Panteion University
Applied Econometrics

Gregory Kordas
Spring 2023

Final Exam

Instructions: Three of the following questions will be on the Final Exam.

1. [**Joint, Marginal and Conditional Probabilities**]. Let

$$f_{X|Y}(x|y) = \begin{cases} c_1 x/y^2, & 0 < x < y < 1, \\ 0, & \text{otherwise}, \end{cases}$$

be the conditional p.d.f of $X|Y$ and

$$f_Y(y) = \begin{cases} c_2 y^4, & 0 < y < 1, \\ 0, & \text{otherwise}, \end{cases}$$

be the marginal p.d.f of Y. Determine

(a) The constants c_1 and c_2.
(b) The joint p.d.f of X and Y.
(c) $\Pr(\frac{1}{4} < X < \frac{1}{2} | Y = \frac{5}{8})$
(d) $\Pr(\frac{1}{4} < X < \frac{1}{2})$
(e) $E(X|Y)$.
(f) The cdf and pdf of $Z = E(X|Y)$, F_Z and f_z, respectively.

2. [**Joint, Marginal and Conditional Probabilities**]. Let

$$f_{XY}(x, y) = cx^3y^2, \qquad 0 < x < 1,\ 0 < y < 2.$$

be the joint probability density function of X and Y.

(a) Find c that makes $f(x, y)$ a valid probability density function.
(b) Find $g_{Y|X}(y|x)$, the conditional probability density function of $Y|X$.
(c) $\Pr(\frac{1}{3} < X < \frac{2}{3} | Y = \frac{2}{3})$.
(d) Find $Cov(X, Y)$, the covariance of X and Y.
(e) Are X and Y stochastically independent? Justify your answer.
(f) Let $Z = X^2 + Y^2$. Find $E(Z)$, the expected value of Z.

3. [**Least Squares Identities**]. Prove that in the linear regression model $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u}$ where $\boldsymbol{X}$ includes an intercept (a column of 1's as the first regressor), the OLS plane $\widehat{\boldsymbol{y}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}$ has the following mathematical properties:

(a)

$$\overline{\boldsymbol{x}}'\widehat{\boldsymbol{\beta}} = \overline{y}.$$

where $\overline{\boldsymbol{x}} = (1, \overline{x}_1, ..., \overline{x}_k)'$ is the $k \times 1$ vector of means of the independent variables x_j, $j = 1, ..., k$. This means that the point $(\overline{y}, \overline{\boldsymbol{x}}) \in \mathbb{R}^{k+1}$ satisfies the normal equations, and therefore the OLS plane always passes through the sample means when the regression includes a conscant term. We say that OLS passes through the "center-of-gravity" $(\overline{y}, \overline{\boldsymbol{x}})$ of the sample.

(b)

$$\overline{\widehat{y}} = \overline{y},$$

that is, the mean of the fitted values $\widehat{\boldsymbol{y}}$ equals the mean of $\boldsymbol{y}$.

(c)

$$\mathbf{1}'\widehat{\boldsymbol{u}} = \overline{u} = 0,$$

that is, the sum and the mean of the OLS residuals is zero.

(d)

$$\widehat{\boldsymbol{y}}'\widehat{\boldsymbol{u}} = 0 \qquad \text{or} \qquad \widehat{\boldsymbol{y}} \perp \widehat{\boldsymbol{u}}.$$

that is, the OLS fitted values $\widehat{\boldsymbol{y}}$ and the OLS residulas $\widehat{\boldsymbol{u}}$ are orthogonal vectors.

[Hint: See Stavrinos, ch.3.]

4. [**Linear Regression Model under Endogeneity**]. Consider the linear regression model

$$y = X\beta + u$$

where, y is an $n \times 1$ vector, X is an $n \times k$ matrix of regressors (including an intercept), β is a $k \times 1$ vector of coefficients, and u is an $n \times 1$ vector of errors.

(a) State the classical assumptions and briefly explain them.

