

# Data Science Campus

## Deriving New Measures of Subjective Wellbeing using the UK Household Longitudinal Study

Silvia Lui (ONS, ESCoE) and  
Ana Rincon Aznar (NIESR, ESCoE)

ESCoE Research Webinar  
23 February 2023

# Introduction

# Introduction

- Improving wellbeing and standards of living across the UK is at the core of the policy agenda which has recently focused on reducing disparities
- Wellbeing and productivity are closely related
- While productivity is key in achieving a higher level of prosperity, there are other social outcomes that need to be considered to understand wellbeing
- Measuring wellbeing is challenging largely because it is difficult to quantify wellbeing experience
- The need to better understand how wellbeing varies across UK and across time, and the need to have a better measure to quantify wellbeing to inform policy prompt us to undertake our study

# Aims

- ❖ Draw a consistent picture of changes in subjective wellbeing (SWB) over time using longitudinal survey data
  - Offer new evidence that cannot be revealed using annual snapshots produced by other types of surveys
- ❖ Provide detailed breakdown of wellbeing trends in the UK and UK regions
  - Useful for policymakers interested in regional disparities in quality of life
- ❖ Explore a method to quantify categorical wellbeing responses accounting for impact of individual-specific and regional factors
  - More informative summary measures compared to simple aggregation

# Research questions

- ❖ How do responses to different SWB metrics relate to each other?
- ❖ What are the most significant drivers of SWB?
- ❖ How does the trend of different SWB metrics change over time?
- ❖ How can one obtain weights that account for heterogeneity in individuals' SWB experience in response to a set of observables?

# Data

# UK Household Longitudinal Study (UKHLS)

- The UKHLS, also known as the Understanding Society, is a longitudinal survey of the members of about 40,000 households in the UK
- UKHLS began in 2009-2010, it followed from the British Household Panel Survey (1991-2008)
- Currently has a total of 12 waves (wave 12: 2020 to mid-2022)
- Asks a comprehensive set of questions that covers different aspects of subjective wellbeing (SWB) as compared to other existing UK surveys
- Fieldwork period is 24 months, those interviews that couldn't be conducted were followed up within a few months after

# Our analysis sample

- We use 10 waves of responses to the individual questionnaires (wave 10: 2018 – mid-2020)
- Focus on the ordinary sample members (OSMs) who are followed through the life of the study
- Unbalanced panel: 327,752 observations across all 10 waves
  - ❑ All OSMs who give valid answers to all survey questions we consider
- Balanced panel: 7321 individuals participated in all 10 waves (73210 observations)
  - ❑ Number of observations by regions are reduced significantly – hinder valid inferences
- Use both the unbalanced panel and balanced for national analysis but focus on the unbalanced panel for regional analysis



# Subjective wellbeing (SWB) variables

- 12 General Health Questions (GHQs) asking respondents' experience in different aspects of subjective wellbeing – categorical responses
- We recode the SWB variables: higher values means higher wellbeing
- We reduce 4 categories into 3 for simplicity (combining the middle 2 categories)

Subject wellbeing questions/variables in the UKHLS			
Concentration	Loss of sleep	Playing a useful role	Capable of making decisions
Constantly under strain	Problem overcoming difficulties	Enjoy day-to-day activities	Ability to face problem
Unhappy or depressed	Losing confidence	Believe worthless	General happiness

# UK regions

- Use information about which region respondents locates (variable Government Office Regions (GORs) in UKHLS)
- Conduct UK-wide analysis as well as regional analysis
- GORs is roughly equivalent to the International Territorial Level 1 (ITL1)

UK Government Office Regions (GORs)			
East Midlands	West Midlands	East of England	London
North East	North West	South East	South West
Yorkshire and the Humber	Wales	Scotland	Northern Ireland

# Individual factors – six categories

- Demographic characteristics
- Household condition
- Employment and income
- Caring responsibility
- Physical health condition
- Location

## Respondent-level dataset contains the following information

Sex	Age	Marital status	Level of education	Ethnicity
Migrant	Household size	Housing ownership	Responsibility for under 16	Job
Income (monthly)	A carer or not	Care load	Physical health condition	Live in urban area or not

# Regional factors

- Use ONS data at International Territorial Level 1 (ITL1) break down
  - ITL1 is roughly equivalent to Government Office Region
- Consider two regional factors:
  - Gross Disposable Household Income (GDHI) per head
  - Gross Valued Added (GVA) per hour
- Merge regional factors to the respondent-level dataset according to the year in which an interview was conducted

# Interview years

Wave	Interview years		
1	2009	2010	2011
2	2010	2011	2012
3	2011	2012	2013
4	2012	2013	2014
5	2013	2014	2015
6	2014	2015	2016
7	2015	2016	2017
8	2016	2017	2018
9	2017	2018	2019
10	2018	2019	2020

Most interviews are conducted in the first 24 months. Majority of them in the first 12 months.

# Methodology

# Polychoric correlation

- We use polychoric correlation to examine pairwise relationships (See Olsson (1979))
- Based on contingency table, it measures the level of agreement between two normally distributed continuous latent variables with observed ordinal values
- The coefficient has values between 0 (no correlation) and  $\pm 1$  (perfect correlation)
- Results inform us about how the different metrics relate to each other unconditionally
- Evidence of whether a possible (unobserved) common factor that underlies SWB exists
  - Expect to observe high correlations if it is the case

# Ordered Probit Model (oprobit)

We estimate relationships between SWB experiences and a set of observable factors.

