***ECOM90003: Applied Microeconometric Modelling – Assignment 1***

|  |  |  |
| --- | --- | --- |
| **Q** | **Question** | **Answer** |
| 1 | [8 marks] What public goods does Martinez-Bravo use as outcomes to evaluate the effect of mass school constructions? What impact (positive, negative, none) do you think/expect the school construction will have on these outcomes? Justify your response. | The author uses five measures of public goods in villages for this paper. They are the presence of: a primary health centre (a *Puskesmas*), at least one doctor, access to safe drinking water, community based healthcare facilities (*Posyandus*) and a garbage disposal system.  I would expect the provision of all these public goods to be positively impacted by the mass school construction program, but for different reasons.  For the health-related public goods, I would expect the impact to be driven significantly by both the general impact of labour market augmentation as well as improved public governance. This is because the labour required to provide to provide these public goods is quite specialised, where labour supply would greatly benefit from improved education. Similarly, improving the education of village leaders should also make public governance more aware of the developmental benefits of these public goods.  Given I would not expect the specialisation of labour to be a limit to administering access to safe drinking water and a garbage disposal system, I assume their increased provision would be largely driven by improving local governance. |
| 2 | **[3 marks] Explain why comparing the average value of outcomes before and after the first election in areas with a positive amount of INPRES schools may be problematic if the aim is to identify the causal impact of the mass school construction.** | By simply comparing average public good provisions before and after the first election in areas with a positive amount of INPRES schools we are lacking a control group for our analysis. Without knowledge of how those villages without any INPRES schools change their public good provision we lack an effective counterfactual of the policy, that is we can’t truly estimate what would have occurred without the treatment. |
| 3 | **[3 marks] Explain why comparing the average outcomes for villages with and without INPRES schools may be problematic if the aim is to identify the causal impact of the mass school construction.** | Because, to derive accurate causal impacts, we need information about outcomes pre- and post-treatment for each village. If we simply compare outcomes for villages that did and did not receive the treatment and estimate causal effects we are effectively assuming that all villages are completely homogenous, and there is no reason apart from the intervention that different outcomes may arise. This is clearly an infeasible assumption. |
| 4 | **[3 marks] What is the empirical challenge in identifying the causal impact of mass education interventions on local governance?** | A mass education intervention improves the level of education in the entire labour force as well as the education level of village heads. Therefore, the key challenge is disentangling to impact of changes in local governance (village head education levels) from this general labour market augmenting effect. |
| 5 | **[6 marks] The author implements a strategy discussed in class using two different specifications (see equations 1 and 2).** |  |
| 5a | **Briefly explain the overall strategy and the two empirical specifications used to implement this strategy.** | The author’s overall strategy is to first show if increases in public good provisions corresponds with when wholly INPRES educated cohorts start contesting elections. She then seeks quantify the overall impact of the INPRES program on public good provision.  Equation 1 seeks to establish if the increase in public good provision corresponds with the first village election post-1992, which is the first year wholly INPRES educated cohorts can contest elections. It is estimated for three different groups of villages, where grouping is determined by when after 1992 the election occurs (groups are limited by the fact census data is published triennially).  Equation 2 the second equation seeks to establish the magnitude of the overall effect of this intervention on public good provision in each village. |
| 5b | **What is the main parameter of interest in each?** | In Equation 1: .  In Equation 2: . |
| 5c | **What is the interpretation of these parameters? What different purposes do they serve? Make sure you address both.** | In Equation 1, measures the impact of improved public governance on public good provision for each village group.  In Equation 2, measures the impact of improved public governance on overall public good provision.  The purpose of the prior parameter is to demonstrate a descriptive relationship between public good provision and improved public governance. The latter’s purpose is to estimate the overall effect of improved public governance on public good provision. |
| 6 | **[2 marks] What is the key assumption that must be met for this strategy to deliver estimates of causal effects of the school construction program?** | It needs to be shown that the first village election post-1992 and its interaction with the school construction program as quasi-random (i.e. as good as random). She considers this assumption holds given that after testing exogeneity with 50 different pairwise correlations, only 3 of them are significant at the 10 per cent level. |
