGST. This means that the more hours you work, the less health problems there should be. When we conduct the endogeneity tests using the Durbin-Wu-Hausman Test and the Wu-Hausman F Test. With the Durbin-Wu-Hausman Test (refer to test 5 in appendix), it yields a chi-squared statistics of 17.9892 and a p-value of 0.000. And the Wu-Hausman F Test yields a F-statistic of 17.9913 with a p-value of 0.000. based on these results, we confirm that the variable HOURPAY is endogenous and it's appropriate to use our chosen instrument, since both tests reject the null hypothesis (which indicates that HOURPAY is exogenous). The purpose of these endogeneity tests is to determine whether HOURPAY is truly endogenous, meaning if HOURPSY is correlated with the error term, to justify the use of BENTYP4 as an instrument for 2SLS.

To further strengthen the instrument BENTYP4, we conducted another test (refer to test 6 in the appendix). Our findings include an F-statistic of 76.51 (above the used threshold of 10), further suggesting that BENTYP4 is a strong instrument for HOURPAY. This reduces concerns of having utilised a weak instrument in the two-stage least squares (2SLS) regression. Furthermore, we receive a minimum eigenvalue statistic of 76.51. In models with a single endogenous regressor and one instrument, this statistic equals the first-stage F-statistic. This underpins the point of having a strong and valid instrument in our model.

## IV. Data

The data that was used for the research was obtained from the Quarterly Labour Force Survey, July - September 2018, which is drawn from Office for National Statistics and is made available through the UK Data Service website. The descriptive statistics are given in Table 1 and the definitions for the variables are given in Table 2. A large sample size of 87,555 observations allowed for more accurate results when using econometrics models the key endogenous variables that are essential for the linear regression test are Ethnicity (9 categories) UK Level, Income Support, Gross Hourly Pay, LNGLST and Current disability status. LNGLST is the main categorical variable that measures health problems lasting or expected to last for 2 years. This variable is critical to assess how health outcome vary between ethnic groups and income. Exogenous variables include; Sex, Nationality and Martial status. The addition of other variables also improves the model as we can infer if income is the main factor impacting health. A limitation is the survey was conducted in 2018 which may not carry the same results in the current year, as improvements may be undertaken. There is no information provided on the geographic location, as the correlation may vary from city to city in the UK. This data would be useful for policy planning to support ethnic minorities that may face disability and income barriers. By providing an understanding of factors that may contribute to inequalities, policies can be adapted to accommodate the barriers of ethnic individuals who face income disparities.

Also, we did some checks to appropriately include categorical variables such as SEX, ETHNICITY, and STAT in our regression model (refer to appendix). We will treat them as categorical variables using the 'i.' prefix in Stata, to create dummy variables for each of the 3 categories.

**Table 1. Descriptive Statistics**

| Variable | Mean | Standard Deviation | Minimum | Maximum |
|---|---|---|---|---|
| Age of Respondent | 39.52016 | 22.88546 | 0 | 99 |
| Income Support | -8.901749 | 0.9863389 | -9 | 1 |
| Sickness or disability benefits | -8.582349 | 2.00053 | -9 | 1 |
| Claiming unemployment | 2.726884 | 7.837829 | -9 | 8 |
| Current Disability | 0.6792076 | 5.203719 | -9 | 4 |
| Ethnicity | 1.53214 | 1.716717 | -8 | 9 |
| Gross Hourly Pay | -6.344483 | 8.386105 | -9 | 369.2 |
| Basic Economic Activity | 15.58567 | 14.98767 | 1 | 34 |
| Whether health problem affects amount of work can do | -6.755275 | 4.327804 | -9 | 2 |
| Whether health problem affects kind of work can do | -6.761585 | 4.312175 | -9 | 2 |

| | | | | |
|---|---|---|---|---|
| Health problems lasting or expected to last more than 1 year | -0.6671345 | 4.385089 | -9 | 4 |
| Marital Status | 1.802175 | 1.033317 | 1 | 6 |
| Nationality | 923.5565 | 69.8754 | -9 | 997 |
| Self-employed Status | -7.718588 | 3.937324 | -9 | 8 |
| Sex of Respondent | 1.512353 | 0.4998502 | 1 | 2 |
| Employment Status | -2.768729 | 4.952992 | -9 | 4 |

