Econometrics
Problem Set 2.

Instructor: Alexei Onatski

1. Consider a random walk with drift

$$x_t = \delta + x_{t-1} + \eta_t,\ \ t = 1, ..., T,$$

where $x_0 = 0$, and η_t are i.i.d. random variables distributed as $N(0, \sigma^2)$.

(a) Prove that for all integer $T \geq 1$,

$$f_{x_T,...,x_1|x_0}\left(x_T, ..., x_1|x_0\right) = \prod\nolimits_{t=1}^{T} f_{x_t|x_{t-1},...,x_0}\left(x_t|x_{t-1}, ..., x_0\right).$$

(b) Using (a), write down the log likelihood function (conditional on $x_0 = 0$) and find the maximum likelihood estimators of δ and σ^2.

(c) Suppose that you know that $\sigma^2 = 1$. What is the smallest possible variance of an unbiased estimator of δ? Does your estimator from (b) achieve this bound?

(d) File SP500Index.xlsx contains data on the level of the Standard & Poor's 500 Composite Index at the end of the trading month. The index does not include dividends. Let us denote the index as SP_t. Compute $x_t = \log\left(SP_t/SP_0\right)$, where $t = 0$ corresponds to January 1960 (the first month in our dataset). Find numerically $\hat{\delta}_{ML}$ and $\sqrt{\hat{\sigma}^2_{ML}}$. Interpret.

(e) Now assume that x_t is related to y_t by the equation

$$y_t = \alpha + \beta x_t + \varepsilon_t,\ \ t = 1, ..., T,$$

where $\varepsilon_1, ..., \varepsilon_T$ are i.i.d. random variables distributed as $N(0, \gamma^2)$, and independent from $x_1, ..., x_T$. Does this model satisfy the Gauss-Markov assumptions? Are (y_t, x_t) i.i.d. across $t = 1, ..., T$? How would you test a hypothesis that $\beta = 1$ if T is very small, say $T = 3$?

2. STATA file ps1.dta contains data from 1990 cross-section of the NLSY (National Longitudinal Survey of Youth). The file contains wage (variable w0),

education (variable ed0), and age (variable a0) variables for 1500 individuals. Create variables lwage=log(w0), educ=ed0, and exper=age-(educ+6). Variable exper is the proxy for experience.

(a) Create dummy variables corresponding for each value of the education variable and to each value of the experience variable. Regress lwage on this dummy variable set using command *xi: regress lwage i.educ i.exper* (Executing this command will automatically create the dummies). We will call the vector of OLS coefficient estimates from this regression $\hat{\alpha}$, and the vector of coefficients being estimated α.

(b) What restrictions on the parameter α would lead to the linear specification of the conditional expectation function $E\,(\text{lwage}|\text{educ,exper}) = \beta_0 + \beta_1\text{educ}+\beta_2\text{exper}$? (i.e. for $H_0 : R\alpha = q$, specify R and q) Now compute F statistic, Do you accept or reject the null hypothesis?

(c) Is the methodology from (a) and (b) a good way of testing the assumption of a linear conditional expectation function? What part of the assumption are we testing?

3. Consider the following model

$$y_i = 1 + 0.5z_i + \varepsilon_i,$$

where $z_i \sim \chi^2(3)$, $\varepsilon_i \sim U\,[-1, 1]$ (uniformly distributed on the segment $[-1, 1]$), and z_i and ε_i are independent. Provide Monte Carlo evidence (using MATLAB or R or any other package of your choice) of the following claims.

(a) OLS estimator of the regression coefficients is unbiased.

(b) OLS estimator of the regression coefficients is consistent.

(c) OLS estimator of the regression slope is asymptotically normal with the asymptotic variance $\sigma^2/Var\,(z_i)$.

(d) $RSS/(n-k)$ is an unbiased estimator of σ^2.

4. Consider the linear regression model

$$y_i = \beta x_i + \varepsilon_i,$$

where $x_i = \begin{cases} i \text{ if } i \text{ is even} \\ 0 \text{ if } i \text{ is odd} \end{cases}$ with $i = 1, 2, ..., n$ and ε_i are i.i.d. $N(0, 1)$

(a) Show that the OLS estimator $\hat{\beta}$ for β is consistent.

(b) Is $\hat{\beta}$ consistent for β if $x_i = \lambda^i$, $i = 1, ..., n$, where $|\lambda| < 1$?

(c) Do the assumptions of the Gauss-Markov theorem hold when $x_i = \begin{cases} i \text{ if } i \text{ is even} \\ 0 \text{ if } i \text{ is odd} \end{cases}$? Do they hold when $x_i = \lambda^i$? Comment.

(d) Find the distribution limit of $\hat{\beta}$ under the conditions described in (b).

5. A random sample y_i, $i = 1, ..., 1000$, is taken from the exponential distribution with density

$$f_\theta(y) = \frac{1}{\sqrt{\theta}} e^{-y/\sqrt{\theta}} \text{ for } y > 0.$$

In particular, $Ey_i = \sqrt{\theta}$ and $Var\,(y_i) = \theta$.

(a) Find the sample's Fisher information, $\mathcal{I}\,(\theta)$.

(b) Find the ML estimator of θ and its bias.

(c) Suppose that, in fact, y_i, $i = 1, ..., 1000$, comes from a random sample from the $\chi^2(1)$ distribution with density

$$g(y) = \frac{1}{\sqrt{2\pi y}} e^{-y/2} \text{ for } y > 0$$

so that our original model is misspecified. What does the ML estimator found in (b) converge to as the sample size goes to infinity?

(d) Consider the Kullbach-Leibler divergence

$$d\,(g, f_\theta) = \int_0^\infty g(y) \log \frac{g(y)}{f_\theta(y)} dy.$$

Which value of θ minimizes it? Compare with your answer in (c). Comment.

(e) Assuming that the correct density of y_i is $g(y)$, find and compare

$$Var\left(\frac{d}{d\theta} \log f_\theta\,(y_i)\right) \text{ and } -E\left(\frac{d^2}{d\theta^2} \log f_\theta\,(y_i)\right).$$