

Data Explained

Using the Earnings and Employees Study 2011 to estimate wage differentials

Authors: Andrew Gillespie, Duncan McVicar, Neil Rowland, Ian Shuttleworth

Queen's University Belfast

Date: November 2023

We used the Earnings and Employees Study (EES) 2011 to estimate Catholic-Protestant wage differentials in Northern Ireland (NI) as of 2011. This dataset has a larger sample size and likely more accurate wage data compared to alternatives such as the Labour Force Survey (LFS), and therefore was particularly valuable for this research.

There has been a long-running concern around labour market differentials between the two main communities in NI, most notably in unemployment. Despite an extensive body of research documenting and seeking to explain such differentials, there is little existing evidence specifically on wage differentials.

An exception to this is the NI [Executive Religion Reports](#), which until 2017 presented annual estimates of the unadjusted median wage differential between Catholics and Protestants using

LFS data. The most recent estimate (for 2017) suggested no median wage differential. It is unclear, however, how much we can learn from this, or previous LFS Religion Report estimates, given their imprecision and volatility over time. This likely reflects the weaknesses of the LFS data on which they are based. Nor do these reports account for differences in relevant measurable characteristics between Catholics and Protestants that may help to explain or even obscure wage differentials. The bottom line is that there remains very little historical and current evidence on this potentially crucial aspect of labour market inequality between the two main communities in NI.

This research project re-estimates Catholic-Protestant wage differentials in NI using the newly-available EES data for 2011. We begin by comparing wage distributions for Catholic and Protestant employees, noting any apparent differences. We then present information on differences in the measured characteristics of the two groups that are themselves likely to be correlated with wages, such as education levels, age, and geography. We also present the characteristics of the jobs that they hold, such as contract type and sector. We then estimate wage regressions, first unadjusted, then adjusted, for these measured individual and job characteristics. Finally, Oaxaca-Blinder decompositions are used to distinguish between the part of any wage differential that is explained by differences in measured individual and job characteristics, and the part which remains unexplained by such differences.

Initial research questions

The study addressed several key research questions:

1. Was there a wage differential between Catholic and Protestant employees as of 2011, and if so, what was it?
2. Was there a wage differential between male Catholic and Protestant employees as of 2011, and if so, what was it?
3. Was there a wage differential between female Catholic and Protestant employees as of 2011, and if so, what was it?
4. How much of any such differentials can be explained by differences in observable characteristics such as education/age/geography (and other relevant differences) as of 2011, and how much cannot be explained by such differences?

Dataset and key variables

The variables available in the EES consist of employee information from the Annual Survey of Hours and Earnings (ASHE) 2011, linked to the personal characteristics of the Census 2011 and capital value data from the Land and Property Services.

In this project we used the following ASHE variables: Sex, Age_topcoded, Ft (full-time/part-time marker), Pt (Permanent/temporary marker), Adr (Adult rate marker), Lop (loss of pay marker), Colag (Collective agreement), Sic07_2_cat (Industry), Occ10_1_cat (Occupation), He (Hourly earnings), Hexo (Hourly earnings excluding overtime), Hpay (Stated hourly rate of pay), Gpay (Gross pay - weekly), Gpox (Gross pay excluding overtime - weekly), Bhr (Basic paid hours), Ovhrs (Paid overtime hours), Thrs (Total paid hours), Pubpriv (Public Private 1997-2013), LGD2014Work (Work NI Local government districts (2014)).

Census variables used: MARSTATP1_AGG (Marital and Same Sex Civil Partnership Status), STUDENTP1 (Student Indicator), COBP1_5cat (Census 2011 - Country of Birth - Five Cat), RELBTP1_AGG (Current Religion), RELBTBUIP1_AGG (Religion - Brought Up In), DISABILITYP1 (Disability), HEACONP1 (Health Conditions Indicator), CARERP1 (Care), HLQP1_AGG (Highest Level of Qualifications (inclusive of those who stated UK equivalents), DPCH1_AGG (Household Dependent Children).

Land and Property Services variables used: CV_NON_EX5 (Capital Value - 2005 Based).

Data limitations encountered

Sample size in the EES is considerably larger than that used for the NI Executive Religion Reports (by a factor of around 17). However, it remains small compared to UK-wide sample sizes for large-scale surveys such as the LFS or Understanding Society, and small relative to the NI population in employment. This limits the precision with which wage differentials can be estimated, although likely not to the same extent as in the case of the NI Executive Religion Reports. It also limits the extent to which the overall EES sample can be split to examine evidence for wage differentials in demographic subgroups.

The variables for the Local Government District (LGD) in which individuals live and work (LGD2014HOME and LGD2014WORK, respectively) have numerous missing values. This likely reflects the voluntary nature of the question relating to LGD2014HOME in the questionnaire and problems with LGD2014WORK associated with large employers with multiple sites. This limits the extent to which we are able to account for spatial differences (which may be confounding) in our analysis.