**Table 2. Variable Definitions**

| Variable Label | Definition |
|---|---|
| Dependent Variable | |
| LNGLST | Health problems lasting or expected to last more than 1 year |
| Independent Variables: Initial Model | |
| HOURPAY | Gross hourly pay for each individual |
| ETHUKEUL | Ethnicity (9 Categories)<br>1 = White<br>0 = Otherwise |
| Independent Variables: Extended Model | |
| HOURPAY | Gross hourly pay for each individual |
| STAT | Employment status: employee, self-employed, government scheme |
| CLAIMS14 | Whether claiming unemployment related benefits |
| AGE | Age in years |
| SEX | Female or Male |

## V. <u>Conclusion</u>

Through conducting an econometrics analysis of our data, our findings seem to contradict our hypothesis. As mixed/multiethnic individuals show the strongest positive effect; a £1 increase in income leads to a 2.98% increase in health problems. Whilst our model predicts that Chinese individuals' health improves by 3.12% as income increases by £1. Although the model may lack accuracy, it could be improved, for example adding more control variables such education, the overarching conclusion is that higher income does not improve health and may in fact worsen health outcomes, as better income may be accompanied by increased occupational stress or discrimination that could not be accounted for in the model. This also implies that there are systemic barriers that limit the health benefits typically associated with financial improvement. Therefore, this highlights the importance of examining how structural and cultural factors interact with socioeconomic status to shape health outcomes, reinforcing the need for equity-focused policy responses that consider both ethnicity and income together.

Possible solutions include health programmes tailored to ethnic minorities, such as multilingual heath navigators and Anti-Discrimination policies that could focus on improving work environments for ethnic minorities, for example, there could be a mandatory ethnic pay gap reports done every year that could track any changes over a period of time.

We believe that our model is a great foundation but lacks the ability to provide a satisfactory explanation for our research question. Regardless of conducting OLS assumption checks, our model may remain underspecified (according to the RESET test). This could be due to the absence of some key variables where the estimates of coefficients may be biased from omitted variable bias. Alternatively, our model may just be too simple. As a response, we have included interaction terms between ethnic groups and hourly pay (as a measure of income) in models 3 and 4 (refer to appendix), an adaptation of initial models 1 and 2. But, we are aware of some presence of bias still.

After conducting further tests e.g., a test for endogeneity we found that the 2SLS regression doesn't follow the same results as our OLS regression model, as the OLS model shows that as income increases health worsens but the 2SLS model shows that income increases health improves,  due to the presence of endogeneity and that residuals not being normally distributed.

For the future, we consider incorporating relevant extra variables if they are available to enhance our previous model, due to the complicating relationship of income, health and ethnicities. Nevertheless, our results do suggest that variables income and ethnicity and their interaction terms are relatively significant factors that do explain the health outcome differences for all ethnic groups in the UK.

# References

Brenda Hayanga, M. S. L. B., 2023. *Ethnic inequalities in multiple long-term health conditions in the United Kingdom: a systematic review and narrative synthesis,* s.l.: BMC Public Health.

Ciphr, 2025. *Workplace discrimination statistics in 2025 | Discrimination at work.* [Online] Available at: https://www.ciphr.com/infographics/workplace-discrimination-statistics [Accessed March 2025].

Institute for Fiscal Studies, 2024. *Health inequality and economic disparities by race, ethnicity, and gender.* [Online]
Available at: https://ifs.org.uk/sites/default/files/2024-10/WP202441-Health-inequality-and-economic-disparities-by-race-ethnicity-and-gender.pdf

Lucinda Platt, A. N., 2020. Ethnic diversity in the UK: new opportunities and changing. *Journal of ethnic and Migration Studies,* p. 18.

Nigel O'Leary, M. J. S. D. D. B., 2023. *Understanding Ethnicity Pay Gaps in the UK Public Sector,* s.l.: UK Government.