Respondent  $i$ 's response to SWB question  $k$  in wave  $t$ ,  $y_{it}^{(k)}$ , is triggered by a continuous latent variable  $y_{it}^{(k)*}$  via the observation rule

$$y_{it}^{(k)} = j \quad \text{if } \mu_{ij} < y_{it}^{(k)*} < \mu_{i(j+1)}; \quad i = 1, \dots, N_t \text{ and } j = 0, 1, 2$$

and  $\mu_{i0} = -\infty$ ,  $\mu_{ij} \leq \mu_{i(j+1)}$ , and  $\mu_{i3} = \infty$ .  $N_t$  denotes # of respondents in wave  $t$ .

$y_{it}^{(k)*}$  is related to individual factors  $x_{it}$  and regional factors  $z_{rt}$  ( $r$  denotes region,  $r = 1, 2, \dots, 12$ ) in the following manner

$$y_{it}^{(k)*} = \beta_i' x_{it} + \delta_i' z_{rt} + \varepsilon_{it} \quad ; \varepsilon_{it} \sim N(0, 1)$$



# Ordered Probit Model (oprobit)

Conditional probabilities of observing  $y_{it}^{(k)} = j$  in wave  $t$  is given by

$$P_{jit}^{(k)} = P\left(y_{it}^{(k)} = j \mid x_{it}, z_{rt}\right) = \Phi\left(\mu_{i(j+1)} - \beta_i' x_{it} - \delta_i' z_{rt}\right) - \Phi\left(\mu_{ij} - \beta_i' x_{it} - \delta_i' z_{rt}\right)$$

Unknown parameters  $\mu_{ij}$ 's,  $\beta_i$ 's and  $\delta_i$ 's can be estimated using maximum likelihood.

The likelihood function is given by

$$l_i(\boldsymbol{\mu}, \boldsymbol{\beta}, \boldsymbol{\delta}) = \sum_{j=0}^2 I\left[y_{it}^{(k)} = j\right] \ln\left[\Phi\left(\mu_{i(j+1)} - \beta_i' x_{it} - \delta_i' z_{rt}\right) - \Phi\left(\mu_{ij} - \beta_i' x_{it} - \delta_i' z_{rt}\right)\right]$$

# Ordered Probit Model (oprobit)

- Estimate pooled national and regional oprobit models with only the individual factors  $x_{it}$  to examine the drivers of SWB
- Estimate respondent-level oprobit models with individual factors  $x_{it}$  and regional factors  $z_{rt}$ 
  - Obtain weights to adjust summary measures of SWB to account for heterogeneity

# Summary measures

We quantify qualitative responses on reported SWB in the form of Balance Statistics. Balance statistics are widely used to summarise economic sentiment (survey-based).

The proportions of positive, neutral (no change) and negative responses are

$$G_t^{(k)} = \frac{\sum_{i=1}^{N_t} (y_{it}^{(k)} = 2)}{N_t} \times 100; \quad E_t^{(k)} = \frac{\sum_{i=1}^{N_t} (y_{it}^{(k)} = 1)}{N_t} \times 100; \quad M_t^{(k)} = \frac{\sum_{i=1}^{N_t} (y_{it}^{(k)} = 0)}{N_t} \times 100$$

Where  $(y_{it}^{(k)} = j)$  is one count of response  $j$  ( $j = 0, 1, 2$  denote negative, neutral/no change and positive responses) given by respondent  $i$  in  $t$ . The balance statistic of SWB metric  $k$  is given by

$$B_t^{(k)} = G_t^{(k)} - M_t^{(k)}$$

# Adjustment for heterogeneity

- We suspect heterogeneity among respondents' SWB experiences due to different individual characteristics and differences in regional economic conditions
- Aggregation method should consider the relevance of these differences to ensure the resulting summary measures are informative if heterogeneity exists
- We demonstrate a way to create summary measures of SWB:
  - Give more weights to respondents whose SWB outcomes are more closely associated with socio-economic factors and regional economic conditions
  - Resulting aggregates should better reflect the direct or indirect influence of policy via the correlations with the observables

# Adjusted summary measures

- Estimate oprobit model (include individual and regional factors) for each  $i$  in the balanced panel for each SWB outcomes,  $k$ ; use pseudo  $R^2$  to construct weights
- These weights take into account the collective impact of  $x_{it}$  and  $z_{rt}$  on  $y_{it}^{(k)*}$  on  $i$
- Weighted balance statistics of SWB metric  $k$  in wave  $t$ ,  $B_t^{Adj,(k)}$  is

$$B_t^{Adj,(k)} = \left[ \sum_{i=1}^N w_i^{(k)} (y_{it}^{(k)} = 2) \times 100 \right] - \left[ \sum_{i=1}^N w_i^{(k)} (y_{it}^{(k)} = 0) \times 100 \right]$$