(b) Which of the above assumptions is violated when a regressor is endogenous? Give an example of a regression in which the problem is likely to arise.

(c) What are the properties of the OLS estimates under endogeneity?

(d) Which estimator should you use in this case, and what are its properties?

5. [**Long and Short Regressions**].

(a) Assume that the true linear regression model explaining y is given by

$$y = X_1\beta_1 + X_2\beta_2 + u$$

where, y is an $n \times 1$ vector, X_1 is a $n \times k_1$ matrix of regressors (including an intercept), X_2 is a $n \times k_2$ matrix of regressors, β_1 is a $k_1 \times 1$ vector of coefficients, β_2 is a $k_2 \times 1$ vector of coefficients, and u is an $n \times 1$ vector of errors. Instead of estimating the true model, we estimate by OLS the *short* model

$$y = X_1\beta_1 + u.$$

What are the properties of the OLS estimate $\widehat{\beta}_1$?
[Hint: Write the OLS estimator for β_1 and compute its expectation using the true model for y. See Stavrinos, section 4.4, p.143-146].

(b) Now consider the opposite situation where the true model for y is given by

$$y = X_1\beta_1 + u$$

we estimate by OLS the *long* model

$$y = X_1\beta_1 + X_2\beta_2 + u$$

What are the properties of the OLS estimate $\widehat{\beta}_1$ in this case?
[Hint: We can write $\widehat{\beta}_1 = (X_1'M_2X)^{-1}X_1M_2y$, where $M_2 = I - X_2(X_2'X_2)^{-1}X_2'$ is an indempotent matrix that projects into the space of X_2 residuals, $S^{\perp}(X_2)$. Now take the expectation using the true model for y. See Stavrinos, section 4.4, p.143-146]

6. [**Structural Change**]. Consider the classical time-series linear regression model

$$y = X\beta + u, \qquad u \sim \text{iid}N(0, \sigma^2 I).$$

where y is an n vector, X is a $n \times k$ matrix of order k (full order), β is a k vector of coefficients, and u is a homoskedastic normal error term.

Recall that the general linear hypothesis may be written as

$$H_0 : R\beta = r$$

where R is a $q \times k$ restriction matrix (with $q < k$), and r is a q vector of known constants.

(a) Starting from the fact that in this model the OLS estimate b is distributed as

$$b \sim N(\beta, \sigma^2(X'X)^{-1})$$

(explain why) show that under the null

$$(Rb - r)'[\sigma^2 R(X'X)^{-1}R']^{-1}(Rb - r) \sim \chi^2(q).$$

(b)) Using the fact that (explain why)

$$\frac{u'u}{\sigma^2} \sim \chi^2(n-k),$$

determine the distribution of the statistic

$$D = \frac{(Rb-r)'[R(X'X)^{-1}R']^{-1}(Rb-r)/q}{u'u/(n-k)}.$$

Now consider OLS estimation under the constraint. The restricted least squares (ROLS) estimator b_* minimizes the Lagrangian

$$(y-Xb)'(y-Xb) - 2\lambda'(Rb-r)$$

where λ is a q vector of Lagrange multipliers.

(c) Show that the ROLS estimator is given by

$$b_* = b + (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(r-Rb).$$

(d) Writing u for the OLS residuals and u_* for the ROLS residuals first show that

$$u_*'u_* = u'u + (b_*-b)'X'X(b_*-b)$$

and then that

$$u_*'u_* - u'u = (r-Rb)'[R(X'X)^{-1}R']^{-1}(r-Rb).$$

Thus, our statistic above may be written as

$$D = \frac{(u_*'u_* - u'u)/q}{u'u/(n-k)}.$$

Explain briefly the intuition for this statistic and give its theoretical distribution under the null.