Peter Matejic, K. S. Y. I. I. T., 2024. *Bangladeshi, Black African and Pakistani households at higher risk of very deep, long-term poverty,* s.l.: Joseph Rowntree Foundation.

Renee Luthra, A. N. M. B., 2020. Unravelling the 'immigrant health paradox': ethnic maintenance, discrimination, and health behaviours of the foreign born and their children in England. *JOURNAL OF ETHNIC AND MIGRATION STUDIES,* 46(5), p. 980–1001.

Sarah Stopforth, D. K. J. N. L. B., 2021. Ethnic inequalities in health in later life, 1993–2017: the persistence of health disadvantage over more than two decades. *Ageing & Society,* 43(8).

Yaojun Li, A. H., 2020. Persisting disadvantages: a study of labour market dynamics of ethnic unemployment and earnings in the UK (2009–2015). *Journal of Ethnic and Migration Studies,* 46(5), pp. 857-878.

# Appendix

## Model 1 – initial model/OLS regression

```
. reg LNGLST ETHUKEUL CLAIMS14 STAT HOURPAY AGE SEX
```

| Source | SS | df | MS |
|---|---|---|---|
| Model | 1403929.06 | 6 | 233988.176 |
| Residual | 281147.258 | 87,626 | 3.20849129 |
| Total | 1685076.31 | 87,632 | 19.2290067 |

| | | |
|---|---|---|
| Number of obs | = | 87,633 |
| F(6, 87626) | = | 72927.79 |
| Prob > F | = | 0.0000 |
| R-squared | = | 0.8332 |
| Adj R-squared | = | 0.8331 |
| Root MSE | = | 1.7912 |

| LNGLST | Coefficient | Std. err. | t | P>\|t\| | [95% conf. | interval] |
|---|---|---|---|---|---|---|
| ETHUKEUL | .03131 | .0035535 | 8.81 | 0.000 | .0243453 | .0382748 |
| CLAIMS14 | .3206722 | .0011628 | 275.77 | 0.000 | .3183931 | .3229513 |
| STAT | .0670011 | .0018737 | 35.76 | 0.000 | .0633287 | .0706735 |
| HOURPAY | .0021481 | .000745 | 2.88 | 0.004 | .0006879 | .0036083 |
| AGE | .0927427 | .0002848 | 325.70 | 0.000 | .0921846 | .0933008 |
| SEX | .0851222 | .012149 | 7.01 | 0.000 | .0613103 | .1089342 |
| _cons | -5.184346 | .0243206 | -213.17 | 0.000 | -5.232014 | -5.136678 |

## Model 2 – initial OLS regression (changed)

```
. reg LNGLST ETHUKEUL CLAIMS14 STAT HOURPAY AGE SEX, robust
```

Linear regression

| | | |
|---|---|---|
| Number of obs | = | 87,555 |
| F(6, 87548) | > | 99999.00 |
| Prob > F | = | 0.0000 |
| R-squared | = | 0.8343 |
| Root MSE | = | 1.7844 |

| LNGLST | Coefficient | Robust std. err. | t | P>\|t\| | [95% conf. | interval] |
|---|---|---|---|---|---|---|
| ETHUKEUL | .0217658 | .003333 | 6.53 | 0.000 | .0152332 | .0282985 |
| CLAIMS14 | .3219781 | .0015772 | 204.14 | 0.000 | .3188868 | .3250695 |
| STAT | .0646373 | .0023994 | 26.94 | 0.000 | .0599345 | .06934 |
| HOURPAY | .002087 | .0006769 | 3.08 | 0.002 | .0007602 | .0034138 |
| AGE | .0927177 | .0003755 | 246.89 | 0.000 | .0919816 | .0934538 |
| SEX | .0833274 | .0121575 | 6.85 | 0.000 | .0594987 | .107156 |
| _cons | -5.17366 | .024014 | -215.44 | 0.000 | -5.220727 | -5.126593 |

