M. Sc. Program in International Economics and Finance
Otto-von-Guericke University, Magdeburg
Prof. Dr. Karl-Heinz Paqué

Advanced Econometrics: a guided tour
(20308 Econometrics)
Final Exam
Winter term 2011/12 - February 15, 2012

Antonio Fidalgo, Ph.D. candidate
Faculty of Business and Economics
University of Lausanne
Lausanne, Switzerland

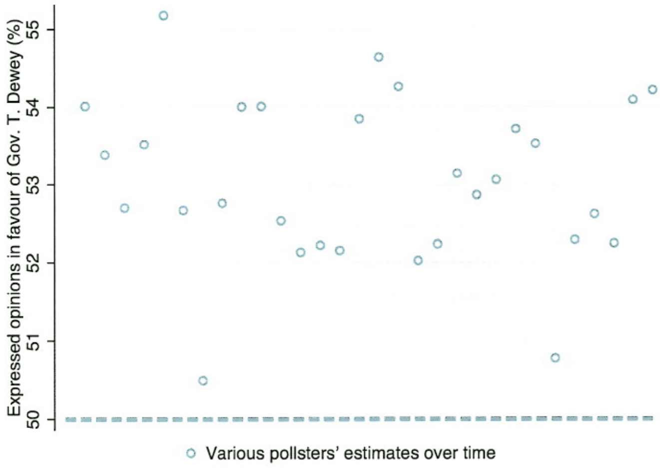
Instructions

- No textbook, lecture note or any other piece of documentation is permitted. A dictionary for language issues is allowed.
- This exam contains 7 problems. You must solve all of them.
- Each problem is worth 12.5 points, except for Problem 3 that has 25 points. Hence, you can get a maximum of 100 points.
- For your answers, use exclusively the working sheets provided. Nothing you would write on these problems' sheets will be considered in the grading.
- The duration of the exam is 120 minutes. Allocate your time wisely.

Good luck!

Problem 1. "Dewey defeats Truman"

Figure 1: Illustrative view of the results obtained by the polls in Problem 1.



In 1948, Harry S. Truman was elected President of the United States of America. This came as a huge surprise to most, if not all, of the political commentators[1]. Indeed, the major pollsters such as Gallup, Crossley and Roper had no doubt when predicting an easy victory of Gov. T. Dewey at the U. S. presidency.
Figure 1 is simply an illustrative view[2]. It depicts the results of the polls obtained via phone from several large samples of randomly selected American citizens who faithfully reported their opinion.
Consider this problem from an econometric point of view. Given the data in Figure 1, you can get the ordinary least squares (OLS) estimate of the percentage of voices in favour of Dewey (as predicted by the polls) by estimating the model

$$y_i = \alpha + \varepsilon_i \tag{1}$$

i.e. by regressing the results of the polls (y_i's) on a constant[3].

Required. Explain with sufficient detail and clarity why estimating the regression model (1) in this case will produced a biased/ inconsistent estimate of the true α. Notice that the lack of an answer to that question explains why all the commentators were fooled by the polls.

[1]The *Chicago Daily Tribune*, for instance, didn't wait for the official results. A famous picture shows President Truman holding that newspaper on the day after the election with the quote above in his front page.

[2]The actual numbers don't really matter for our purpose here. It turns out I generated the data in Figure 1 with the command: `gen y=53+rnormal()`.
It is important to emphasize, however, that the expected percentage of voices for Gov. T. Dewey as predicted by the pollsters was way above 50.

[3]Obviously, taking the mean of the observations would produce the same estimate.

Problem 2. Where Stata has it wrong

The Stata command `kdensity` offers to produce a nonparametric, kernel-based estimate of a univariate density. Figures 2 and 2 (cont'd) reproduce the help-file for that command.

Required. Based on that document, explain with sufficient detail and clarity why this software provides unreliable estimates of a univariate density.

