Term paper information

Econ 444, de Jong, Fall 2004

Choose one of the projects described below for your paper.

Try different specifications for your model, for example, you may want to try a double-logarithmic model as well as a model in levels. Consider including squares of variables, including interaction terms, or creating additional dummy variables if that should be a good idea. Examine your dataset carefully; for example, perhaps you will want to delete observations that you think do not make sense. Try to be creative and critical. Try to do things like add extra regressors if that should be possible. Sometimes relationships between variables in your dataset are very unlikely to be linear. Multiple regressions are very often more informative than simple regressions for the reasons discussed. Be aware that your instructor will only be able to judge your work on what you hand in, and not on all the work you may have done but cannot be found back in your final product. Also, take great care not to misinterpret output or draw conclusions that are verifiably incorrect.

For reporting your regression results, follow the following format (example):

\[
\hat{WGE} = 1234 + 56.3 \times EXP - 901 \times F + 341 \times W - 344 \times E - 861 \times S
\]

(231) (26) (540) (562) (668) (659)

se in parentheses; \( R^2 = 0.63 \), \( n = 1031 \)

Do not include Eviews output in your paper.

If you can find literature or references about the topic of your choice, mention these in the appropriate section; it will improve your paper’s quality. Try different model specifications, and try to work towards a "final model" if possible. Take care to choose a sensible model specification; for example, do not use a regressor that is coded "1" for train, "2" for bus, "3" for airplane.

The projects described below are different, and you may want to be careful when you choose one. You should choose the project that you can feel you can write the most competently
about. *The choice of project is an integral part of this project*. You cannot excuse a poor paper by claiming that the dataset or your project was not very fertile.

*Low t-values can still make an interesting project.* The point here is to do economic or social science style research; it is your job to objectively report what is suggested by the data. If you think there ought to be a relationship between y and x, but you don’t find one, then that is a perfectly fine outcome of a piece of research. Scientific research should ideally be free of biases, and reporting unbiased conclusions in a verifiable and correct way is what science should ideally be all about.

*Your paper should not exceed four double-spaced typed pages.* Your paper should be typed. The four-page limit is not there to make your life difficult, but instead, the upper limit serves to limit the amount of work that you can usefully spend on this project.

For your paper, *use the following section format*. You are explicitly allowed and encouraged to make minor changes to this format if you think this is appropriate.

1. Statement of the problem
2. Review of the literature
3. The economic model
4. The econometric model
5. The data
6. The estimation and inference procedures
7. The empirical results and conclusions
8. Possible extensions and limitations of your study
9. Acknowledgements
10. References

*Your paper should be self-contained as much as is reasonably possible.* That is, the paper should be readable without prior knowledge of the dataset, and you should
introduce any nonstandard symbols or notation. An undergraduate econometrics student from say, University of Indiana should be able to read your paper and make sense of it.

The deadline to hand in your paper is Monday November 29, 2004. If you hand in your paper late without valid excuse, in order to be fair to the students who make a big effort to hand their paper in on time, I will take points away from the grade awarded for your paper. If you feel you should have a valid excuse for turning in your paper late, contact me before Monday November 29, at the first instance that you know you will not be able to meet the deadline.
Project descriptions

1. Math scores in Michigan high schools

The file

http://www.econ.ohio-state.edu/dejong/meap93.wf1

contains data for 408 Michigan high schools on performance on the MEAP tenth grade math test, school enrollment, teacher compensation, number of staff per 1000 students, poverty as measured by the number of students enrolled in the school lunch program, and expenditure per student. A description of this datafile can be found in

http://www.econ.ohio-state.edu/dejong/meap93.des

The question that we are interested in is whether students at smaller schools will really do better than students at larger schools. This is what is often suggested in the press, but it may be that this is due to not controlling for certain factors.

For this project, assess whether there is statistical evidence that students at larger schools will do worse than students at smaller schools, all else being equal. Regress the math score variable on the other variables and interpret your results. Also, use a specification that includes total compensation and staff size in logarithms rather than in levels and see if your conclusions change.

2. The Kuznets curve

The file

http://www.econ.ohio-state.edu/dejong/k.wf1

contains 107 observations on the variables "couname", "nyrs", "gini", "gnp", "icp", and "year". "couname" is a numeric code for the country name; a list of the countries used can be found at

http://www.econ.ohio-state.edu/dejong/k.des
Missing observations are coded as "NA". This datafile should be loaded using the simple command "infile using k.txt" (assuming that you put the file k.txt into the right directory). We can use this datafile to research a phenomenon that is known in the literature as the "Kuznets curve". Kuznets hypothesized that as a country develops, income inequality will first go up in the earlier stages of development, and will go down in later stages of the development of the country. We want to find out whether or not we can find empirical evidence for the Kuznets curve.

The variable "gini" contains observations on the so-called Gini coefficients of the country in a certain time period. The Gini coefficient is a measure of income inequality that is large when inequality is large. The variable "country" contains the name of the country that is listed for that particular observation. "nyrs" contains the number of observations that we have in our dataset on that particular country. "ie" stands for the "income concept" that was used; this variable is mostly equal to "i" (meaning "individual"), but for some countries different income concepts were used. "gnp" contains data on gnp of the country. "icp" is an alternative measure of prosperity of a country that - unlike gnp - takes into account exchange rates fluctuations. Both "gnp" and "icp" are corrected for inflation.

For this project you should do regressions of the variable "gini" on the variables "gnp" and on the square of gnp (a variable that you will have to create yourself). If a Kuznets curve is present, the coefficient for the square of gnp should be negative (why?). Small p-values for this variable will therefore be evidence in favor of the presence of a Kuznets curve. Try different specifications, take a good look at your dataset (perhaps you will want to delete observations that you think do not make sense), and try the variable "icp" instead of "gnp" and see whether or not your results improve.

3. Wage discrimination

At the internet location


you will find a description a datafile and a zipped data file. This data is about wages and characteristics of 8748 people living in Malaysia of different ethnicity (Malay, Indian, and Chinese). An article in a professional journal (Journal of Applied Econometrics) has appeared that analyzed these data. We are interested in analyzing wage discrimination in Malaysia, and casual conversations with Malaysian grad students suggests
the potential presence of such discrimination. Examine the end of the datafile carefully - it looked to me as though there was an obsolete number at the end of this datafile.

This datafile is too big to fit into the student edition of Eviews. Therefore, you should make as large a selection of datapoints that will fit into Eviews.

There is a substantial literature in Labor Economics about the estimation of wage equations, and for this project it should not be too difficult to write a decent overview of the literature.