## Model 3 – OLS regression (extended)

```
. reg LNGLST i.ETHUKEUL##c.HOURPAY CLAIMS14 STAT AGE SEX
```

| Source | SS | df | MS |
|---|---|---|---|
| Model | 1403899.81 | 21 | 66852.3717 |
| Residual | 278482.274 | 87,533 | 3.1814547 |
| Total | 1682382.08 | 87,554 | 19.2153651 |

| | | |
|---|---|---|
| Number of obs | = | 87,555 |
| F(21, 87533) | = | 21013.15 |
| Prob > F | = | 0.0000 |
| R-squared | = | 0.8345 |
| Adj R-squared | = | 0.8344 |
| Root MSE | = | 1.7837 |

| LNGLST | Coefficient | Std. err. | t | P>\|t\| | [95% conf. | interval] |
|---|---|---|---|---|---|---|
| ETHUKEUL | | | | | | |
| Mixed/Multiple ethnic groups | .098133 | .0667347 | 1.47 | 0.141 | -.0326663 | .2289324 |
| Indian | .159082 | .0470227 | 3.38 | 0.001 | .0669179 | .251246 |
| Pakistani | .2086391 | .0727247 | 2.87 | 0.004 | .0660993 | .3511789 |
| Bangladeshi | .5288243 | .1613026 | 3.28 | 0.001 | .2126726 | .8449761 |
| Chinese | .52851 | .0959 | 5.51 | 0.000 | .3405469 | .7164731 |
| Any other Asian background | .2016766 | .0802157 | 2.51 | 0.012 | .0444545 | .3588986 |
| Black/African/Caribbean/Black British | .1288041 | .0505543 | 2.55 | 0.011 | .0297181 | .2278901 |
| Other ethnic group | .2509352 | .0725243 | 3.46 | 0.001 | .1087881 | .3930823 |
| HOURPAY | .0009923 | .0007803 | 1.27 | 0.204 | -.0005371 | .0025216 |
| ETHUKEUL#c.HOURPAY | | | | | | |
| Mixed/Multiple ethnic groups | .0287611 | .0066279 | 4.34 | 0.000 | .0157703 | .0417518 |
| Indian | .0051492 | .003956 | 1.30 | 0.193 | -.0026046 | .012903 |
| Pakistani | .017494 | .0075078 | 2.33 | 0.020 | .0027788 | .0322093 |
| Bangladeshi | .0319896 | .0177026 | 1.81 | 0.071 | -.0027073 | .0666866 |
| Chinese | -.0032124 | .0081715 | -0.39 | 0.694 | -.0192284 | .0128036 |
| Any other Asian background | .0103346 | .0080357 | 1.29 | 0.198 | -.0054153 | .0260845 |
| Black/African/Caribbean/Black British | .0115059 | .0051421 | 2.24 | 0.025 | .0014275 | .0215843 |
| Other ethnic group | .009523 | .0073901 | 1.29 | 0.198 | -.0049615 | .0240076 |
| CLAIMS14 | .3217509 | .0011606 | 277.23 | 0.000 | .3194761 | .3240256 |
| STAT | .064615 | .0018717 | 34.52 | 0.000 | .0609464 | .0682836 |
| AGE | .0926614 | .0002852 | 324.87 | 0.000 | .0921024 | .0932205 |
| SEX | .0836348 | .0121054 | 6.91 | 0.000 | .0599083 | .1073614 |
| _cons | -5.157403 | .0237238 | -217.39 | 0.000 | -5.203901 | -5.110904 |

