Helpful Hints for Problem Set #1 Professor Morris M Kleiner

PA 5033 Multivariate Techniques Spring 2022

**1. When to interpret coefficients in Part A:** You do NOT need to interpret coefficients in parts A.2, A.3, or A.4, unless you are using the coefficient values as part of your pros/cons discussion in any of those parts. You DO need to have coefficient interpretations for Part A.1. (Additionally, since Part B is your best model, you will want to incorporate coefficient interpretations in some way in Part B as you will be explaining Part B with fine detail, and coefficient interpretation will likely feed directly into how you make a policy decision.)  
  
**2. What to include in A.1:** Be sure to show write out your models, show your STATA output, and explain your results per the items listed in A.1 in the instructions.  
  
**3. What to include in A.2:** Write out your models, show your STATA output (including dwstat and VIF). Test for serial correlation and discuss potential multicollinearity. Discuss the statistical/theoretical benefits and costs relative to the model in A.1  
  
**4. What to include in A.3:** Write out your models, show your STATA output (including VIF). Discuss multicollinearity and explicitly test for serial correlation using the Lagrange Multiplier Serial Correlation Test (LMSC). Explicitly show and calculate the Long Run Impact Multiplier, and discuss the statistical/theoretical benefits and costs relative to the model in A.1  
  
**6. The approach to A.4:** A.4 asks to CREATE a simultaneous system of equations and CREATE instrumental variables for the endogenous variables. The only constraint on A.4 is that you must use **employment to population ratio** and **average wage equations** as the endogenous variables.  
  
Thus, you are constrained to the following structural equations:

**employment to population ratio = β0 + β1 average wage equations + … (predetermined variables of your choosing)**

**average wage equations** **= α0 + α1 employment to population ratio + … (predetermined variables of your choosing)**

In this case you have many options for what to create for your structural equations. This is where some freedom comes into play, and there is no correct answer! If you want to borrow the models/variables from Part A.1, 2., or 3., or even take the models from lab, that is fine. You can also start from scratch and decide which variables to include based on theory/logic, or if you have no idea what variables to include you can run a correlation matrix on the data (just literally type "corr" in STATA) and see if that gives you any hints.  
  
Once you find add some variables that are appropriate, then your model for A.4 is complete!   
  
The next step is to create the equations for the instrumental variables for **employment to population ratio and average wage equations**:

= π0 + … (predetermined variables of your choosing or reduced form method)

= πn + … (predetermined variables of your choosing or reduced form method)

(where the 'n' is just some number following from the last coefficient subscript from the previous reduced form equation)  
  
You can create instrumental variables two ways here:  
  
1. Simply run the reduced form equation for both **employment to population ratio and average wage equations**: (i.e., run both both employment to population ratio and average wage equations against all predetermined variables in your system of equations).  
2. Take a free-form approach and choose the predetermined variables from scratch just as you did for creating your model.  
  
Method 1 is preferable if you have no idea what to include making the instrumental variables, and if you just want a direct approach to making your instrumental variables in short order. Method 2 is good if you want to finesse your instrumental variables a bit more to try to make them as “best” as possible. That means it is even appropriate to just have ONE VARIABLE serve in place of the endogenous variable if they correlate well enough. Either approach is fine!  
  
Regardless of method, make sure the reduced form / instrumental variable regression **has an F score greater than 10!** This ensures you are creating an acceptable instrumental variable.

Remember, make sure you run the full 2SLS model.  
  
**Regardless of how you create your instrumental variables in A.4, you MUST take a paragraph or two to logically/theoretically justify and explain why the predetermined variables you use are indeed good variables to use to create the instrumental variable.** So, even if you use Method 1 (which involves directly including all predetermined variables in the system for EACH reduced form equation), you still should explain why each of the predetermined variables should logically/theoretically be included in that reduced form equation. You should ATTEMPT to logically explain your predetermined variable choices.  
  
**7. What to include in A.4, more or less taken from point 6 above:** Write out your system of equations, write out your equations for your instrumental variables (e.g., methodical reduced form equations, or whatever equations you created to make your instrumental variables), show your STATA output for running 2SLS (use the "ivreg" shortcut but also show the F-stat for the first stage) and the dwstat (test for serial correlation), take a couple paragraphs to justify the predetermined variables used to create your instrumental variables, and explain the theoretical/statistical benefits and costs of 2SLS relative to A.1.  
  
**8. What to include in Part B:** You can pick ANY model from A.1, A.2., A.3, or A.4, OR you can create ANOTHER model using any of the approaches of A.1 (typical regression), A.2 (ad hoc), A.3 (dynamic), or A.4 (simultaneous system).  
  
**If, for example, you use your A.4 model (or A.3 model, or something else using a simultaneous system or dynamic lag), you should discuss the following:**  
  
- Justify your approach! (e.g., why a simultaneous system and why correct with 2SlS? Or, why use a dynamic model?)  
  
- Justify the variables in your model (so, in A.4 justify them) Also, if you do use your A.4 model and keep the SAME predetermined variables to make your instrumental variable, you do NOT have to re-justify them in Part B, as you already did in A.4  
  
- Show your model (and instrumental variable equations if using a simultaneous system)  
  
- Show your STATA output (just copy and paste it from your previous section, if borrowing a previous model)  
  
- **Justify why your model is best!** This means you should discuss your model's results and compare the results to the other models you didn't use (e.g., if you used your A.4 model, compare it to models A.1, A.2, and A.3.) So, how do the adj-R values compare? Significance of variables? Is there multicollinearity? Is there serial correlation? Are the coefficient magnitudes useful/large enough? Are their tradeoffs between the model you chose and the other models you didn't choose in part A? This may seem like a lot, but you can efficiently take one sentence per issue/comparison and cover things in a paragraph or two.  
  
- When the problem asks for the pro and cons of your model with respect to A.1, if picked A.4 or A.3, then you've already done this in answering A.4 and A.3, as well as discussing your full results, so don't worry about adding more to this. Otherwise, compare the pro/cons of your new model with respect to A.1.  
- Use the results of your model to make a policy recommendation regarding whether or not a lower minimum wage benefits Puerto Rico (recall the first page of the problem set for full details on the case)  
This policy recommendation should at least be a couple paragraphs but could also be a half-page to 1.5 pages depending on how much you think you need to defend your policy decision. As a hint, your coefficients and their magnitude will be play a vital role in determining how you make your policy recommendation. Also, think about both the impacts of minimum wage on employment to population ratio and average wage public policies.