$N$ : # of respondents in the balanced panel

$w_i^{(k)}$ : weight assigned to respondent  $i$ . It is the rescaled pseudo  $R^2$  from respondent  $i$ 's

oprobit such that  $\sum_{i=1}^N w_i^{(k)} = 1$

# Adjustment – an alternative method

- Assign more weights to respondents whose SWB experiences are less correlated with the socio-economic factors and regional economic conditions
  - i.e., Those whose oprobit models gives a lower pseudo  $R^2$
- Resulting aggregates are less correlated with observables that could be influenced (directly or indirectly) by current policy
  - More informative about the level of wellbeing which current policy do not have an impact on
- Both ways of aggregation can be considered

# Composite index of SWB

- We summarise the sentiments of the 12 SWB metrics to reflect overall wellbeing experience using a composite index
- We assign equal weights to each of the SWB metrics
  - No apparently good reason to believe one SWB metric is more important than the other in determining overall wellbeing experience (Laplace's Principle of Insufficient Reason)
- Composite index: rescale the series of average standardised balance statistics of the SWB metrics to an index values with mean equals to 100 and standard deviation equals 10
- Compute composite index with and without adjustment for heterogeneity

# UK results

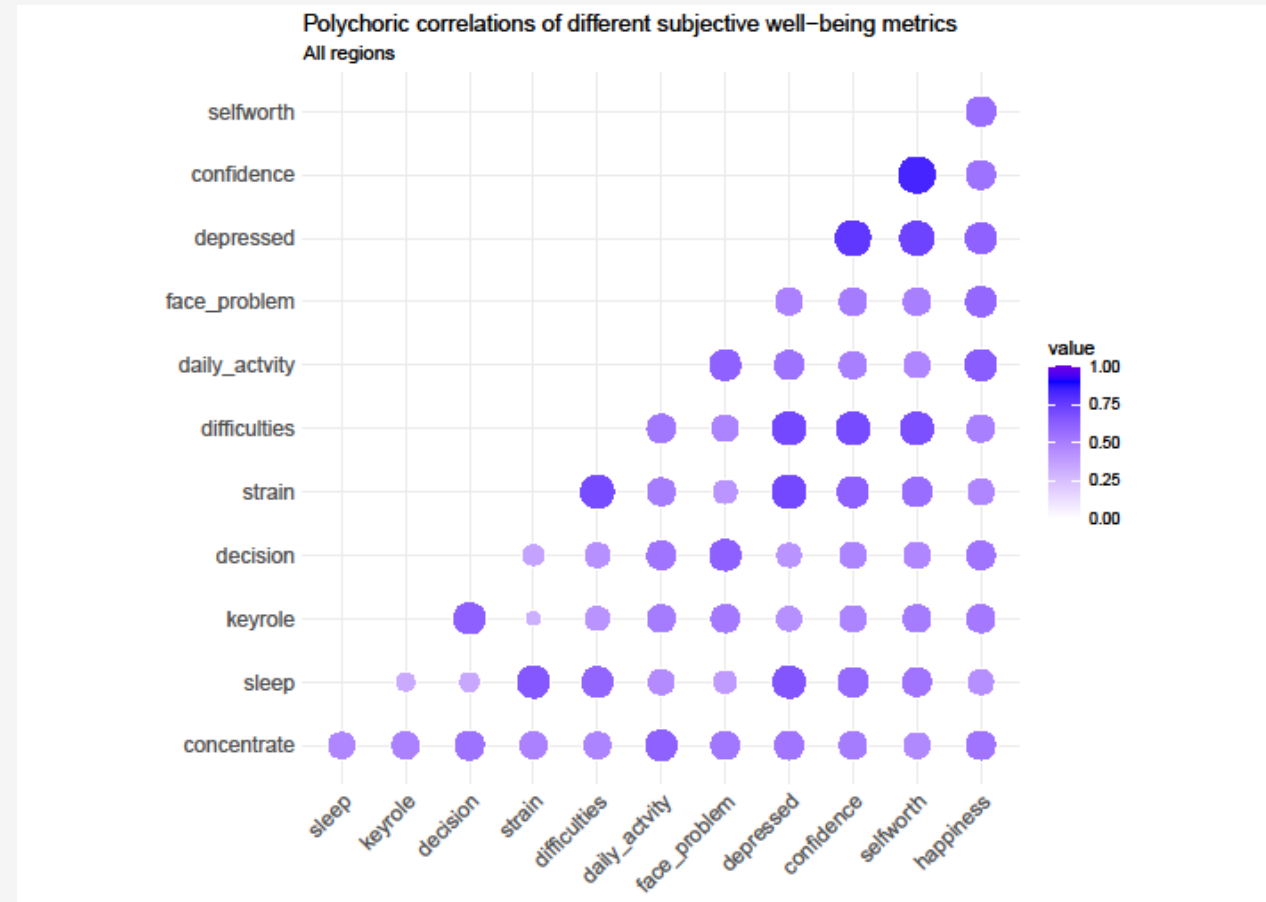


# Correlation analysis do not show strong evidence that SWB experiences are driven by an unobserved common factors

- Correlations range between 0.3 and 0.8 with an average of 0.54
- General happiness has an average correlation of 0.54 (with all other SWBs)

