

Outline of Progress Report C: Data and Modeling

The main purposes of this third type of progress report is to get you to think clearly about the data at your disposal, and the econometric modeling you will use to answer your question.

To wit, I want you to prepare a 15-20 page progress report with the following template. This is a long assignment.

Goal: I want you to describe a **hypothetical** dataset which you will use to answer your question. Don't make up an "ideal" dataset, make up a dataset with shortcomings, data problems, etc. You will write a short paper which present an econometric model and empirical "results" *using the hypothetical dataset*.

There are *three parts* to this progress report:

1. **Data** Describe in great detail the dataset you plan to use to answer your question. Include information on:
 - The data source(s)
 - The sample period
 - The data structure (i.e., the sampling units, the frequency of sampling, cross-section vs. time series vs. panel, etc.)
 - What variables are in the dataset
 - Discuss the data requirements in order to answer your question, and why this is the "right" data: which variables are particularly important "y" variables, which are important "x" variables, and which are important control variables?
 - Discuss any data shortcomings (sample selection, attrition, measurement errors, etc.)
2. **Model** Formulate an econometric framework which you will use to answer the question. *This section should contain lots of equations!*

It is often useful to write down, first, a "baseline" estimating equation (this could, for example, be a simple linear regression $y_{it} = X'_{it}\beta + \epsilon_{it}$). Be explicit about what variable(s) each of the symbols stands for: what are the y variables, the x variables, what types of unobservables are captured in ϵ , etc. Also, be very explicit what the subscripts stand for (eg. i labels households, t labels states, etc).

Then, discuss potential complications with the baseline equation, and suggest remedies (for example, you might wish to include fixed effects, or one of the

right-hand variables may be endogenous). Include *at least three* alternative specifications which can be considered “robustness checks” of the baseline equation.

Discuss how you plan to estimate the parameters of the model. For example, if you plan to estimate a given model using IV, describe what the instruments are, etc.

For more complicated models, such as limited-dependent variable (logit, probit, duration) models, or simultaneous equation models, etc., you should also discuss what your assumptions on the unobservables are, and derive the likelihood function if it is relevant.

3. **Empirical results** In the final section, you should present your empirical results. Clearly, two weeks is not enough time to obtain actual empirical results, so I want you to *make up hypothetical empirical results*. This section should contain at least the following:

- A table of summary statistics of the variables which you will use in your analysis. Discuss at least three noteworthy features of the summary statistics.
- A table (or tables) containing parameter estimates for the various specifications of your model. Be sure to include the standard errors, standard goodness-of-fit statistics (R^2 , log-likelihood function), and information on number of observations, etc. Have footnotes explaining any notation in the tables which is not obvious. Take as a model the results tables from the “mystery” paper which you wrote several weeks ago.
- You should interpret your results, and discuss how they answer your question. What differences do you find between the alternative specifications, and what are possible reasons for these differences? Which are your “preferred” estimates?
- Consider some policy implications of your results. For example, if you are looking at the effect of mergers on market power, you should consider any implications that your results may have on antitrust policy. Do a “back of the envelope” calculation: how much less (or more) would consumers be paying in the absence of the merger?

4. You should end with a short conclusion. Briefly summarize your main findings. What are the outstanding questions? What is the logical next step to take?