Empirical Etiquette Survey
Goldfarb, King & Simcoe (Fall 2013)

Introduction

This survey is designed to measure the use of certain norms or best practices in empirical research, especially for papers that attempt to make causal inferences using observational data. The survey is **NOT** meant to assess the overall quality of any piece of published research.

Although we have tried to write each question clearly, some may seem ambiguous or hard to answer in the context of a particular paper. This is expected. In such cases, we ask that you read the paper closely and do your best to select the most appropriate answer to the question as it is written.

Survey

1) Please **identify the paper's main hypothesis**. If the paper does not state hypotheses, select the empirical model that provides the best measurement or test of the paper's thesis. If you cannot decide on a single model or estimate, please use the first set of empirical results presented in the paper.
   a. Hypothesis: <<example: "H1a" or "no hypotheses">>
   b. Table and column number where estimates reported

2) Does the paper **discuss the magnitude or economic significance** of the results that you selected in step 1?
   a. No. The paper only refers to the sign and statistical significance of results.
   b. Yes, the paper describes the absolute or percentage change in the predicted outcome (Y) for a given absolute or percentage change in the explanatory variable (X).
   c. Yes, the paper provides the information described in (b) AND puts these results in context by comparing them to either the total variance in outcome (R-squared) or to the results of other studies of the same or similar outcomes.

3) Does the paper **place a causal interpretation on the observed relationship** between X and Y in the data? (This is a subjective question. Papers that

describe the "impact" or "effect" of changing X, as opposed to the "association" or "correlation" between X and Y, are making causal claims.)

    a. Yes

    b. No

4) Does the paper **explicitly address any of the following threats to causal inference**? Please select yes/no for each "threat" based on (a) a word search of the document, and (b) your own interpretation of any text that describes the authors' choice of empirical methods.

                      Word Search   Your Opinion

    1. Endogeneity

    1.a   Omitted Variables

    1.b   Reverse Causality

    1.c   Selection / Missing Data

    1.d   Simultaneity

5) **Does the paper use any of the following methods** (check "yes" for all that apply):

    a. Regression control (i.e. adding X variables)

    b. Matching methods (propensity score, CEM)

    c. Instrumental variables or exclusion restrictions

    d. Regression discontinuity

    e. Panel Data (Fixed Effects / Differencing)

    f. Panel Data with (b), (c) or (d)

6) If you answered yes to any options in question 5, does the paper **discuss and/or test the methods' maintained assumptions**? (Note: The intent of this question will become more clear as you become familiar with particular methods over the course of the semester. Please answer to the best of your ability.)

    a. Yes

    b. No

7) Which of the following options **best describes the empirical model or specification** used to produce the estimates you selected in step 1?

    a. Comparison of means (e.g. T test)

    b. Linear regression (OLS) in logs or levels

c. Exponential regression: Poisson or Negative Binomial

d. Binary choice: Logit or Probit

e. Discrete choice: Multinomial logit/probit

f. Ordered choice: Ordered Logit/Probit

g. Censored outcome: Tobit

h. Sample selection: Heckman

i. Discrete-time hazard: Cox, proportional

j. Continuous-time hazard: Weibull, Logistic, etc.

k. Rare events: Zero-inflated or "rare event" models

l. Other (describe)

8) Approximately **how many pages of the manuscript fall into each of the following categories**. (Please round to the nearest page, include Figures and Tables in your page counts, and use your best judgment to decide whether the Introduction and Conclusion are properly included in one of the categories):

a. Hypothesis / theory development

b. Describing data and measures

c. Describing empirical methods

d. Discussing empirical results

9) On a scale of 1-5, **please rate the expositional clarity of the manuscript** with respect to:

a. The logic underpinning theory or hypotheses

b. The motivation behind selection of empirical methods

c. The interpretation of empirical results