## Implications:

- ❖ Any surveys that aims to use a consensus of happiness as a measure of overall wellbeing could risk missing important aspects
- ❖ Different aspects of SWB could be driven by different factors



# Drivers of subjective wellbeing

- We look at the factors that are linked to the probability of observing a particular outcome
- We use marginal effects to understand the magnitude of the effects emerging from the oprobit regressions
  - E.g., being divorced/separated reduces (increase) the likelihood of reporting highest levels of wellbeing by x%

# Drivers of subjective wellbeing – summary of findings (I)

- **Gender:** we find that men are more likely to score high in all 12 aspects of wellbeing
- **Marital status:** Being married or in a partnership increases the likelihood of scoring higher in most aspects of wellbeing
- **Education:** Having a degree is associated with a higher probability of scoring lower generally across all aspects of wellbeing, except for the ability to play a useful role
- **Ethnicity:** People of black ethnicity and those of Asian origin tend to score higher in the majority of wellbeing measures
- **Migrant:** Being a migrant is associated with a higher probability of scoring higher in wellbeing (but the effect is small)

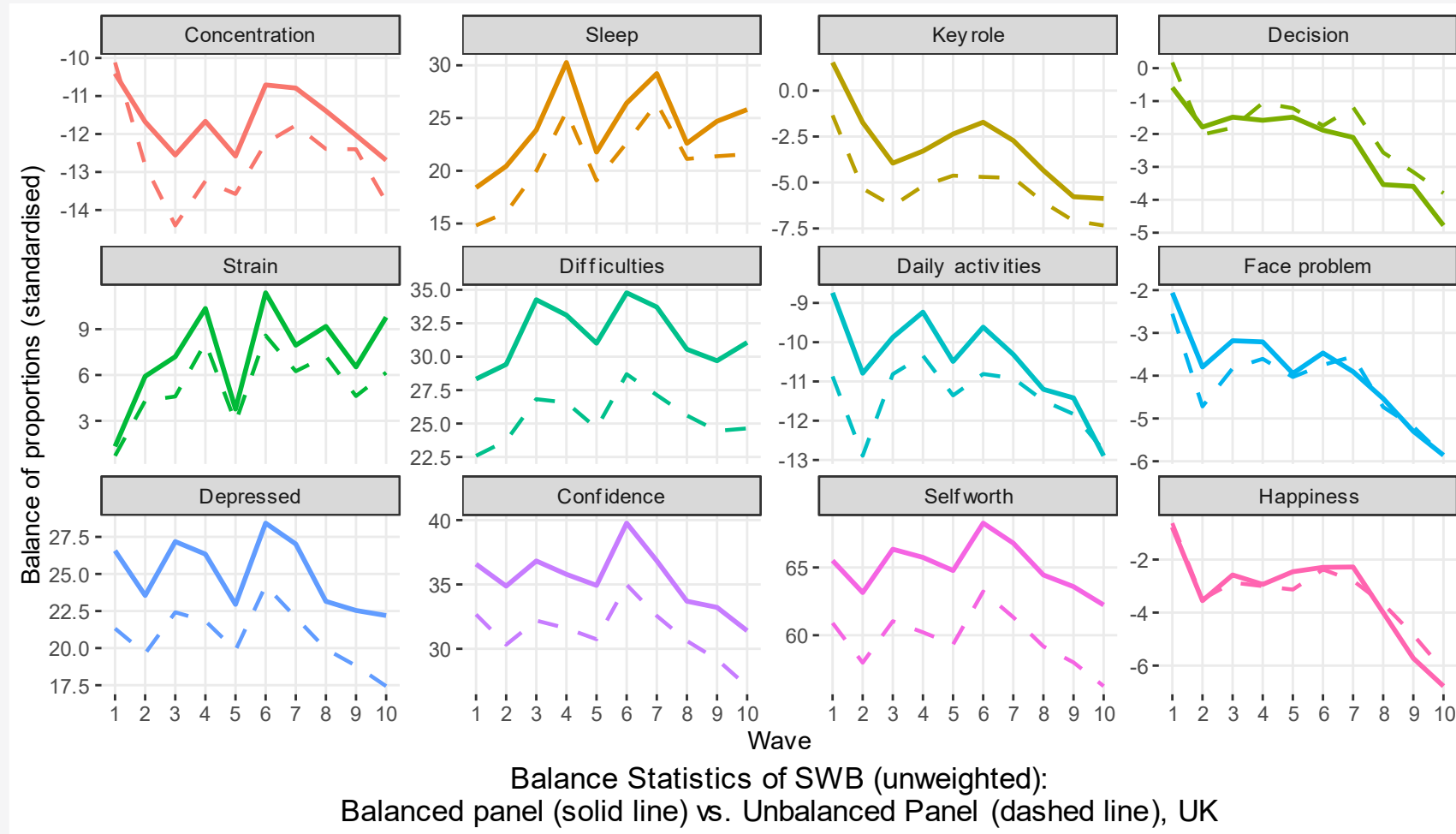
# Drivers of subjective wellbeing – summary of findings (II)

