UNIVERSITY OF BIRMINGHAM | DEPARTMENT OF ECONOMICS

Estimates of labour supply elasticities for Transport Analysis and other guidance

Report to the

Department for Transport

March 2025

By

Marco G. Ercolani[*†]

David Hearne[**]

Yi Liu[*]

With the research assistance of:

Nicolaus Teng[***]

[*] Department of Economics, University of Birmingham, England

[**] Department of Management, University of Birmingham, England

[***] University of Birmingham, University of Cambridge and University of East Anglia

† We extend our thanks to Joe Blackwell and Chris Wanzala-Ryan at the Department for Transport for their invaluable advice and guidance throughout this research project. Ethics approval for this research was granted by the University of Birmingham (ERN_2835-Jul2024).

Executive summary

- **Updated Labour Supply Elasticity Estimates**: The report provides new estimates of labour supply elasticities (LSEs) in the UK from 1997 to 2024, informing the Department for Transport's (DfT) employment impact assessments of transport investments.

- **Recommendation to Increase Extensive Margin LSE**: The analysis suggests that the current 0.10 extensive margin LSE used by DfT is an underestimate and should be updated to approximately 0.60, reflecting long-term employment responses to changes in transport infrastructure. The 0.60 estimate is based on individual-level data, while we infer the 0.10 estimate had been based on regionally aggregated data.

- **Declining Intensive Margin LSE**: The responsiveness of existing workers to wage changes (intensive margin LSE) has declined over time, from around 0.20 in 1997 to 0.12 in 2024.

- **Potential Impact of Declining Response Rates**: The response rates for the UK Labour Force Survey (LFS) have dropped sharply, particularly since 2020, raising concerns about data reliability. The report cross-validates findings using the UK Household Longitudinal Survey (HLS) data to mitigate potential LFS response rate biases.

- **Policy Implications and Future Recommendations**: The findings suggest that transport infrastructure investments have a greater impact on employment than previously estimated, supporting higher LSE values in policy assessments. Regular updates to LSE estimates using high-quality data are recommended for continued accuracy in economic modelling.

Table of Contents

Contribution:
All authors contributed to all sections.

- Marco G. Ercolani managed the project, was the main writer, provided the formulae, derived the proof for extensive margin labour supply elasticity, carried out estimates on the UK LFS data, provided the Stata code, and combined GB LFS response rate data.

- David Hearne carried out estimates on the UK HLS data and provided the description for this, provided the R code, reviewed survey weighting and population-weighted estimation procedures.

- Yi Liu cross-validated the LFS estimates, was in charge of staff hiring and management, and managed project finances.

- Nicolaus Teng wrote Python code to scrape websites for UK LFS response rates, and carried out the literature review for survey response rates.

1 Introduction

This report presents estimates of labour supply elasticities (LSEs) for the United Kingdom over the period 1997-2024 using individual-level data. These estimates were prepared for the Department for Transport (DfT) to aid in appraising the employment effects of transport infrastructure investments. These elasticities are important because changes in transport infrastructure can change effective wages if travel time is included in the working day, thereby influencing employment decisions. These LSE estimates may also be applied in other appraisal contexts.

The goal is to produce LSE estimates that are parsimonious, rigorous, and reproducible. To this end, we have used UK Labour Force Survey (LFS) and UK Household Longitudinal Survey (HLS) data as these offer individual-level population-representative data that include both the employed and the non-employed (unemployed or inactive). The appendix includes examples of software commands.

The DfT (2018) currently uses an extensive margin LSE of 0.10 in its Transport Analysis Guidance (TAG) A2.3 (equation 2) to evaluate employment effects of transport investments that alter travel times. Our analysis indicates 0.10 is an underestimate and suggests a more appropriate value is approximately 0.60 (see Figure 8). Although 0.60 may appear high, it should be viewed as a long-run elasticity, reflecting the period after infrastructure projects are completed and individuals have fully adjusted. Section 5 discusses how the 0.10 estimate may have arisen and why it is may be an underestimate.

Though not directly relevant, we also estimated ‘intensive margin’ LSEs, which measure how responsive existing workers are to wage changes in their supply of work hours. The DfT may choose to include these intensive margin LSE estimates in future reformulations of TAG to model the work hours of those who remain in employment. Our estimates indicate that the current intensive margin LSE should be around 0.12, which the DfT and other departments might find relevant for future policy analyses.

During the analysis, we found that LFS response rates have declined, while those in the HLS have remained stable. Accordingly, we collated this data and present it in Section 6. The recommended 0.60 extensive margin LSE reflects these response rates.

1.1 Core principles

A LSE is the proportionate change in labour supply relative to the proportionate change in wages that may have caused it to change. Ercolani et al. (2024) provide a comprehensive literature review of LSEs in the UK and other economies. In the literature on labour market analysis, two elasticities are typically considered:

- The *intensive margin* LSE is the proportionate change in work hours divided by the proportionate change in the wage(s).
- The *extensive margin* LSE is the proportionate change in the number of workers (or probability of employment) divided by the proportionate change in the mean wage.

Intensive margin LSEs are relatively easy to estimate, as one needs only to use a representative sample of workers. The functional form used to estimate it is also easy to specify, as it involves a simple regression of log-wages as a regressor on log-hours as the dependent variable. The resulting estimated coefficient on log-wages is the estimated intensive margin elasticity. Our view is that intensive margin LSEs are not directly relevant in TAG A2.3. This is because the former is an estimate of the responsiveness of those who remain in work while the latter models people moving into and out of employment. However, intensive margin LSEs could become relevant if the TAG A2.3 analysis of existing workers 'moving to more or less productive jobs' (M2MLPJ) were extended to include changes in work hours for those already in employment. Our estimates suggest that the UK intensive margin LSE was approximately 0.20 during the period 1997-2005 and has gradually declined to around 0.12 by 2024.

The more relevant measure is the extensive margin LSE, which is a measure of how employment changes as wages change. This seems to be the relevant LSE used in labour supply (employment) equation 2 of TAG A2.3, published by the DfT. Extensive margin LSEs are relatively complicated to estimate for several reasons. One is that it is also necessary to estimate the potential wage of the non-employed in order to model binary employment outcomes. Another complication is that employment is a binary dependent variable, and some sophisticated algebra is needed to accommodate its discrete nature. The analysis can also be extended in other ways, for instance, to account for possible 'sample selection' issues. Our basic estimates of the extensive margin LSE suggest that it is 0.40 (i.e. 40%) and that it peaked at 0.48 in 2010 and at 0.52 in 2024. Though these estimates seem high, we believe they are realistic in the context of long-run marginal responses.

1.2 Caveats

On balance, we feel that the use of a single extensive margin LSE by the DfT in TAG to model changes in employment is a reasonable approach, as it provides transparency in the assessment of transport proposals. Given this use of a single elasticity, we suggest that it be re-estimated on a regular, perhaps annual, basis. This should be done using the best available UK survey data and statistical techniques that are well understood and widely accepted. This is our aim in the present report.

Apart from the reduced-form regression equations, no attempt was made to impose theoretical restrictions, such as assuming full employment or perfectly elastic or inelastic labour demand functions. Anything that might increase or decrease labour productivity, and consequently wages, may alter the long-run level of employment and the number of work hours, as indicated by the empirical evidence. The only attempt at structural modelling was sample selection (Heckman) correction in wage regressions for the probability of being in employment. Imposing this selection correction did not noticeably alter the results.

Though we also present some disaggregated results, we caution against using a 'family' of elasticities to account for various socio-economic characteristics such as industrial sector, educational status, or geographical location. On balance, relying on several elasticities might obfuscate the TAG and may not be robust to structural changes brought about by the proposed transport infrastructure. In other words, the use of several elasticities may not be robust to what is sometimes referred to as the Lucas (1976) critique, which argues that policy evaluations based on past data are unreliable because agents' behaviour (and thus model parameters) change in response to policy shifts. One possible exception to this advice is the use of different elasticities for females and males, if this is deemed relevant, given the LSEs are substantially different between the sexes. However, the implications for TAG of using separate elasticities by sex warrant closer inspection, which is beyond the scope of this report. Presently, labour force participation seems relatively equal and stable between the sexes, and therefore a single LSE seems a reasonable simplification.

There might also be a temptation to use increasingly sophisticated statistical models to estimate the LSEs. Though it is true that omitted variable bias or simplified functional forms might bias estimates of the LSE, we cannot say ex ante the direction of this bias. Hence, we feel that simple statistical models based on population-representative data are just as likely to yield population-representative LSEs.

An important caveat is that, due to measurement error, our LSEs and those of others are likely to be at least slightly underestimated. It is well known that measurement error in explanatory variables, such as the wage rate, in linear models is likely to lead to underestimated parameters in a process known as attenuation. This is therefore likely to bias downward the estimates of the intensive margin LSEs, which are based on log-linear models. Though it is harder to show a systematic bias in the non-linear models used to estimate the extensive margin LSEs, this attenuation is also likely to be present if the estimated parameters fall within a reasonably narrow range. Measurement error in the hourly wage might arise from the way it is constructed in survey data. This is done by dividing pay, as reported on a payslip, by the number of self-reported usual work hours, which might be subject to recall error. Another source of measurement error might stem from the use of wages estimated from regression equations when modelling people's decision on whether to work. In this case, wages are based on regression estimates of latent wages.

A final caveat is that response rates in some national surveys have been declining. In many surveys, response rates have been gradually decreasing since the early 1990s but fell more sharply in 2020 with the onset of the Covid-19 pandemic. Response rates have since recovered, but only slightly, and in October 2023 variables derived from the LFS lost their 'official' status. We have adopted the standard approach of addressing this decline by using population-weighted regressions when estimating elasticities. However, this issue is sufficiently significant that we have dedicated Section 6 to a detailed discussion of the decline in response rates.

All these caveats are well known to researchers and policy advisers. Therefore, they should be seen as cautionary notes on our analysis rather than fundamental flaws. We have been transparent about these limitations, the steps we have taken to address them, and the measures we have implemented to assess their impact.

2 Technicalities

An elasticity is the proportionate change in one variable divided by the proportionate change in another variable that caused the change.[1] If no relationship exists, the elasticity will be reasonably close to zero allowing for statistical noise. Although each individual at any point in time could have their own elasticity, we seek the average population elasticity for each year in order to operationalise the analysis.

2.1 Extensive margin LSE in TAG A2.3

Here we replicate the DfT (2018) TAG A2.3 equations 1 and 2 as presented in Ercolani et al. (2024, pp. 7-8). We interpret the elasticity ε^{LS} in TAG equation 2 as an extensive margin LSE given it is used to generate forecasts of employment changes resulting from changes in generalised 'time and money' commuting costs.

TAG Equation 1 Average Round-trip Generalised Commuting Cost
$$G_{i,j}^{S,f} = \frac{\sum_m \left(g_{i,j}^{S,m,f} + g_{j,i}^{S,m,f}\right) T_{i,j}^{S,m,f}}{\sum_m T_{i,j}^{S,m,f}}, \forall i, f, S$$
$G_{i,j}^{S,f}$ average round-trip generalised commuting cost, including all expenses & time.
S is either the alternative transport scenario A or the baseline scenario B.
$g_{i,j}^{S,m,f}$ general commuting costs from i to j under scenario S in mode m for year f.
$g_{j,i}^{S,m,f}$ are the general commuting costs of returning from j to i under scenario S.
$T_{i,j}^{S,m,f}$ is the *annualised* number of commuting trips form i (home) to j (workplace).

TAG Equation 2 Labour Supply (Employment) Impact
$$\Delta E^f = \sum_i \left[\varepsilon^{LS} \left(\frac{\sum_j \left(G_{i,j}^{B,f} - G_{i,j}^{A,f}\right) W_{i,j}^f \Omega}{(1-\tau_1) \sum_j y_j^f W_{i,j}^f} \right) \sum_j W_{i,j}^f \right], \forall f$$
ΔE^f are the national changes in employment in each forecast year f.
ε^{LS} is the labour supply elasticity, currently assumed to be 0.10.
$G_{i,j}^{A,f}$ and $G_{i,j}^{B,f}$ are the average round-trip costs estimated in equation (1).
$W_{i,j}^f$ number of workers living in area i, employed in area j in the baseline scenario.
Ω average number of round-trip commuting journeys per worker.
τ_1 tax wedge required to convert gross earnings (y_j^f) into net earnings for new workers, based on average tax revenue from income tax, NI contributions, corporation tax and mixed income.
y_j^f average gross annual pay rates for workers employed in area j.

[1] Often, elasticities are described as the percentage change in one variable divided by the percentage change in another variable. However, since a percentage is merely a proportion multiplied by 100, including 100 in both the numerator and the denominator results in them cancelling each other out.

2.2 Intensive margin LSE

Although the intensive margin LSE is not directly relevant to the parameter in TAG A2.3, we present it first because it is simpler to estimate and therefore provides a good starting point for understanding the extensive margin LSE. The basic algebraic formula for the intensive margin LSE (iLSE) is the proportionate change in working time divided by the proportionate change in wage that caused it:

$$iLSE = \frac{\frac{\partial H}{H}}{\frac{\partial W}{W}} = \frac{\partial H}{\partial W}\frac{W}{H}$$

where H is work-hours, W is wages[2] and ∂ denotes partial derivatives.

Estimating intensive margin LSEs on data is relatively straightforward. Typically, the following log-log regression equation is estimated, possibly with other covariates:

$$lnH_i = \beta_0 + \beta_1 lnW_i + e_i \qquad (1)$$

where $\ln$ denotes natural logarithms of the variable, i indexes each individual, and the resulting β_1 estimate is the mean intensive margin LSE. The reason why β_1 is a direct estimate of the elasticity can be shown by using the chain rule and the rule for differentiating logarithms to partially differentiate equation (1):

$$\partial lnH_i = \partial(\beta_0 + \beta_1 lnW_i + e_i)$$

$$\frac{\partial H_i}{H_i} = \beta_1 \frac{\partial W_i}{W_i}$$

$$\frac{\frac{\partial H_i}{H_i}}{\frac{\partial W_i}{W_i}} = \beta_1$$

Note that the single value of β_1 implies a single elasticity is estimated across the whole sample given the log-log specification.

Note also that in this and other regressions the natural logarithm (ln) of wages is typically used also because it transforms the left-skewed wage distribution to a symmetric ln(wage) distribution which more closely approximates a normal distribution. Regressions with normally distributed variables are more likely to generate normally distributed residuals which is desirable for inference statistics. Though using ln(wages) simplifies the computation of the extensive margin LSE, it actually complicates the calculation of the extensive margin LSE, as shown below and in the Appendix.

---

[2] This analysis uses gross (pre-tax) wages, as net (post-tax) wages are unavailable. Elasticity, which measures proportionate change, is unaffected by whether wages are pre- or post-tax, provided that marginal tax rates remain constant. While most countries have progressive tax systems, tax rates adjust at discrete thresholds rather than continuously. If a wage change does not move the average worker into a different tax bracket, elasticity estimates based on gross or net earnings should yield the same results. Here is a simple algebraic proof. Let τ be a constant tax rate on earnings and Δ denote the change in any variable. It is evident that the two elasticities below are the same whether calculated on the gross wage W or the net wage $(1\text{-}\tau)W$, given the $(1\text{-}\tau)$ cancel out:

$$\frac{\Delta H/H}{\Delta W/W} = \frac{\Delta H/H}{\Delta(1-\tau)W/(1-\tau)W}$$

2.3 Extensive margin LSE

The *extensive margin* LSE is the proportionate change in the probability of employment divided by the proportionate change in wage that caused it, where E_i is the probability of being employed:

$$eLSE = \frac{\frac{\partial E}{E}}{\frac{\partial W}{W}} = \frac{\partial E}{\partial W}\frac{W}{E}$$

Estimating extensive margin LSEs is more challenging than estimating the intensive one. One difficulty is that wages are not observed for non-workers. Hence wage regressions are estimated first and used to compute fitted wages for non-workers (and workers too). Usually these are specified as log-wage regressions:

$$lnW_i = \beta_0 + \beta_1 X_{1j} + \cdots + \beta_n X_{ni} + e_i \qquad (2)$$

In equation (2), the dependent variable $\ln W_i$ is the natural logarithm of the 'derived' LFS variable HOURPAY, described in subsection 3.1. For the HLS data regressions, we constructed HOURPAY using the LFS definition, as the ratio of usual gross weekly pay to usual work hours and paid overtime. The regressors in equation (2) are described in subsections 3.1 and 3.2, and include binary variables for qualifications, age groups, and region.

A greater difficulty is that computing the elasticity is complicated by the fact that the dependent variable is discrete while the explanatory wage variable is in logarithms. Fitted values $\widehat{\ln W_i}$ from regression (2) are used as estimated wages for non-workers and workers in Probit[3] employment regression (3) given the binary nature of the dependent variable:

$$\Pr(E_i = 1) = \Phi\big(\beta_0 + \beta_1 \widehat{lnW_i}\big) \quad (3)$$

where $\Phi(.)$ is the cumulative density function (CDF) of the normal distribution. Equation (4) is the resulting extensive margin LSE for each individual where $\phi(z_i)$ is the probability density function (PDF) of the normal distribution and $z_i = \beta_0 + \beta_1 \widehat{\ln W_i}$ is the linear predictor for each individual:

$$eLSE_i = \beta_1 \frac{\emptyset(z_i)}{\Phi(z_i)} \qquad (4)$$

Equation (4) corresponds to equation 5 in Detilleux and Deschacht (2021); however, given that we could not find a derivation for it anywhere in the literature, we provide one in Appendix A1.

Alternative interpretations are available to operationalise equation (4). The more common interpretation is that the Inverse Mills Ratio (IMR) for each individual is

$$IMR_i = \frac{\emptyset(z_i)}{\Phi(z_i)}$$

hence computing the mean for the IMR and multiplying it by β_1 produces the mean extended margin LSE (eLSE). Alternatively, one can use the fact that

[3] Logit regression cannot be used in this instance because the normal distribution would not emerge as it does from the estimation of the Probit regression.

$$\frac{\partial E_i}{\partial lnW_i} = \beta_i \phi(z_i)$$

is the estimated marginal effect (slope of the probability curve) of $\ln(W_i)$ on E_i, hence use the econometric software to compute the mean marginal effect for each respondent. Then for each respondent scale (divide) the marginal effect by their $\Phi(z_i)$ to obtain each individual's extensive margin LSE.

2.4 Sample selection in the extensive margin LSE

Various refinements to the analysis are possible, here we outline how to implement sample selection correction in what it commonly called the Heckman correction.

The Heckman correction (Heckit) seeks to account for selection bias in econometric models, particularly in cases where the dependent variable is observed only for a non-randomly selected subsample. In this analysis, the wage equation (2) is only estimated for those who work because, by definition, only observed wages are available. An issue arises when sample selection is correlated with the outcome of interest. In this case, employment is correlated with the wage because only those who are offered higher wages are likely to gain employment. This correlation may lead to biased regression estimates. For example, a positively correlated employment probability and wage will lead to upwardly biased parameter estimates in a wage equation.

