Replication of the paper *From policy to practice: The role of national policy instruments and social barriers in UK energy efficiency adoption in households*

Article Summary

The core potential additions of this article to the scientific literature is first the open publication of a new dataset of energy policy survey data from the United Kingdom and second an initial analysis of the survey results. The sector focused on is home energy efficiency upgrades, such as installation of insulation in the loft/walls, upgrading hot water boilers, installation of solar PV systems, and much more. The first aspect, the new dataset, is best summed up simply by a quote from section 3.1 *"To the best of our knowledge, this survey is the most recent one performed in order*`import warnings` *to try to understand attitudes and behaviors related to energy efficiency adoption and energy efficiency policy in the UK in the post pandemic context."* In the context of climate change adaptation and mitigation, the goal of this study is clearly highly relevant and could be helpful for policymakers to design more effective energy efficiency policies in the UK and perhaps further afield. The survey results therefore certainly merit publishing, as well as, in principle, analysis of the results presented alongside the questionnaire.

The brief development of the questionnaire in the text did not do justice to the complexity of survey data and what I presume is a rich literature of survey methods (Writing Survey Questions). The article did include information about the sample: summary statistics in the data section, the appendix (S1) provides a comparison with You Gov data of subgroups, and a correlation matrix of the key independent and dependent variables was presented in the annex Figure 1. However, they were not put into a useful context. There was no review of these control variables over time and no investigation into some of the core subgroups (poor renters vs. rich owners for example) that could complicate interpretations. There was also no consistent external validation of the raw survey results. For example, the UK government publishes quarterly public opinion polling on attitudes of households regarding green policy awareness and includes similar questions regarding awareness (DESNZ Spring Public Attitudes Tracker). This would have been very useful: A 2014 report from the DECC Public Attitudes Tracker shows that some 70% of households had either already implemented loft insulation or were in the process, a strong contrast to the 30% of households in this sample (Gilchrist & Craig).

The analysis of the questionnaire involves using logit and clogit regression models to evaluate three hypotheses laid out in the literature review:

1. *(H1) The implementation of public policies that facilitates the access to information, the adoption and the reduction of upfront costs, increases the probability of adopting EE measures for heating at the household level.*
2. *(H2) Policies oriented to tackle informational and administrative barriers are the most relevant, i.e. soft instruments oriented to the provision of information about schemes and savings and to easing administrative barriers increase the probability of adopting EE measures for heating at the household level to a greater extent than other policies.*
3. *(H3) Effective strategies to phttps://github.com/thorerismann/replication_taskromote energy efficiency must adopt a holistic approach, addressing multiple barriers and leveraging diverse motivators.*

To evaluate H1 and H2 the authors conduct a series of regressions using a *dichotomous logit model (logit)*, a *cloglog model* and a *Poisson model* to identify the marginal probability of a household adopting any energy efficiency measure based on the independent variables (awareness of policies) and a series of control characteristics. The controls are as expected: region, gender, house type (renter or owner), number of people in the household, physical size of the house, economic status, and more. The independent variable is a binary variable that equals 1 if a household has adopted any of the energy efficiency options, and 0 otherwise. The results, along with sub group analysis of expensive options versus cheaper options, and policy combinations, are presented in long regression tables with many coefficients and significance asterisks. The same is done with policy combinations to address H3 with the regression tables reported in the annex and some highlights reported in the main text.

The robustness checks, jackknifing and multivariate interpolation, do not in my mind alleviate potential representativity issues, especially among subgroups, that can be found in survey data. Resampling from the survey data for jackknifing or filling with a multivariate distribution d*o*es not feel adequate to address the criticism that the results are due to the survey data itself and may not represent adequately the entire population, especially given the lack of external validation discussed above. Some of the purported effects highlighted in the text, in Figure 6 a. for example move the probability of adoption of any energy efficiency of the marginal effect of the knowledge of tax credits and the information about potential savings, from 0.01 to 0.02. It is hard to say what to do with such results, given the number of p-values calculated and reported and the magnitude of such values. Restricting more appropriately for the number of variables in the regressions might eliminate much of the subgroup significance. The lack of specificity in the stated hypotheses leaves one open to interpretation as to whether they are met and what they really mean in this context. Similarly, pseudo R2 values for the model hover around the meaningful cutoff of 0.2, making one wonder about the replicability of this study were it to be conducted again using a new survey sample.

