ID#:

EXAM TWO

K Foster, Statistics and Introduction to Econometrics, Eco B2000, CCNY, Fall 2013

The questions are worth 120 points. You have 120 minutes to do the exam, one point per minute.

All answers should be put into the blue books or submitted electronically. Please submit computer files via Blackboard and email. No need to put your name, just last 4 digits of ID to identify yourself, so grading is blind.

You may refer to your books, notes, calculator, computer, or astrology table. The exam is "open book." However, **you must not refer to anyone else, either in person or electronically**! You must do all work on your own. Cheating is harshly penalized. Please silence all electronic noisemakers such as mobile phones. Good luck. Stay cool.

Exam has 3 pages.

- 1. (15 points) You might find it useful to sketch the distributions.
 - a. If a variable has a Normal Distribution with mean 9 and standard deviation 9, what is area to the right of -8.3?
 - b. For a Normal Distribution with mean 5 and standard deviation 0.4, what is area to the left of 4.7?
 - c. For a Normal Distribution with mean 6 and standard deviation 0.3, what is area in both tails farther from the mean than 5.7?
 - d. For a Normal Distribution with mean -2 and standard deviation 3.8, what is area in both tails farther from the mean than 2.9?
 - e. For a Normal Distribution with mean 6 and standard deviation 7.5, what is area in both tails farther from the mean than -2.3?
 - f. For a Normal Distribution with mean 14 and standard deviation 3.4, what values leave probability 0.292 in both tails?
 - g. For a Normal Distribution with mean 8 and standard deviation 2.6, what values leave probability 0.253 in both tails?
 - h. For a Normal Distribution with mean -11 and standard deviation 2.6, what values leave probability 0.420 in both tails?
 - i. For a Normal Distribution with mean 2 and standard deviation 4.7, what values leave probability 0.007 in both tails?
 - j. For a Normal Distribution with mean -10 and standard deviation 7.9, what values leave probability 0.156 in both tails?
- 2. (15 points)
 - a. A regression coefficient is estimated to be equal to 1.902 with standard error 1.5; there are 26 degrees of freedom. What is the p-value (from the t-statistic) against the null hypothesis of zero?
 - b. A regression coefficient is estimated to be equal to 12.942 with standard error 9.6; there are 8 degrees of freedom. What is the p-value (from the t-statistic) against the null hypothesis of zero?
 - c. A regression coefficient is estimated to be equal to 3.647 with standard error 2.6; there are 15 degrees of freedom. What is the p-value (from the t-statistic) against the null hypothesis of zero?
 - d. A regression coefficient is estimated to be equal to -5.130 with standard error 3.5; there are 17 degrees of freedom. What is the p-value (from the t-statistic) against the null hypothesis of zero?

- e. A regression coefficient has standard error 2.40; there are 14 degrees of freedom. The tstatistic is 2.5994. What is the coefficient?
- f. A regression coefficient has standard error 3.40; there are 28 degrees of freedom. The tstatistic is -1.4877. What is the coefficient?
- g. A regression coefficient has standard error 2.30; there are 12 degrees of freedom. The tstatistic is -1.0175. What is the coefficient?
- h. A regression coefficient is estimated to be equal to 11.219; there are 7 degrees of freedom. The t-statistic is 1.6259. What is the standard error?
- 3. (20 points) A recent research paper, looking at how much attractiveness and personal grooming affects wages, used data from The National Longitudinal Study of Adolescent Health in 2001-2.
 - a. Are there gender differences? Among the 6074 people (48.4% female), 38.8% of the males were rated as being well groomed or very well groomed; 50.6% of the females were rated that way. Is this a statistically significant difference?
 - b. The study considers interrelations between physical attractiveness and grooming. People were ranked on a 4-point scale (where 1 is below average, 2 is average, 3 is above average, and 4 is very much above average) for each attribute. The full details are:

	Physically					
	4 Very Attractive	3 Attractive	2 Average	1 Less Attractive		
4 Very well groomed	297	199	57	30		
3 Well groomed	290	1169	607	54		
2 Average grooming	75	788	2013	167		
1 Less than average grooming	1	25	164	138		

Conditional on a person being ranked physically 3 or 4 in attractiveness (above average), what is the chance that they are above average (3 or 4) in grooming as well. Conditional on being above average physically, what is the chance that they are average or below average (1 or 2) in grooming? Are these statistically significantly different?

c. The study also considers the attractiveness of someone's personality (charisma), with the same 4-point scale. These data are:

	Personality					
	4 Very Attractive	3 Attractive	2 Average	1 Less Attractive		
4 Very well groomed	326	171	60	26		
3 Well groomed	416	1186	467	51		
2 Average grooming	212	966	1729	136		
1 Less than average grooming	11	49	184	84		

Conditional on having an above-average personality, what is the chance that someone has above-average grooming? Conditional on having an above-average personality, what is the chance that their grooming is at or below average? Is there a statistically significant difference?

d. Comment on the study. If overall attractiveness is a combination of these 3 factors, is there evidence that they are gross substitutes or complements in production?

PK Robins, JF Homer, MT French (2011). "Beauty and the Labor Market: Accounting for the Additional Effects of Personality and Grooming," Labour, 25(2), pp 228-251.

The next two questions ask you to use the dataset, scf2010 data exam2. This is the Federal Reserve's Survey of Consumer Finances (SCF), which is oddly weighted (so don't worry that the averages seem a bit off! They oversample rich people since they have interesting financial portfolios) but has interesting data on the financial situation of households.

- 4. (25 points) Consider a regression to explain household liquidity (the variable LIQ, which adds up checking account, saving account, and other liquid assets very much what we might consider M1 in macro). Carefully designate the range of data that you will explore, then show and describe simple statistics, and only then create one (or more) interesting regressions. (Interesting regressions might include polynomial terms, interactions, quantiles, nonparametric, etc.) Explain why each variable is in the model and consider whether it is endogenous. Describe what you learn from the regression model.
- 5. (25 points) With the same dataset, look at the ratio of monthly house debt payments to income (PIRMORT) and the ratio of consumer debt payments to income (PIRCONS). For housing debt, many financial planner suggest a ratio of higher than 33% is worrisome. Create a dummy o/1 variable for whether a household has mortgage payments more than 33% of income. Show descriptive statistics. Then estimate interesting probit and/or logit regressions to explain this variation. ("Interesting" as defined in previous question.) Again explain your variable choices. Explain what you learn from these models.

For the next question, I've combined the dataset on beer consumption and taxes with data^{*} from Prof Nagler on car accidents, fatalities, and social capital (thanks to him for the generosity!). Now we have data on the fraction of population over 65, with high school or bachelor's degree, fraction African American, rates of divorce and suicide, traffic fatalities (number and rate) overall and in summer, miles of road and fraction unpaved, population density and per road mile, gas price and gas stations per population, attendance at church and if the fraction who pray daily, and many responses to 'values' questions: is it OK to cheat on taxes, is life dull, are most people trusted or fair, etc. These are in R (use ls(all = TRUE) to see the variables) and SPSS; if you use SPSS note that you can save predicted values with the button "Save..." then choose to save the unstandardized predicted values.

6. (20 points) Use instrumental variables techniques to try some of these new variables as instruments and/or controls in interesting regression specification(s) about economic growth rates across US states. Again ensure that you designate the range of data and show simple statistics. Explain about the rationale for the instrument(s).

* Actually his data was a bit more complicated with some missing values, etc – so I filled in those with random numbers, so this is really not quite his data, I've messed with it. But there are still some variables with missing values, which might trip you up – be careful.