

Final Exam – Monday December 12, 2004

3 hours – 40 % of the final grade – scores add up to 40

Instructions:

- Answer the question in the space provided in this exam.
- You may use your books and notes
- Show your work
- Be careful when you make calculations!
- Answer all the different parts of the questions
- Explain your answers
- Sign the below statement

I understand the Honor Code and pledge to abide by it. I will not talk with anybody about the exam until the 14th of December, 2005.

Print Name: _____

Sign Name: _____

Date: _____

A. Discrimination and (French) fries (20 points)

Fast-food restaurants have been accused of many “bad” practices over the years. One of the accusations involved practices of price discrimination against African Americans. In order to analyze these accusations, a researcher collected information on the price of a small portion of French fries and the price of a medium soda from a random sample of fast-food restaurants in New Jersey (NJ) and Pennsylvania. He found a substantial price difference between the prices of fries in different restaurants and wanted to analyze whether there is evidence of higher prices in restaurants located in neighborhoods with a higher proportion of African Americans.

He collected data on 410 restaurants from 4 different chains: Burger King, Kentucky Fried Chicken, Roy Rodgers, and Wendy’s. He also collected information that characterizes the neighborhood¹ in which the fast food restaurant is located: specifically he knows the proportion of black population (measured between 0 and 1) median family income (in US\$), and the proportion of population without a car (between 0 and 1). He also has information on the crime rate (between 0 and 1) and the population density of the town where the restaurant is located (nr of people per square mile). He presents the regression output in table A.1. (see attach)

- 1) Give a careful interpretation of the coefficients in the second column (sign, magnitude, significance). (3 points)

¹ Neighborhood’s were defined as areas with the same zip code

6) If you had to choose one of the 5 specifications, which one would you chose? Explain in detail which tests you would conduct before making your decision. (5 points)

7) Explain why the researcher might want to present 5 different specifications, as opposed to just showing us one estimation result. (1 point)

8) The author also knows in which county each restaurant is located. Specifically, the sampled restaurants are from a total of 29 different counties. He has a variable *county* that assigns a number to each of these counties.

- a. Explain why it would be a good idea to incorporate this information in the regression. (2 points)
- b. Explain how he can incorporate this information in the regression. (1 point)

- 3) Do you think it was a good idea to use robust standard errors in the first regression? And is there a more efficient way to estimate the model ? Explain. (3 points)
- 4) Do you think it was a good idea to use robust standard errors in the second regression? Explain carefully why or why not. (2 points)
- 5) Give a careful interpretation of the coefficient of *usemeth* in regression 2. Discuss whether you find the result intuitive. (2 points)
- 6) The researcher learns that there has been a family planning program in Botswana in the year before the survey. Because of lack of funds, the program could not cover the entire country, and therefore went only to a subset of randomly selected locations, and distributed free contraceptives to all women in these locations. The researcher knows which of the women in her sample live in the selected locations. Explain how she could use this information to improve the estimated regressions. Write out the model(s) she would estimate and explain. (3 points)

- 7) Imagine the researcher wants to predict what type of women are likely to have more than 2 children. Discuss the potential problems she might face when she uses the estimates in regression
2. Discuss what she can do to solve these problems. (2 points).

C. True or False (4 points)

- 1) Econometricians often refer to violations of the zero conditional mean assumptions as problems of identification. T/F? Explain. (2 points)
- 2) Adding lagged independent variables in a time-series model, can help reduce the bias of the coefficient estimates but might also lead to a decline in precision of the coefficient estimates. T/F? Explain. (2 points)

D. Grade your teammates:

On scale from 1(=a little) to 5(=a lot), how much did they contribute to the problem sets?

