Econometrics
Problem Set 3.

Instructor: Alexei Onatski

1. Show rigorously that $y_n \xrightarrow{d} y, x_n \xrightarrow{p} c$, where y is a random variable and c is a constant implies $\begin{pmatrix} y_n \\ x_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} y \\ c \end{pmatrix}$. Briefly argue that your finding justifies Slutsky's lemma as a corollary to the continuous mapping theorem.

2. This problem is based on Nerlove (1963) "Returns to Scale in Electricity Supply," in C. Christ (ed.), *Measurement in Economics: Studies in Mathematical Economics and Econometrics in Memory of Yehuda Grunfeld*, Stanford: Stanford University Press. It is given as Exercise 23.10 in Hansen's textbook.

   Nerlov estimates total cost function for electricity-producing firms. His specification is

$$\log TC_i = \beta_1 + \beta_2 \log Q_i + \beta_3 \log PL_i + \beta_4 \log PK_i + \beta_5 \log PF_i + \varepsilon_i, \quad (1)$$

   where $\log TC_i$ is the log of total cost of production for firm i, $\log Q_i$ is the log of quantity produced by firm i, and $\log PL_i$, $\log PK_i$, and $\log PF_i$ are the logs of the prices of labour, capital and fuel, respectively, faced by firm i. Nerlov's data can be downloaded from https://www.ssc.wisc.edu/~bhansen/econometrics/

   (a) Following Nerlov, add the variable $(\log Q_i)^2$ to the regression. Assess the merits of this new specification using a hypothesis test. Do you agree with this modification?

   (b) Now try a nonlinear specification. Consider model (1) plus the extra term $\beta_6 z_i$, where

$$z_i = \log Q_i \left(1 + \exp\left(\beta_7 - \log Q_i\right)\right)^{-1}.$$

   In addition, impose the restriction $\beta_3 + \beta_4 + \beta_5 = 1$. This model is called a smooth threshold model. For values of $\log Q_i$ much below β_7, the variable $\log Q_i$ has a regression slope of β_2. For values much above β_7, the regression slope is $\beta_2 + \beta_6$, and the model imposes a smooth transition between these regimes. The model is non-linear because of the parameter β_7.

The model works best when β_7 is selected so that several values (in this example, at least 10 to 15) of $\log Q_i$ are above β_7, and several values (again, at least 10 to 15) are below β_7. Examine the data and pick an appropriate range for β_7.

(c) Estimate the model by non-linear least squares. I recommend the concentration method: Pick 1000 values of β_7 in the range chosen in (b). For each value of β_7, calculate z_i and estimate the model by OLS. Record the sum of squared residuals, and find the value of β_7 for which this sum is minimized. It would be convenient to use MATLAB or R or any similar programme to do these calculations, but STATA funs can do this in STATA if they want. However, I do not want you just to use STATA command 'nl' to compute the non-linear least squares estimate.

(d) Calculate standard errors for all parameter estimates $\left(\hat{\beta}_1, ..., \hat{\beta}_7\right)$. Again, try to do this directly, using your knowledge of theory, rather than using STATA's 'nl' command.

3. This problem is based on Box (1953) "Non-normality and tests on variances," *Biometrika* 40, 318-335. In that paper, Box coined the term "robustness". Let $Y_{11}, ..., Y_{1n_1}$ and $Y_{21}, ..., Y_{2n_2}$ be two independent samples, each sample being i.i.d. with cumulative distribution function $G_j(y)$, mean μ_j and variance σ_j^2, $j = 1, 2$. The sample means and variances are $\bar{Y}_j = n_j^{-1} \sum_{i=1}^{n_j} Y_{ji}$ and $s_i^2 = (n_j - 1)^{-1} \sum_{i=1}^{n_j} \left(Y_{ji} - \bar{Y}_j\right)^2$. Suppose that you would like to test for $H_0 : \sigma_1^2 = \sigma_2^2$. The usual test (based on the assumption that the data are normally distributed) is to compare s_1^2/s_2^2 to an F distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom.

(a) Consider the logarithm of the normal theory test statistic s_1^2/s_2^2, standardized by sample sizes:

$$T = \left(\frac{n_1 n_2}{n_1 + n_2}\right)^{1/2} \left[\log s_1^2 - \log s_2^2\right].$$

Prove that asymptotically as n_1 and n_2 go to infinity, the usual test is equivalent to comparing T to a $N(0, 2)$ distribution. (Hint: as $n_1, n_2 \to \infty$, an F distribution with $n_1 - 1$ and $n_2 - 1$ would put all mass at 1 because both its 'numerator' and 'denominator' converge in probability to one. But what about small deviations of the 'numerator' and 'denominator' from unity? Can you use a CLT to figure out how these deviations behave, and therefore derive an approximate distribution of the lograithm of F, multiplied by $\left(\frac{n_1 n_2}{n_1+n_2}\right)^{1/2}$?)