Figure 2: Stata help-file for `kdensity`, pp.1-2

Stata 12 help for kdensity

```
Title

    [R] kdensity -- Univariate kernel density estimation


Syntax

        kdensity varname [if] [in] [weight] [, options]

    options                           Description
    ----------------------------------------------------------------------
    Main
      kernel(kernel)                  specify kernel function; default is
                                        kernel(epanechnikov)
      bwidth(#)                       half-width of kernel
      generate(newvar_x newvar_d)     store the estimation points in newvar_x
                                        and the density estimate in newvar_d
      n(#)                            estimate density using # points; default
                                        is min(N, 50)
      at(var_x)                       estimate density using the values
                                        specified by var_x
      nograph                         suppress graph

    Kernel plot
      cline_options                   affect rendition of the plotted kernel
                                        density estimate

    Density plots
      normal                          add normal density to the graph
      normopts(cline_options)         affect rendition of normal density
      student(#)                      add Student's t density with # degrees of
                                        freedom to the graph
      stopts(cline_options)           affect rendition of the Student's t
                                        density

    Add plots
      addplot(plot)                   add other plots to the generated graph

    Y axis, X axis, Titles, Legend, Overall
      twoway_options                  any options other than by() documented in
                                        [G-3] twoway_options
    ----------------------------------------------------------------------

    kernel                            Description
    ----------------------------------------------------------------------
    epanechnikov                      Epanechnikov kernel function; the default
    epan2                             alternative Epanechnikov kernel function
    biweight                          biweight kernel function
    cosine                            cosine trace kernel function
    gaussian                          Gaussian kernel function
    parzen                            Parzen kernel function
    rectangle                         rectangle kernel function
    triangle                          triangle kernel function
    ----------------------------------------------------------------------

    fweights, aweights, and iweights are allowed; see weight.


Menu

    Statistics > Nonparametric analysis > Kernel density estimation


Description

    kdensity produces kernel density estimates and graphs the result.


Options

        +------+
    ----+ Main +-----------------------------------------------------------

    kernel(kernel) specifies the kernel function for use in calculating the
        kernel density estimate.  The default kernel is the Epanechnikov
        kernel (epanechnikov).

    bwidth(#) specifies the half-width of the kernel, the width of the
        density window around each point.  If bwidth() is not specified, the
        "optimal" width is calculated and used; see [R] kdensity.  The
        optimal width is the width that would minimize the mean integrated
        squared error if the data were Gaussian and a Gaussian kernel were
        used, so it is not optimal in any global sense.  In fact, for
        multimodal and highly skewed densities, this width is usually too
        wide and oversmooths the density (Silverman 1992).

    generate(newvar_x newvar_d) stores the results of the estimation.
        newvar_x will contain the points at which the density is estimated.
        newvar_d will contain the density estimate.

    n(#) specifies the number of points at which the density estimate is to
        be evaluated.  The default is min(N,50), where N is the number of
        observations in memory.

    at(var_x) specifies a variable that contains the values at which the
        density should be estimated.  This option allows you to more easily
        obtain density estimates for different variables or different
        subsamples of a variable and then overlay the estimated densities for
        comparison.

    nograph suppresses the graph.  This option is often used with the
```

Figure 2 (cont'd): Stata help-file for **kdensity**, pp.3-4

generate() option.

```
    +-------------+
----+ Kernel plot +--------------------------------------------------------
```

cline_options affect the rendition of the plotted kernel density estimate. See [G-3] *cline_options*.

```
    +---------------+
----+ Density plots +------------------------------------------------------
```

normal requests that a normal density be overlaid on the density estimate for comparison.

normopts(*cline_options***)** specifies details about the rendition of the normal curve, such as the color and style of line used. See [G-3] *cline_options*.

student(#**)** specifies that a Student's t density with # degrees of freedom be overlaid on the density estimate for comparison.

stopts(*cline_options***)** affects the rendition of the Student's t density. See [G-3] *cline_options*.

```
    +-----------+
----+ Add plots +----------------------------------------------------------
```

addplot(*plot***)** provides a way to add other plots to the generated graph. See [G-3] *addplot_option*.

```
    +----------------------------------------+
----+ Y axis, X axis, Titles, Legend, Overall +-------------------------
```

twoway_options are any of the options documented in [G-3] *twoway_options*, excluding **by()**. These include options for titling the graph (see [G-3] *title_options*) and for saving the graph to disk (see [G-3] *saving_option*).

Examples

Setup
```
. sysuse auto
```

Graph kernel density estimates for **length**
```
. kdensity length
```

Same as above, but use 20 for the half-width of the kernel
```
. kdensity length, bw(20)
```

Obtain kernel density estimates for **weight** using the Parzen kernel function, store these results in **x2**, and suppress the graph
```
. kdensity weight, kernel(parzen) gen(x2 parzen) nograph
```

Saved results

kdensity saves the following in **r()**:

Scalars

r(bwidth)	kernel bandwidth
r(n)	number of points at which the estimate was evaluated
r(scale)	density bin width

Macros

r(kernel)	name of kernel

Reference

Silverman, B. W. 1992. *Density Estimation for Statistics and Data Analysis*. London: Chapman & Hall.

Problem 3. Do da do-file

Read the following do-file. You should understand it produces a graph with four sampling distributions (for beta, gamma, delta and theta).