Name 1: _____ Grade: _____

Name 2: _____ Grade: _____

Regression results

Part A:

Table A1: Determinants of fries prices in fast food restaurants:

Dependent variable: log (price) of small fries

	(1)	(2)	(3)	(4)	(5)
proportion black population	0.039 (1.23)	0.138 (3.96)	-0.024 (0.62)	-0.017 (0.51)	-0.327 (1.47)
log(median family income)		0.130 (5.71)	0.166 (5.42)	0.114 (4.20)	0.118 (4.33)
crime rate			0.256 (1.53)	0.186 (1.27)	0.175 (1.19)
population density			0.000 (3.65)	0.000 (3.29)	0.000 (3.23)
proportion of population without car			0.153 (1.46)	0.135 (1.46)	0.153 (1.64)
New Jersey dummy			0.083 (5.94)	0.090 (7.33)	0.078 (5.09)
(New Jersey dummy)*proportion black pop.					0.312 (1.41)
Burger King dummy				0.034 (2.53)	0.033 (2.43)
Kentucky Fried Chicken dummy				-0.029 (1.87)	-0.030 (1.95)
Roy Rogers dummy				0.116 (7.68)	0.115 (7.65)
Constant	-0.092 (13.42)	-1.493 (6.08)	-1.980 (6.00)	-1.461 (4.97)	-1.492 (5.07)
Observations	392	392	391	391	391
R-squared	0.00	0.08	0.25	0.43	0.43

Absolute value of t statistics in parentheses

Part B: Stata output

* VARIABLE DEFINITION AND DESCRIPTIVE STATISTICS

Contains data from C:\karen\teaching\metrics_teaching\metrics_data\FERTIL2.DTA

obs: 4,361
vars: 7 17 Aug 1999 15:26
size: 61,054 (94.2% of memory free)

variable name	storage type	display format	value label	variable label
age	byte	%8.0g		age in years
educ	byte	%8.0g		years of education
children	byte	%8.0g		number of living children
usemeth	byte	%8.0g		=1 if ever use birth control
heduc	byte	%8.0g		husband's years of education
urban	byte	%8.0g		=1 if live in urban area
kids2	float	%9.0g		=1 if more than 2 kids

Sorted by:

. sum

Variable	Obs	Mean	Std. Dev.	Min	Max
age	4361	27.40518	8.685233	15	49
educ	4361	5.855996	3.927075	0	20
children	4361	2.267828	2.222032	0	13
usemeth	4290	.5776224	.4939956	0	1
heduc	1956	5.144683	4.803028	0	20
urban	4361	.5166246	.4997808	0	1
kids2	4361	.3728503	.4836182	0	1

*** REGRESSION 1**

. regress children age educ heduc urban usemeth, robust

Linear regression

Number of obs = 1908
 F(5, 1902) = 270.88
 Prob > F = 0.0000
 R-squared = 0.4424
 Root MSE = 1.6999

children	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
age	.1618161	.0058674	27.58	0.000	.150309	.1733233
educ	-.102144	.0116235	-8.79	0.000	-.1249403	-.0793478
heduc	-.0499712	.0099902	-5.00	0.000	-.0695641	-.0303782
urban	-.3694232	.082796	-4.46	0.000	-.5318037	-.2070427
usemeth	1.099197	.093251	11.79	0.000	.9163123	1.282082
_cons	-1.469215	.1797136	-8.18	0.000	-1.821671	-1.116758

*** REGRESSION2**

. regress kids2 age educ heduc urban usemeth, robust

Linear regression

Number of obs = 1908
 F(5, 1902) = 253.35
 Prob > F = 0.0000
 R-squared = 0.3104
 Root MSE = .40652

kids2	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
age	.0278685	.0012127	22.98	0.000	.0254901	.0302469
educ	-.0187484	.0029311	-6.40	0.000	-.0244969	-.013
heduc	-.0107396	.0025848	-4.15	0.000	-.0158089	-.0056703
urban	-.0730406	.0201459	-3.63	0.000	-.1125511	-.0335302
usemeth	.2431552	.0204449	11.89	0.000	.2030584	.283252
_cons	-.2504485	.0452028	-5.54	0.000	-.3391008	-.1617961