| 7 | **[4 marks] Is the error term in the data likely to be identically and independently distributed (iid)? If not, why not? Is this a problem for inference? How should the standard errors be estimated? Justify your answer (hint: think of the structure of the datasets).** | No, given this is panel data set, it is highly unlikely errors will be iid. This is primarily because of the unobserved characteristics in each village independent of the intervention which would impact the provision of public goods over time.  These unobserved characteristics are likely to reflect common geographical and cultural difference common across Java. Therefore, it makes sense to cluster the standard errors at the group level. In this dataset, that corresponds with the district level variable. |
|  | Questions 8 and 9 ask you to look at variation in the data that will be used to identify the impact of the school construction program on local governance. |  |
| 8 | [8 marks] This question asks you to produce summary statistics using the data provided before we estimate the main treatment effects in Table 3 of the paper by Martinez-Bravo. |  |
| 8a | What is the unit of observation in the dataset? How many villages and years are available in it? | The unit of observation are villages. There are 9855 unique villages in this dataset observed over 6 different years: 1986, 1990 1993, 1996, 2000, 2003. |
| 8b | How many schools were constructed nationwide as part of the INPRES program between 1974 and 1978? What is the average number of schools in the villages studied by Martinez-Bravo? What is the age range of the children that could attend these schools? | 61000 schools were constructed between 1974 and 1978. On average, 0.89 schools were constructed in each village. Children between the ages of 7 and 12 could attend these new schools. |
| 8c | What proportion of villages had 0, 1 and 2 INPRES schools by the end of the school construction program? (Hint: look at variable *num\_PSINPRES1980*). | 32.14% had zero, 46.33% had 1 and 21.53% had two schools. |
| 8d | In addition to the variation introduced by the INPRESS school construction program, what other sources of variation help the author discern between public good increases driven by a more educated labor force versus more educated politicians? | The timing of village elections is the other key source of variation used by the author to isolate the impact of improved public governance on public good provision. Because village elections are not synchronised, their staggering means their timings are quasi-random. |
| 8e | What is the minimum age requirement for candidates of village head in Indonesia? In what year does the first INPRES cohort meet this requirement? | Village heads need to be at least 25 years old in Indonesia. The first cohort educated entirely within the INPRES program schools reached age 25 in 1992. |
| 9 | We will now create a table where we will explore the variation in timing of first elections after 1992. Label this table as “Table 1” in your assignment and refer to Table 1 when necessary.  Step 1. Using the data provided to you, create one column with the *proportion* of villages in the data who had their first post 1992 election in each year between 1992 and 2000. These values should match those in Appendix Table 2, column 2. (hint: look at variable ele1v\_post92).    Step 2. Create a second column with the *number* of villages who had who had their first post 1992 elections in each year between 1992 and 2000. (hint: remember that each village is repeated 6 times in the dataset, search for commands such as drop and collapse to help you get the right values but remember that you will need the full data for the rest of the questions). |  |
| 9a | Based on the Table 1 you created, is there variation in the timing of the first election across villages? Which years had the highest and lowest proportion of villages with their first post 1992 election? Is there a pattern in the data (increasing, decreasing, neither)? | Yes, there is variation. Around three quarters of villages had their first election between 1997 and 1999. 1996 had the lower proportion of villages with their first post-1992 election (1.08%), whereas 1999 had the highest (36.00%). There is no constant/linear pattern in the data: the proportion of each year is pretty flat beforeincreasing significantly from 1997 and 1999 and then crashes again in 2000. |
| 9b | Now refer to Table 2 in the paper by Martinez-Bravo, is this variation in timing significantly correlated with changes in the number of doctors, health centers or health posts in the pre-treatment period? Why would this be a threat to identification? | This variation is not significantly correlated with changes in the number of doctors, health centres or health posts in the pre-treatment period up to the 10% significance level.  If we could demonstrated significant correlation, it would imply the timing of village election was not random. This would be a probably as it would suggest the timing of an election was influenced endogenously by village characteristics and therefore could not be used to estimate causal impacts of the INPRES program. |