When estimating the extensive margin LSE, the process can be started by estimating an initial employment equation that should include at least one variable that effects the probability of employment but does not affect the wage:

$$\Pr(E_i = 1) = \Phi(\beta_0 + \mathbf{\beta_1 Z_i}) \qquad (5)$$

where $\mathbf{Z}_i$ is a matrix of variables that excludes the wage and $\boldsymbol{\beta_1}$ is a vector of slope parameters that explain the probability of employment. In our LFS and HLS estimates the same explanatory variables as in equation (2) are used, with the addition of separate binary variables for each sex for the presence of a child younger than 16. These two binary variables are included as they are believed to affect the probability of employment but not the wage rate once in employment.

The equation (5) estimates are used to compute the IMR for each individual:

$$IMR_i = \frac{\phi(\zeta_i)}{\Phi(\zeta_i)} \qquad (6)$$

where $\zeta_i = \beta_0 + \mathbf{\beta_1 Z_i}$ is the linear prediction for each individual. Thereafter, the individual IMR values are included in the log-wage regression with the addition of the regressors already described in regression equation (2):

$$lnW_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_n X_{ni} + \gamma IMR_i + e_i \qquad (7)$$

The results of regression equation (7) are used to compute sample-selection adjusted fitted log-wages $\widehat{\ln W_i}$ for the employed and the non-employed. This fitted log-wage is then used to estimate employment regression (3) and the extensive margin elasticity as the mean value of equation (4).

3 The Data

The estimated elasticities are derived from regression models using individual-level LFS and HLS data for respondents aged 18 to 64. Annual estimates span the period 1997-2024 and may use population weights. No adjustments are made for price inflation, as each estimate corresponds to a separate year, and inflation is unlikely to have a significant impact within any one year. We are not permitted to distribute the data and are required to destroy it once this project is complete. Anyone attempting to reproduce our results is likely to obtain very similar, but not identical, findings. This is because LFS and HLS surveys are frequently revised, particularly in early releases, with older waves sometimes in their seventh or eighth revision.[4]

3.1 UK Labour Force Survey (LFS) data

The main dataset used in the analysis is the quarterly UK Labour Force Survey (LFS), which is administered by the Office for National Statistics (ONS, 1997-2024). Quarters 1997q1 to 2024q4 were combined into a single file, and the elasticities are estimated by calendar year. The year is determined based on the Wednesday in the Reference Week during which respondents were surveyed.

As mentioned in the final part of the Introduction, we are also conscious that, in October 2023, statistics derived from the LFS lost their 'official' status and reverted to 'experimental', but issues with the data began in 2020 or earlier. We are therefore confident in the LFS results only up to 2019. This is discussed in detail in Section 6.

The overall LFS sample for 1997q1 to 2024q4 comprises just over 2.5 million respondents of working age (18 to 64). The analysis is conducted on a year-by-year basis, with the annual sample size averaging around 130,000 in earlier years and approximately 45,000 in later years. Exact sample sizes are reported in the Appendix regression tables. Variables in uppercase are directly available in the LFS and, in almost all cases, are either based on self-reported responses or derived from them.

- HOURPAY is an LFS 'derived variable' representing hourly wages measured in the first and last (fifth) interview (wave).[5] It is based on the ratio of gross pre-tax weekly pay (GRSSWK)[6] divided by the number of self-reported usual weekly paid-for work hours including paid overtime.
- TTACHR are actual weekly work hours including (paid or unpaid) overtime.

---

[4] Surveys that include only workers, such as the Annual Survey of Hours and Earnings (ASHE), were not used because estimates of the extensive margin LSE require both workers and non-workers to be present. Another potential issue with ASHE is that it records contracted work hours from employers. In contrast, the LFS and HLS record workers' recollections of usual work hours, which may be more useful in capturing work patterns, even allowing for recall error. The Annual Population Survey (APS) could have been another data source, but the publicly available APS files do not typically include earnings data, and hours data, if present, are often less detailed. Thus, APS data are designed more for population estimates and regional analysis, not for modelling wages or work hours.

[5] From 1992q1 to 1996q4 LFS pay data was collected only in Wave 5 (i.e. the last interview). Starting in 1997q1 pay data started being collected in both Wave 1 and Wave 5.

[6] GRSSWK is also a derived variable based on evidence from a payslip converted into weekly pay.

- TTUSHR are usual weekly work hours including (paid or unpaid) overtime.
- BACTHR are actual weekly work hours excluding overtime.
- BUSHR are usual weekly work hours excluding overtime.

The following variables are also used:

- AGE is used to select the working age group 18-64.
- SEX is used when estimating separate regressions for Males and Females.
- PWT is a person weight variable where each value represents the number of people that respondent represents in the UK. PWT is generated by combining PWT variables with various suffixes meant to indicate the base year being used. Most regressions use these population weights.

Regression equation (2), which is used to generate fitted hourly wages, includes binary explanatory variables based on the following LFS variables:

- HIQUALD[7] measures highest education qualification based on these categories and their equivalent: (University) Degree, (College) Higher Education, A-levels, GCSE grades C and above and Other qualification. Instead of deleting Don't know and No Answer responses a separate category is created for them.
- URESMC records in which of twenty UK regions the respondent is resident. Inner and Outer London are recorded as separate locations.
- AYFL19 records the "Age youngest child in family under 19". This variable is used to generate binary variables, for youngest child aged less than 16, for each sex among the respondents.
- AGE is used to generate five-year age-band binary variables for the respondents. The final category spans ages 63 and 64. Other age categories are used in some graphs.

3.2 UK Household Longitudinal Survey (HLS) data

Given concerns regarding the falling LFS response rates, it is important to determine whether results remain comparable when using alternative data. To this end, the results are cross-validated using the UK HLS data, which, in combination with the earlier British Household Panel Survey (BHPS), constitutes the Understanding Society dataset. The HLS is independent of the LFS, being administered by the University of Essex Institute for Social and Economic Research (UoE ISER, 2024) rather than the ONS. Hence, it should provide an entirely independent view of relative elasticities.

The HLS has strengths that make it a good complement to our core results from the LFS. In particular, it contains information on a considerably greater breadth of variables than the LFS included, crucially for our purposes, hours worked and profits for the self-employed. Moreover, HLS has not experienced the precipitous fall in response rates that the LFS has (although like any panel it is subject to ongoing attrition). Whilst we do not make use of its longitudinal nature, this is an important facet of these data that future work should seek to exploit.

[7] This LFS variable changes suffix on occasion, for example when GCSEs replaced O-levels.

Notwithstanding this, the HLS remains unsuitable as a primary data source for our purposes for two major reasons. Firstly, the achieved sample size (particularly of adults of working age) remains much smaller than that of the LFS, despite the fall in response rates for the latter. Additionally, because of its nature as a longitudinal panel, the annual samples are not independent. This may complicate efforts to estimate changes over time and should be considered, especially when comparing intertemporal patterns with those of the LFS.

Like the LFS, HLS pay variables for employees are checked against payslips "if possible". If the most recent payslip differs from "usual" (for example due to absence or overtime) then the respondent is pressed as to "usual" pay. This is subject to recall bias and is potentially unknown with any degree of accuracy. Self-employment incomes is similarly subject to a degree of uncertainty, although this again checked where possible. Pay is a derived variable, being calculated from responses to several variables. We utilise the following variables from the HLS:

- JBHRS – the (estimated) number of hours per week usually worked in the respondent's main job.
- JSHRS – the (estimated) number of hours per week usually worked for those self-employed in their main job.
- PAYGU_DV – usual gross monthly pay (employees)
- SEEARN_DV – self-employment earnings per month (self-employed)
- SEX – as stated (this is coded as binary, although in more recent surveys a small number of "don't know" responses have been included, which we discard due to the extremely small number of responses).
- AGEGR13_DV – age category (13 categories, although we include only individuals between 18 and 64).
- GOR_DV – Government Office Region
- MARSTAT – marital status
- HIQUAL_DV – highest qualification
- RACH16_DV – responsibility for a child under 16.

Hourly wages are calculated from earnings and reported hours, as outlined in section 4.2. There are some subtle differences in certain variables compared to the LFS. In particular, GOR_DV (3 nations plus 9 English regions) is not equal to URESMC (20 regions), although the impact of using one relative to the other should be relatively modest. There are some differences in terms of the coding of HIQUAL_DV relative to HIQUALD in the LFS, although again the impact should be quite modest. Pay data was not collected for individuals who were not present (proxy respondents) and a small number of individuals either refused or reported positive hours worked but no income. All were excluded from the analysis (since employed individuals with no pay reported have the potential to bias the IMR). Self-employed individuals reporting a loss or breakeven (zero income after expenses) were likewise excluded.

4 Results: Intensive margin Labour Supply Elasticities

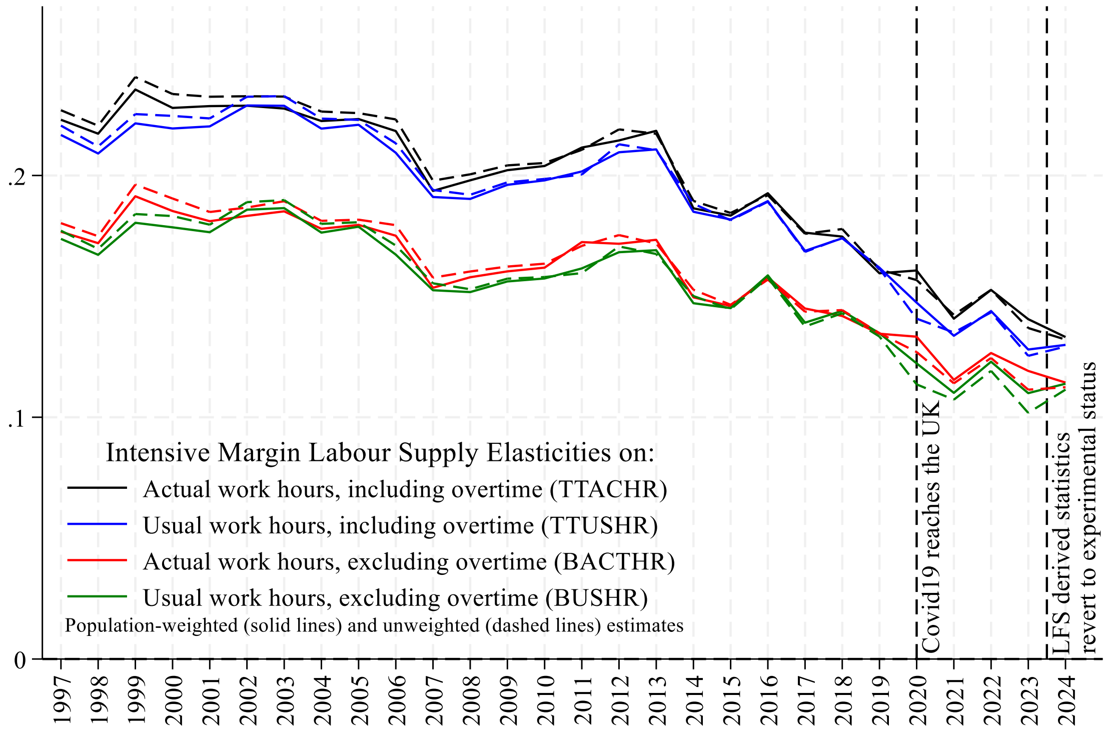
Intensive margin LSEs for workers who may change their work hours but remain employed are presented first, as they are simpler to estimate and provide initial benchmarks. Although they are not directly relevant to the DfT TAG A2.3, they could be included in other analyses and future TAGs.

4.1 Overall intensive margin LSEs using LFS data

Figure 1 illustrates intensive margin elasticities for each year and the four main measures of labour work hours in the UK LFS. The elasticities in Figure 1 are based on yearly regression equation (1) estimates, with $\ln(\text{HOURPAY}_i)$ as the explanatory variable $\ln W_i$. The dependent variable, $\ln H_i$, is the natural logarithm of any one of the four LFS hours variables: TTACHR, TTUSHR, BACTHR, and BUSH. As described in the data section, these four hours variables are based on either usual or actual work hours and may include or exclude paid and unpaid overtime hours. Individual regressions are reported in Appendix A2 tables.

Solid lines represent estimates that use the person weight (PWT), while dashed-line estimates are based on unweighted regressions. We observe that person weights do not substantially alter the results; hence, all subsequent estimates include person weights for completeness.

Figure 1: Intensive margin Labour Supply Elasticities (LSE)



One can see in Figure 1 that, early in the sample period (1997-2005), the elasticities are higher, at around twenty per cent (0.22 when including overtime and 0.18 when excluding it). Twenty per cent implies that a 10 per cent increase in wages would lead to a 2 per cent increase in work hours. Elasticities based on work hours that include overtime (paid or unpaid) are higher than those that exclude overtime, possibly due to workers having greater flexibility over their overtime hours. Apart for a sharp fall in 2007, these elasticities show a general gentle fall in the period 1997 to 2013. After 2013, they enter a prolonged period of steeper decline and by 2024 the elasticities are around twelve per cent (0.13 including overtime and 0.11 excluding it). Twelve per cent implies that a 10 per cent increase in wages would lead to a 1.2 per cent increase in work hours. The LSE estimates in Figure 1 are at the upper end of past estimates for the UK, as presented in the literature survey Ercolani et al. (2024, Table 3.3.1), and are similar to those in Blundell et al. (2011, 2016). These results suggest that existing workers have become progressively less sensitive to wage changes.

Figure 2 illustrates the same person-weighted intensive margin LSEs but with the addition of 95% confidence bands. These confidence bands are calculated by multiplying the standard error of the estimated elasticity by ±1.96. We see that the 95% confidence bands are narrow across the entire sample period. These bands widen slightly towards the end due to the smaller sample size in later years.

Figure 2: Intensive margin LSEs with 95% confidence bands

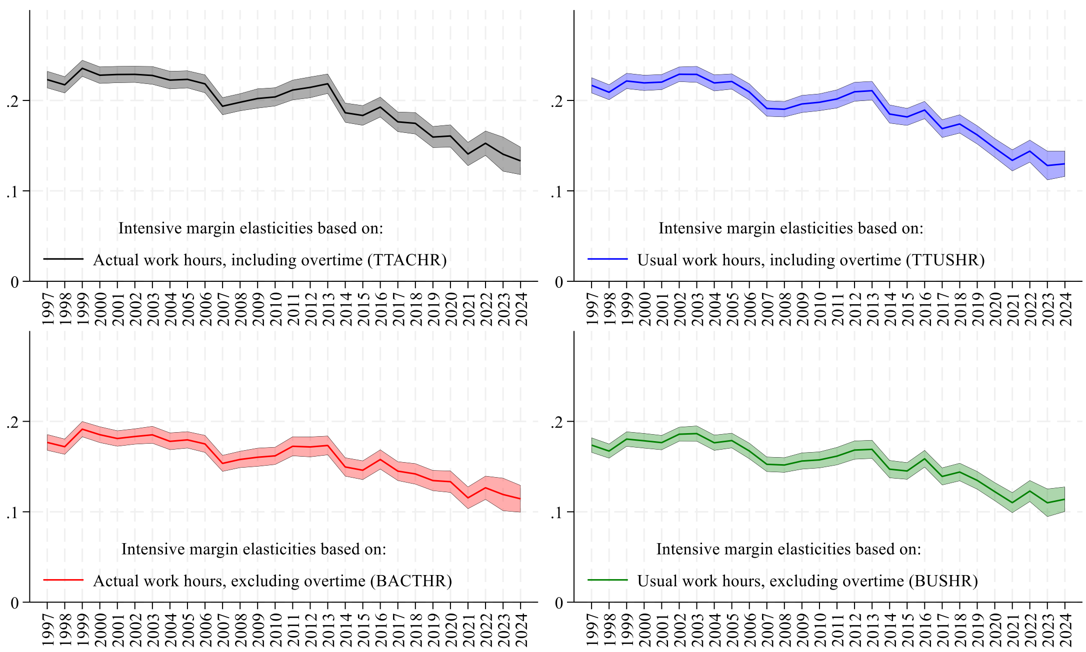
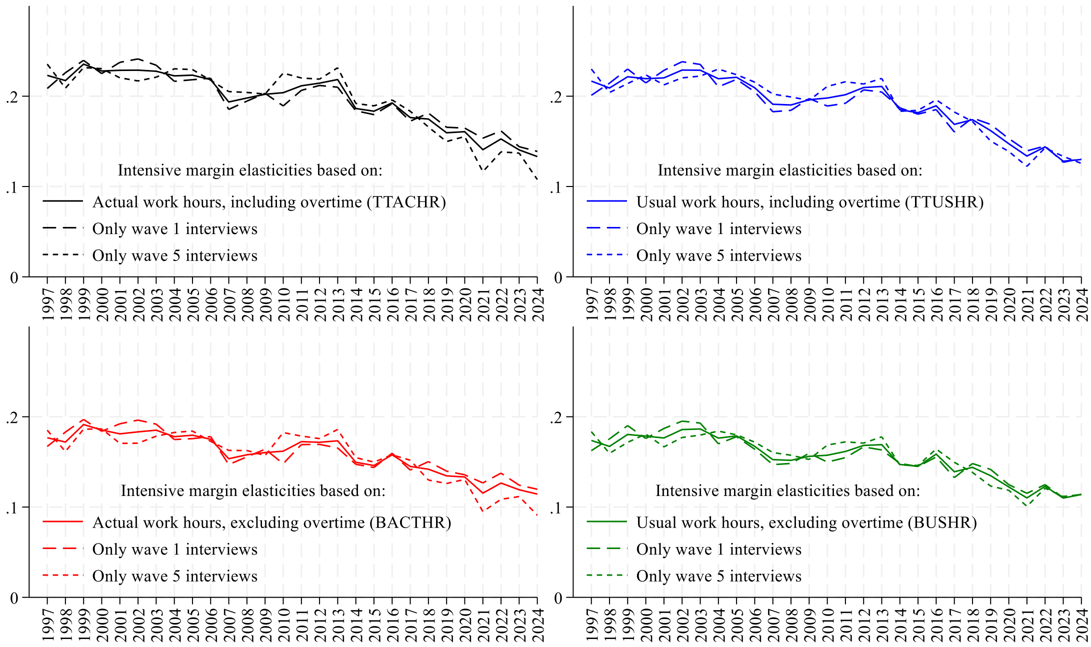


Figure 3 illustrates the same person-weighted intensive margin LSEs that include both Wave 1 and 5 interviews (solid lines), and estimates based on either only Wave 1 interviews (long-dash lines) or only Wave 5 interviews (short-dash lines). We observe that the choice of interview wave does not substantially affect the estimated intensive margin LSE. However, for the two dependent variables that measure actual work hours (TTACHR and BACTHR) based on Wave 5 interviews, there seems to be slight downward deviation in recent years.