Required. Draw the resulting graph ("graph_samp_dist"). Be as precise as possible and provide relevant explanations about the way you construct it.

```
/* do-file with four
sampling distributions */

clear all
#delimit ;
set mem 1G ;
capture log close ;
set more off ;
log using mylogfile.log,
replace ;

set obs 1000;
set seed 9809;
local samplesize=100;
local nsamples=1000;

gen beta=.;
gen gamma=.;
gen delta=.;
gen theta=.;
gen z=10*rnormal();

foreach kk of numlist 1(1)
`nsamples'{ ;
gen shock`kk'=20*rnormal();
gen y`kk'=80+5*z+shock`kk' in
1/`samplesize';
reg y`kk' z in 1/`samplesize';
matrix A`kk'=e(b);
matrix A`kk'=A`kk'[1,1];
svmat A`kk', names(A`kk') ;
egen B`kk'=max(A`kk');
replace beta=B`kk' in `kk';
drop A* B* ;

gen x`kk'=z- shock`kk'/5 in
1/`samplesize' ;
replace y`kk'=80 + 5*x`kk' +
shock`kk' in 1/`samplesize';
reg y`kk' x`kk' in 1/
`samplesize' ;
matrix A`kk'=e(b);
matrix A`kk'=A`kk'[1,1];
svmat A`kk', names(A`kk') ;
egen B`kk'=max(A`kk');
replace gamma=B`kk' in `kk';
drop A* B* ;

replace shock`kk'=20*rnormal()
in 1/`samplesize' ;
replace shock`kk'=40*rnormal()
in 50/`samplesize' ;
replace y`kk'=80 + 5*z + shock`kk'
in 1/`samplesize';
reg y`kk' z in 1/ `samplesize';
matrix A`kk'=e(b);
matrix A`kk'=A`kk'[1,1];
svmat A`kk', names(A`kk') ;
egen B`kk'=max(A`kk');
replace delta=B`kk' in `kk';
drop A* B* ;

reg y`kk' z in 1/`samplesize',
robust;
matrix A`kk'=e(b);
matrix A`kk'=A`kk'[1,1];
svmat A`kk', names(A`kk') ;
egen B`kk'=max(A`kk');
replace theta=B`kk' in `kk';
drop A* B* ;
} ;

twoway (kdensity beta, lcolor(blue)
legend(on order(1 "beta" 2
"gamma" 3 "delta" 4 "theta"))
(kdensity gamma, lcolor(red))
(kdensity delta, lcolor(yellow))
(kdensity theta, lcolor(green)),
ytitle(Densities) xtitle(Values);
graph2tex, epsfile(graph_samp_dist);

log close;
 /* end of current do-file */
```

Problem 4. Master of Divinity degree effect on income

In their paper[4], sociologists Chang and Perl "explore how Protestant denominations use education to stratify their pastors among lower and higher income jobs and how this use of education intersects with gender". Table 1 reproduces Table 3 of that paper. It gives the estimates for key variables explaining *income*, namely *gender* and *education*.

Table 1: OLS regression of logged annual income on gender and education, denominations not requiring a Master of Divinity degree

Indep. Variable	β Coeff.	Std. Error
Gender and Education		
Female $(0, 1)$	-0.523***	(0.090)
Highest College/ Seminary Degree $(0-5)$	0.054**	(0.016)
Female X Highest Degree	0.099***	(0.024)
Control Variables		
Racial Minority	0.071	(0.089)
Age at Ordination	-0.009***	(0.002)
Prior Religious Work Experience	0.0008***	(0.0002)
Religious Career has been Interrupted	-0.118**	(0.044)
Constant	3.434	

$R^2 = 0.292$, $N = 642$

$*p < .05$; $**p < .01$; $***p < .001$

The *gender* variable takes the value 1 if the individual is a female whereas *education* is "a measure with the following six levels: less than a college degree (coded zero); a college degree; some seminary but no degree; a lower-track seminary degree; the Master of Divinity degree; and any seminary degree higher than the M.Div. (coded five)."

Required. As far as the effect of education[5] on income is concerned, clearly state which constraints this regression model imposes on the data generating process. Argue that those constraints are likely to be too restrictive.
Call Z the set of control variables. Write down the same regression model but with a less restrictive, hence potentially more correct way of dealing with the *education* variable.

[4] Perl, P. and P. Chang (2000), "Credentialism Across Creeds: Clergy Education and Stratification in Protestant Denominations", *Journal for the Scientific Study of Religion*, Vol. 39(2).
[5] Leave aside the interaction term.

Problem 5. Maximum likelihood estimate of a Poisson distribution

Following Ladislaus Bortkiewicz, you think that the number of soldiers of the Prussian army killed accidentally by horse kick follows a Poisson distribution of some parameter λ.
In that case, remember that the probability of observing exactly k times such a rare event in a given period of time is given by

$$f(k|\lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{2}$$

where e is the base of the natural logarithm and the '!' sign stands for the factorial operator of k.
Suppose you have a sample of n independent values of k, *i.e.* you observe the k_i of n units of the Prussian army.