- **Age:** Older individuals are more likely to report better concentration, more able to overcome difficulties, more positive feelings of confidence and self-worth; but they are less likely to feel they can play a useful role, less capable of making decisions, and to enjoy day to day activities. But these effects are small.
- **Job:** Having a job is associated with scoring lower in many aspects of wellbeing; except for the ability of playing a useful role, and feelings of confidence and self-worth
- **Income:** Higher income is mostly associated with higher levels of wellbeing in several aspects, but the size of the effects can be small. Higher income is also associated with feelings of being constantly under strain

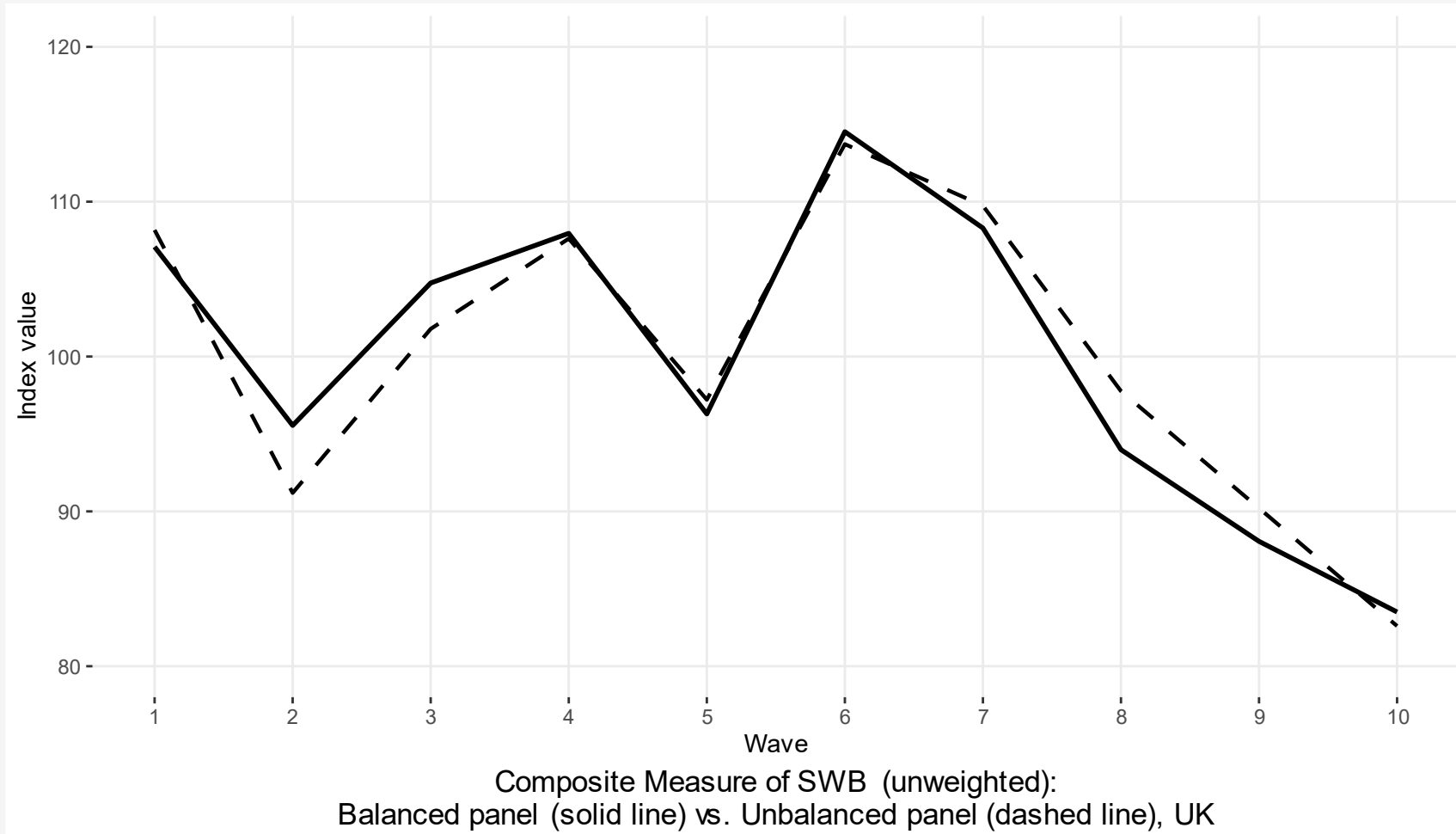
# Drivers of subjective wellbeing – main findings (III)

- **House ownership:** respondents who own their house outright are more likely to score higher in wellbeing
- **Urban or rural:** respondents who live in urban areas are more likely to report worse wellbeing, compared to those living in rural areas, but the differences are quite small
- **Physical health:** The relationship between health and wellbeing is the strongest among the drivers considered
  - Those with excellent health are up to 30% more likely to report the highest SWB

# Trends of (unadjusted) balance statistics show decline in most metrics of subjective wellbeing since wave 6 (since 2014)



# Overall level of subjective wellbeing has declined in the UK since wave 6 (year 2014) as shown by the (unadjusted) composite index