Figure 3: Intensive margin LSEs by interview wave 1 (first) or 5 (last)



4.2 Intensive margin LSEs by sex, age, region and controlling for covariates using LFS data

Figure 4 illustrates the intensive margin LSEs by sex. As is well known in the research literature, LSEs tend to be lower for males than for females, indicating that males are less responsive to wage changes. In Figure 4, the intensive margin LSEs for males are much lower, with a clear separation between hours measures that include or exclude overtime. The intensive margin LSEs for females tend to be higher, and the choice of hours measure appears to have a greater impact. In the UK, the labour force participation rate for both females and males is slightly over 70 per cent. The overall joint values presented in the previous graphs therefore provide viable single values for the workforce, unless TAG A2.3 is expanded to model females and males separately.

Figure 5 illustrates intensive margin LSEs based on the four work-hours measures across four age groups. The youngest group (aged 18-28) has a comparatively stable elasticity, whereas the other three groups exhibit a declining elasticity throughout the period. The oldest age group (aged 53-64) has a slightly lower elasticity than the others. Discrepancies in past research might, therefore, be due to how the working-age range has been previously defined.

Figure 4: Intensive margin LSEs for Females and Males

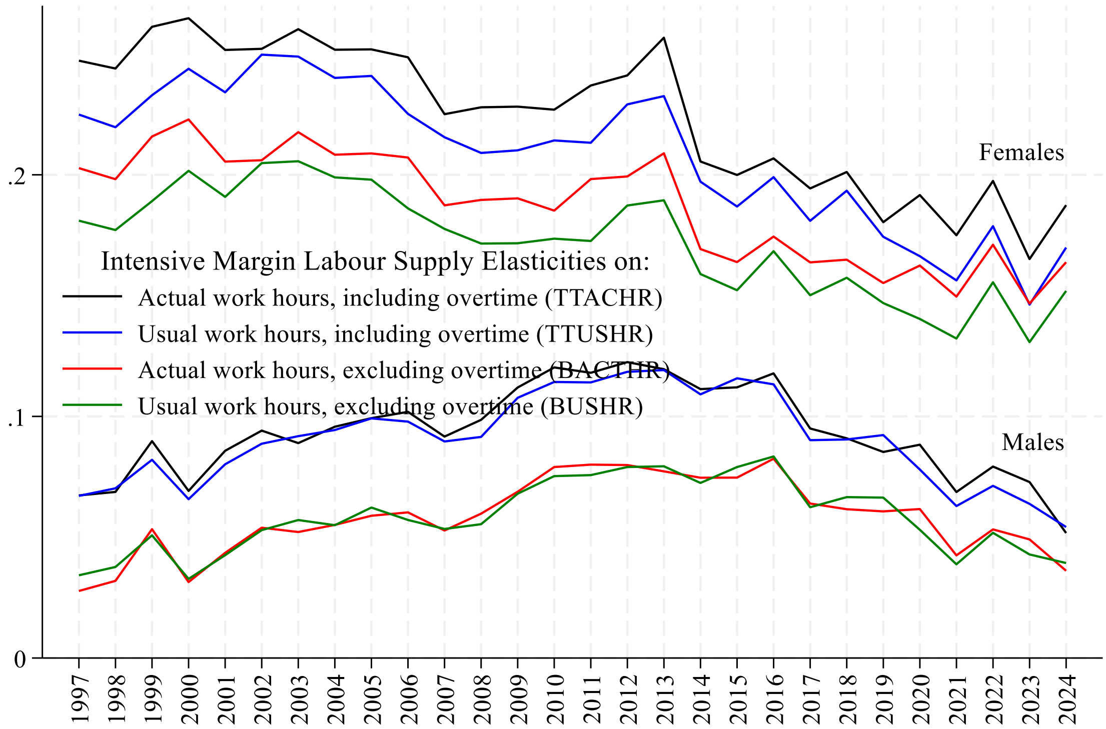


Figure 5: Intensive margin LSEs by age group

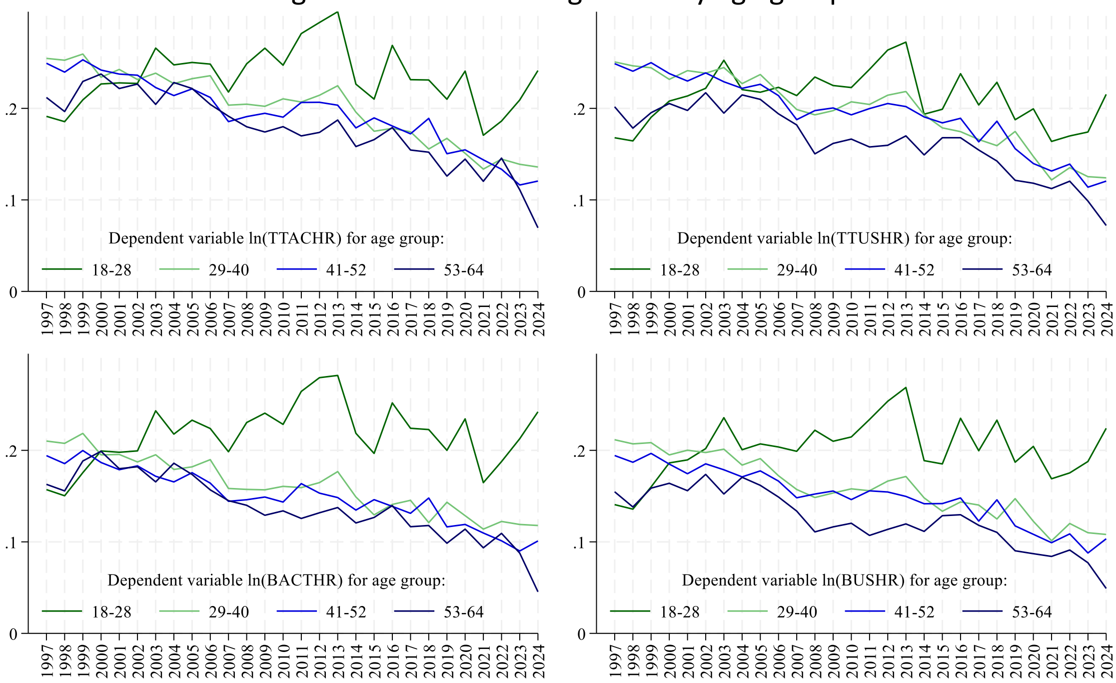
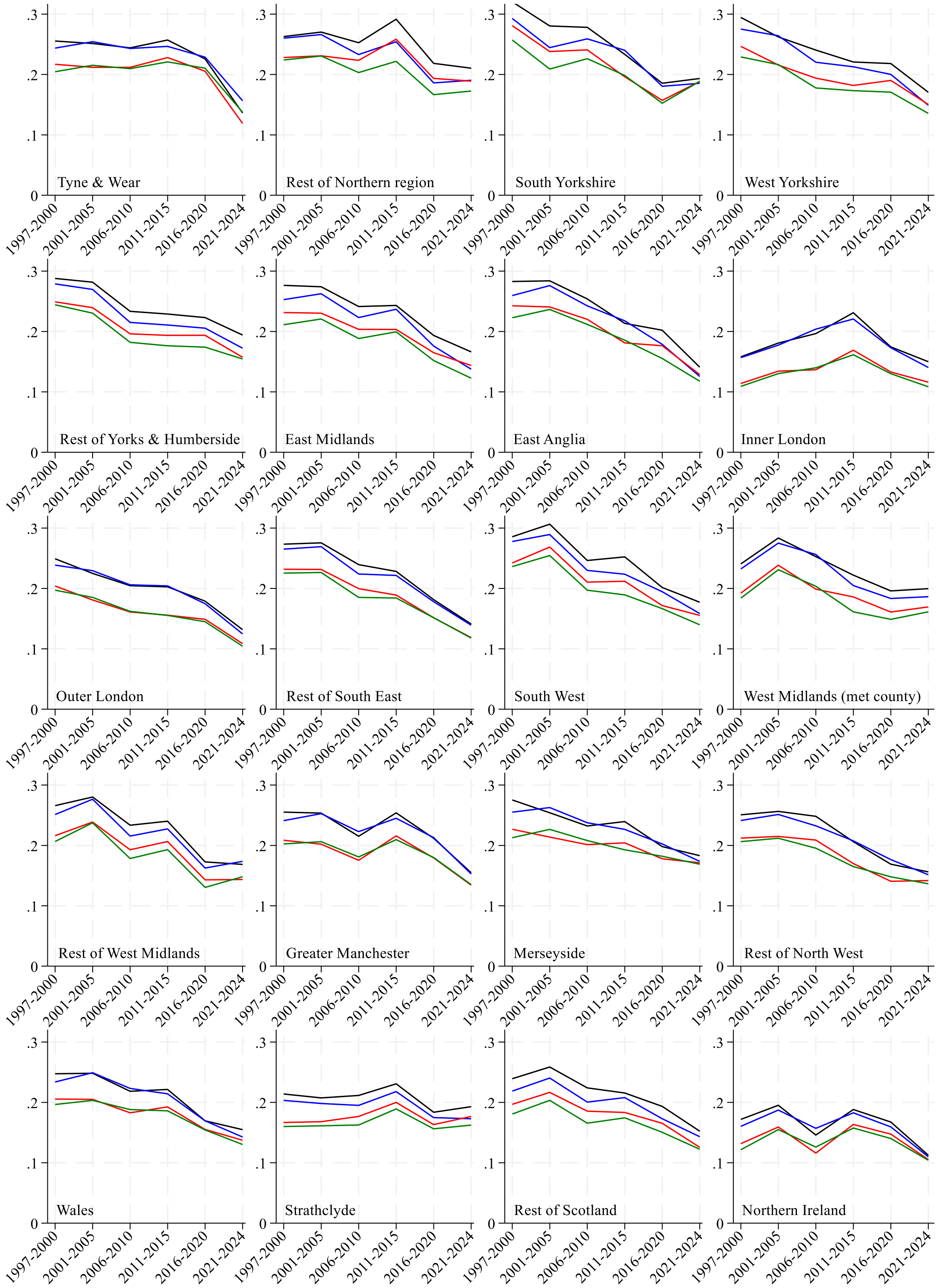


Figure 6 illustrates the intensive margin LSEs estimated separately for the twenty UK regions of residence recorded in the LFS variable URESMC. To maintain large sample sizes, equation (1) is estimated across five-year intervals, except for 1997-2000 and 2021-2024, which span four years. The regions are presented in the same order as in the LFS documentation. Although there appears to be some geographic variation, the intensive margin LSE has been declining in most regions. One exception is Inner London, where this elasticity increased until 2011-2015 before declining thereafter. Another exception is Northern Ireland, where the elasticity remained relatively stable and only declined noticeably in 2021-2024. Inner London and Northern Ireland also share the characteristic of having lower intensive margin LSEs compared to other regions. Conversely, regions that end the period with high elasticities, at around 0.30, include South Yorkshire, West Yorkshire, the South West, and the Rest of the Northern region.

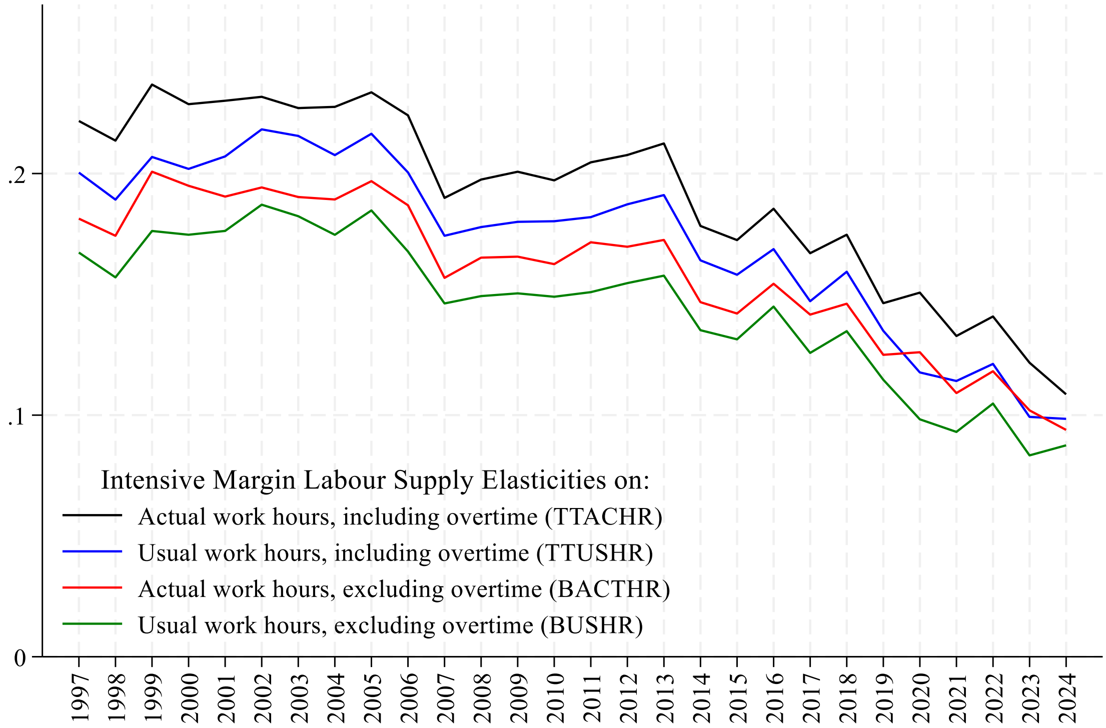
Figure 6: Intensive margin LSEs across twenty UK regions



Note: See other figures for legends describing four dependent variables used in the log-log regressions.

Figure 7 illustrates intensive margin LSEs when regression equation (1) also includes regressors in the form of binary variables for highest educational qualification, age of the youngest child in the household, region of usual residence, and age group by five-year age band. These are described in the notes to Figure 6 and in the Data section. In principle, including covariates might reduce omitted variable bias, but we cannot determine with certainty in advance the direction of this bias. From Figure 6, it appears that adding this rich set of covariates changes the estimated elasticities only slightly. At the start of the sample period, these elasticities seem slightly higher, and at the end of the sample period, they appear slightly lower than the elasticities illustrated in Figure 1.

Figure 7: Intensive margin LSEs for multivariate regressions



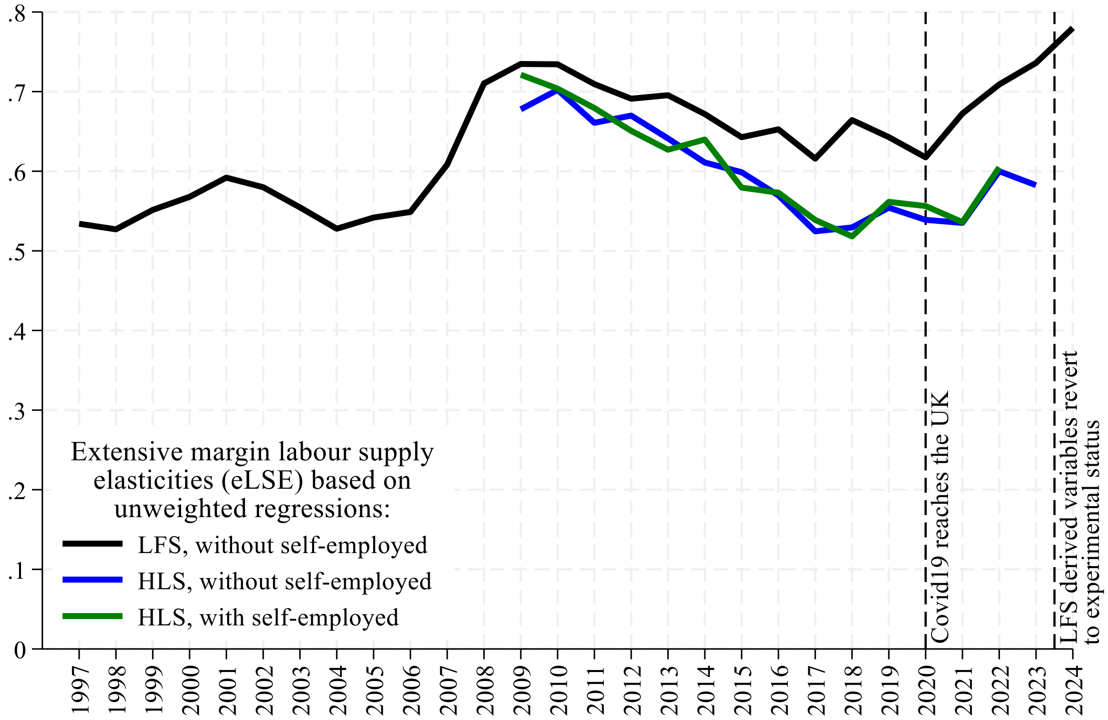
Notes: In addition to ln(HOURPAY), binary regressors based in these variables are included: HIQUALD (highest qualification), URESMC (region of usual residence), AYFL19 (used to indicate age of youngest child in household aged 0 to 15), AGE is used to generate binary regressors at five year intervals.

5 Results: Extensive margin Labour Supply Elasticities

Extensive margin LSE estimates for changes into and out of employment are presented in this section. These estimates are less straightforward than the intensive margin LSEs but are directly relevant to equation 2 in the DfT TAG A2.3 modelling. Before presenting all the permutations for these estimates, Figure 8 shows a simple comparison supporting our recommended extensive margin LSE of 0.60. The following subsections present estimates for various socio-economic sub-groups.

In Figure 8, estimates based on LFS data are around 0.55 for 1997-2006, rising to over 0.70 by 2008 before gradually falling to 0.62 in 2020. Although LFS estimates increase after the onset of the Covid19 pandemic, it is uncertain whether this is due to real labour market changes or the substantial drop in LFS response rates described in Section 6. Estimates using HLS data provide additional evidence, utilising smaller but potentially more reliable sample response rates. The fall in HLS extensive margin LSE since during 2009-2020 mirrors that in the LFS but is always lower. We find the difference in LFS and HLS estimates is small but systematic, and we cannot find specific reasons for this in the data or the literature. The HLS data estimate of 0.58 in 2023 supports our recommended extensive margin LSE of 0.60.

Figure 8: Extensive margin Labour Supply Elasticities, LFS UKLHS comparison



Although we favour the estimates based on LFS data due to their much larger sample size, these omit the self-employed, as their earnings are not recorded in the LFS. We therefore carry out estimates using HLS data, both with and without the self-employed, for cross-validation. Though we hypothesised that their elasticities might differ, the HLS results indicate that the extensive margin LSEs are very similar, whether or not the self-employed are included.