Now consider the situation where a researcher is worried that at some specified moment of time a *structural change* has occurred, that resulted in a shift in β. Let $y_i, X_i, i = 1,2$ indicate the partitioning of the data into the two subperiods, which we will call *peace time* and *war time*, and consider the model

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

where $\beta_i, i = 1,2$ are the relevant k vectors of coefficients for the subperiods and $u_i, i = 1,2$ are also iid with common variance σ^2. We also assume that the X_i's are of full order too. We are interested in testing the null hypothesis

$$H_0 : \beta_1 = \beta_2$$

(e) Specify R and r for this hypothesis.

(f) Describe the process you would use to test this hypothesis given a sample of $n = n_1+n_2$ observations, and give the test statistic and its theoretical distribution under the null.

[Hint: This the Chow Test for Structural Change. See Stavrinos, sec. 3.13, p. 104]

7. Consider the logit model for the survival of the passengers on the Titanic, as we discussed it in class.

TABLE 1. Logit Model 1

Variable	Coefficient	Std. Error	Odds Ratio	Std. Error
Child	1.062	.277	2.8908	.705
Female	2.420	.136	11.247	1.579
1st Class	-0.376	.126	0.6864	.093
2nd Class	-1.394	.129	0.2480	.039
3rd Class	-2.154	.144	0.1160	.015
Crew	-1.234	.080	0.2912	.023

(a) Based on the model in the lecture notes, compute the survival odds of a passenger traveling 1st class relative to a passenger traveling 3rd class. Prove any formulas you use.

(b) Give a 95% CI for the survival odds estimate in (a) (Hint: Use the bootstrap.)

8. Let $X \sim U[0, 1]$ be uniformly distributed on the interval $[0, 1]$.

(a) Find the probability distribution function, the cumulative distribution function, and the quantile function of $Y = -b \log X$.

(b) Find the median of the distribution in (a).

(c) Find the moment generating function of the distribution in (a).

(d) Let $(Y_1, ..., Y_n)$ be a random sample from the distribution in (a). Find the mle of b and its asymptotic distribution.

9. Consider a random variable X from the Pareto(a, c) distribution with pdf

$$f(x) = \frac{ca^c}{x^{c+1}}, \qquad x \geq a$$

where, $a > 0$ is a location parameter, and $c > 0$ is a shape paramater.

(a) Plot the pdf for $(a, c) = (1, 1)$, $(a, c) = (1, 2)$, and $(a, c) = (1, 3)$.

(b) Find the cdf and quantile function (qf) of X.

(c) Find $E(X)$ and $Var(X)$. Show that $E(X)$ exists only for $c > 1$, and $Var(X)$ exists only for $c > 2$.

(c) Justify your findings in (c) in terms of the fatness of the right tail (see Lecture 3)

(d) Let $X_1, ..., X_n$ be random sample from the Pareto(a, c) distribution. Find the mles for c and a. Is the asymptotic distribution of these mles normal? Justify your answer.

10. Consider the IV model used in THOMAS G. HANSFORD and BRAD T. GOMEZ, "Estimating the Electoral Effects of Voter Turnout", *The American Political Science Review*, Vol. 104, No. 2 (May 2010), pp. 268-288. The paper examines the electoral consequences of variation in voter turnout in the United States. The authors examine several hypotheses about the behavior of US voters but we will focus in the:

Partisan Effect Hypothesis: Increases in turnout lead to increases in the Democratic candidate's vote share.

A simplified model of their analysis is given by

$$\text{DemoShare}_{it} = \beta_0 + \beta_1 \text{Turnout}_{it} + \mu_t + u_{it}$$

where,

- Demoshare_{it} : Two-party vote share for Democratic candidate in county i in the presidential election in year t.
- Turnout_{it} : Turnout rate in county i in the presidential election in year t.
- μ_t : Year fixed effects. Time dummies for each presidential election year.
- u_{it} : iid error term.

(a) What would you expect about the coefficients in this regression if the *Partisan Effect Hypothesis* is true?