Other data limitations encountered on the project include:

- The qualification variable - HLQP1_AGG - has limited granularity. In particular, the highest level identified is level four, so we cannot distinguish e.g. degrees and above (or equivalent) from any post-secondary. Again, this limits the extent to which we are able to account for education-related differences (which may be confounding) in our analysis.
- There is no young dependent child variable, despite this information being in the Census. The current variable - DPCH1_AGG - does not distinguish dependent children by age. Labour force participation, particularly among women, is strongly associated with the presence of dependent children under school age.
- There is no anonymised firm marker in the EES, despite this information being available in ASHE. Such a marker would allow analysis to control for differences between firms, which in our case may be confounding.
- He (Hourly earnings) and Hexo (Hourly earnings excluding overtime) were rescaled (divided by 100) to fix a formatting error in these variables. These variables could be properly scaled in the next or current EES release.

Necessary modifications to initial research questions or research design

There were no necessary modifications to the initial research questions or research design.

Necessary modifications to the data

It was not necessary to make any major modifications to the data. We constructed binary dummy variables from categorical variables where these were required for the analysis (including 'missing' categories for LGD2014HOME and LGD2014WORK).

Recommendations to data owners

- Consider extending the current linkage to also link in ASHE 2010 and 2012 (and perhaps additional years). This would increase sample size by adding individuals not in employment in the reference week in 2011, but otherwise in the sampling frame and in employment in one or more years either side of 2011. This would have the bonus of adding a longitudinal dimension to the EES (i.e. employees could be tracked over time between jobs), which would substantially increase the potential of the EES to inform policy. This is the approach being taken by the sister dataset for England and Wales: the Wage and Employment Dynamics (WED) dataset.
- Increase granularity for the highest qualification variable.
- Include variable for dependent children aged 0-4 in household and/or split current dependent children variable by age.
- Include an anonymised firm marker in the EES. Enterprise reference number is already in the ASHE, and this will allow researchers to control for firm fixed effects.
- Take steps to minimise measurement error in LGD WORK information in the case of large multi-site employers.
- Greater cooperation between the NISRA team responsible for EES and the WED team would likely be mutually beneficial for both data providers and data users.

Additional data which would help to further develop the research

Although the changes suggested above would help to further develop this research, at this point they are intended rather as suggestions to increase the potential usefulness of the EES data for further research and policy purposes. This is in particular with the mooted EES 2021 in mind.

Please include code files used in your analysis

Please contact NISRA Research Support Unit: rsu@nisra.gov.uk for code files.

Feedback on metadata, synthetic data and other documentation provided

There are several minor improvements which can be made to the documentation of this project:

- Labels for SIC industry codes and SOC occupation classifications (1 or 2 digit) could be included
- Labels for LGD2014WORK and LGD2014HOME codes could be included (e.g. instead of / as well as N09000001 write Antrim and Newtownabbey)
- Note that (and clarify the reasons why) there are a high number of missing observations for LGD2014HOME and LGD2014WORK in the documentation
- Combine documentation into a single folder.

Any other feedback

From 2017 a question on No Guaranteed Hours Contracts (NGHC) has been included in the ASHE questionnaire. This, coupled with Census information from the EES-link, potentially offers researchers and policymakers valuable insights into the prevalence, distribution, antecedents, and consequences of Zero Hours Contract employment in NI. We recommend the inclusion of a NGHC variable in the EES 2021.

Acknowledgements

Administrative Data Research Northern Ireland (ADR NI) takes privacy protection very seriously. All information that directly identifies individuals will be removed from the datasets by trusted third parties, before researchers get to see it. All researchers are trained and accredited to use sensitive data safely and ethically, they will only access the data via a secure environment, and all of their findings will be vetted to ensure they adhere to the strictest confidentiality standards. The help provided by the staff of Administrative Data Research Centre Northern Ireland (ADRC NI) and the Northern Ireland Statistics and Research Agency (NISRA) Research Support Unit is acknowledged. ADR NI is funded by the Economic and Research Council (ESRC). The authors alone are responsible for the interpretation of the data and any views or opinions presented are solely those of the author and do not necessarily represent those of the ADR NI. NISRA's data has been supplied for the sole purpose of this project.

About ADR UK

ADR UK (Administrative Data Research UK) is a partnership transforming the way researchers access the UK's wealth of public sector data, to enable better informed policy decisions that improve people's lives. By linking together data held by different parts of government and facilitating safe and secure access for accredited researchers to these newly joined-up and de-identified data sets, ADR UK is creating a sustainable body of knowledge about how our society and economy function – tailored to give decision makers the answers they need to solve important policy questions. ADR UK is funded by the Economic and Social Research Council (ESRC), part of UK Research and Innovation.