The extensive margin LSE estimates in Figure 8 are at the upper end of past estimates for the UK, as summarised in the literature survey by Ercolani et al. (2024, Table 3.3.1) and somewhat higher than those in Blundell et al. (2011, 2016). Nonetheless, they are well within some of the higher elasticities listed in these surveys.

5.1 Overall extensive margin LSEs using LFS data

Extensive margin labour LSEs presented in Figure 9 for the period 1997-2020 show it fluctuates around 0.60. This implies a 10 per cent increase in hourly wages leads to a 6 per cent increase in the number of people employed. In 2021, this elasticity began to increase, reaching 0.84 by 2024. Recommending a long-run extensive margin LSE is problematic given the low recent LFS response rates, but we are inclined to suggest a steady-state value of 0.60 despite the 0.84 estimate in 2024. Using a steady-state value matters because infrastructure projects can take many years to come to completion. The high recent values may be cyclical or a by-product of low LFS response rates. We also recommend estimating the 2025 elasticity as soon as the LFS data are available.

The extensive margin elasticities in Figure 9 are based on mean values from equation (4), which results from regression equations (2) and (3). Estimates based on unweighted regressions (dashed lines) are similar to those based on population-weighted regressions (solid lines), though they begin to deviate in 2020. We could not identify the reasons for this but speculate that it is related to falling LFS response rates. Estimates (in blue) that include Heckit selection correction based on equations (5) to (7) are not noticeably different from those that do not (in black).

Figure 9: Extensive margin Labour Supply Elasticities

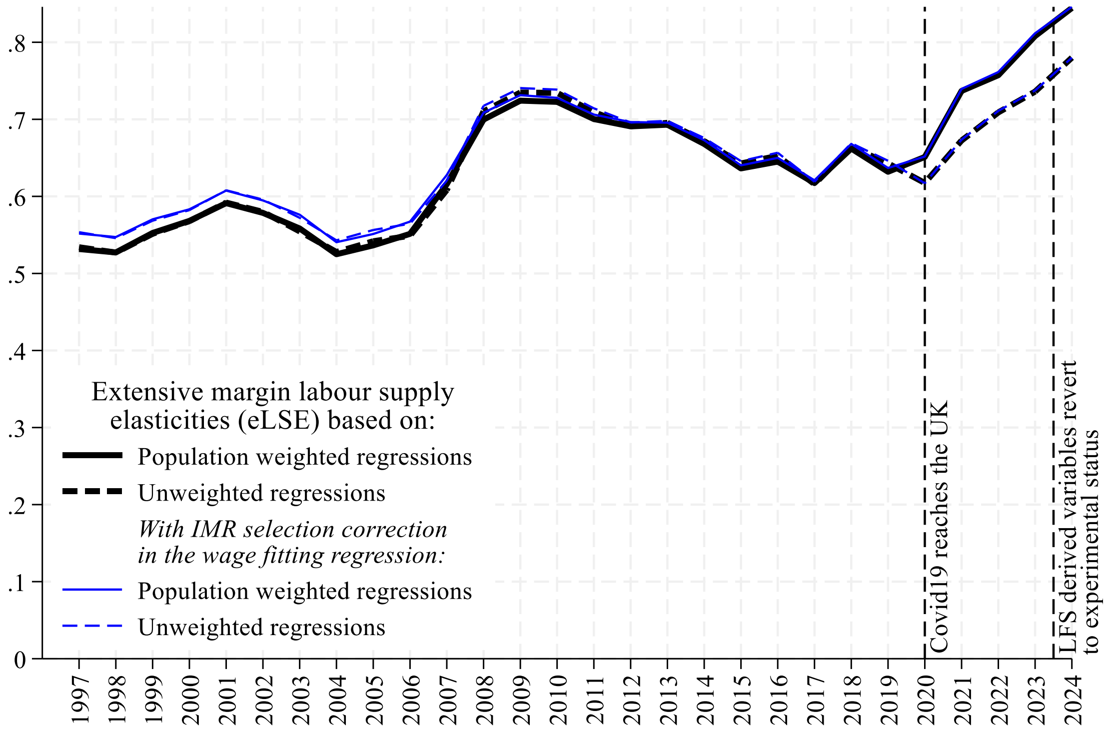
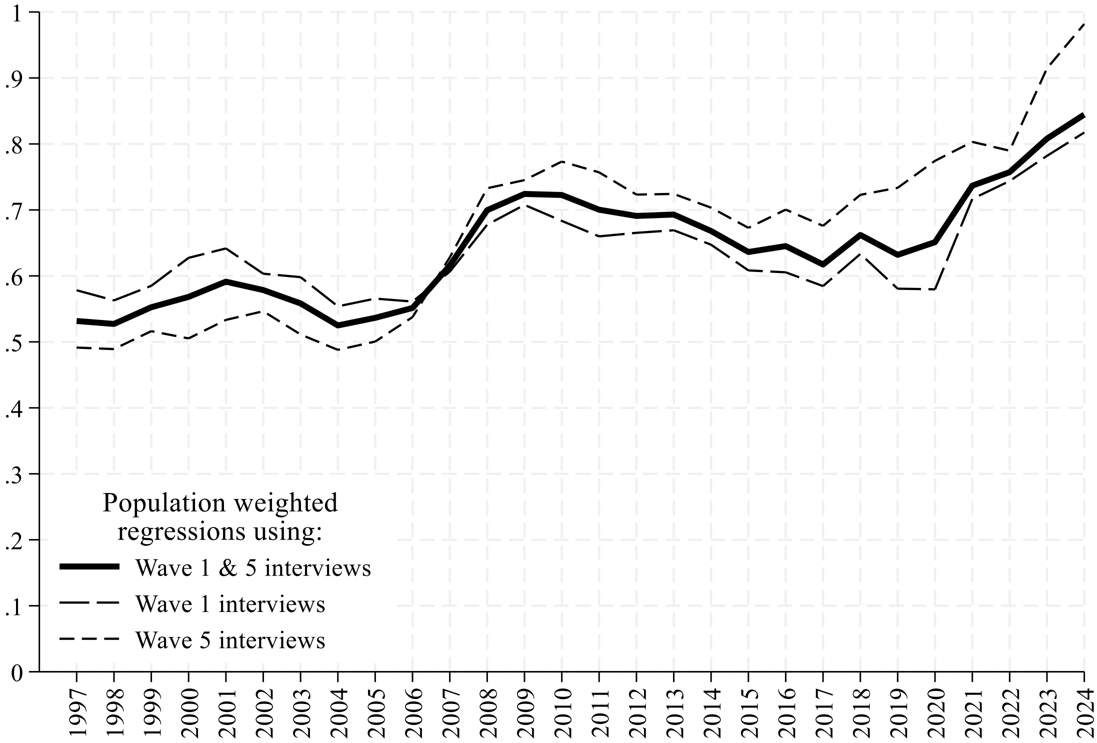


Figure 10 compares the person-weighted extensive margin LSEs that include both Wave 1 and 5 interviews (solid lines), to those based on either only Wave 1 interviews (long-dash lines) or only Wave 5 interviews (short-dash lines). As it did for the intensive margin estimates in Figure 3, the choice of interview wave does not substantially affect the estimated extensive margin LSE. The largest deviation between Wave 1 and 5 values occurs in 2020.

Figure 10: Extensive margin Labour Supply Elasticities, by interview wave 1 or 5



5.2 Extensive margin LSEs by sex, age, region and controlling for covariates using LFS data

Figure 11 illustrates the extensive margin LSEs by sex. As is standard in the research literature, the LSEs are found to be higher for females than for males. In Figure 10 this difference is about 0.10 but the gap narrows towards the end of the sample period when both series exceed 0.80. Given the work participation rate for both sexes is slightly over 70 per cent, the joint mean values in Figure 8 are good approximation for the working age population as a whole. The differences in Figure 11 suggest that employment-augmenting transport investment might affect females more than males. Conversely, changes that reduce employment might also have a greater effect on females.

Figure 11: Extensive margin Labour Supply Elasticities, for Females and Males

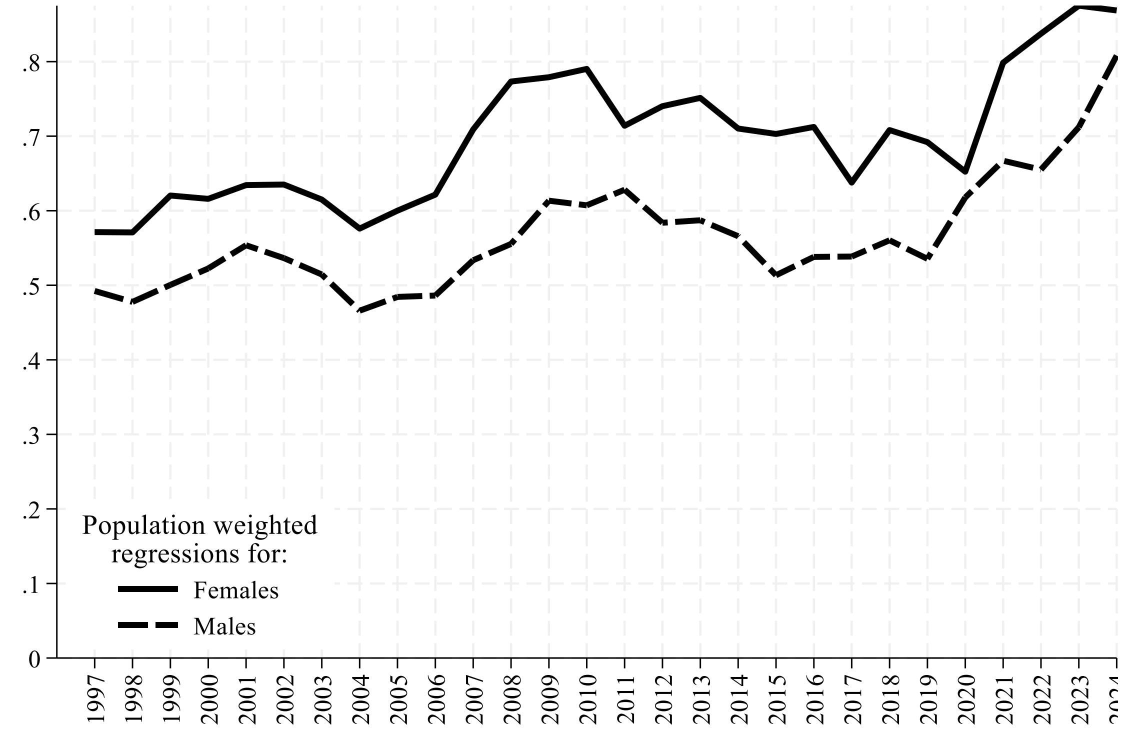


Figure 12 illustrates the extensive margin LSEs across four age groups. The youngest group (aged 18-28) has a noticeably higher elasticity, which has risen substantially, peaking at over 1.7 in 2024 and drives the overall rise in elasticity shown in Figure 8. The oldest group (aged 53-64) has a slightly higher elasticity than the rest, peaking in 2008-2010. The other two groups (aged 29-40 and 41-52) have relatively stable elasticities, ranging from around 0.40 to 0.65, and tending to rise slightly toward the end of the period.

Figure 12: Extensive margin LSEs by age group

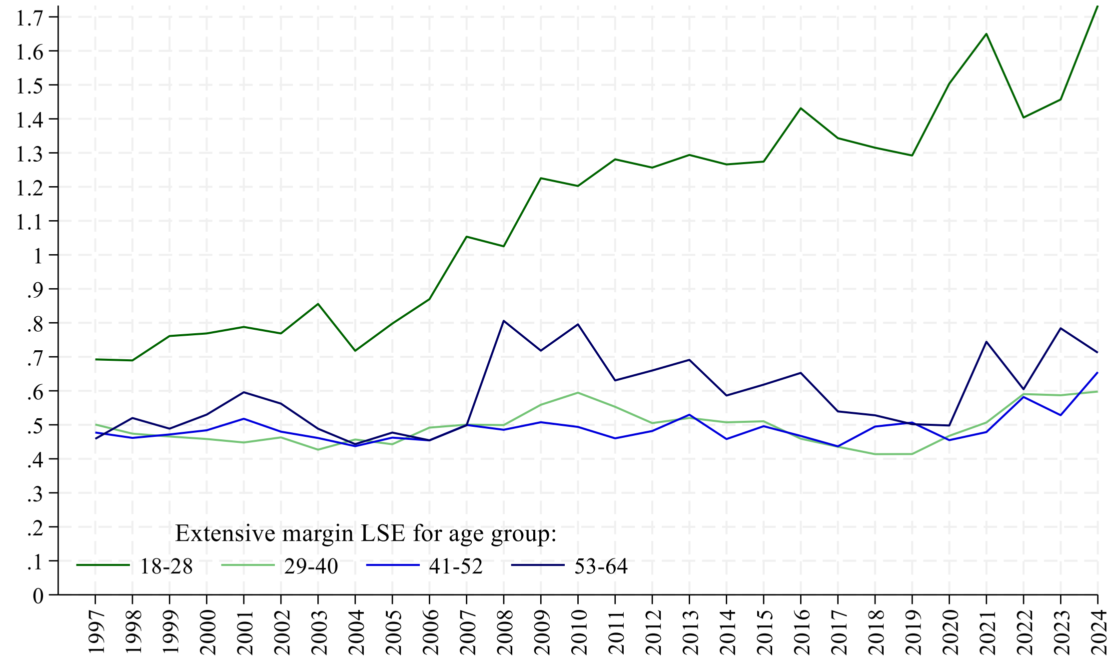
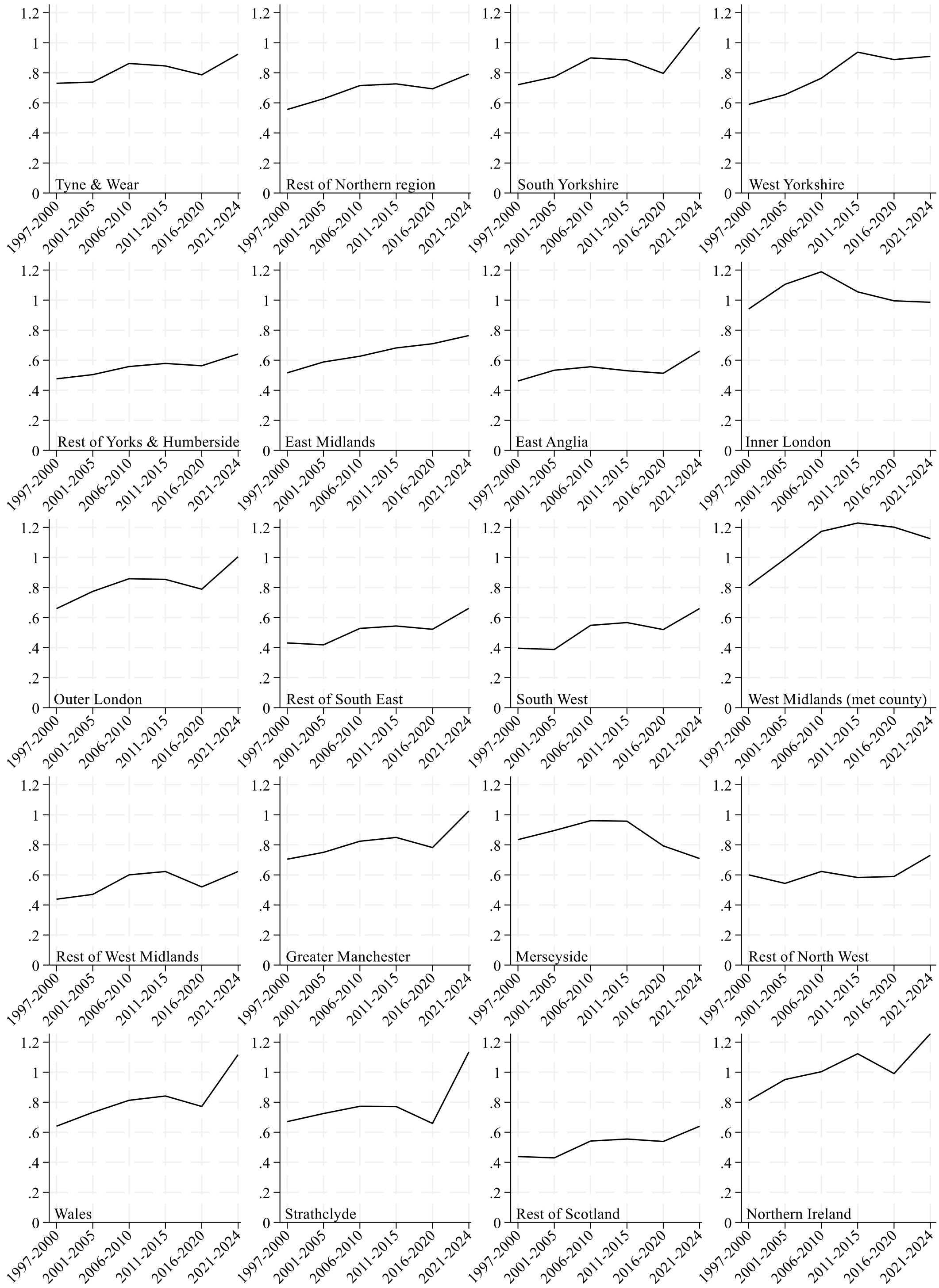


Figure 13 illustrates the extensive margin LSEs estimated separately for the twenty UK regions of residence identified in the LFS variable URESMC. All regions seem to illustrate an upward trend. The ‘Rest of the South East’ (excluding London) has the lowest mean elasticities while the West Midlands (metropolitan area) has the highest.

Figure 13: Extensive margin LSEs across twenty UK regions



5.3 Extensive margin LSEs using HLS data

The estimates in this section are based on HLS data. On the one hand these estimates might be more accurate than LFS derived ones because response rates are better in the HLS, as shown in the next section. On the other hand, the LFS has a much larger sample size, and is likely to provide more reliable estimates when these are sub-divided by into categories with small sub-samples.

Figure 14 presents estimates of the overall extensive margin LSE based on HLS data. These values are slightly lower than the LFS based ones in Figure 9 but they follow the same trend since 2009 and they are consistent with our recommended value of 0.60.