In conclusion, this study has an important contribution to the scientific literature: meaningful questions about energy efficiency attitudes and adoption conducted by a well-reputed organization and made available to the public. A much stronger version of this paper would have highlighted the dataset itself, in great respect by moving some of the appendix into the body of the article, conducted a more rigorous external validation, and drawn more tentative conclusions from the statistical analysis of the survey results. The general arguments put forth 20 years ago by John Ionnidis in his classic 'Why most published research findings are false.' would be hard to refute in this case (Ionnidis 2005). However, the authors do find the p-values they are looking for, which do suggest the plausible relationships that underlie their hypotheses are true.

Replication Results

The paper is overall **highly replicable** due to the provision of data and code to reproduce the results. This is in contrast to most studies found during the search for replication studies in this course, therefore two facts alone set it in a rare class of studies from recent edition of the Journal *Energy Policy*. On the other hand, the results **are not able to be replicated.**

The summary data is replicated exactly following the data preparation steps. However, the first step of the analysis (Table 3) as well as Table 3a (in the appendix) fails to be replicated with the same level of significance. For Table 3, where the differences are major, this could be due to the differential calculation of marginal effects using weighted data in STATA and Python. In most Python packages, these marginal effects are calculated against the original dataset even if if the data is weighted in the modeling processing, while in STATA this may or not may not be the case. Calculating the marginal effects would then require a manual calculation, presumably with a finite difference method, but such a calculation is beyond the scope of this replication work. The weighted coefficients, agree much more closely than the marginal effects, but are still not a perfect replication. Each of the *.csv* and *.png* files mentioned below can be found as either reference data in the repo/data or as code outputs in the repo/outputs folder. The code can be run via the *main.py* file after installing the needed python packages.

The diagnostics are fairly similar for the models. It could be defaults in STATA for the regression function are different from those of the *statsmodels* package in Python, although this was not immediately clear from reading through the documentation of both of the packages (Statacorp 2023 and Seabold and Skipper 2023). An attempt to verify using the *pylogit* package was not succesful as there is too much multicolinearity in the data for that package to calculate the logistic (or cloglog) regressions using the provided variables (Brathwaite and Walker 2018).[1] Using scikit-learn for the logistic regression (it does not have an immediately accessible cloglog function), led to different results from both STATA and from statsmodels. Part of this likely comes from the fact that the magnitude of the values under consideration are quite small and therefore small differences in processing can push results in and out of significance. The high BIC values in the Python statsmodels implementation speaks to potential instability of results.

Table 1:
Reproduced as *table_1.png* and *table_1.csv* and *table_1_comparison.csv*

This table is reproduced almost exactly. There is a discrepancy of 10 between the number of households that have someone aged less than 16 and households with someone with more than age 75, all of the other indicators line up as can be seen in table_1_comparison.csv. The small discrepancies are due to machine rounding of floating decimals.

Figure 1 (appendix):
Reproduced as *corrplot.png* and *corrplot.csv*

This correlation matrix is not able to be reproduced from their exact code and it cannot easily be compared numerically as in **Table 1.** but can be done visually. The provided text and color scheme of the image do not allow it to be easily processed with an OCR to convert the image to numbers and it is too many small numbers to accurately process. The overall color scheme looks similar between the reproduced and the original plot.

Data preparation steps are missing from the section of their file as they manage to produce output that requires numerical data with categorical data. By reading through table A1 in the appendix, we discover the recoding that was used to accomplish it. Address time, Region, and Energy Rating (all of which are named different things in their plots, raw data, labels file and tables) are encoded as strings which cannot be converted to float or integer values by any statistical package. Even the time spent at the address looks like "> 10" or "> 5 & < 7"). These are later assigned numbers 1–6, 1-9 and 1-8 respectively, but this is not in the code, again implying the STATA code provided was not actually used to produce their results. If they did it with the provided data, it would provide a unique dummy variable for each of the separate categories in the correlation matrix (e.g. "> 10" or "> 5 & < 7" would each have a row and a column in the correlation matrix), but this is not the case.

I added the code to the data preparation steps in order to be able to replicate their findings, which is not entirely straight forward as the labels with_suffin their table are not the same as the labels in the

1 The error message of colinearity prints to the screen...remove the try, except block in the main.py code that wraps around the pylogit implementation to see the error.

dataset, and there were character encoding difficulties in the Don't Know answers. This is also necessary to reproduce the later marginal effects.

