Checklist for Empirical Papers

J.R. Walker

September 2, 2013

In econometric classes and in many discussions of paper we focus on a single primary issue — is the model identified? Is the estimator consistent? For learning such concentration is useful and many times necessary. Yet, doing research our objective is to answer the target research question. What is the effect on employment and earnings of the JTPA labor market training programs? Do house prices fully capitalize local fiscal and climate amenities? Are households saving adequately for retirement? etc. Instead of one we have seemingly countably many questions to address. We face questions on which theoretical framework to apply, which data to use, and related, what known flaws does the selected data possess? Are there econometric procedures able address the data deficiencies? If not are there available data sources that can fill in (partially or completely) for the missing data? Are these problems so insurmountable that makes any attempted solution to the research question not credible?

Unfortunately, minimizing one problem makes another one worse. For example, to overcome unobserved, time–invariant person heterogeneity within a panel data setting, a fixed effect specification is commonly assumed. This isn't a NP control, but it's a natural place to start. However, first differencing usually amplifies measure error should the treatment be measured with error. Thus, to minimize bias from unobserved person–specific differences a fixed–effect estimator may increase measurement error bias.

So the argument isn't whether a procedure is "right" or "wrong" but why the focus on these particular issues? And why were they addressed within the chosen framework? It is trite to say that all empirical analysis is an approximation. What's needed is a motivation for *the particular approximation* made.

I've assembled this checklist to help you anticipate possible questions at workshops and from referees.

What's the purpose of the paper?

- Descriptive?

- Causal?

- Policy focused?

Use of Theory (connected with purpose)

- To generate endogenous/exogenous labels for variables?

- To identify instruments?

- What (if any) are the testable implications? (e.g., covariance?)

- (If causal) What are the counterfactual settings you wish to evaluate?

Model Specification

- What is the nature of the approximation? What is the 80 to 90 percent of the problem this approach seeks to cover?

- What motivates the functional form? (Theory? Previous Lit?, Math simplicity?

- What generates the disturbance (error) terms? Alternatively, we are the error terms introduced in the model in this way?

- What are the stochastic assumptions (e.g., distributional assumptions, correlation or independence of disturbances) are imposed? What restrictions and implications follow from these assumption? For example, to estimate a discount rate, β, which you parameterize as $\beta = \frac{e^\mu}{1+e^\mu}$ and estimate (the time rate of preference) μ. That is fine, but it's no surprise that β is in the unit interval.

Sampling Plan: Where do the data come from?

- If survey data, what is the mode of the survey (telephone, mail, in–person)?

- What is known about the quality of the data? Known deficiencies?

- Are measures harmonized? Wording of questions can and do change over time. Need to check that a consistent coding scheme used across time and across data sources.

- If sampling plan is exogenous is it stratified/clustered? Will sample weights improve efficiency properties of estimators? Adjustment for stratification/clustering in calculation of standard errors?

- What is the assumed beginning of the agent's decision problem? What is the youngest age observed in the data? If the two do not agree, is the outcome or covariates observed in the first sample period assumed to be endogenous or exogenous? Why?

- If Endogenous how is the sampling plan incorporated into the analysis?

Homogeneity

- Are differences across individuals constant over time? Are differences fully observed or are some differences unobserved?

- If the disturbances are heteroskedastic, what is the effect on the estimator? Standard Errors? Are there easy adjustments?

Measurement Error

- Are key variables measured with error?

- Remedies available? E.g., Instrumental variables

- Error symmetric or is it likely to be biased in one direction?

- Estimates of magnitude in literature?

- Make sense to conduct simulation study to gauge bias?

Empirical Findings

- What are the primary results?

- What's the evidence for each finding?

- What are the implications of the estimates? Implied elasticities? Important differences by subpopulation?

- What oddities appear? What's the source of the oddity? Evidence of outliers?

Statistical Significance

- If multi–stage estimation procedure: are standard errors correct?

- What level of significance makes sense? 1%, 5%, 10%

Economic Significance

- How to place the estimates in context?

- Are the estimates substantively significant?

- Do relative effects make sense (across subpopulations or cohorts)?

Goodness–of–Fit and Robustness

- What methods or procedures are available to test the fit of the specification? Over–identification tests? Out of sample predictions? What is the power of such tests or predictions?

- How well does the specification fit the data?

- If specifications are non–nested, BIC, AIC may be useful.

- What are alternative dimensions by which to evaluate model?

- How robust are the estimates to alternative (reasonable) definition of variables?

- Have you mapped out a "neighborhood" of alternative assumptions, specifications "close" to the preferred specification?

- If results are sensitive, obligated to report. What are the estimates nevertheless useful?