## Model 4 – extended OLS model (changed)

```
. reg LNGLST i.ETHUKEUL##c.HOURPAY CLAIMS14 STAT AGE SEX, robust
```

Linear regression

| | | |
|---|---|---|
| Number of obs | = | 87,555 |
| F(21, 87533) | = | 31173.51 |
| Prob > F | = | 0.0000 |
| R-squared | = | 0.8345 |
| Root MSE | = | 1.7837 |

| LNGLST | Coefficient | Robust std. err. | t | P>\|t\| | [95% conf. | interval] |
|---|---|---|---|---|---|---|
| ETHUKEUL | | | | | | |
| Mixed/Multiple ethnic groups | .098133 | .0503121 | 1.95 | 0.051 | -.0004782 | .1967442 |
| Indian | .159082 | .0362514 | 4.39 | 0.000 | .0880296 | .2301343 |
| Pakistani | .2086391 | .0478943 | 4.36 | 0.000 | .1147667 | .3025114 |
| Bangladeshi | .5288243 | .102134 | 5.18 | 0.000 | .3286426 | .729006 |
| Chinese | .52851 | .0764598 | 6.91 | 0.000 | .3786495 | .6783705 |
| Any other Asian background | .2016766 | .0624817 | 3.23 | 0.001 | .0792129 | .3241402 |
| Black/African/Caribbean/Black British | .1288041 | .0363618 | 3.54 | 0.000 | .0575352 | .2000729 |
| Other ethnic group | .2509352 | .0497645 | 5.04 | 0.000 | .1533973 | .3484731 |
| HOURPAY | .0009923 | .0007126 | 1.39 | 0.164 | -.0004045 | .002389 |
| ETHUKEUL#c.HOURPAY | | | | | | |
| Mixed/Multiple ethnic groups | .0287611 | .004689 | 6.13 | 0.000 | .0195707 | .0379514 |
| Indian | .0051492 | .0024026 | 2.14 | 0.032 | .0004401 | .0098584 |
| Pakistani | .017494 | .004402 | 3.97 | 0.000 | .0088661 | .026122 |
| Bangladeshi | .0319896 | .0109652 | 2.92 | 0.004 | .0104979 | .0534813 |
| Chinese | -.0032124 | .0051409 | -0.62 | 0.532 | -.0132885 | .0068637 |
| Any other Asian background | .0103346 | .0057378 | 1.80 | 0.072 | -.0009114 | .0215807 |
| Black/African/Caribbean/Black British | .0115059 | .003301 | 3.49 | 0.000 | .005036 | .0179758 |
| Other ethnic group | .009523 | .004663 | 2.04 | 0.041 | .0003836 | .0186625 |
| CLAIMS14 | .3217509 | .0015789 | 203.79 | 0.000 | .3186563 | .3248454 |
| STAT | .064615 | .0024019 | 26.90 | 0.000 | .0599072 | .0693228 |
| AGE | .0926614 | .0003766 | 246.03 | 0.000 | .0919233 | .0933996 |
| SEX | .0836348 | .0121549 | 6.88 | 0.000 | .0598113 | .1074583 |
| _cons | -5.157403 | .0235358 | -219.13 | 0.000 | -5.203533 | -5.111273 |

## Model 5 – Marginal Effects

```
. margins ETHUKEUL, dydx(HOURPAY)

Average marginal effects                          Number of obs = 87,555
Model VCE: OLS

Expression: Linear prediction, predict()
dy/dx wrt:  HOURPAY
```

| | dy/dx | Delta-method std. err. | t | P>\|t\| | [95% conf. | interval] |
|---|---|---|---|---|---|---|
| **HOURPAY** | | | | | | |
| ETHUKEUL | | | | | | |
| White | .0009923 | .0007803 | 1.27 | 0.204 | -.0005371 | .0025216 |
| Mixed/Multiple ethnic groups | .0297533 | .0065901 | 4.51 | 0.000 | .0168368 | .0426698 |
| Indian | .0061415 | .0038846 | 1.58 | 0.114 | -.0014723 | .0137553 |
| Pakistani | .0184863 | .0074739 | 2.47 | 0.013 | .0038376 | .033135 |
| Bangladeshi | .0329819 | .0176892 | 1.86 | 0.062 | -.0016888 | .0676526 |
| Chinese | -.0022201 | .0081373 | -0.27 | 0.785 | -.0181692 | .013729 |
| Any other Asian background | .0113269 | .0080025 | 1.42 | 0.157 | -.004358 | .0270118 |
| Black/African/Caribbean/Black British | .0124982 | .0050914 | 2.45 | 0.014 | .002519 | .0224773 |
| Other ethnic group | .0105153 | .0073544 | 1.43 | 0.153 | -.0038992 | .0249298 |