(b) Why would one suspect the variable Turnout to be endogenous (i.e., correlated with the error term)? [Hint: see paper]

(c) In the paper, the authors instrument Turnout with the variable Rain (DNorm-Prcp_KRIG) which measures the precipitation above the expected (average) amount for the day of the election. Justify this choice of instrument. [Hint: see paper]

(d) Run the OLS and IV regression to obtain the results below. Describe what we find.

```
> # Load packages we will use (install first if not already installed)
> # install.packages("AER")
> # install.packages("readr")
> # install.packages("stargazer")
> library(AER)
> library(readr)
```

```
> library(stargazer)

> # Read csv datafile
> HGdata <- read_csv("HansfordGomez_Data.csv")

> # Inspect the data - sample stats
> stargazer::stargazer(as.data.frame(HGdata), type="text")

=========================================================================
Statistic           N      Mean       St. Dev.      Min        Max
-------------------------------------------------------------------------
Year             27,401  1,973.972     16.111      1,948      2,000
FIPS_County      27,401 29,985.500   13,081.250    4,001      56,045
Turnout          27,401   65.562       10.514     20.366     100.000
Closing2         27,401   23.053       13.042      0.000     125.000
Literacy         27,401    0.058        0.234        0          1
PollTax          27,401    0.001        0.023        0          1
Motor            27,401    0.211        0.408        0          1
GubElection      27,401    0.434        0.496        0          1
SenElection      27,401    0.680        0.467        0          1
GOP_Inc          27,401    0.501        0.500        0          1
Yr52             27,401    0.071        0.258        0          1
Yr56             27,401    0.071        0.258        0          1
Yr60             27,401    0.071        0.258        0          1
Yr64             27,401    0.071        0.258        0          1
Yr68             27,401    0.071        0.258        0          1
Yr72             27,401    0.071        0.258        0          1
Yr76             27,401    0.071        0.258        0          1
Yr80             27,401    0.071        0.258        0          1
Yr84             27,401    0.072        0.258        0          1
Yr88             27,401    0.072        0.258        0          1
Yr92             27,401    0.072        0.258        0          1
Yr96             27,401    0.072        0.258        0          1
Yr2000           27,401    0.070        0.256        0          1
DNormPrcp_KRIG   27,401    0.005        0.208     -0.419      2.627
GOPIT            27,401   33.282       34.066      0.000     100.000
DemVoteShare2_3MA 27,401  44.250       10.606     10.145      88.982
DemVoteShare2    27,401   43.622       12.415      6.420      97.669
RainGOPI         27,401    0.007        0.142     -0.407      2.234
TO_DVS23MA       27,401  2,886.877     792.530    473.161   8,526.616
Rain_DVS23MA     27,401    0.355       10.188    -25.054    144.257
dph              27,401    0.021        0.145        0          1
dvph             27,401    0.018        0.133        0          1
rph              27,401    0.025        0.155        0          1
rvph             27,401    0.025        0.155        0          1
state_del        27,401    0.037        0.187     -0.821      0.619
dph_StateVAP     27,401 77,525.150   597,474.000     0      6,150,988
dvph_StateVAP    27,401 63,138.400   663,707.600     0      12,700,000
rph_StateVAP     27,401 243,707.900 1,720,659.000 0.000  18,300,000.000
rvph_StateVAP    27,401 142,166.500 1,071,445.000    0      12,800,000
State_DVS_lag    27,401   46.896        8.317     22.035      80.872
State_DVS_lag2   27,401  2,268.381     786.199    485.533   6,540.244
-------------------------------------------------------------------------
```

```
> # OLS regression
> hg_ols <- lm( DemVoteShare2 ~ Turnout + factor(Year)   ,  data = HGdata)
> #coeftest(hg_ols, vcov = vcovHC, type = "HC1")
>
> # Iv regression
> hg_ivreg <- ivreg( DemVoteShare2 ~ Turnout + factor(Year) |
+ factor(Year) + DNormPrcp_KRIG, data = HGdata)
> #coeftest(hg_ivreg, vcov = vcovHC, type = "HC1")
>
> # Show result
> stargazer(hg_ols, hg_ivreg, type ="text")

=========================================================================
                                             Dependent variable:
                                  ---------------------------------------
                                                DemVoteShare2
                                             OLS              instrumental
                                                                variable
                                             (1)                 (2)
-------------------------------------------------------------------------
Turnout                                   -0.157***            0.363**
                                           (0.007)             (0.175)

factor(Year)1952                         -10.215***          -15.832***
                                           (0.345)             (1.928)

factor(Year)1956                          -8.756***          -13.656***
                                           (0.343)             (1.692)

factor(Year)1960                          -3.862***          -11.094***
                                           (0.350)             (2.464)

factor(Year)1964                          10.851***            6.837***
                                           (0.341)             (1.402)

factor(Year)1968                          -6.477***           -8.514***
                                           (0.338)             (0.780)

factor(Year)1972                         -13.749***          -16.473***
                                           (0.338)             (0.989)

factor(Year)1976                           -0.367             -2.111***
                                           (0.337)             (0.694)

factor(Year)1980                         -10.346***          -11.696***
                                          (0.337)             (0.586)

factor(Year)1984                         -13.134***          -13.515***
                                           (0.336)             (0.391)

factor(Year)1988                          -5.712***           -4.951***
                                          (0.337)             (0.450)

factor(Year)1992                           -0.327             -1.008**
```

```
                                             (0.337)          (0.435)

factor(Year)1996                            -1.193***          0.811
                                             (0.337)          (0.770)

factor(Year)2000                            -9.013***        -8.130***
                                             (0.338)          (0.476)

Constant                                    59.085***         26.910**
                                             (0.487)          (10.843)

------------------------------------------------------------------------
Observations                                 27,401           27,401
R2                                            0.281            0.130
Adjusted R2                                   0.280            0.130
Residual Std. Error (df = 27386)             10.533           11.582
F Statistic                         763.153*** (df = 14; 27386)
========================================================================
Note:                                          *p<0.1; **p<0.05; ***p<0.01
```