Figure 14: Extensive margin Labour Supply Elasticities

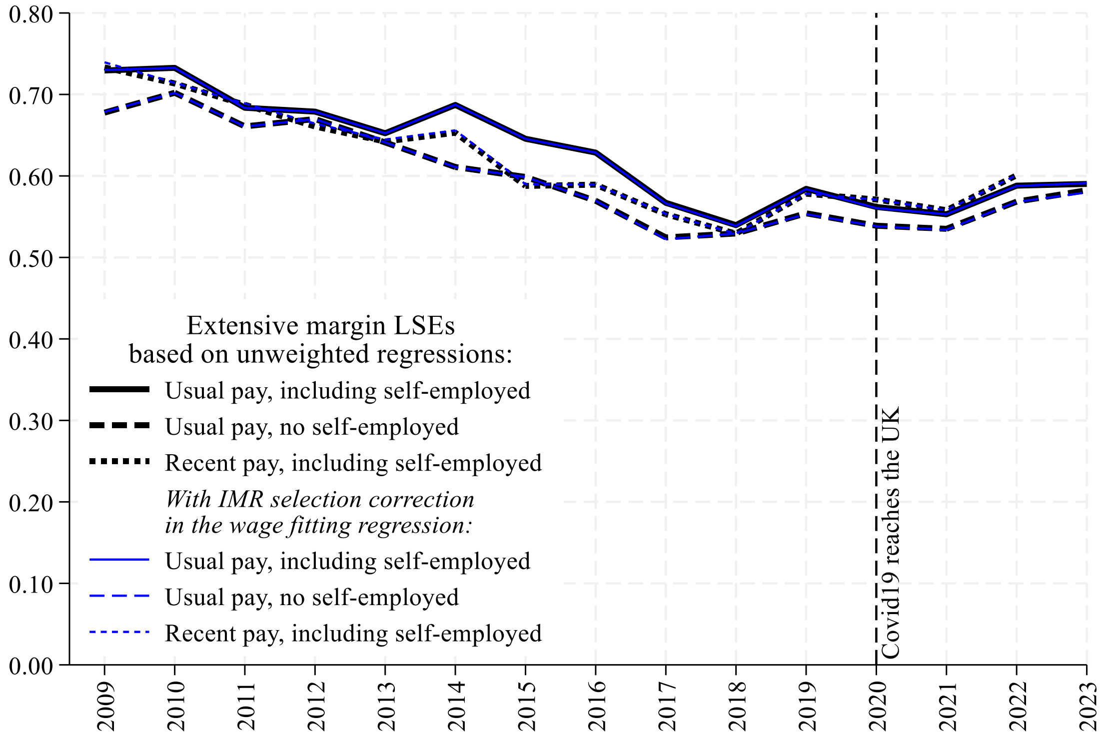


Figure 15 illustrates the extensive margin LSEs by sex, estimated using HLS data, with results similar to those obtained using LFS data in Figure 11 over the period 2009-2020. However, after 2020, the LFS-based estimates increase for both sexes, while the HLS-based estimates increase to around 0.70 for female but remain stable at around 0.50 for males.

Figure 16 illustrates the extensive margin LSE by age group. These show a similar pattern to those in Figure 12, with elasticities remaining relatively stable for the older age groups at around 0.50 to 0.70. Similarly to Figure 12, the youngest age group (18-28) has higher elasticities. However, the youngest group in Figure 16 has elasticities around 0.9 by the end of the sample period, whereas in Figure 12, they rise to 1.7 by 2024. Another feature of the Figure 16 plots is that the series for the youngest group appears more volatile, perhaps due to the smaller sample size in the HLS.

Figure 15: Extensive margin Labour Supply Elasticities, for Females and Males

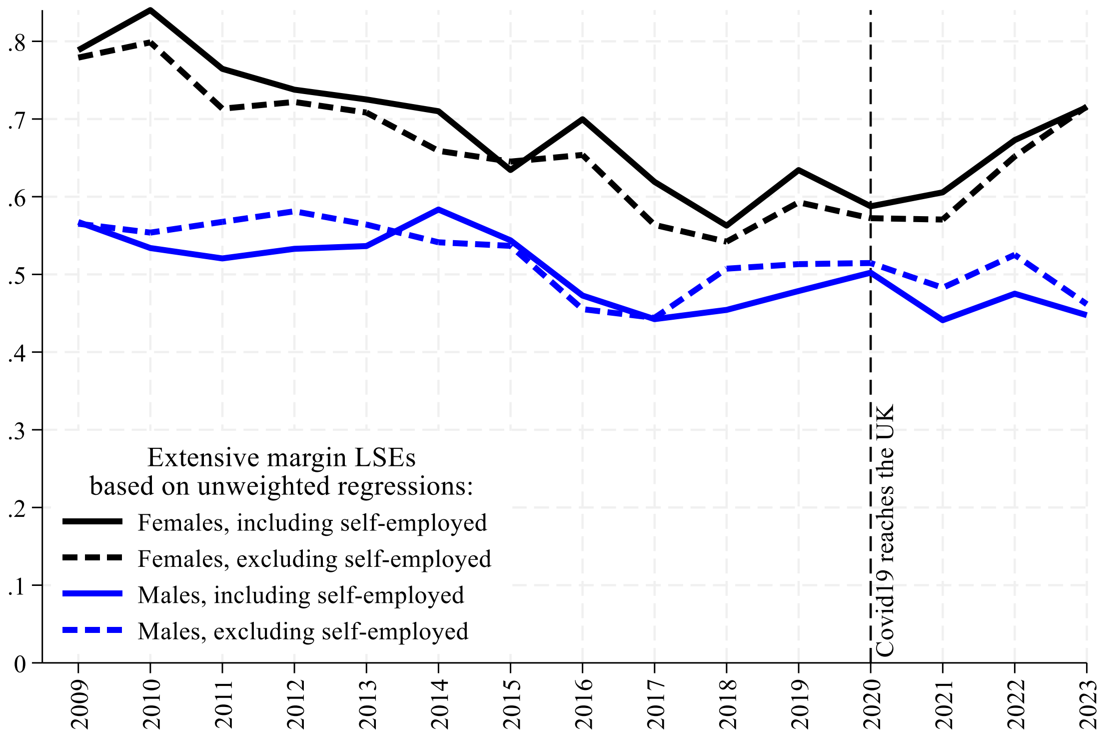
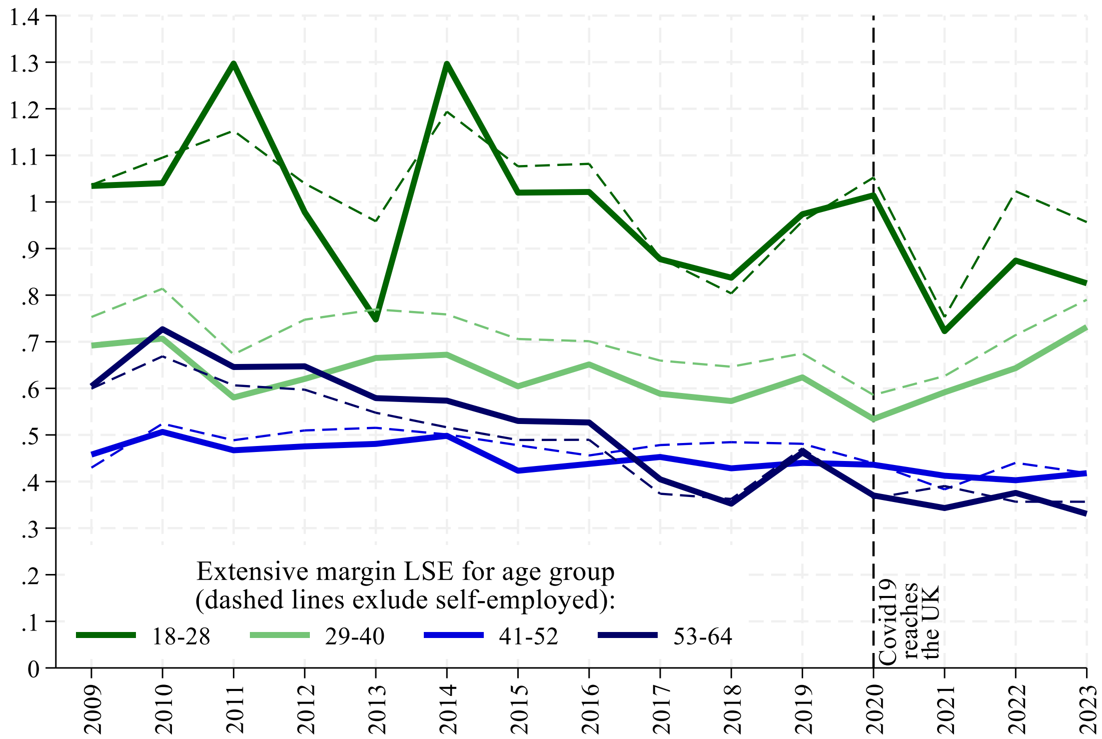
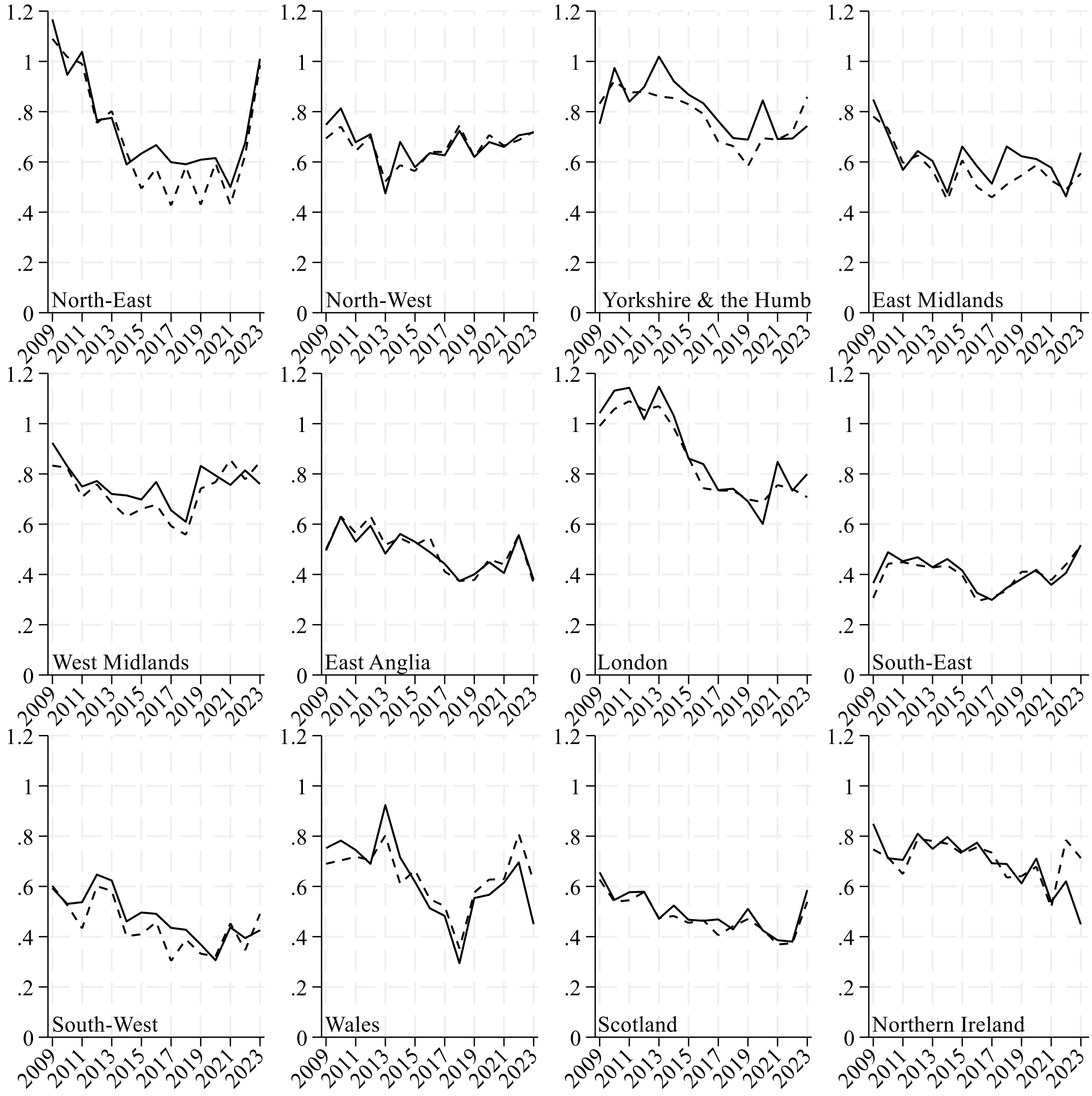


Figure 16: Extensive margin Labour Supply Elasticities, by age



In Figure 17, extensive margin LSEs across twelve UK regions are plotted. Though the regions and years do not exactly match those in Figure 13 for the LFS, they show a similar pattern: stable and slightly falling elasticities during 2009-2021. The South East (excluding London) has lower extensive margin LSEs than other regions, while the West Midlands has high extensive margin LSEs. One notable difference is the high extensive margin LSE for London as a whole, compared to Inner and Outer London in Figure 13. The aggregation of London is unlikely to be the cause of this difference; rather, any differences might be due to sampling differences.

Figure 17: Extensive margin Labour Supply Elasticities, by region (dashed lines exclude the self-employed)



5.4 Biased extensive margin LSEs from using aggregated LFS data

In this section we offer an empirical example to show why the present TAG A2.3 extensive margin LSE of 0.10 might be an underestimate, and why it might come about if one uses aggregated rather than individual-level data. The problem is one of aggregation where most of the variation present in the individual level data is attenuated, leading to attenuated parameter estimates.

The 0.10 elasticity was estimated in late 1999 or early 2000 by the Department for Work and Pensions (DWP), in consultation with the DfT, using National Online

Manpower Information System (NOMIS) data.[8] NOMIS data are only available at a national or regional level, not at the individual level. We speculate that this bias in DWP estimates might reasonably have arisen if the analysis was based on mean regional values, which NOMIS generates by aggregating individual-level data.[9]

To verify the extent to which aggregation might attenuate the elasticity estimates we 'collapsed' the LFS data by Year, SEX, and URESMC which identifies twenty regions across the UK. Starting with the data we had already pre-processed for the individual-level analysis, we used the Stata 'collapse' command to generate mean values for mean employment rates, the mean hourly wage rate, and the sum of the population weights (PWT):

```
rename REFWKY Year
keep if Year>=1997 & Year<=2024
replace HOURPAY = . if HOURPAY<=0
ge  E = (ILODEFR==1)
collapse (mean) E HOURPAY (sum) PWT, by(Year URESMC SEX)
ge lnE = ln(E)
ge lnW = ln(HOURPAY)
```

Thereafter we carried out female, male and joint pooled regressions using these Stata commands using unweighted and population-weighted OLS regressions:

```
reg lnE lnW if SEX==2
reg lnE lnW if SEX==1
reg lnE lnW // Both females and males
reg lnE lnW if SEX==2 [pweight=PWT]
reg lnE lnW if SEX==1 [pweight=PWT]
reg lnE lnW [pweight=PWT] // Both females and males
```

The extensive margin LSEs are the estimated coefficients on `lnW`. The resulting estimates reported in Table 1 are much lower than the estimates based on individual-level data and similar to the 0.10 estimate obtained by DWP in 1999/2000.

[8] This is confirmed in footnote 1 of the TAG A2.3 which states "Estimate based on DWP calculations and wider literature review" and in DfT (2005, pages 52-53, paragraphs 329-330) which states "239. Numerous studies have been undertaken to examine the effect wages have on the labour market. Summaries are available in Blundell (1992), and Ashenfelter and Card (1999). Using these sources, and data from ONS (Nomis), we have calculated an overall elasticity of 0.1 for men and 0.4 for women. Weighting these according to the national claimant count leads to an overall estimate of 0.15. It may be appropriate to vary this estimate if any of the above splits are significantly different from the national average in the study area. 240. Labour model runs by DWP presents an alternative source of information on labour supply response. These runs suggest a somewhat lower elasticity of about 0.05. 241. Considering the above evidence we recommend using a range for the labour supply elasticity of 0.05 to 0.15, with a best estimate of 0.1."

[9] These are the ASHE, the LFS, the Annual Population Survey (APS) and DWP benefit statistics.

Table 1: Extensive margin LSE estimated from pooled regressions on aggregated LFS data

	Females	Males	Females & Males
Unweighted regressions	0.0781	0.0439	0.130
t-statistics	(6.15)	(3.46)	(12.78)
Population-weighted regs	0.0621	0.0359	0.128
t-statistics	(4.39)	(2.80)	(12.97)
Observations	560	560	1120

Notes: The observations are across 28 years, 20 regions and 2 sexes.

Note that in these aggregated models, the elasticity estimates are very sensitive to the exact regression model specification. For instance, running separate regression for each year leads to very different estimates across time. Using random effects panel regression does not substantially alter the results but including region or year fixed effect can lead to extremely high or even negative elasticity estimates.

6 LFS and HLS response rates

The issue of declining LFS response rates is relevant for the estimation of extensive and intensive margin LSEs. The data in this section were retrieved from sub-webpages of the Office for National Statistics (ONS, 2025).[10] Plots are followed by an extensive but somewhat inconclusive literature review of the implications.

Figure 18 illustrates LFS response rates across the four nations of the UK for first-contact (Wave 1) interviews. Initial response rates of around 55%, declined during the Covid19 pandemic and have remained at around 30%. In Northern Ireland, these fell to below 20% during the pandemic. In response, the NISRA (Northern Ireland Statistics and Research Agency, 2024) undertook a "Knock to Nudge" campaign. This involved visiting sampled addresses to encourage participation and arranging a later telephone interview. The campaign had some initial success but response rates have since fallen again. Most values in Figure 18 were obtained by writing a Python script to scrape data from sub-webpages of the ONS (2025). Early values for 1995q4-1996q2 were retrieved manually from PDF files on the same webpage.