**Model 6 – 2SLS**

```
. ivregress 2sls LNGLST (HOURPAY = BENTYP4) ETHUKEUL CLAIMS14 STAT AGE SEX, first
```

First-stage regressions

| | | |
|---|---|---|
| Number of obs | = | 87,555 |
| F(6, 87548) | = | 977.89 |
| Prob > F | = | 0.0000 |
| R-squared | = | 0.0628 |
| Adj R-squared | = | 0.0627 |
| Root MSE | = | 8.1203 |

| HOURPAY | Coefficient | Std. err. | t | P>\|t\| | [95% conf. | interval] |
|---|---|---|---|---|---|---|
| ETHUKEUL | -.0466657 | .0163395 | -2.86 | 0.004 | -.078691 | -.0146403 |
| CLAIMS14 | .0945218 | .005317 | 17.78 | 0.000 | .0841004 | .1049431 |
| STAT | .31547 | .0084988 | 37.12 | 0.000 | .2988123 | .3321276 |
| AGE | -.0136251 | .0012906 | -10.56 | 0.000 | -.0161546 | -.0110956 |
| SEX | -.1473557 | .0551222 | -2.67 | 0.008 | -.2553947 | -.0393167 |
| BENTYP4 | -.2458146 | .0281035 | -8.75 | 0.000 | -.3008972 | -.190732 |
| _cons | -7.083032 | .2800641 | -25.29 | 0.000 | -7.631955 | -6.534109 |

Instrumental-variables 2SLS regression

| | | |
|---|---|---|
| Number of obs | = | 87,555 |
| Wald chi2(6) | = | 356945.76 |
| Prob > chi2 | = | 0.0000 |
| R-squared | = | 0.7953 |
| Root MSE | = | 1.983 |

| LNGLST | Coefficient | Std. err. | z | P>\|z\| | [95% conf. | interval] |
|---|---|---|---|---|---|---|
| HOURPAY | -.104418 | .0279195 | -3.74 | 0.000 | -.1591392 | -.0496968 |
| ETHUKEUL | .0167542 | .0042009 | 3.99 | 0.000 | .0085205 | .0249878 |
| CLAIMS14 | .3313896 | .0027827 | 119.09 | 0.000 | .3259357 | .3368435 |
| STAT | .0991984 | .0092912 | 10.68 | 0.000 | .080988 | .1174089 |
| AGE | .0912551 | .0004963 | 183.87 | 0.000 | .0902823 | .0922278 |
| SEX | .066152 | .0141888 | 4.66 | 0.000 | .0383424 | .0939616 |
| _cons | -5.687781 | .1373854 | -41.40 | 0.000 | -5.957052 | -5.418511 |

Endogenous: **HOURPAY**
Exogenous: **ETHUKEUL CLAIMS14 STAT AGE SEX BENTYP4**

**Test 1 - VIF**

```
. vif
```

| Variable | VIF | 1/VIF |
|---|---|---|
| ETHUKEUL | | |
| 2 | 1.94 | 0.515297 |
| 3 | 1.36 | 0.735517 |
| 4 | 2.67 | 0.374913 |
| 5 | 4.82 | 0.207606 |
| 6 | 1.26 | 0.791967 |
| 7 | 1.73 | 0.578626 |
| 8 | 1.91 | 0.524800 |
| 9 | 2.08 | 0.481786 |
| HOURPAY | 1.18 | 0.848287 |
| ETHUKEUL# c.HOURPAY | | |
| 2 | 1.95 | 0.513603 |
| 3 | 1.39 | 0.717751 |
| 4 | 2.68 | 0.373715 |
| 5 | 4.82 | 0.207506 |
| 6 | 1.27 | 0.787966 |
| 7 | 1.74 | 0.576241 |
| 8 | 1.93 | 0.519240 |
| 9 | 2.09 | 0.479585 |
| CLAIMS14 | 2.28 | 0.439053 |
| STAT | 2.36 | 0.422835 |
| AGE | 1.17 | 0.852633 |
| SEX | 1.01 | 0.992456 |
| Mean VIF | 2.08 | |