(b) Now, suppose that

$$G_1(y) = F_0\left(\left(y-\mu_1\right)/\sigma_1\right) \text{ and } G_2(y) = F_0\left(\left(y-\mu_2\right)/\sigma_2\right),$$

where F_0 is a non-Gaussian cdf with $\int y \mathrm{d}F_0(y) = 0$ and $\int y^2 \mathrm{d}F_0(y) = 1$. Show that as n_1 and n_2 go to infinity, under the null hypothesis, T converges in distribution to $N\left(0, \kappa - 1\right)$, where κ is the kurtosis of F_0, that is

$$\kappa = \int y^4 \mathrm{d}F_0(y)$$

is the i-th central moment of F_0. (Hint: you might want to use the fact that $Var\left\{\left(\frac{Y_{ji}-\mu_j}{\sigma_j}\right)^2\right\} = \kappa - 1$)

(c) Using the result from (b), demonstrate that, if the populations have kurtosis greater than 3, comparison of s_1^2/s_2^2 to an F distribution is asymptotically equivalent to comparing an $N\left(0, \kappa - 1\right)$ random variable to an $N(0,2)$ distribution. What would the true asymptotic level of a nominal $\alpha = 0.05$ one-sided test would be if $\kappa = 5$?

(d) The file wage.xlsx contains data on hourly wages for 3296 working individuals. Variable "male" equals 1 for males and 0 for females. Suppose that we would like to test a hypothesis that the population variance of the logarithm of wage for males equals that for females against the alternative that the variance for females is larger than the variance for males.The above discussion suggests that a test robust to non-normality of the population would compare $T/\sqrt{\hat{\kappa}-1}$ to $N(0,1)$, where

$$\hat{\kappa} = \frac{(n_1+n_2)\sum_{j=1}^{2}\sum_{i=1}^{n_j}\left(Y_{ji}-\bar{Y}_j\right)^4}{\left[\sum_{j=1}^{2}\sum_{i=1}^{n_j}\left(Y_{ji}-\bar{Y}_j\right)^2\right]^2}$$

is an estimate of κ. Conduct such a test, then perform the standard (normal theory) test based on s_1^2/s_2^2 and compare the results.

4. Consider the following joint probability distribution function

X\Y	0	1	2
0	1/6	1/6	0
1	1/6	0	1/6
2	1/6	0	1/6

Clearly, the conditional expectation function

$$E(Y|X) = \begin{cases} 1/2 & \text{for } X = 0 \\ 1 & \text{for } X = 1 \text{ or } X = 2 \end{cases}$$

is nonlinear. Therefore, the linear regression

$$Y = \alpha + \beta X + \varepsilon \tag{2}$$

estimates the BLP $\alpha + \beta X$ rather than CEF.

(a) Find α and β.

(b) Now find $Var(Y|X)$. Observe that $Var(\varepsilon|X)$ does depend on X.

(c) Suppose that you have observations (Y_i, X_i) with $i = 1, ..., n$ which are i.i.d. draws from the joint distribution of Y and X. Further, suppose that someone revealed the explicit form of $Var(Y|X)$ (found in (b)) to you. Thus, you decide to estimate α and β by GLS. That is, you run a regression of $Y_i/\sqrt{Var(Y_i|X_i)}$ on $1/\sqrt{Var(Y_i|X_i)}$ and $X_i/\sqrt{Var(Y_i|X_i)}$. What is the probability limit of the so obtained $\left(\hat{\alpha}_{GLS}, \hat{\beta}_{GLS}\right)$ as the number of observations goes to infinity?

(d) Compare plim $\left(\hat{\alpha}_{GLS}, \hat{\beta}_{GLS}\right)$ to plim $\left(\hat{\alpha}_{OLS}, \hat{\beta}_{OLS}\right)$. Which of these two estimators would you prefer, and why?