|  | Questions 10—12 and 15 ask you to estimate various models using the data provided. Please report (all) results for questions 10-12 in a single table labelled Table 2, and use the format specified in the questions pdf file.  Report the standard errors in parentheses under each coefficient. [Hint: See outreg2 or esttab to output Stata results into excel. Make sure to adjust your standard errors following your response to question 4].  Step 1. Create a variable called *post92* which should take a value of 1 if the year of the survey is greater or equal to the year of the first election, and zero otherwise. To check that you’re doing this correctly, make sure that the mean of this variable is equal to 0.3852  Step 2. Create a variable called *interaction* where you will interact the *post92* dummy and the demeaned number of INPRES schools per village (this variable is called *num\_dev* in the dataset). (Tip: when interacting a dummy and a continuous variable, its useful to subtract the mean off each unit so that we can interpret the interactions as an increase with respect to the mean).  Hint: Stata has a number of commands to help us deal with regressions that include a large number of fixed effects such as “areg” and “reg2hdfe”. I suggest using the “areg” command instead of our usual reg command for question 12 in particular. |  |
| 10 | [5 marks] Regress the 3 binary outcomes in panels A, B, and C of Table 3 on the indicator for the first post 1992 election, the number of INPRES schools and the interaction of both variables that you created earlier (without any other controls). Report these coefficients in columns 1, 4 and 7. What is the interpretation (sign, size and significance) of the estimated coefficients? | The first row of coefficients in Table 2 reflects the difference in probability the first election post 1992 has on public good provision. Depending on the public good, this election increases the odds of having these public goods by 5-6 percentage points. This impact is significant at the 1% significance level in all cases.  The second row of coefficients reflects the difference in the probability of public good provision owing to the school construction program after the first election post 1992. This impact is only significant positive for having a primary health centre in the village, increasing the probability by around 1 percentage point. For doctors and safe drinking water, this program increases the probability by 0.4 and 0.5 percentage points respectively. However, these two coefficients are insignificant up to the 10% level. |
| 11 | [4 marks] Next, add indicators for year of survey to the above specifications. Report these coefficients in columns 2, 5 and 8. What happens to our estimated coefficients on *interaction* (sign, size and significance)? In light of this, do you think the specifications from Question 10 can be used to obtain the causal impact of the school construction program? Why? | After making this amendment, there is no impact on our coefficient estimates (except increasing doctors in the village by 0.001) or their significance levels.  Therefore, the specification from Question 10 cannot be used to obtain the causal impact of the school construction program. We need to account for the two dimensions of fixed effects (time and village) to sufficiently estimate the causal impact of this intervention. |
| 12 | [5 marks] Next, add indicators for village to the above specifications. Report these coefficients in columns 3, 6 and 9. What happens to our estimated coefficients for the *interaction* variable (sign, size and significance)? | In general, all interaction coefficients increase in size and become significant at least to the 10% significance level. Therefore, at this significance level, they are all significantly positive.  All interaction coefficient terms increase. The chance of having a primary health centre, doctor or access to safe drinking water in the village increases by 1.2, 0.6 and 1.8 percentage points respectively. |
| 13 | [2 marks] Are these sizable effects? What is the change in the outcomes relative to the sample mean? | Yes, they are large. For primary health centre, doctor and access to safe drinking water, these estimates show a 7%, 6% and 1.7% increase in outcomes relative to the sample mean respectively. |
| 14 | [3 marks] Martinez-Bravo finds that these effects are heterogenous. When are the estimated interaction effects larger? | The author finds the interaction effects larger for those villages with bad quality of service to begin with, meaning the effects are a function of villagers’ demand for these public goods. This suggests the more intense the school construction program was in a village, the more receptive public governance became to what public goods were needed most. |
| 15 | [6 marks] Re-estimate the above specification but now, instead of including the post dummy, the demeaned number of schools and the *interaction* variable you created earlier, use the double hash (##) function to interact post92 and num\_PSINPRES1980 (the non-demeaned variable indicating whether a village has 0, 1 or 2 INPRES schools). Report these coefficients in a separate table labelled Table 3 using the format below (Tip: combine the ## and i. functions to achieve the desired result, e.g., i.post##i.treatment). What do you conclude from these latest results? | These results tell us that, for most public goods, the more intense the school construction program the higher the chance there is of having good public good provision in a village. |
| 16 | [3 marks] How does the author address the parallel trends assumption? What does she find? (Hint: refer to question 5). | The parallel trends assumption is satisfied by exploiting the unsynchronised timing of village election, and showing that these elections are not correlated with public goods provisions.  This implies that without the mass school construction, villages would have adhered to similar trends in public good production over time. |

