

ECON 250A Fall 2011
Very Short Paper I: Data Collection
 Go find some interesting data.

Email Prof. Antonovics an outline of the dataset and the research question by Tuesday, Oct 4 at noon.
Email Prof. Antonovics a table of means and related information, in a format to be explained in the first lecture, by Tuesday, October 11 at noon.

Your data should include at least 50 observations of values for at least two variables in a cross-section or a panel ($N \geq 50$, $T > 0$).

Report a table of means, which includes for each variable to be used in the analysis: the mean, standard deviation, number of observations, minimum and maximum value (i.e., `summ` in Stata*). If weights are to be used in analysis, report the weighted means and standard deviations as well. At the bottom of the table you should document the source, including enough information to make replication possible. An example follows.

The very short paper will be marked on the econometric method alone, with no marks deducted for even the most ludicrous economic analysis, so feel free to have fun. On the other hand, you will spend many intimate hours with this project, so you may as well construct it in a way that will make it interesting for yourself.

Feel free to contact any of us (preferably by Oct 3) with questions about data choices.

Note: This example includes the number of observations in the note. The min and max are often excluded in the published version of a paper due to space constraints. A good first draft should have them.

TABLE 3.—DESCRIPTIVE STATISTICS FOR OIL REFINERIES PACE, LRD, AND REGULATORY DATA

	Mean	Standard Deviation
Value of shipments*	1,707,848	2,890,197
Value added	118,772	231,349
Employment	372	500
Air pollution abatement investment	2,096	7,618
Net abatement investment	1,495	7,475
Depreciation of abatement capital	601	1,796
Abatement operating costs	6,586	16,607
Change in abatement operating costs	141	6,951
New regulation adoption dates	0.053	0.369
New regulation compliance dates	0.041	0.267
New increased stringency dates	0.012	0.136
South coast indicator	0.055	0.228
California indicator	0.129	0.335
Texas indicator	0.208	0.406
Louisiana indicator	0.094	0.292

* Thousands of 1991 dollars deflated by the Producer Price Index.

Source: Pollution Abatement Costs and Expenditures microdata.

The sample contains 1,914 observations weighted by PACE sampling weights to represent 2,425 plant-years in the population. Sampled from 1979–1991, excluding 1983 and 1987. Data from 1992 and 1993 were excluded due to errors. Change in operating costs is from year to year and is defined only for plants observed for two consecutive sampled years. Employment is measured in persons.

* If you are thinking of using some software other than Stata please contact student health services for referral.

VSP II: Guidelines for Very Short Papers

First Draft Due Wednesday October 26, in class

(submit hard copy to Prof. Dahl, but email to Prof. Berman as well)

The Very Short Paper (VSP) is an exercise in empirical analysis designed to give you some practice in using econometric methods responsibly and communicating the results efficiently (without the added burden of having a brilliant research idea). We will grade you on following the rules on content and the general advice below. If you discover something along the way, that's great. Who knows? Maybe you will get a quick start on your thesis.

The VSP should include:

- a) an introduction, which generates interest in the exercise,
- b) a derivation or motivation for the estimating equation,
- c) a table of means and as many tables of estimation results as you like,
- d) a written discussion of the table of means and of estimation results, and
- e) a conclusion.

The text (everything but (c)) should be between 3 and 5 pages long. You can include as many tables, graphs and diagrams as you like, as long as the entire VSP is no more than 10 pages long. We will stop reading at the 5th page of text or the 10th page overall, and write down a grade.

While the first draft will be graded, only the grade on the final version will count. That is, the 5 points listed on the syllabus will be given for submitting a rough draft on time.

*The second and final draft will be due Nov. 23 in class **and** by email (elib@ucsd.edu), along with the graded hard copy of the first draft. The first draft grade should be a lower bound on the final grade (provided you don't remove anything good.)*

Presentations will be given in our final class meeting.

Good Luck

General Advice for Very Short Papers

DO

1. Write down the estimating equation.
2. Declare what object you are estimating: i.e., CEF (Conditional Expectation Function), BLP (Best Linear Predictor), causal effect, linear causal effect.
3. Discuss the economics underlying an assumption that a variable is uncorrelated with a residual, if there is any.
4. Discuss the economic interpretation of structural parameter estimates. Speculate on the behavior that may be causing interesting BLP coefficients you have found.
5. Include a table of means. It should describe the same variables and observations as your table of regression results (below).
6. Document the table of means in a way that would allow replication.
7. Report estimation results in a table that includes at least the following: coefficient estimates, standard errors, root mean squared error, R-squared, number of observations. (IV estimates should include the F-statistic on omitted instruments in the first stage, the partial R^2 of omitted instruments in the first stage and an overidentification test if applicable).
- (optional) 8. Estimate the confidence level of a reported confidence interval by a bootstrap procedure. Use enough iterations to get a std. error of your estimate less than .005.
9. Document the table of regression results.

DON'T

20. Declare a BLP (Best Linear Predictor) or CEF (Conditional Expectation Function) and then interpret the coefficients as reflecting causal relationships (i.e., as if they are estimates of a causal effect of a linear causal effect) without some explanation.
21. Include a printout of all the data.
22. Include computer output of regressions.
23. Include graphs that are not referred to in the text.
24. Include a discussion of models or estimating equations that you do not address empirically.
25. Express disappointment if the results are not (or are) statistically insignificant. (Remember that your job is to test a hypothesis or estimate a parameter as precisely as possible.)
26. Report second stage instrumental variable estimates without reporting first stage estimates.

DO (cont.)

10. Refer to variables by their English names or by the symbols given to them in the estimating equation. Choose symbols that suggest the content of variables
11. Explain the method of estimation if it is other than OLS.
12. Report the heteroskedasticity consistent (i.e., White) standard errors (or explain why not).
13. If you are replicating some other research, explain in the introduction how your work differs (or doesn't).
14. Either type the paper, or make sure that your handwriting is legible, your spelling is perfect, and your grammar is efficient and concise.
15. Spell check and *grammar check* your work.
16. Include page numbers.
17. Read the final copy out loud to yourself to check grammar and clarity.
18. Have a classmate read your work.
- 19. Keep the comments or graded first draft to hand in with the completed paper.**

Enjoy yourself. You're doing economic research. This is what you dreamed about.

Start early and set a reasonable timetable so you can enjoy a calm assignment.

DON'T (cont.)

27. Refer to variables by names (like "dcdgpl") that they have in the computer program, or by names like x_1 , x_2 , x_3 .
 28. Use the word "obviously," especially in the sentence "z is obviously a valid instrument." It is a red flag for nitpicking applied econometricians.
 29. *Confuse economic and statistical significance.* E.g., a regression of life expectancy on (randomly assigned) cigarette use of individuals yields $b = -10$ years and $se(b) = (7)$. The effect is statistically insignificant but may well be large. A 95% confidence interval includes 23 less years of life!
 30. Waste lots of precious time wordprocessing equations. Handwritten equations are fine, even in pencil, as long as they are legible.
- Work when you should be sleeping. It tends to decrease accuracy, though perceived accuracy remains constant.*
- Expect an extension because your time-management didn't work out.*