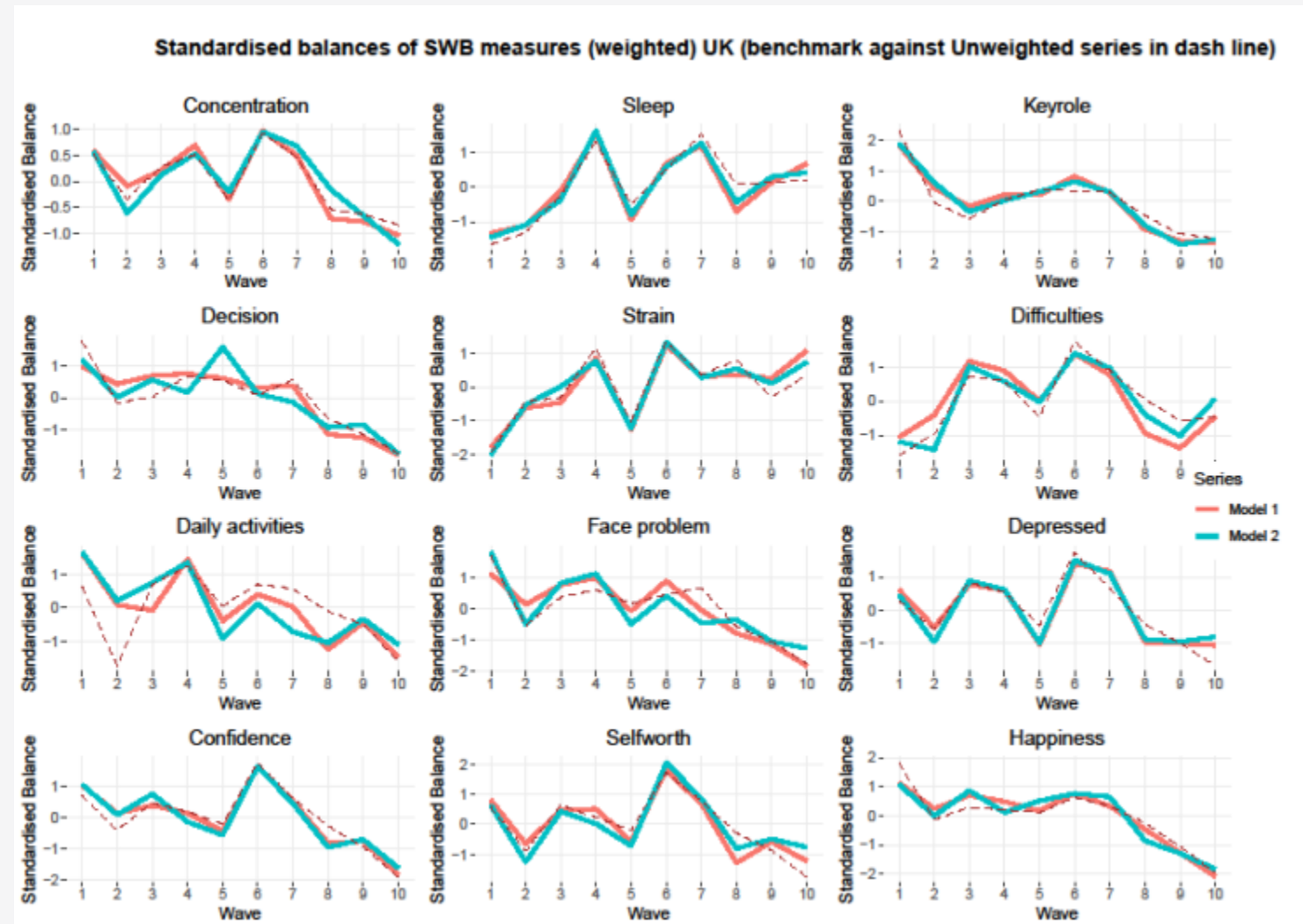
# Adjustment for heterogeneity

- Estimate 7321 respondent-level oprobit models using data from the balanced panel to obtain pseudo  $R^S$  values to construct weights
- We use small-scale oprobit models with a few explanatory variables for adjustment
  - This decision is based on our findings from SWB drivers, findings in the literature, and the concern of result accuracy – there are only 10 time observations for each respondent
- Two parsimonious representations of oprobit model with less regressors:
  - Model 1: the 2 regional factors and personal income
  - Model 2: the 2 regional factors and health
- Larger scale respondent-level oprobit models can be used if there are more time observations e.g. extend the dataset to include those also participated in BHPS