Required. Write $L(\lambda) = f(k_1, ..., k_n|\lambda)$, the log-likelihood function for that sample. Derive $\hat{\lambda}^{MLE}$, the maximum likelihood estimate of λ to show

$$\hat{\lambda}^{MLE} = \frac{1}{n}\sum_{i=1}^{n} k_i \tag{3}$$

Problem 6. Criteria for estimators and tests

Required. For the following questions, provide a short yet sufficient, appropriate answer.

a. When choosing among estimators, what would you consider to buy against some bias of your estimates? Explain.

b. If two tests are asymptotically equivalent, which criteria will you consider in order to make a choice among them? Give and explain two of them.

Problem 7. Your classic authors in *Econometrica*

Required. The following quote was copied from the 1978 paper "Specification Tests in Econometrics", *Econometrica*, 46(6), pp. 1251-1271. Guess who the author is[6]. With or without that information remind thanks to an example in which context this paper has proved pivotal.

"The theory underlying the proposed specification tests rests on one fundamental idea. Under the (null) hypothesis of no misspecification, there will exist a consistent, asymptotically normal and asymptotically efficient estimator, where efficiency means attaining the asymptotic Cramer-Rao bound. Under the alternative hypothesis of misspecification, however, this estimator will be biased and inconsistent. To construct a test of misspecification, it is necessary to find another estimator which is not adversely affected by the misspecification; but this estimator will not be asymptotically efficient under the null hypothesis. A consideration of the difference between the two estimates, $\hat{q} = \hat{\beta}_1 - \hat{\beta}_0$ where $\hat{\beta}_0$ is the efficient estimate under H_0 and $\hat{\beta}_1$ is a consistent estimator under H_1, will then lead to a specification test. If no misspecification is present, the probability limit of $\hat{q}$ is zero. With misspecification plim $\hat{q}$ will differ from zero; and if the power of the test is high, $\hat{q}$ will be large in absolute value relative to its asymptotic standard error. Hopefully, this procedure will lead to powerful tests in important cases because the misspecification is apt to be serious only when the two estimates differ substantially.
In constructing tests based on $\hat{q}$, an immediate problem comes to mind. To develop tests not only is the probability limit of $\hat{q}$ required, but the variance of the asymptotic distribution of $\sqrt{T}\hat{q}$, $V(\hat{q})$, must also be determined. Since $\hat{\beta}_0$ and $\hat{\beta}_1$ use the same data, they will be correlated which could lead to a messy calculation for the variance of $\sqrt{T}\hat{q}$. Luckily, this problem is resolved easily and, in fact, $V(\hat{q}) = V(\hat{\beta}_1) - V(\hat{\beta}_0) = V_1 - V_0$ under the null hypothesis of no misspecification. Thus, the construction of specification error tests is simplified, since the estimators may be considered separately because the variance of the difference $\sqrt{T}\hat{q} = \sqrt{T}(\hat{\beta}_1 - \hat{\beta}_0)$ is the difference of the respective variances. The intuitive reasoning behind this result is simple although it appears to have remained generally unnoticed in constructing tests in econometrics. The idea rests on the fact that the efficient estimator, $\hat{\beta}_0$, must have zero asymptotic covariance with $\hat{q}$ under the null hypothesis for any other consistent, asymptotically normal estimator $\hat{\beta}_1$. If this were not the case, a linear combination of $\hat{\beta}_0$ and $\hat{q}$ could be taken which would lead to a consistent estimator $\hat{\beta}_*$ which would have smaller asymptotic variance than $\hat{\beta}_0$ which is assumed asymptotically efficient. To prove the result formally, consider the following lemma:

LEMMA 2.1: *Consider two estimators $\hat{\beta}_0$, $\hat{\beta}_1$ which are both consistent and asymptotically normally distributed with $\hat{\beta}_0$ attaining the asymptotic Cramer-Rao bound so $\sqrt{T}(\hat{\beta}_0 - \beta) \overset{A}{\sim} N(0, V_0)$ and $\sqrt{T}(\hat{\beta}_1 - \beta) \overset{A}{\sim} N(0, V_1)$ where V_0 is the inverse of Fisher's information matrix. Consider $\hat{q} = \hat{\beta}_1 - \hat{\beta}_0$. Then the limiting distributions of $\sqrt{T}(\hat{\beta}_0 - \beta)$ and $\sqrt{T}\hat{q}$ have zero covariance, $C(\hat{\beta}_0, \hat{q}) = 0$, a zero matrix.* "

[6] Notice that you don't need to read it all to make that guess: Lemma 2.1 should be enough.