**Group Task 4: Fixed Effects with Propensity Score Weighting**

**Section 1 Answers**

8/8 pts

**Question 3**

1. **Is there evidence of imbalance? Defend your claims by using the LOGIT output AND the Quick Means Comparison output.**

Yes, there is evidence of imbalance. The quick means t-test results show a difference in means for 12 variables, including electricity consumption, how many people are in the home, how many occupants are over 15 years old, level of education, and internet access.

The logit results showed statistically significant differences between control and treatment groups for 4 different variables: electricity consumption in September, electricity consumption in December, how many occupants are over 15 years old, and D\_4701.2.

1. **Compare what variables are considered significant (if any) in the Logit imbalance test versus the "Quick Means Comparison". Are the results different? If so, what do you think is driving the differences?**

Yes, the two found different results, shown below.

Logit Significant Results (3):

* Electricity consumption in September, December 2009
* How many people over 15 years of age live in your home? 3
* D\_4701.2\_1 -- Not in edited document

Different Means Results (12):

* Electricity consumption in August
* Electricity consumption in September
* Electricity consumption in October
* Electricity consumption in November
* Electricity consumption in December
* What best describes the people you live with? 3
* How many people over 15 years of age live in your home? 2
* How many people over 15 years of age live in your home? 4
* Level of Education - Primary
* Internet Access (yes/no)
* 42111\_2 -- Not in edited document \*\*Not in the data either!\*\*
* 3521\_4 -- Not in edited document \*\*Also not in the data!\*\*

The differences in the two can likely be attributed to the fact that quick means tests are only observing differences in the group average for each variable on their own, whereas a regression model takes into account the other independent variables.

1. **What do you think are the pros and cons of each imbalance check (Logit and Quick Means)?**

Logistic regression is more appropriate for comparing categorical variables, which is what we have in our regression. A quick means t-test is likely more suitable for checking balances in continuous variables. \*\*Not true! I would argue the opposite/reverse in both cases!\*\*

Multicollinearity could be an issue with logit regression. In our C4 logit, there are a few variables with high standard error (D\_240\_5.0, D\_43111\_4.0, D\_470.1\_1), which is an indication of high multicollinearity. In this case, quick means test is a good way to check for imbalances.

1. **Comment on the survey questions selected. Are there any you think are irrelevant or redundant? What variables do you are missing but should have been included?**

43521 and 5414 may capture some of the same information relating to household perceptions of how it can/will change its energy consumption. Question 410 is not needed as that information is captured in the questions on the number of individuals over and under age 15 in the household.

Additional information which we would have included are the variables on housing type, floor size (square meters), and whether they own or rent their home. \*\*Indeed.\*\*

**Section 3 Answers**

4/8 pts

**Question 7**

1. **Compare the coefficient estimates of the treatment-trial interaction variable. How did it change after using the weights?**

In the model with weights, the magnitude of the absolute value of the coefficients for TP and Trial are larger, and p-value become smaller. Therefore using the weights makes it more statistically significant and magnitude of its impact on electricity consumption larger. \*\*What are the numbers? What is the difference in magnitudes? Being 2x larger vs 10x larger is important to note! -1 pt\*\*

1. **Interpret the coefficient estimate for the *first* regression *without* weights. If you were consulting CER on the effectiveness of the C4 treatment, what would you conclude?**

The coefficient for the trial and treatment (TP) interaction variable is -0.0080(p=0.537). This means if you’re in the the treatment group during the trial period, you would reduce your kWh consumption by 0.8 percent. However, the estimate does not appear to be statistically significant at the 5 percent level.

~~The coefficient for trial is -0.0201(p=0.104). This means during the trial period, kWh consumption reduces by 2 percent. Similar with the TP variable, the trial variable is not statistically significant at the 5 percent level.~~

\*\*This variable doesn’t matter, as we care only about the TREATMENT effect. -1 pt\*\*

Based on these results, our conclusion to CER is that the trial and treatment seemed to have reduced kWh consumption, but we could not reject the null hypothesis that there is no difference. \*\*You should conclude that it seems to have had NO effect! -1 pt\*\*

1. **Interpret the coefficient estimate for the *second* regression *with* weights. If you were consulting CER on the effectiveness of the C4 treatment, what would you conclude?**

The coefficient for the trial and treatment (TP) interaction variable in the regression with weights is -0.0253(p=0.009). According to this result, the interaction of being treated and being in the trial period was associated with a statistically significant reduction in electricity use of 2.5 percent.

~~The coefficient for the trial variable is -0.0222(p=0.052). According to this result, being in the trial period was associated with a reduction in electricity consumption of 2.2 percent. This coefficient is not statistically significant at the 5 percent level, but it is at the 10 percent level.~~

If weighting and fixed effects regression were successful in correcting for imbalance--and approximating randomization--the difference between TP and trial (0.3 percent) can be interpreted as the Average Treatment Effect (Harding 2013, 3-7). \*\*I do not see this on 3-7 of Harding (2013). Furthermore, this is not the correct interpretation. The trail variable (P) tells you the average difference in consumption before and during the trial period WITHIN each group. The variable TP then tells you the difference attributed to the treatment effect. -1 pt\*\*

Based on these results, our conclusion is that the combination of trial and treatment did result in a slight reduction in electricity consumption, over the trend in electricity use reflected in the trial variable. However, this effect is very small and not statistically significant. \*\*The interpretation should be of TP, which is significant, as you noted above. As a consultant, you would have said there is evidence to support the treatment was effective, reducing consumption by an average of 2.5%. -1 pt\*\*

1. **Do you think, given how biased the data was, that the weighted regression coefficient estimate on treatment-trial are believable? Please be concise.**

The experiment was constructed to correct two issues: bias based on imbalance in observable characteristics captured by the pre-trial like income and number of people in the household, and bias from unobservable differences between treatment in control.

The first part of the analysis, propensity score weighting, calculated weights based on observable characteristics that treatment assignment may have been conditioned on, such as income. Though this step is helpful in correcting for the imbalances we found in our logistic regression in Section 0, it is only helpful insofar as the “correct” independent variables were selected.

In order to accurately estimate the ATE, unobservable characteristics will also need to be accounted for. For this reason, fixed effects regression--which averages out time-invariant, unobservable trends--is appropriate for this setting.

Using both propensity score weighting and fixed effects regression, we can be more confident that the ATE estimated here is valid. However, there is always the risk that, in using fixed effects, we may also net out some beneficial time invariant characteristics.

\*\*All true. A robustness check would make the results more believable. This could be done by trying different variables in the weighting stage and seeing if there was a significant shift in the TP variable.\*\*