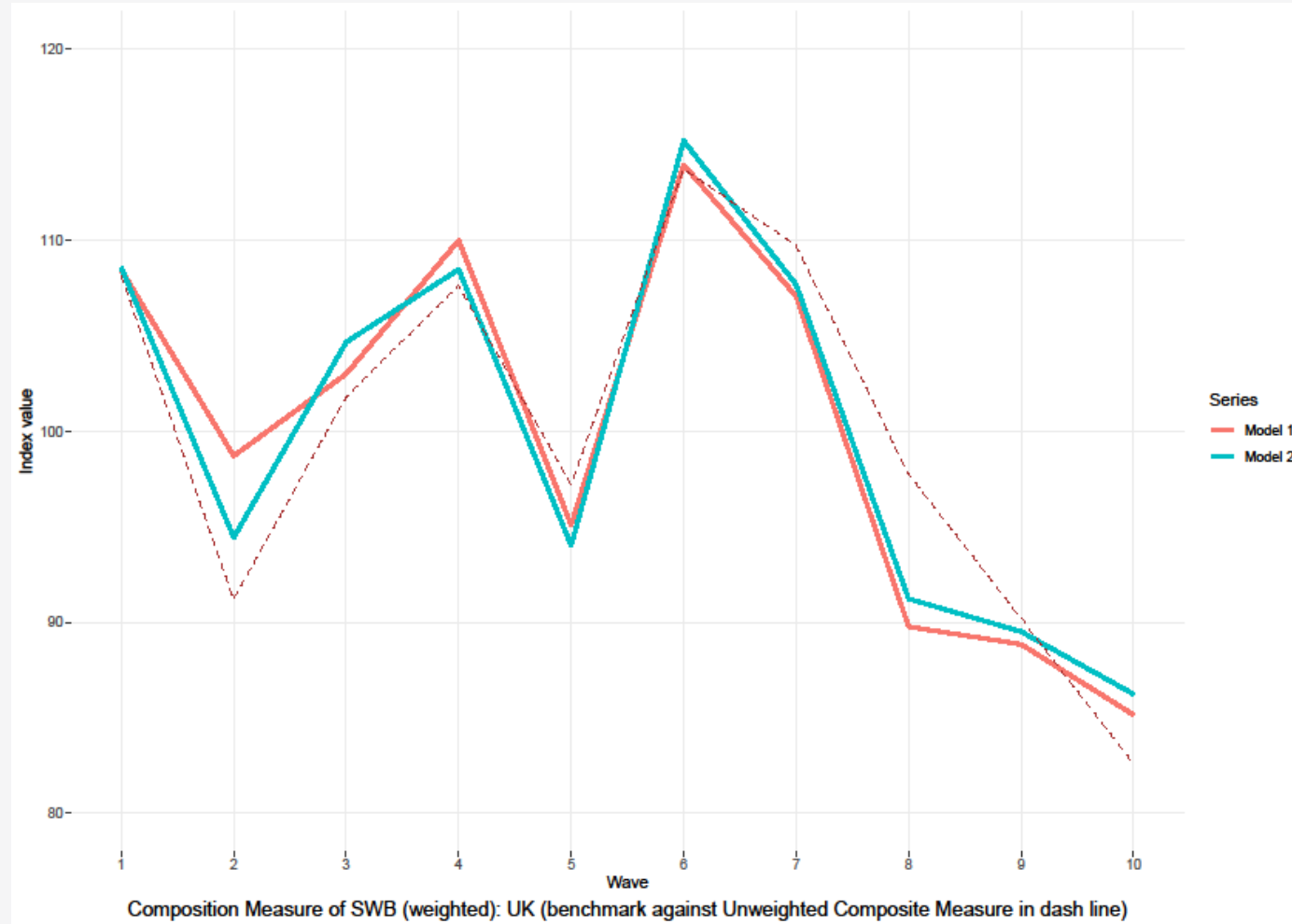


Similar decreasing trend in the adjusted balance statistics of SWB since 2014, more variation before then

- The differences between the unadjusted and adjusted series are not large but there are some year on year variations