***Appendix A: Requested tables***

**Table 1: Year of first election post 1992**

|  |  |  |
| --- | --- | --- |
| **Year** | **Proportion** | **Frequency** |
| 1992 | 4.09 | 403 |
| 1993 | 6.89 | 679 |
| 1994 | 5.72 | 564 |
| 1995 | 2.34 | 231 |
| 1996 | 1.08 | 106 |
| 1997 | 12.62 | 1244 |
| 1998 | 28.88 | 2846 |
| 1999 | 36.00 | 3548 |
| 2000 | 2.37 | 234 |

**Table 2: The effects of school construction in public good provision (Q10-12)**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Primary health centre in the village** | | | **Doctors in the village** | | | **Access to safe drinking water** | | |
|  | **(1)** | **(2)** | **(3)** | **(1)** | **(2)** | **(3)** | **(1)** | **(2)** | **(3)** |
| **Post first 1992 election** | 0.049\*\*\* | 0.005 | -0.005\* | 0.064\*\*\* | 0.021\*\* | 0.001 | 0.052\*\*\* | 0.027\* | -0.005 |
|  | (0.003) | (0.007) | (0.003) | (0.004) | (0.010) | (0.003) | (0.015) | (0.015) | (0.005) |
| **Post x num INPREs schools** | 0.010\*\*\* | 0.010\*\*\* | 0.012\*\*\* | 0.004 | 0.005 | 0.006\* | 0.005 | 0.005 | 0.018\*\* |
|  | (0.003) | (0.003) | (0.003) | (0.004) | (0.004) | (0.003) | (0.014) | (0.014) | (0.007) |
| **Observations** | 59130 | 59130 | 59130 | 59130 | 59130 | 59130 | 39420 | 39420 | 39420 |
| **Mean of dependent variable** | 0.10 | 0.10 | 0.10 | 0.12 | 0.12 | 0.12 | 0.77 | 0.77 | 0.77 |
| **Year fixed effects** | No | Yes | Yes | No | Yes | Yes | No | Yes | Yes |
| **Village fixed effects** | No | No | Yes | No | No | Yes | No | No | Yes |

*Note: standard errors in parentheses. \* = p < 0.1, \*\* = p < 0.05, \*\*\* = p < 0.01*

**Table 3: Intensity of school construction (Q15)**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Primary health centre in the village** | **Doctors in the village** | **Access to safe drinking water** |
|  | **(1)** | **(2)** | **(3)** |
| **Post first 1992 election** | -0.017\*\*\* | -0.006 | -0.017\*\* |
|  | (0.004) | (0.005) | (0.007) |
| **Post x INPRES schools = 1** | 0.013\*\*\* | 0.009\* | 0.010 |
|  | (0.005) | (0.005) | (0.009) |
| **Post x INPRES schools = 2** | 0.024\*\*\* | 0.011 | 0.036\*\* |
|  | (0.006) | (0.007) | (0.014) |
| **Observations** | 59130 | 59130 | 39420 |
| **Mean of dependent variable** | 0.10 | 0.12 | 0.77 |
| **Year fixed effects** | Yes | Yes | Yes |
| **Village fixed effects** | Yes | Yes | Yes |

*Note: standard errors in parentheses. \* = p < 0.1, \*\* = p < 0.05, \*\*\* = p < 0.01*

**Appendix B: Stata code**