5. Consider the following model

$$y_i = x_i'\beta + \varepsilon_i, \; i = 1, ..., n$$

where $E(\varepsilon_i|x_i) \neq 0$. However, you have an $n \times m$ matrix of instruments, Z, where $Z' = [z_1, ..., z_n]$. Let X be the matrix of the regressors, that is, $X' = [x_1, ..., x_n]$. Suppose that X is $n \times k$ with $k < m$ (there are more instruments than regressors). Finally, suppose that (y_i, x_i, z_i) is an i.i.d. sequence.

(a) State the assumptions that Z must satisfy to be valid instruments. Suppose these assumptions hold. Let $\hat{W} = Z\hat{A}$, where $\hat{A} = (Z'Z)^{-1} Z'X$. Prove that $\hat{\beta}_{2SLS} = \left(\hat{W}'X\right)^{-1} \hat{W}'Y$ is consistent.

(b) Suppose that heteroskedasticity is present, with $E\left(\varepsilon_i^2|z_i\right) = \sigma_i^2$. Calculate the asymptotic variance of $\hat{\beta}_{2SLS}$, that is, the variance of the limiting distribution of $\sqrt{n}\left(\hat{\beta}_{2SLS} - \beta\right)$.

(c) Take the residuals $\hat{\varepsilon} = Y - X\hat{\beta}_{2SLS}$ and form the matrix $\hat{Q}$ so that

$$\hat{Q}' = \left[z_1\hat{\varepsilon}_1, ..., z_n\hat{\varepsilon}_n\right].$$

Then form $\hat{R} = \hat{Q}'\hat{Q}$. Consider

$$\hat{\beta}_{MCS} = \arg\min_{\beta} (Y - X\beta)' Z\hat{R}^{-1}Z' (Y - X\beta).$$

Prove that $\hat{\beta}_{MCS}$ (minimum chi-square) is consistent, and calculate its asymptotic variance.

(d) Prove that under hetroskaedasticity $\hat{\beta}_{MCS}$ is asymptotically more efficient than $\hat{\beta}_{2SLS}$.

6. Robert Hall showed (JPE 1978, p.971-987) that permanent-income hypothesis implies that the marginal utility of consumption is a first-order autoregressive process and the lagged values of such variables as disposable income and consumption do not have additional predictive power for the marginal utility of consumption. For a simple quadratic preferences Hall's hypothesis can be formulated as

$$E\left(C_t|I_{t-1}\right) = \beta_0 + \beta_1 C_{t-1},$$

where C_t is consumption and I_{t-1} is information available up to and including time $t-1$.

(a) Using data in PS4data.xls file, run regression

$$C_t = \beta_0 + \beta_1 C_{t-1} + \beta_2 C_{t-2} + \beta_3 C_{t-3} + \beta_4 C_{t-4} + \varepsilon_t$$

Test hypothesis that $\beta_2 = \beta_3 = \beta_4 = 0$. Does your result supports or rejects Hall's hypothesis? Use sum of real consumption of nondurables per capita and real consumption of services per capita as your measure of consumption in the above regression. To get the per capita values, divide the totals by population.

(b) Campbell and Mankiw (NBER Macroeconomic Annual, 1989) argued that Hall's regression has little power against standard Keynesian consumption function according to which people consume out of current income and not out of permanent income. Suppose, for example, that a fraction of consumers γ_1 simply spends its current income, while the remainder follows Hall's model. Then

$$C_t - C_{t-1} = \gamma_0 + \gamma_1 (Y_t - Y_{t-1}) + \varepsilon_t,$$

where ε_t represents news about the permanent income as before. Argue that standard OLS of changes in consumption on changes in the current income and a constant will give inconsistent estimate of γ_1.

(c) Campbell and Mankiw compute 2SLS estimator for the equation in (b) after making one adjustment. They replace the levels of consumption and income with logs of these variables. Using logs instead of levels, estimate the regression in (b) by 2SLS and find the estimates of the corresponding standard errors (using White's method). Use the same measure of consumption per capita as in (a). Measure Y as real disposable income per capita. To get the per capita values, divide the totals by population. Use $\log (C_{t-2}/C_{t-3})$, $\log (C_{t-3}/C_{t-4})$, $\log (C_{t-4}/C_{t-5})$, and $\log (C_{t-5}/C_{t-6})$ as your instruments. Interpret your results.

(d) Test the hypothesis that variable $\log (Y_t/Y_{t-1})$ in the regression

$$\log(C_t/C_{t-1}) = \gamma_0 + \gamma_1 \log(Y_t/Y_{t-1}) + \varepsilon_t$$

is exogenous. Also, test the over-identified restrictions in the 2sls from (c). What do you conclude?