# Overall trend of adjusted measure is broadly similar but some of the annual changes are larger



# UK regional results

## (Main findings)

# Polychoric correlation analysis

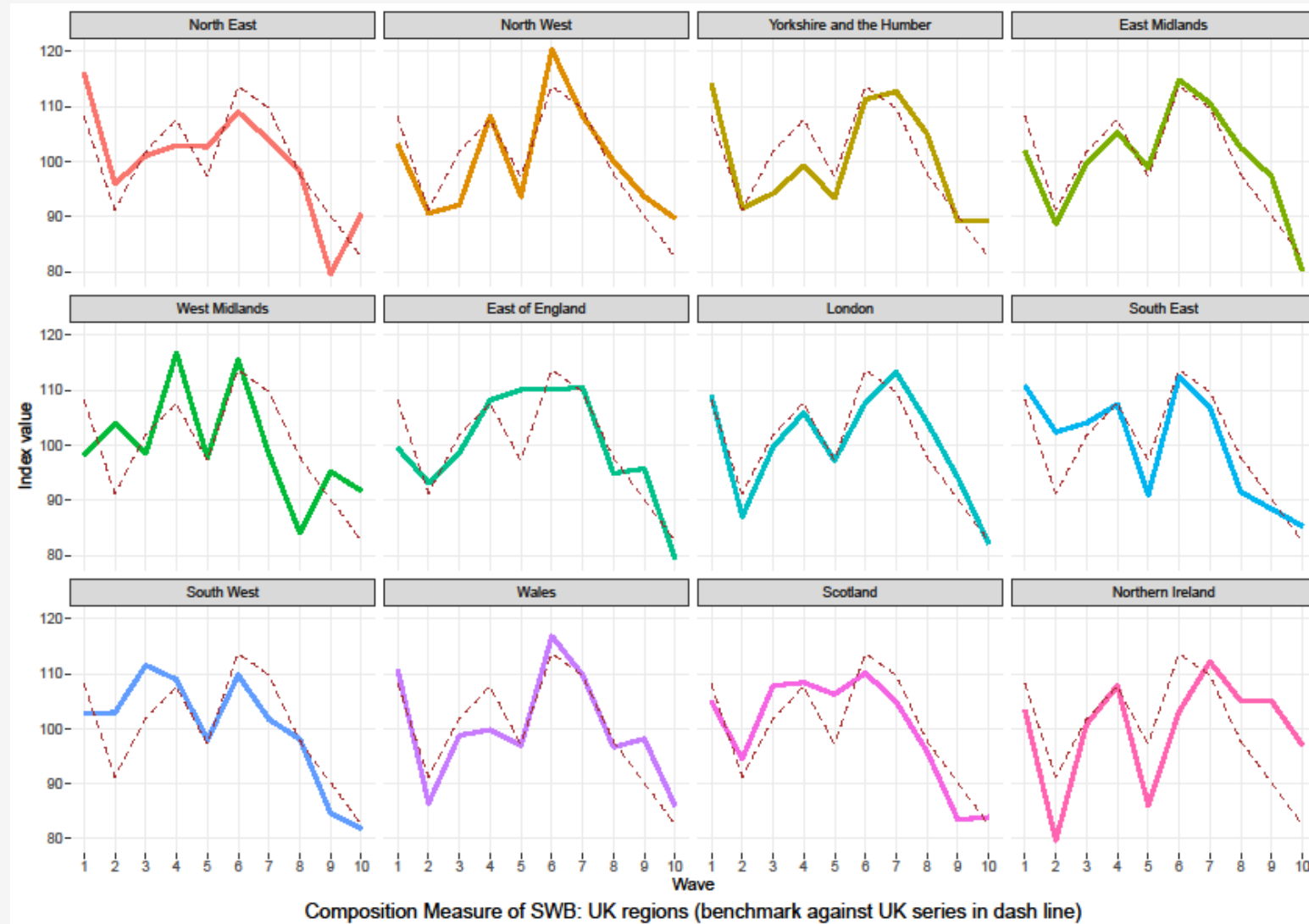
- Findings from unconditional correlations analysis at the regional level show a consistent picture with the national level results
- Pairwise correlations between different SWB metrics do not vary much across regions
- Only in a few cases correlation coefficients have values  $>0.7$ 
  - Further confirms that the evidence of a common factors underlying the twelve different metric is not strong

# Drivers of subjective wellbeing: key highlights

- Being divorced or separated is associated with lower probability of positive wellbeing outcomes in regions such as North East, North West and Northern Ireland
- Divorced or separated individuals in London do not necessarily report lower level of wellbeing, in contrast to what is observed in the majority of other regions
- Respondents with a degree are more likely to feel they can play a useful role in most regions, but the link to other aspects of wellbeing is less clear
- Respondents of black ethnicity are more likely to score high in wellbeing in London and the East of England, but not in other regions
- We observe a stronger relationship between income and the likelihood of higher wellbeing in London and the West Midlands

# Regional composite measures of SWB follows similar trend as the UK as a whole

- We observe some regional differences in the pattern of the balance statistics of the 12 wellbeing metrics
- But these differences do not translate into major regional differences when we aggregate the twelve SWB measures.



# Conclusion

# Conclusion

- We explore the use of individual data on self-reported SWB experience to construct national and sub-national SWB summary measures for the UK
- We investigate the drivers of SWB focusing on a set of individual-specific factors
- We illustrate how one can obtain weights to produce aggregate SWB measures that accounts for heterogeneity
- Findings from polychoric correlation analysis confirm that wellbeing is a complex concept and cannot be easily represented by a common factor
- Relationships between different SWB variables appear similar across regions, but there are some regional variations in the extent to which some individual socio-demographic factors impact on wellbeing



# Conclusion

- We find overall decreasing levels of wellbeing in the UK since 2014 at the national level and in most regions
- We do not observe a unique pattern of changes across all aspects of SWB but there are similarities; e.g. different timing when the decline started
- Our analysis provides a range of results that looks at differences in the aspects of SWB and in different regions; as well as showing changes over time
- Our findings can be informative for policymakers interested in understanding regional disparities in quality of life

# Caveats and Future research

# Caveats

- We have not examined the reasons that cause changes in the trend of SWB
- Our weighted analysis assumes all individuals are being affected by the same set of individual and regional factors – this is a testable assumption and may not be true across all respondents
- Our weighting scheme assigns those whose SWB that are less correlated or orthogonal to observable factors with small or zero weights
- We use parsimonious specifications to construct weights to adjust for heterogeneity, but the models can be enriched if we have longer respondent-level time series

# Future research

- Future research could consider extending the data to include those from the British Household Panel Survey (BHPS) to obtain longer respondent-level time series
- We illustrate one way to obtain weights, but there are also other ways of aggregation
- Future work could consider e.g., giving more weights to the respondents whose wellbeing is less correlated with observables (i.e., more correlated with unobservable factors)
  - This allows one to capture the part of wellbeing that is more random and not so closely related to socio-economic and regional factors
- Future work could also explore changes in subjective wellbeing during the pandemic to understand whether there are any changes in their drivers

# Disclaimers

\* This work was produced using statistical data from ONS. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates.

\* The views expressed are those of the authors and may not reflect the views of the Office for National Statistics or the wider UK Government.

# Thank you