Data description:

Name	Description
Year	Election Year
FIPS_County	FIPS County Code
Turnout	Turnout as Pcnt VAP
Closing2	Days b/w registration closing date and election
Literacy	Literacy Test
PollTax	Poll Tax
Motor	Motor Voter
GubElection	Gubernatorial Election in State
SenElection	U.S. Senate Election in State
GOP_Inc	Republican Incumbent
Yr52	1952 Dummy
Yr56	1956 Dummy
Yr60	1960 Dummy
Yr64	1964 Dummy
Yr68	1968 Dummy
Yr72	1972 Dummy
Yr76	1976 Dummy
Yr80	1980 Dummy

Yr84	1984 Dummy
Yr88	1988 Dummy
Yr92	1992 Dummy
Yr96	1996 Dummy
Yr2000	2000 Dummy
DNormPrcp_KRIG	Election day rainfall - differenced from normal rain for the day
GOPIT	Turnout x Republican Incumbent
DemVoteShare2_3MA	Partisan composition measure = 3 election moving avg. of Dem Vote Share
DemVoteShare2	Democratic Pres Candidate's Vote Share
RainGOPI	Rainfall measure x Republican Incumbent
TO_DVS23MA	Turnout x Partisan Composition measure
Rain_DVS23MA	Rainfall measure x Partisan composition measure
dph	=1 if home state of Dem pres candidate
dvph	=1 if home state of Dem vice pres candidate
rph	=1 if home state of Rep pres candidate
rvph	=1 if home state of Rep vice pres candidate
state_del	avg common space score for the House delegation
dph_StateVAP	= dph*State voting age population
dvph_StateVAP	= dvph*State voting age population
rph_StateVAP	= rph*State voting age population
rvph_StateVAP	= rvph*State voting age population
State_DVS_lag	State-wide Dem vote share, lagged one election
State_DVS_lag2	State_DVS_lag squared

I was gratified to answer promptly. I said I don't know.

— *Mark Twain*, *Life on the Mississippi.*