Figure 18: LFS Wave 1 response rates in the Four Nations of the UK

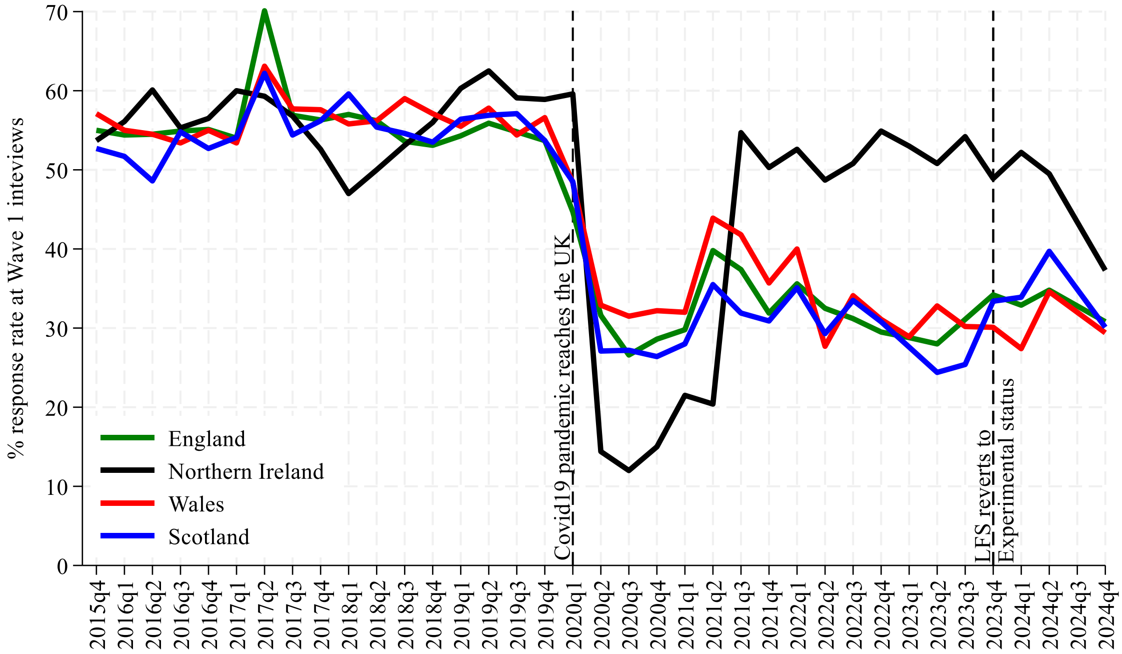
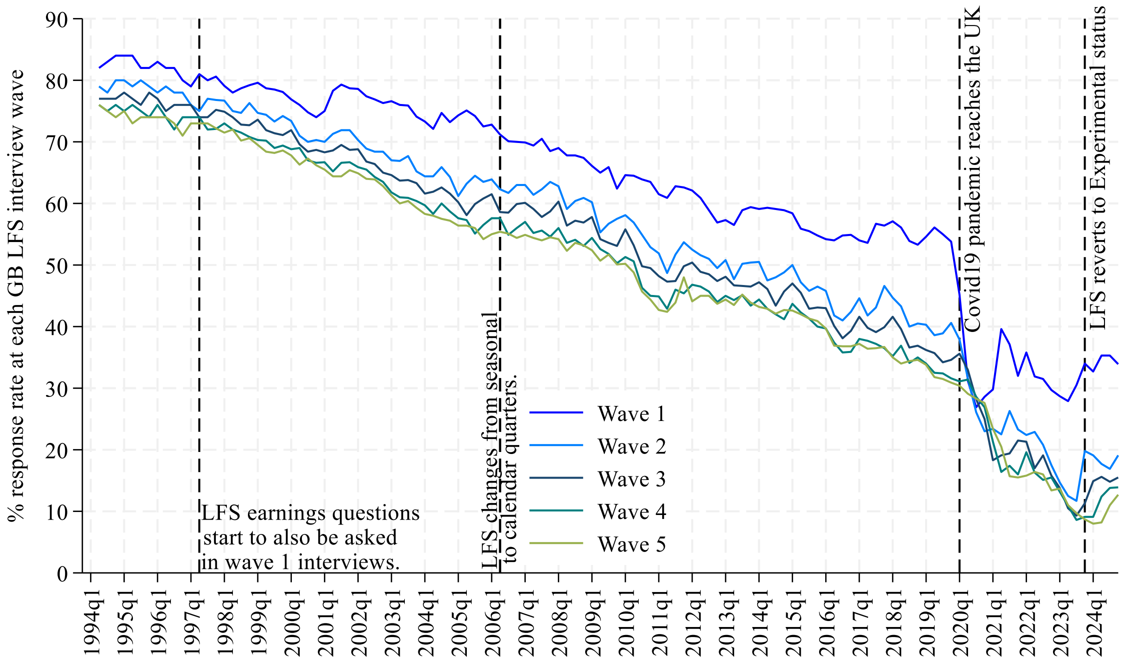


Figure 19 illustrates response rates for Great Britain back to 1994q2 that exclude Northern Ireland but include responses across all five interview waves. Early quarters have high response rates of around 80% that decline thereafter. Wave 1 interviews show a small recovery in 2021q2 and do not decline as much as other wave. During the Covid19 pandemic there is a substantial decline in all response rates. Thereafter, Wave 1 rates have settled at around 30%, while Waves 2 to 5 remain below 20%. The

[10] Sub-pages in https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/labourforcesurveyperformanceandqualitymonitoringreports

values in Figure 19 were simpler to retrieve than those for Figure 18. They involved downloading the early (October-December 2016, Figure 4.2 data) and latest (October-December 2024, Figure 3 data) data sheets from ONS (2025) and concatenating them. Earlier response rates were retrieved from ONS (2003, Volume 1, pp. 21-22). All overlapping values across these three data files matched exactly.

Figure 19: LFS Wave 1 to Wave 5 response rates in Great Britain



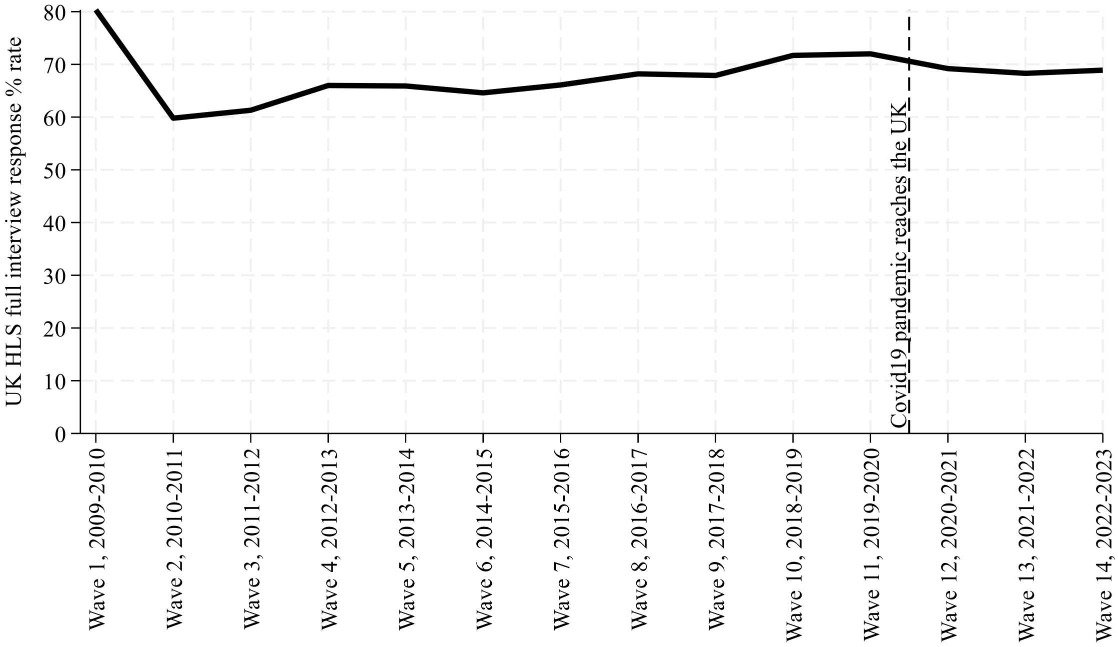
In response to declining LFS response rates, the ONS initiated field trials of the Transformed LFS (TLFS) in October 2023, incorporating online data collection, a revised questionnaire, updated weights, and improved imputation methods. The first TLFS data are now available to users via the Secure Research Service (SRS). A transition period, with both the 'old' LFS and TLFS running in parallel, is expected to continue into 2027, see ONS (2024).

Figure 20 illustrates the full interview response rates for the UK HLS. Though not directly comparable to the LFS rates because the samples are smaller, they indicate that the HLS response rates have held up well to the effects of the Covid-19 pandemic and have remained between 60% and 70%. Based on the evidence in this and the previous sections, we follow the simple premise that pre-2020 LFS data are likely to provide more reliable estimates, albeit at the expense of not including the most recent data. For more recent estimates, we rely on HLS data while allowing for any systematic deviation from LFS estimates prior to 2020, as illustrated in Figure 8.

More broadly, it well known that household survey response rates have been declining in many countries (Statistics Canada, 2023; United States Census Bureau; Flodberg and Wasén, 2024; de Leeuw, Hox and Luiten, 2018). Francis-Devine (2023) noted that UK LFS response rates fall sharply from about 47.9% in mid-2013 to about 14.6% by mid-2023. The rapid decline during the Covid19 pandemic was due to lockdowns which forced a switch from in-person interviews to telephone calls. The long-term decline in survey participation has been attributed to a variety of factors that include issues of privacy and public trust (e.g. rising phone scams), survey fatigue (e.g. marketing bombardment), information overload (e.g. long, complicated surveys), lifestyle changes (e.g. busier modern work arrangements), and cultural shifts in civic participation (e.g. a diminished sense of obligation).

Falling response rates may negatively impact data quality by increasing sampling uncertainty and non-response bias. These considerations mean that, from October 2023, statistics derived from the LFS (employment, unemployment, and economic inactivity rates) have reverted to an ‘experimental’ status from their ‘accredited’ status (ONS, 2023). The impact of low response rates on labour market statistics may be manifold. Measurement errors and bias in key variables, such as employment and wages, may result in unreliable and misleading coefficient estimates. Low response rates may skew a sample’s composition, violating the statistical fundamentals of a representative distribution. Unlike random sampling error, non-response bias is a systematic error that might not be addressed by increasing sample size and repeated sampling (Greene, 2018). The ONS has addressed the issue of low LFS response rates by implementing, like us, weight adjustments. However, this can only correct for bias in the observable characteristics that are used in the weighting method.

Figure 20: UK HLS full interview response rates



Notes: Each wave includes a few interviews that occur in the first months after the second listed year.

The issue of whether recent falls in LFS response rates introduce bias affecting estimates has not been definitively resolved, but it is not a new issue in the literature. Regarding labour supply elasticities, Heckman (1979) showed that non-response bias may lead to biased wage coefficients in labour supply models. For instance, women with low potential wages are less likely to work, so a simple wage-hours regression on working women may overstate their LSE. This issue led to the development of the Heckman two-step correction estimation (Heckit).

Bollinger et al. (2019) suggest that earnings non-response is non-random. They find that non-response is U-shaped over the earnings distribution, being highest for low-income and high-income earners. This means that labour surveys under-represent the lowest- and highest-wage workers. High earners typically have smaller labour supply responses, so missing many high earners might inflate the estimated elasticity. Conversely, missing low earners could deflate the estimated elasticity. Bollinger and Hirsch (2012) suggest that understated mean wages due to selective survey participation can impact wage elasticity estimates. These biases can lead to both underestimation and overestimation of how strongly labour supply reacts to wage changes. These two biases are unlikely to exactly cancel each other out.

7 Summary and implications

This report provides updated estimates of the UK workforce's labour supply elasticity (LSEs). This elasticity is crucial for understanding the responsiveness of labour supply to wage changes and is included in the DfT (2018) TAG for evaluating transport investment proposals. Given the evidence, we recommend that the DfT use an extensive margin LSE of 0.60 in place of 0.10 for its TAG equation 2.

For the associated intensive margin LSE, which measures changes in hours worked by existing employees in response to wage variations, results indicate that it has declined from approximately 0.20 in 1997 to 0.12 in 2024 (see Figure 1). This suggests that UK workers are becoming less responsive to wage changes in their work hours.

Of more relevance for the DfT TAG, the extensive margin LSE, which captures the probability of individuals entering or exiting employment in response to wage changes, was approximately 0.60 in 2019 (see Figure 8). This is the last year for which we are confident in the LFS response rate levels (see Figure 18). During 1997-2006, the extensive margin LSE fluctuated around 0.55, before rising to around 0.70 during 2008-2014 and gradually falling again until 2020. These changes highlight that employment decisions have fluctuated, possibly due to changes in labour market dynamics, economic conditions, or policy reforms. We are less confident in the substantial increase in the estimated extensive margin LSE during 2020-2024 from 0.60 to 0.84, given the low LFS response rates. Estimates based on HLS data are slightly lower, but they too suggest its value was 0.60 in 2022.

Any estimates are subject to caveats, including ours. Though we are confident in the rigour of the statistical techniques, we urge some caution with respect to the data. We found that adjustments to the statistical models, such as accounting for sample selection (Heckit estimation) or using multivariate versus bivariate regressions, did not noticeably alter the estimated elasticities. This is shown in our comparisons, and we believe it is due to the population representativeness of the data. Of greater importance to the results was the care taken in pre-processing the data to omit unresponsive respondents in order to avoid generating fitted wages for them. Failure to do this carefully could lead to estimates that were as much as 0.05 above or below our central estimate of 0.60. Another caveat relates to the falling LFS response rates since the onset of the Covid19 pandemic in 2020q1. We, therefore, remain uncertain as to the extent to which the rising extensive margin LSE is real or a sampling issue. The LFS is taking steps to improve response rates, and we remain of the opinion that the UK LFS and UK HLS (Understanding Society) are the best available population-representative datasets outside of the censuses.

The evolution of these labour market elasticities holds implications for policy analysis and the research used to inform it. The high responsiveness of employment to wages (extensive margin LSE) suggests that wage adjustments or transport-related policy changes (such as improvements in commuting infrastructure) could have a significant impact on employment levels. This should be reflected in the DfT TAG, particularly in its employment equation 2. Beyond this, the declining intensive margin LSE implies that existing workers are less likely to adjust their working hours in response to wage changes. This could reflect increased work rigidity, changing job structures, or the extent to which minimum wages alter labour market outcomes. This report highlights the need for regular re-estimation of LSEs, especially in light of structural labour market changes.

In conclusion, this report underscores the dynamic nature of labour supply elasticities in the UK. The findings have important implications for labour market policies, wage-setting mechanisms, and economic planning, particularly in relation to transport infrastructure and broader employment strategies.

Bibliography

Blundell, R., Bozio, A., & Laroque, G. (2011). Labor supply and the extensive margin. American Economic Review, 101(3), 482-486.

Blundell, R., Costa Dias, M., Meghir, C., & Shaw, J. (2016). Female labor supply, human capital, and welfare reform. Econometrica, 84(5), 1705-1753.

Bollinger, C.R., Hirsch, B.T., Hokayem, C. and Ziliak, J. (2019). Trouble in the Tails? What We Know about Earnings Nonresponse 30 Years after Lillard, Smith, and Welch. *Journal of Political Economy*, 127(5)

Bollinger, C.R. and Hirsch, B.T. (2012). Is Earnings Nonresponse Ignorable? University of Kentucky Center for Poverty Research Discussion Paper Series. 32

Department for Transport (2005), Transport, Wider Economic Benefits and Impacts on GDP, discussion paper, http://webarchive.nationalarchives.gov.uk/+/http:/www.dft.gov.uk/pgr/economics/rdg/webia/webmethodology/sportwidereconomicbenefi3137.pdf

Department for Transport (2018), Transport Analysis Guidance A2.3, https://www.gov.uk/government/publications/webtag-tag-unit-a2-3-employment-effects-may-2018

Department for Transport (2024), Transport Analysis Guidance, www.gov.uk/transport-analysis-guidance-webtag

Detilleux, C., & Deschacht, N. (2021). The causal effect of the number of children on gender-specific labour supply elasticities to the firm. Industrial Relations Journal, 52(1), 2-24. https://doi.org/10.1111/irj.12314

Ercolani, M., Bromfield, J., & Selim, S. (2024). A review of TAG AA2.3 employment parameters: labour supply elasticity and labour tax wedges. (Transport appraisal and strategic modelling (TASM) research reports). Department for Transport. https://www.gov.uk/government/publications/transport-analysis-guidance-and-employment-parameters

Francis-Devine, B. (2023). Has labour market data become less reliable? Available at: https://commonslibrary.parliament.uk/has-labour-market-data-become-less-reliable/ [Accessed: 28 February 2025]

Flodberg, C and Wasén, P. (2024). Economic Commentary Lower response rates implies challenges for monetary policy in several countries. Available at: https://www.riksbank.se/globalassets/media/rapporter/ekonomiska-kommentarer/engelska/2024/economic-commentary---lower-response-rates-implies-challenges-for-monetary-policy-in-several-countries.pdf#:~:text=some%20statistically%20significant%20bias%2C%20including,the%20Public%20Employment%20Service%20on [Accessed: 28 February 2025]

Greene. W.H. (2018). Econometric Analysis (8th ed.). Harlow: Pearson

Heckman, J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, 47(1)

de Leeuw E., Hox J. and Luiten A. (2018), International Nonresponse Trends across Countries and Years: An analysis of 36 years of Labour Force Survey data. Survey Insights: Methods from the Field. Available at: https://surveyinsights.org/?p=10452#:~:text=Force%20Survey%20in%2016%20countries%3A,and%20de%20Heer%202002%2C%20Appendix [Accessed: 28 February 2025]

Lucas, R. E., Jr. (1976). **Econometric policy evaluation: A critique**. *Carnegie-Rochester Conference Series on Public Policy, 1*(1), 19-46. https://doi.org/10.1016/S0167-2231(76)80003-6

Northern Ireland Statistics and Research Agency (NISRA) (2024) The transformation of labour market statistics in Northern Ireland: NISRA user information paper, April 2024. NISRA. https://www.nisra.gov.uk/publications/labour-force-survey-transformation-northern-ireland-user-guidance

Office for National Statistics. (2003). Labour Force Survey user guide - Volume 1: Background and methodology. [Available with the LFS data package.]

Office for National Statistics. (2023). *Review of Experimental Labour Market Statistics*. Available at: https://osr.statisticsauthority.gov.uk/wp-content/uploads/2023/11/Review_of_experimental_labour_market_statistics_produced_by_the_office_for_national_statistics.pdf [Accessed: 29 February]

Office for National Statistics and Northern Ireland Statistics and Research Agency, (1997-2024). Labour Force Survey, UK Data Service. https://ukdataservice.ac.uk

Office for National Statistics (2024) Labour market transformation: Update on progress and plans. Available at: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/articles/labourmarkettransformationupdateonprogressandplans/december2024 (Accessed: 21 March 2025).

Office for National Statistics. (2025). Labour Force Survey performance and quality monitoring report: October to December 2024, February 18. https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/labourforcesurveyperformanceandqualitymonitoringreports

United States Census Bureau. (2024). Current Population Survey (CPS) 2023 Modernization Efforts. Available at: https://www.census.gov/programs-surveys/cps/about/modernization.html [Accessed: 28 February 2025]

University of Essex Institute for Social and Economic Research (2024). Understanding Society: Waves 1-14, 2009-2023 and Harmonised BHPS: Waves 1-18, 1991-2009.

(SN 6614.; Version 19th Edition) [Dataset]. UK Data Service. https://doi.org/10.5255/UKDA-SN-6614-20

Statistics Canada. (2023). Labour Force Survey Response Rates, September 2023. Available at: https://www150.statcan.gc.ca/n1/pub/75-005-m/75-005-m2023001-eng.htm [Accessed: 28 February 2025]

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). “Welcome to the tidyverse.” *Journal of Open Source Software*, 4(43), 1686. doi:10.21105/joss.01686.