Figure 1
Reproduced as *figure_1.png*

The reproduced histogram of adoption of energy efficiency measures matches the original image visually in terms of layout, structure, and content. For a more rigorous comparison, directly verifying the underlying data used to generate each subplot would ensure complete accuracy but since the original do not provide the actual numbers, this is not possible to do.

Figure 2:
Not relevant, it is descriptions of policy changes and checking the dates of UK energy policy implementations is out of the scope of this reproduction work as it is not related to the data but rather the regulatory context.

Figure 3:
reproduced as *figure_3.png*

The reproduced visualization matches the original image visually in terms of layout, structure, and proportions of "Yes" and "No" responses for each policy. For a more rigorous comparison, verifying the underlying data used to generate each subplot would ensure complete accuracy, but labels are not available in the original image so this is not possible.
results_df_coeff.reset_index().to_csv(output_path_annex.with_suffix('.csv'), index=False)

Table 2:
This table categorizes policy preferences but does not involve any data or analysis to reproduce. As it simply provides a conceptual framework for policy types, it falls outside the scope of this reproduction work.

Table 3:
reproduced as *table_3.png, table_3.csv* (main results) and *table_3_diag.png, table_3_diag.csv* (the results of the test of model fit)

The original results are not able to be exactly replicated via the same model using the statsmodel Python. Beginning with the model diagnostics, the Poisson model fits much better in Python than in stata and then Bayesian Information Criterion gives a strong warning sign in Python, (value of some -8000), but behaves normally in Stata. The logit model is similar between both and the pseudo-R2 values are also similar for logit and clogit.

However, the main variables that the authors look at it in the text, loans and grants, are not significant at the 5% level in the replication, whereas they are significant to the 1% level in the original paper. This could be due to the non-weighting of marginal effects calculations in statsmodels compared to STATA.

Table 3a:
reproduced as *table_3_a.csv*

The results are close to being reproduced but not exactly. Some of the variables are signficant that should not be at the 5% level and vice versa. However, the results match up more closely in the statsmodels reproduction. When scikit-learn is used, however, the logistic regression gives quite different results. Again, this is likely due to the high multicolinearity in the underlying data which can lead results to being unstable.

Remaining figures

Figures 4 and 6, as well as tables 4, 5, 6 and 7, and their analogs in the appendix, are also all reproduced following the same naming pattern as above. They also do not agree with the replication results so they are note examined in detail.

Extension

As mentioned in the article summary, I think the best extension of the article would be a more detailed analysis of the survey results. Ideally, situating them within more thoroughly within the results of past and current surveys and conducting a similar analysis on other data to confirm the relationships. The analysis presented in the annex can be brought into the body of the article and the background and meaning of the questions can be better explored to lay the foundations for later interpretation of quantitative models. This can be done using some of the sources mentioned in the annex of the article, including the English Housing Survey[2] and YouGov data[3] used to verify representivity of the sample and create weights for the data, respectively. In my mind, the article could be split into two articles, the first focused on the data set and the second on the development of the theory behind the proposed hypotheses and a deeper quantitative exploration of the proposed hypotheses.

In this spirit, the proposed extension is a basic analysis of the colinearity of the results, as revealed by looking at the correlation plot presented in the appendix of the original paper, and in the reproduction under outputs/corrplots.png. One effective was of testing for colinearity is by using the variance inflation factor (VIF) (Murray et al 2014). This can be calculated using the *statsmodels.stats.outliers_influence.variance_inflation_factor* function in the same package used to implement the logit, clogit and poisson regressions.[4] Filtering only for variables used in the regressions, potential issues immediately appear in the responses to the survey questions regarding energy efficiency information and applied measures.[5] This is already seen in the correlation plot in the annex of the original paper, and the one reproduced in outputs/corrplot.png, but the magnitude of the factors (greater than 20), suggests that they should best be interpreted in aggregate rather than tested individually. Repeating the exercise on a smaller subset of variables, just the controls, we see that problems of colinearity remain in Table 2. Merging squarefootage and bedrooms into a composite household size index might be one start.

Table 1: Variance Control Factor calculation for variables in the main regression (Table 3).