Test 2 – Breusch Pagan Test

```
Breusch–Pagan/Cook–Weisberg test for heteroskedasticity
Assumption: Normal error terms
Variable: Fitted values of LNGLST

H0: Constant variance

    chi2(1) = 367.68
Prob > chi2 = 0.0000
```

### Test 3 – Ramsey Reset Test

```
. estat ovtest

Ramsey RESET test for omitted variables
Omitted: Powers of fitted values of LNGLST

H0: Model has no omitted variables

F(3, 87530) = 142454.03
   Prob > F =    0.0000
```

### Test 4 – IM test on final model

```
. estat imtest

Cameron & Trivedi's decomposition of IM-test
```

| Source | chi2 | df | p |
|---|---|---|---|
| Heteroskedasticity | 13108.27 | 94 | 0.0000 |
| Skewness | 332.74 | 21 | 0.0000 |
| Kurtosis | 71.38 | 1 | 0.0000 |
| Total | 13512.40 | 116 | 0.0000 |

### Test 5 – endogeneity tests

```
. estat endog

  Tests of endogeneity
  H0: Variables are exogenous

  Durbin (score) chi2(1)          =  17.9892  (p = 0.0000)
  Wu-Hausman F(1,87547)           =  17.9913  (p = 0.0000)
```

**Test 6 - First-Stage F-Statistic**

```
. estat firststage
```

First-stage regression summary statistics

| Variable | R-sq. | Adjusted R-sq. | Partial R-sq. | F(1,87548) | Prob > F |
|---|---|---|---|---|---|
| HOURPAY | 0.0628 | 0.0627 | 0.0009 | 76.5059 | 0.0000 |

Minimum eigenvalue statistic = 76.5059

| Critical Values | # of endogenous regressors: 1 |
|---|---|
| H0: Instruments are weak | # of excluded instruments: 1 |

| | 5% | 10% | 20% | 30% |
|---|---|---|---|---|
| 2SLS relative bias | (not available) | | | |
| | **10%** | **15%** | **20%** | **25%** |
| 2SLS size of nominal 5% Wald test | 16.38 | 8.96 | 6.66 | 5.53 |
| LIML size of nominal 5% Wald test | 16.38 | 8.96 | 6.66 | 5.53 |

Some checks done on SEX, STAT and ETHNICITY

```
. tab SEX
```

| Sex of respondent | Freq. | Percent | Cum. |
|---|---|---|---|
| Male | 42,691 | 48.76 | 48.76 |
| Female | 44,864 | 51.24 | 100.00 |
| Total | 87,555 | 100.00 | |

```
. tab ETHUKEUL
```

| Ethnicity (9 categories) UK level | Freq. | Percent | Cum. |
|---|---|---|---|
| White | 76,900 | 87.83 | 87.83 |
| Mixed/Multiple ethnic groups | 1,409 | 1.61 | 89.44 |
| Indian | 2,002 | 2.29 | 91.73 |
| Pakistani | 1,635 | 1.87 | 93.59 |
| Bangladeshi | 593 | 0.68 | 94.27 |
| Chinese | 439 | 0.50 | 94.77 |
| Any other Asian background | 863 | 0.99 | 95.76 |
| Black/African/Caribbean/Black British | 2,440 | 2.79 | 98.54 |
| Other ethnic group | 1,274 | 1.46 | 100.00 |
| Total | 87,555 | 100.00 | |

```
. tab STAT
```

| Employment status | Freq. | Percent | Cum. |
|---|---|---|---|
| Does not apply | 33,811 | 38.62 | 38.62 |
| Employee | 45,844 | 52.36 | 90.98 |
| Self-employed | 7,714 | 8.81 | 99.79 |
| Government scheme | 23 | 0.03 | 99.81 |
| Unpaid family worker | 163 | 0.19 | 100.00 |
| Total | 87,555 | 100.00 | |