Appendices

A1 Algebraic derivation of the extensive margin elasticity

Though some econometrics textbooks provide derivations of elasticities for Probit regression, we could not find any that allow for the regressors to be in logarithms. We therefore provide a derivation of such elasticities here, where the wage is in logarithms in the regression but the elasticity is not in logarithms. For simplicity, we omit the hat on $\widehat{W}_i$ and $\widehat{\ln W_i}$ to indicate these are fitted values.

Given equation (3) the extensive margin LSE is defined as a derivative with respect to the probability of being in employment in the numerator:

$$eLSE_i = \frac{\partial \Pr(E_i = 1) / \Pr(E_i = 1)}{\partial W_i / W_i}$$

$$= \frac{\partial \Pr(E_i = 1)}{\partial W_i} \times \frac{W_i}{\Pr(E_i = 1)} \qquad \text{(A1.1)}$$

Operationalising equation (A1.1) is complicated by the fact that in regression equation (3) the dependent variable is binary and that the wage is a regressor in logarithms. Focusing on the partial derivative in (A1.1), it is trivial to specify it as:

$$\frac{\partial \Pr(E_i = 1)}{\partial W_i} = \frac{\partial \Pr(E_i = 1)}{\partial W_i} \times \frac{\partial lnW_i}{\partial lnW_i}$$

$$= \frac{\partial \Pr(E_i = 1)}{\partial lnW_i} \times \frac{\partial lnW_i}{\partial W_i} \qquad \text{(A1.2)}$$

Substituting $z_i = \beta_0 + \beta_1 lnW_i$. into equation (3) gives the more compact notation for the Probit function (3) that is differentiated by (A1.2):

$$\Pr(E_i = 1) = \Phi(z_i) \qquad \text{(A1.3)}$$

The first derivative on the right-hand side of (A1.2) is given by the chain rule when differentiating the Probit function with respect to lnW_i, not W_i:

$$\frac{\partial \Pr(E_i = 1)}{\partial lnW_i} = \beta_1 \phi(\beta_0 + \beta_1 lnW_i)$$

$$= \beta_1 \phi(z_i) \qquad \text{(A1.4)}$$

where the 'inner' derivative is $\partial(\beta_0 + \beta_1\, lnW_i\,)/\partial lnW = \beta_1$, and the 'outer' derivative is $\partial\Phi(z_i\,)/\partial lnW = \phi(z_i)$. This 'outer' derivative comes about because the derivative of the normal CDF:

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\,(-y^2/2)\, dy$$

is the normal PDF $\Phi(z_i)$ where the derivative can be with respect to any one of the variables, such as lnW_i,

$$\phi(z) = \frac{\partial \Phi(z)}{\partial z} = \frac{1}{\sqrt{2\pi}} \exp\,(-z^2/2)$$

The second of the derivatives on the right of equation (A1.2) results from differentiating the logarithm:

$$\frac{\partial lnW_i}{\partial W_i} = \frac{l_i}{W_i} \qquad \text{(A1.5)}$$

Substituting equations (A1.4) and (A1.5) into (A1.2) gives:

$$\frac{\partial \Pr(E_i=1)}{\partial W_i} = \beta_1 \phi(z_i) \times \frac{l_i}{W_i}. \qquad \text{(A1.6)}$$

Substituting equations (A1.6) and (A1.3) into (A1.1) and cancelling out the W_i gives:

$$eLSE_i = \beta_1 \phi(z_i) \times \frac{l}{W_i} \times \frac{W_i}{\Phi(z_i)}$$

$$= \beta_1 \phi(z_i) \times \frac{l_i}{\Phi(z_i)} \qquad \text{(A1.7)}$$

Equation (A1.7) corresponds to equation (4) in the main text.

A2 Intensive margin LSE estimates, bivariate regression model (1)

The tables presented here are for the OLS estimates using LFS data of the basic log-log model (1) used to estimate the intensive margin LSEs (iLSE) estimated in the parameter β_1. These are illustrated in Figure 1.

Table A2.1: ln(TTACHR) regressed on ln(HOURPAY)

Year	β_0	s.e.(β_0)	β_1 **(iLSE)**	s.e.(β_1)	R^2	Obs
1997	3.06	0.009	**0.22**	0.005	0.06	58118
1998	3.07	0.009	**0.22**	0.005	0.06	63437
1999	3.02	0.010	**0.24**	0.005	0.07	60759
2000	3.02	0.010	**0.23**	0.005	0.06	56921
2001	3.01	0.010	**0.23**	0.005	0.06	55915
2002	2.99	0.010	**0.23**	0.005	0.06	55279
2003	2.98	0.012	**0.23**	0.005	0.06	53337
2004	3.00	0.012	**0.22**	0.005	0.06	48794
2005	2.98	0.012	**0.22**	0.005	0.06	48141
2006	2.99	0.012	**0.22**	0.005	0.06	47044
2007	3.04	0.012	**0.19**	0.005	0.05	47571
2008	3.02	0.012	**0.20**	0.005	0.05	45950
2009	3.00	0.014	**0.20**	0.006	0.05	42666
2010	2.99	0.013	**0.20**	0.005	0.06	40698
2011	2.96	0.014	**0.21**	0.006	0.06	38110
2012	2.96	0.015	**0.21**	0.006	0.06	38125
2013	2.95	0.014	**0.22**	0.006	0.06	37351
2014	3.03	0.014	**0.19**	0.006	0.05	38157
2015	3.05	0.014	**0.18**	0.006	0.05	35934
2016	3.02	0.015	**0.19**	0.006	0.05	34264
2017	3.06	0.015	**0.18**	0.006	0.05	35237
2018	3.06	0.016	**0.17**	0.006	0.05	34093
2019	3.09	0.016	**0.16**	0.006	0.04	33499
2020	3.08	0.018	**0.16**	0.006	0.04	26247
2021	3.15	0.019	**0.14**	0.007	0.03	30276
2022	3.10	0.020	**0.15**	0.007	0.04	26612
2023	3.12	0.028	**0.14**	0.010	0.04	19190
2024	3.13	0.023	**0.13**	0.008	0.03	24502

Table A2.2: ln(TTUSHR) regressed on ln(HOURPAY)

Year	β_0	s.e.(β_0)	β_1 **(iLSE)**	s.e.(β_1)	R^2	Obs
1997	3.13	0.009	**0.22**	0.004	0.06	65564
1998	3.14	0.009	**0.21**	0.004	0.06	71001
1999	3.10	0.009	**0.22**	0.004	0.07	67939
2000	3.09	0.009	**0.22**	0.004	0.07	63790
2001	3.08	0.009	**0.22**	0.004	0.07	62758
2002	3.06	0.009	**0.23**	0.004	0.07	62161
2003	3.04	0.010	**0.23**	0.005	0.07	60491
2004	3.05	0.011	**0.22**	0.005	0.07	55117
2005	3.04	0.010	**0.22**	0.004	0.07	54267
2006	3.06	0.011	**0.21**	0.005	0.07	53201
2007	3.10	0.011	**0.19**	0.004	0.06	53520
2008	3.09	0.011	**0.19**	0.004	0.06	52109
2009	3.06	0.012	**0.20**	0.005	0.06	48276
2010	3.05	0.012	**0.20**	0.005	0.06	45787
2011	3.03	0.013	**0.20**	0.005	0.06	42961
2012	3.01	0.014	**0.21**	0.005	0.07	42518
2013	3.01	0.014	**0.21**	0.005	0.07	41763
2014	3.08	0.013	**0.18**	0.005	0.06	42874
2015	3.09	0.013	**0.18**	0.005	0.06	40341
2016	3.07	0.013	**0.19**	0.005	0.06	38419
2017	3.11	0.013	**0.17**	0.005	0.05	39368
2018	3.10	0.014	**0.17**	0.005	0.06	38096
2019	3.13	0.014	**0.16**	0.005	0.05	37470
2020	3.17	0.015	**0.15**	0.005	0.04	31670
2021	3.20	0.017	**0.13**	0.006	0.04	34413
2022	3.16	0.018	**0.14**	0.006	0.04	29437
2023	3.20	0.024	**0.13**	0.008	0.04	21178
2024	3.18	0.021	**0.13**	0.007	0.03	27037

Table A2.3: ln(BACTHR) regressed on ln(HOURPAY)

Year	β_0	s.e.(β_0)	β_1 **(iLSE)**	s.e.(β_1)	R^2	Obs
1997	3.06	0.009	**0.18**	0.005	0.04	58140
1998	3.08	0.009	**0.17**	0.004	0.04	63450
1999	3.03	0.009	**0.19**	0.004	0.05	60799
2000	3.03	0.010	**0.19**	0.005	0.04	56956
2001	3.03	0.010	**0.18**	0.004	0.04	55968
2002	3.01	0.010	**0.18**	0.004	0.04	55325
2003	3.00	0.011	**0.19**	0.005	0.04	53382
2004	3.02	0.011	**0.18**	0.005	0.04	48824
2005	3.01	0.011	**0.18**	0.005	0.04	48192
2006	3.02	0.012	**0.18**	0.005	0.04	47106
2007	3.07	0.011	**0.15**	0.005	0.04	47623
2008	3.05	0.012	**0.16**	0.005	0.04	45981
2009	3.04	0.013	**0.16**	0.005	0.04	42695
2010	3.03	0.013	**0.16**	0.005	0.04	40738
2011	2.99	0.014	**0.17**	0.005	0.04	38146
2012	3.00	0.015	**0.17**	0.006	0.04	38150
2013	3.00	0.014	**0.17**	0.005	0.04	37376
2014	3.06	0.014	**0.15**	0.005	0.03	38170
2015	3.07	0.014	**0.15**	0.005	0.04	35951
2016	3.04	0.015	**0.16**	0.006	0.04	34279
2017	3.08	0.014	**0.14**	0.005	0.03	35254
2018	3.08	0.015	**0.14**	0.006	0.03	34133
2019	3.10	0.016	**0.13**	0.006	0.03	33497
2020	3.10	0.017	**0.13**	0.006	0.03	26246
2021	3.16	0.018	**0.12**	0.006	0.03	30273
2022	3.12	0.019	**0.13**	0.007	0.03	26620
2023	3.14	0.027	**0.12**	0.009	0.03	19201
2024	3.14	0.022	**0.11**	0.008	0.02	24496

Table A2.4: ln(BUSHR) regressed on ln(HOURPAY)

Year	β_0	s.e.(β_0)	β_1 **(iLSE)**	s.e.(β_1)	R^2	Obs
1997	3.12	0.008	**0.17**	0.004	0.05	65715
1998	3.13	0.008	**0.17**	0.004	0.04	71153
1999	3.10	0.009	**0.18**	0.004	0.05	68140
2000	3.09	0.009	**0.18**	0.004	0.05	64057
2001	3.09	0.009	**0.18**	0.004	0.05	63004
2002	3.06	0.009	**0.19**	0.004	0.05	62385
2003	3.05	0.010	**0.19**	0.004	0.05	60768
2004	3.07	0.010	**0.18**	0.004	0.05	55364
2005	3.06	0.010	**0.18**	0.004	0.05	54522
2006	3.08	0.011	**0.17**	0.004	0.05	53441
2007	3.11	0.010	**0.15**	0.004	0.04	53772
2008	3.11	0.010	**0.15**	0.004	0.04	52323
2009	3.08	0.012	**0.16**	0.005	0.04	48488
2010	3.07	0.012	**0.16**	0.005	0.04	45996
2011	3.06	0.013	**0.16**	0.005	0.04	43152
2012	3.04	0.013	**0.17**	0.005	0.05	42723
2013	3.04	0.013	**0.17**	0.005	0.05	41956
2014	3.10	0.013	**0.15**	0.005	0.04	43057
2015	3.11	0.012	**0.15**	0.005	0.04	40526
2016	3.07	0.013	**0.16**	0.005	0.05	38607
2017	3.12	0.013	**0.14**	0.005	0.04	39581
2018	3.11	0.014	**0.14**	0.005	0.04	38309
2019	3.13	0.014	**0.13**	0.005	0.04	37615
2020	3.17	0.014	**0.12**	0.005	0.03	31755
2021	3.21	0.016	**0.11**	0.006	0.03	34505
2022	3.16	0.017	**0.12**	0.006	0.03	29532
2023	3.20	0.023	**0.11**	0.008	0.03	21223
2024	3.18	0.021	**0.11**	0.007	0.03	27084

A3 Extensive margin LSE estimates, equation (4)

Table A3.1 lists the extensive margin LSE estimates illustrated in figures 8, 9 and 14. Estimates on LFS without population weights are not reported as they are very similar to the weighted estimates. Estimates on HLS data are unweighted as splitting the data by year means the provided population wave weights are not directly applicable. Estimates with the 'Heckit IMR' suffixes indicate the estimates are carried out using Heckman (1979) sample selection correction, which uses the Inverse Mills Ratio (IMR) as described in equations (5) to (7).

Table A3.1: Extensive margin LSE (eLSE), excluding the self-employed

Year	Estimates using LFS data using population weights (PWT) eLSE	eLSE Heckit IMR	Obs	Estimates using HLS data without population weights eLSE	eLSE Heckit IMR	Obs
1997	0.53	0.55	98517			
1998	0.53	0.55	105224			
1999	0.55	0.57	101149			
2000	0.57	0.58	95385			
2001	0.59	0.61	94004			
2002	0.58	0.59	92916			
2003	0.56	0.58	90403			
2004	0.52	0.54	89219			
2005	0.54	0.55	90524			
2006	0.55	0.57	85992			
2007	0.62	0.63	79517			
2008	0.70	0.71	81520			
2009	0.72	0.73	77999	0.73	0.73	18962
2010	0.72	0.73	74383	0.73	0.73	40211
2011	0.70	0.71	70685	0.69	0.69	38130
2012	0.69	0.70	69096	0.68	0.68	33736
2013	0.69	0.70	67011	0.65	0.65	31460
2014	0.67	0.67	67608	0.69	0.69	29888
2015	0.64	0.64	62119	0.64	0.64	30389
2016	0.65	0.65	59222	0.63	0.63	29838
2017	0.62	0.62	59056	0.57	0.57	26750
2018	0.66	0.67	57241	0.54	0.54	25144
2019	0.63	0.64	55868	0.58	0.58	23882
2020	0.65	0.65	46801	0.56	0.56	22445
2021	0.74	0.74	51247	0.55	0.55	20407
2022	0.76	0.76	44323	0.59	0.59	20871
2023	0.81	0.81	31787	0.59	0.59	11801
2024	0.84	0.85	39517			

A4 LFS response rates

Tables A4.1 and A4.2 list the LFS response rates illustrated in Figures 18 and 19. The values in Table A4.1 were scraped by running a python script on the sub-pages of https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/labourforcesurveyperformanceandqualitymonitoringreports and values in three PDF files on the same page were retrieved manually.

The values in Table A4.2 were obtained by concatenating the earliest and latest available data on the same sub-pages. The earliest is the October-December 2016,

Figure 4.2 data and latest is the October-December 2024, Figure 3 data. Earlier response rates were retrieved from ONS (2003, Volume 1, pp. 21-22).

Table A4.1: LFS Wave 1 interview response rates for the UK

Quarter	England	Northern Ireland	Scotland	Wales
2015q4	55.0	53.7	52.7	57.1
2016q1	54.4	56.1	51.7	55.0
2016q2	54.5	60.1	48.6	54.5
2016q3	54.9	55.3	54.8	53.4
2016q4	55.1	56.5	52.7	55.0
2017q1	54.0	60.0	54.1	53.4
2017q2	70.1	59.3	62.2	63.1
2017q3	56.9	56.8	54.4	57.7
2017q4	56.3	52.6	56.2	57.6
2018q1	57.0	47.0	59.6	55.8
2018q2	56.2	50.0	55.4	56.2
2018q3	53.6	53.1	54.6	59.0
2018q4	53.1	56.0	53.5	57.1
2019q1	54.3	60.3	56.4	55.5
2019q2	55.9	62.5	56.9	57.8
2019q3	54.8	59.1	57.1	54.4
2019q4	53.7	58.9	53.7	56.6
2020q1	44.7	59.6	48.5	48.4
2020q2	31.6	14.4	27.1	32.9
2020q3	26.6	12.0	27.2	31.5
2020q4	28.6	15.0	26.4	32.2
2021q1	29.8	21.5	28.0	32.0
2021q2	39.8	20.4	35.5	43.9
2021q3	37.4	54.7	31.9	41.8
2021q4	31.9	50.3	30.9	35.7
2022q1	35.6	52.6	35.1	40.0
2022q2	32.5	48.7	29.3	27.7
2022q3	31.2	50.8	33.5	34.1
2022q4	29.5	54.9	30.8	31.1
2023q1	28.8	53.0	27.6	28.9
2023q2	28.0	50.8	24.4	32.8
2023q3	31.1	54.2	25.4	30.2
2023q4	34.2	48.9	33.4	30.1
2024q1	32.9	52.2	33.9	27.4
2024q2	34.8	49.5	39.7	34.6
2024q4	30.8	37.3	30.1	29.4