Variable	VIF

2 Cited in the article as a link to this spreadsheet file. https://assets.publishing.service.gov.uk/media/65785e250467eb000d55f5ea/2022-23_EHS_Headline_Report_Chapter_1_Profile_of_households_and_dwellings_Annex_Tables.ods

3 Cited in the article as a link to this pdf: https://d3nkl3psvxxpe9.cloudfront.net/documents/TheTimes_VI_Merge_240627_w.pdf

4 https://www.statsmodels.org/dev/generated/statsmodels.stats.outliers_influence.variance_inflation_factor.html

5 This can be found in the basic_results.check_variance_inflation_index function in the accompanying code.

AdoptionLikelihood_1	15.662251
AdoptionLikelihood_2	6.771909
AdoptionLikelihood_3	21.143133
AdoptionLikelihood_4	29.342389
AdoptionLikelihood_5	13.262856
CircumstancesLikelihood_1	28.624348
CircumstancesLikelihood_2	34.098053
CircumstancesLikelihood_3	32.484219
CircumstancesLikelihood_4	25.203975
CircumstancesLikelihood_5	25.667447
awarenessum	1.840475
tenure	8.526448
whenbuilt2	5.667077
house_type2	4.880092
squarefootage	6.967713
bedrooms	23.154944
hhsize	8.361809
income	6.865015
old75	1.206403
minor16	2.040031
address_change_time	18.556647
EPCrating	10.117684
profile_GOR	5.334304

Table 2: Variance Inflation Factor for control variables only.

Variable	**VIF**
awarenessum	1.77
tenure	8.19
whenbuilt2	5.54
house_type2	4.69
squarefootage	6.89
bedrooms	21.66
hhsize	8.23
income	6.64
old75	1.18
minor16	2.03
address_change_time	17.85
EPCrating	9.942
profile_GOR	5.27

In conclusion, a full extension would calculate some hybrid variables to reduce the number of coefficients in the regression, allowing us hopefully to have more robust results. This would allow for a stronger interpretation of the model presented by the authors and would give deeper insights into the meaning of the relationships uncovered. Regardless of the exact significance replication, this deeper issue of multicolinearity means that some work on the independent and control variables needs to be done before the results can be properly interpreted.

"I hereby declare that I have completed this work independently and have not used any sources or aids other than those indicated. All passages in the work that have been taken from other works in terms of wording or meaning have been marked as borrowed, indicating the source in each case. The same applies to any drawings and illustrations included. I am aware that I have otherwise committed plagiarism, that this will be punished with a grade of 1 and that I will receive a reprimand from the dean."

Biel/Bienne 10.02.2025 Gabrial Erismann

Bibliography

Brathwaite, T., & Walker, J. L. (2018). Asymmetric, closed-form, finite-parameter models of multinomial choice. Journal of Choice Modelling, 29, 78–112. https://doi.org/10.1016/j.jocm.2018.01.002

DESNZ Public Attitudes Tracker: Spring 2024. United Kingdom Government. https://www.gov.uk/government/statistics/desnz-public-attitudes-tracker-spring-2024

Gilchrist, Kathryn & Craig, Tony. Home energy efficiency - review of evidence on attitudes and behaviours. ClimateXchange. The James Hutton https://doi.org/10.1016/j.enpol.2024.114308.Institute. 2014.

Ioannidis JPA (2005) Why Most Published Research Findings Are False. PLOS Medicine 2(8): e124. https://doi.org/10.1371/journal.pmed.0020124

Murray, Leigh; Nguyen, Hien; Lee, Yu-Feng; Remmenga, Marta D.; and Smith, David W. (2012). "VARIANCE INFLATION FACTORS IN REGRESSION MODELS WITH DUMMY VARIABLES," Conference on Applied Statistics in Agriculture. https://doi.org/10.4148/24757772.1034

Peñasco, Cristina. From policy to practice: The role of national policy instruments and social barriers in UK energy efficiency adoption in household. Energy Policy. Vol 194. 2024. https://doi.org/10.1016/j.enpol.2024.114308.

Seabold, Skipper, and Perktold, Josef. "statsmodels: Econometric and statistical modeling with python." *Proceedings of the 9th Python in Science Conference.* 2010.

StataCorp. 2023. *Stata 18 Base Reference Manual*. College Station, TX: Stata Press.

Writing Survey Questions. Pew Research Center. https://www.pewresearch.org/writing-survey-questions/. Accessed 30.01.2025