Table A4.2: LFS response rates for GB

	Wave						Wave				
Quarter	1	2	3	4	5	Quarter	1	2	3	4	5
1994q2	82	79	77	76	76						
1994q3	83	78	77	75	75						
1994q4	84	80	77	76	74						
1995q1	84	80	78	75	75	2010q1	64.6	58.1	55.8	51.3	50.2
1995q2	84	79	77	76	73	2010q2	64.5	56.9	53.2	50.6	48.8
1995q3	82	80	76	75	74	2010q3	63.9	54.9	49.8	46.3	45.7
1995q4	82	79	78	74	74	2010q4	63.5	52.9	49.5	45.0	44.4
1996q1	83	78	77	76	74	2011q1	61.5	51.9	48.2	44.9	42.7
1996q2	82	79	75	74	74	2011q2	60.9	48.7	47.3	42.9	42.4
1996q3	82	78	76	72	73	2011q3	62.8	51.7	47.4	46.0	43.9
1996q4	80	78	76	74	71	2011q4	62.6	53.7	49.8	45.4	48.0
1997q1	79	76	76	74	73	2012q1	62.1	52.5	50.4	46.8	44.1
1997q2	81	75	74	74	73	2012q2	60.9	51.6	48.9	46.5	45.0
1997q3	80	77	74	72	73	2012q3	58.9	51.0	48.5	45.7	45.0
1997q4	80.6	76.8	75.2	72.1	72.2	2012q4	56.9	49.5	47.4	44.0	43.7
1998q1	79.1	76.7	74.9	73.0	71.5	2013q1	57.3	50.8	48.1	45.0	44.4
1998q2	78.0	75.0	74.0	72.0	72.0	2013q2	56.5	47.7	46.7	44.3	43.5
1998q3	78.7	74.7	72.8	71.5	70.2	2013q3	58.9	50.2	46.6	45.1	45.2
1998q4	79.2	76.3	72.7	70.8	70.6	2013q4	59.4	50.4	46.5	43.4	43.9
1999q1	79.6	74.7	73.6	70.3	69.5	2014q1	59.1	50.5	47.2	44.4	43.2
1999q2	78.7	74.4	71.9	70.2	68.4	2014q2	59.3	47.5	46.1	42.9	42.9
1999q3	78.5	73.3	71.4	69.0	68.2	2014q3	59.1	48.0	43.4	42.0	42.1
1999q4	78.1	74.2	71.1	69.4	68.6	2014q4	58.9	48.8	45.7	41.2	42.7
2000q1	76.9	73.4	71.9	68.8	67.8	2015q1	58.4	50.0	47.0	43.7	42.6
2000q2	76.0	71.0	69.6	69.0	66.3	2015q2	55.9	47.2	45.5	42.3	42.0
2000q3	74.8	70.0	68.4	67.0	67.3	2015q3	55.5	45.8	42.9	41.3	41.3
2000q4	74.0	70.3	68.7	66.6	66.2	2015q4	54.8	46.5	43.1	40.0	40.9
2001q1	75.0	70.0	68.3	66.7	65.5	2016q1	54.2	45.8	43.0	39.7	39.7
2001q2	78.3	71.3	68.6	65.2	64.4	2016q2	54.0	41.8	40.1	37.4	36.9
2001q3	79.3	71.9	69.5	66.6	64.4	2016q3	54.8	41.0	38.1	35.8	36.8
2001q4	78.7	71.9	68.7	66.7	65.4	2016q4	54.9	42.4	39.3	35.9	36.8
2002q1	78.6	70.3	68.8	65.9	64.9	2017q1	54.0	44.6	41.6	38.0	37.2
2002q2	77.4	68.9	66.8	65.5	64.0	2017q2	53.6	41.8	39.8	37.7	36.4
2002q3	76.9	68.4	66.4	64.3	63.9	2017q3	56.7	43.1	39.1	37.2	36.5
2002q4	76.3	68.4	65.0	63.5	62.8	2017q4	56.4	46.6	39.9	36.5	36.7
2003q1	76.6	67.0	64.6	61.8	61.3	2018q1	57.1	44.7	41.6	35.2	35.0
2003q2	76.0	66.9	63.7	61.0	60.0	2018q2	56.1	43.3	39.6	36.9	34.0
2003q3	75.9	67.7	63.8	60.9	60.4	2018q3	53.9	40.0	36.6	34.1	34.4
2003q4	74.1	65.2	63.3	60.4	59.3	2018q4	53.3	40.5	36.9	35.0	34.6
2004q1	73.3	64.4	61.6	59.7	58.3	2019q1	54.6	40.3	36.2	34.0	33.8
2004q2	72.1	64.4	61.9	58.3	58.0	2019q2	56.1	38.6	35.7	32.5	31.8
2004q3	74.7	65.9	62.6	60.0	57.5	2019q3	55.0	38.9	34.2	32.4	31.5
2004q4	73.2	64.3	61.6	58.7	57.2	2019q4	53.8	40.6	34.6	31.6	30.9
2005q1	74.3	61.2	60.2	57.6	56.4	2020q1	45.2	37.9	35.6	31.1	30.4
2005q2	75.1	63.2	58.1	57.3	56.4	2020q2	31.2	31.6	33.0	31.4	29.1
2005q3	74.2	64.5	59.9	55.1	56.0	2020q3	26.9	26.1	28.2	28.6	28.3
2005q4	72.5	63.5	60.8	56.5	54.2	2020q4	28.6	23.0	25.0	26.9	27.6
2006q1	72.8	63.9	61.5	57.6	55.0	2021q1	29.8	23.4	18.3	21.3	23.4
2006q2	71.2	62.3	58.6	57.6	55.4	2021q2	39.6	22.5	19.1	16.4	20.5
2006q3	70.1	61.7	58.5	54.9	55.1	2021q3	37.1	26.3	19.4	17.4	15.7
2006q4	70.0	63.0	59.9	56.0	54.4	2021q4	32.0	23.3	21.5	16.0	15.5
2007q1	69.9	63.0	60.1	57.0	54.9	2022q1	35.8	22.4	21.3	19.6	15.8
2007q2	69.4	61.4	59.1	55.2	54.4	2022q2	31.9	22.9	17.0	16.3	16.4
2007q3	70.5	62.3	57.8	55.6	54.0	2022q3	31.5	20.8	19.1	15.1	16.0
2007q4	68.5	63.5	58.7	54.6	54.5	2022q4	29.7	17.5	15.8	15.5	13.4
2008q1	69.0	62.8	60.3	56.0	54.2	2023q1	28.7	14.7	13.8	13.2	13.7
2008q2	67.8	59.1	56.4	53.6	52.3	2023q2	27.9	12.5	10.5	10.9	11.0
2008q3	67.8	60.4	57.2	54.1	53.6	2023q3	30.5	11.7	9.3	8.6	9.7
2008q4	67.4	60.9	56.9	53.1	53.2	2023q4	34.0	19.8	11.3	9.1	8.7
2009q1	66.1	60.2	57.8	54.4	52.4	2024q1	32.7	19.1	14.9	9.1	8.0
2009q2	65.0	55.3	54.2	52.6	50.7	2024q2	35.3	17.7	15.6	12.4	8.2
2009q3	65.9	56.7	53.6	51.8	51.7	2024q3	35.3	16.9	14.8	13.8	11.0
2009q4	62.4	57.5	53.1	50.3	50.1	2024q4	33.9	19.1	15.5	13.9	12.7

A5 Software commands

Stata and R software commands are presented in this appendix. They do not represent complete, self-contained scripts that can be run as they are. Rather, they offer code for the sometimes-complicated procedures used to reach the results.

A5.1 Stata commands

The Stata commands show how the LFS data were processed. It is important to pre-process (clean) the LFS data to account for missing data points and avoid generating fitted wages for inappropriate sub-groups. When estimating intensive margin LSE pre-processing matters less, as only employees with valid earnings are retained thanks to using the natural logarithm of hourly pay. However, when estimating the extensive margin LSE one has to very careful to keep only valid respondents. An additional caveat is that in some LFS waves “Does not apply” is correctly coded as -9 while in some waves it is coded as missing. The Stata code needs to account for both circumstances:

```
keep if AGE>=18 & AGE<=64       // Keep working age respondents
tabulate STATR ILODEFR, missing // Cross tabulate to check responses
drop if  STATR== 2       // Drop the self-employed as wage is missing
drop if  STATR== 4       // Drop unpaid family workers
// Drop employees without a recorded wage:
drop if  ILODEFR==1 & (HOURPAY==-9|HOURPAY==.)
// Drop employees with a zero wage:
drop if  ILODEFR==1 & (HOURPAY==0)
// Drop those without any information:
drop if  (STATR==-9|STATR==.) & (ILODEFR==-9|ILODEFR==.)
egen PWT = rowtotal(PWT*) // Combine population weights
```

Estimating intensive margin LSEs in the log-log regressions is relatively straightforward for two main reasons. One is that the logarithm of the hourly wage and the logarithm of work hours will automatically generate missing cases if these are coded using negative values to flag any potential issues. The other is that given the log-log specification, the elasticity is given directly by the estimated coefficient on the logarithm of hourly pay. At the regression stage, one can restrict the estimation to any subset of individuals or time period one chooses using the ‘if’ statement. One can also choose to omit the weighted regression option `[pweight=PWT]`:

```
** CREATE REGRESSION VARIABLES:
ge lnHOURPAY= ln(hourpay)
drop if lnHOURPAY==. // This is unnecessary as they are missing.
ge lnBUSHR  = ln(bushr)
ge lnTTUSHR = ln(ttushr)
ge lnBACTHR = ln(bacthr)
ge lnTTACHR = ln(ttachr)
** CARRY OUT THE REGRESSION(S):
regress lnBUSHR  lnHOURPAY  if Year==2024  [pweight=PWT]
regress lnTTUSHR lnHOURPAY  if Year==2024  [pweight=PWT]
regress lnBACTHR lnHOURPAY  if Year==2024  [pweight=PWT]
regress lnTTACHR lnHOURPAY  if Year==2024  [pweight=PWT]
```

The bivariate log-log regressions above can be estimated as multivariate regression with the addition of other socio-economic covariates. For example, in the regression based on the BUSHR hours measure this could be:

```
regress lnBUSHR lnHOURPAY Female ib(6).HIQUALD /// Line split
                i.AGE_GROUP i.URESMC if Year==2024 [pweight=PWT]
```

where 'ib(6)' uses six as the omitted reference category, 'i' uses the first category as the omitted reference, and the variables are described in the data section. In some instances regressors might need to be omitted, such as omitting Female when estimating separate regressions by SEX or omitting URESMC when estimating separate regression by region.

The regressions above can be corrected for sample-selection using the exact same procedure described below for sample-selection correction in the extensive margin LSE estimates.

Estimating the extensive margin LSEs is more complex and based on equations (5) to (7). First one has to estimate a wage equation using the notation described above, which is used to generate the fitted log wages:

```
regress lnHOURPAY ib(6).HIQUALD i.AGE_GROUP Female /// Line break
                  i.URESMC if Year==2024 [pweight=PWT]
predict  FITlnHOURPAY if Year==2024 // Also fits non-workers
```

The fitted wage is used as the regressor in the Probit regression on whether the respondent is employed or not 'E', and then the IMR for each individual is generated, where RefYear refers to a data sub-section where results are stored for each year's estimates, in this case 2024:

```
probit E FITlnHOURPAY if Year==2024 [pweight=PWT]
* IMR:
predict z, xb  // Linear prediction z
generate IMR_i = normalden(z) / normal(z) if e(sample)==1
drop z
```

With these results, one can compute the estimate the mean extensive margin LSE for the year:

```
summarize IMR_i if e(sample)==1, meanonly
replace meanIMR = r(mean)  if RefYear==2024
replace eLSE= _b[FITlnHOURPAY]*meanIMR if RefYear==2024 // Equat. (4)
```

The regressions above can be corrected for sample-selection by first estimating a Probit employment regression equation (5), where in this case, the variables `ChildOfFemale` and `ChildOfMale` indicate if a female or male respondent has a child aged less than 16 in the household which serve as identifying regressors:

```
probit E ib(6).HIQUALD i.AGE_GROUP ChildOfFemale ChildOfMale ///
         i.URESMC if Year==2024 [pweight=PWT] // Equation (5)
```

An IMR given by equation (6) is then calculated with the purpose of correcting for sample selection in the wage regression:

```
predict z, xb
generate IMR_i = normalden(z) / normal(z) // Equation (6)
drop z
```

This IMR variable is then included as a regressor in the wage regression equation (7) used to generate the log fitted wages:

```
regress lnHOURPAY ib(6).HIQUALD i.AGE_GROUP Female /// Line break
                 i.URESMC IMR_i if Year==2024 [pweight=PWT]
predict  FITlnHOURPAY if Year==2024 // Also fits non-workers
```

Thereafter the fitted log-wage is used in regression equation (4) and the extensive margin LSE estimated computed as described above.

A5.2 R commands

The R commands show how the HLS data were processed. For convenience, we define a simple R function calculating the Inverse Mills Ratio (IMR) from a Probit regression (a binomial GLM with a Probit link).

```
mills <- function(x) {
  probit_lp <- predict(x)
  imr <- dnorm(probit_lp)/pnorm(probit_lp)
  return(imr)}
```

In addition to Base R, we make heavy use of the tidyverse packages of convenience functions (Wickham et al., 2019). Load this and then make a vector of file names pointing to each year of the BHPS and HLS respectively.

```
library(tidyverse) #load tidyverse
#Paste together the directory and file names for the BHPS and HLS
bhps.names <- paste("~/UKDA-6614-tab/tab/bhps/", list.files("~/UKDA-6614-
tab/tab/bhps/"), sep = "")

ukhls.names <- paste("~/UKDA-6614-tab/tab/ukhls/", list.files("~/UKDA-6614-
tab/tab/ukhls/"), sep = "")
```

This results in a vector of the form:

```
## [1] "~/UKDA-6614-tab/tab/ukhls/a_adopt.tab"
## [2] "~/UKDA-6614-tab/tab/ukhls/a_callrec.tab"
## [3] "~/UKDA-6614-tab/tab/ukhls/a_child.tab"
## [4] "~/UKDA-6614-tab/tab/ukhls/a_cohab.tab"
## [5] "~/UKDA-6614-tab/tab/ukhls/a_egoalt.tab"
## [6] "~/UKDA-6614-tab/tab/ukhls/a_empstat.tab"
```

Given the size of the dataset, we only wish to import individual respondent data for each year (labelled “indresp”). The *lapply()* function in R permits us to do this parsimoniously avoiding the need to write a 'for' loop (applying a function to each element).

```
#Iterate over the names of files containing the term "indresp", reading each into R.
bhps <- lapply(bhps.names[grep("indresp", bhps.names)], read_tsv)

ukhls <- lapply(ukhls.names[grep("indresp", ukhls.names)], read_tsv)
```

The result is a list of dataframes (each year being a separate dataframe within the list). This is a convenient form in which to read the data because it permits us to make use of the *lapply()* family of functions to repeat the set of operations on each year (in effect automatically looping over the years).

We then make a list of the variables of interest. Frustratingly, each variable is prefixed by a letter (corresponding to the wave of the survey) and so the names differ for each year ("a_jbhrs" in the first year, "b_jbhrs" in the second etc.) We therefore use a regular expression to extract the variables of interest. Note that these differ slightly between the BHPS and the HLS.

```
#Variables of interest for the UKHLS
v.names <-
"jbhrs$|paygu_dv$|_sex$|agegr13_dv$|gor_dv$|rach16_dv$|marstat_dv$|jshrs$|hiqual_dv$|marstat$|seearngrs_dv$"

#Variables of interest for the BHPS
v.names.bh <-
"jbhrs$|paygu_dv$|sex$|age$|age_dv$|gor_dv$|rach16_dv$|marstat_dv$|jshrs$|seearngrs_dv$"
```

The following code is a little more involved, making use of tidyverse functionality, which enhances readability and parsimony.

1. The first two lines subset each dataframe in the list, extracting only the variables of interest.
2. The second function takes the result of this and adds an additional variable called "year" (with each year being the letter prefix of the variables, such that "a" is the first, "b" the second and so on.
3. For each year, rename all columns (except "year") by removing the first 2 characters (so "a_jbhrs" becomes "jbhrs" etc.)
4. Filter the results to include only individuals aged 18-64.
5. Filter to exclude gender "unknowns" (there are fewer than 5 in each year leading to singular estimates in the Probit regressions).
6. Create a new variable "hours", combining usual weekly hours worked as an employee and usually weekly hours worked as a self-employed individual.

7. Create a new variable “pay”, combining usual monthly pay as an employee and usual monthly pay as a self-employed individual.
8. Calculate hourly pay.
9. Calculate the natural logarithm of hourly pay.
10. Create a binary variable defining employment as an individual with positive usual weekly hours.
11. Create a simplified version of marital status.
12. Convert relevant variables to factors (the equivalent of the ‘i’ prefix in STATA).
13. Take the results and exclude any individuals with negative employment income.

```
ukhls.small <- lapply(ukhls, function(x)
    x[,grep(v.names, names(x))]) %>%
  lapply(function(x)
    mutate(x, year = substring(names(x)[[1]], 1, 1))) %>%
  lapply(function(x) rename_with(x, .fn = substr,
      start = 3, stop = 100L, .cols = -year)) %>%
  lapply(function(x) filter(x, agegr13_dv > 2 & agegr13_dv < 13) %>%
    filter(sex > 0) %>%
    mutate(hours = ifelse(jbhrs > 0, jbhrs,
      ifelse(jshrs > 0, jshrs, 0))) %>%
    mutate(pay = ifelse(paygu_dv > 0, paygu_dv,
      ifelse(seearngrs_dv > 0, seearngrs_dv, 0))) %>%
    mutate(hourpay = 12*pay/(52*hours)) %>%
    mutate(l.hourpay = log(hourpay)) %>%
    mutate(employed = ifelse(hours > 0, 1, 0)) %>%
    mutate(married = ifelse(marstat == 2, 1, 0)) %>%
    mutate(across(c(gor_dv, hiqual_dv, sex, marstat, agegr13_dv),
      as.factor))) %>%
  lapply(function(x) filter(x, employed == 0 | pay > 0))
```

The remainder of the code is a straightforward application of the algebraic results of Section 2.2. First, run a Probit regression of employment against region, gender, marital status, age category and highest qualification. Again, this is looped over all years.

```
probit.1 <- lapply(ukhls.small, function(x)
  glm(employed ~ gor_dv + sex + married + agegr13_dv + hiqual_dv,
    family = binomial(link = "probit"), data = x))
```

Next, apply the “mills” function defined in the beginning to each Probit regression and bind to the source dataset (looping across years).

```
imr.1 <- lapply(probit.1, mills) %>% #Calculates the IMR
  lapply(as_tibble) #For convenience.

#Bind these results into the source data.
ukhls.small <- lapply(seq_along(imr.1), function(x)
    bind_cols(ukhls.small[[x]], imr.1[[x]]))
```

The next stage is to find fitted wages using a Mincer-type equation (but including the saved IMRs as a correction, per the Heckit procedure).

```
#Run a Mincer-type regression for each year on workers
mincer.type <- lapply(ukhls.small, function(x)
  lm(l.hourpay ~ gor_dv + sex + agegr13_dv + hiqual_dv + value,
    data = x[x$employed == 1,]))

#Generate fitted wages for worked & non-workers
ukhls.small <- lapply(seq_along(ukhls.small), function(x)
  predict(mincer.type[[x]], newdata = ukhls.small[[x]]) %>%
  add_column(.data = ukhls.small[[x]], w.hat = .))
```

Finally, run a second Probit regression on employment (using the fitted wages as a regressor) and multiply the coefficients by the (mean) IMR in order to estimate the extensive margin.

```
probit.2 <- lapply(ukhls.small, function(x)
  glm(employed ~ w.hat, data = x,
    family = binomial(link = "probit")))

eLSE <- lapply(seq_along(probit.2), function(x)
  mills(probit.2[[x]])*coef(probit.2[[x]])["w.hat"])
```

Overall values can then be plotted over time, albeit with careful consideration of the caveats outlined in section 3.2

```
#Calculate average extensive labour supply elasticity
average.eLSE <- sapply(eLSE, mean) %>% as_tibble()
average.eLSE$Year <- 2009:2021 #Add years

#Code to recreate Figure 10
ggplot(average.eLSE) +
  geom_line(aes(Year, value), colour = "blue") +
  ylim(0, 0.8) +
  theme_light(base_size = 24) +
  theme(axis.text.x = element_text(angle = 90)) +
```

```
scale_x_continuous(breaks = average.eLSE$Year,
  minor_breaks = NULL)
```