PhD in Economics
1st year Econometrics test (2018-07-09 — resit)

Name: ____

1. Say if the following statements are unambiguously true (TRUE), unambiguously false (FALSE) or impossible to classify the way they are stated (NOT NECESSARILY). Write the motivations to your answers **only** in the space provided. A "Not necessarily" answer with no motivations will be considered wrong.

   (a) If A and B are positive definite square matrices of the same dimension, then $A - B$ is also positive definite.
   True ◯ False ◯ Not necessarily ◯

   (b) If X and Y are two random variables and $E(X) = 0$, then $\text{Cov}(X, Y) = E(X \cdot Y)$.
   True ◯ False ◯ Not necessarily ◯

   (c) If $X_n \xrightarrow{\text{p}} 0.5$, then $\ln(X_n) - \ln(1 - X_n) \xrightarrow{\text{p}} 1$.
   True ◯ False ◯ Not necessarily ◯

   (d) If $\sqrt{n}(X_n - 0.5) \xrightarrow{\text{d}} N(0, 0.25)$, then $\sqrt{n}\,[\ln(X_n) - \ln(1 - X_n)] \xrightarrow{\text{d}} N(0, 4)$.
   True ◯ False ◯ Not necessarily ◯

   (e) You have a sample of n iid realisations of a random variable X, with $E(X) = V(X) = \mu$; consider the statistic $\hat{\mu} = \sqrt{\bar{X} \cdot V}$, where $\bar{X}$ is the sample mean of x_i and V is the sample variance. Then, $\hat{\mu}$ is a consistent estimator of μ.
   True ◯ False ◯ Not necessarily ◯

2. You are reading a paper about the impact of attending (at least) a training course in the last 12 months on the probability of being employed. The author of the paper estimated a linear probability model (LPM) where the dependent variable is a dummy indicator equal to 1 if the individual is employed at the moment of the interview and 0 otherwise (*emp*). The regressor of primary interest is also a dummy indicator, equal to 1 if the individual attended (at least) a training course in the 12 months before the interview and 0 otherwise (*tr*). The estimated equation is therefore

$$emp_i = \delta \cdot tr_i + \mathbf{x}_i \boldsymbol{\beta} + u_i, \tag{1}$$

where u_i is the error term and $\mathbf{x}_i$ are further individual characteristics.

Table 1 reports descriptive statistics of the dependent variable, the training indicator and further individual characteristics used in the analysis.

Table 1: Summary statistics of the variables used in the analysis

Variable	Mean	Std. Dev.	Minimum	Maximum
Employment status	0.6921	0.4616	0	1
Training indicator	0.0666	0.2493	0	1
Female	0.5309	0.4991	0	1
Age	43.498	10.047	26	64
Primary education	0.0872	0.2821	0	1
Secondary education	0.2084	0.4062	0	1
Tertiary education	0.7045	0.4563	0	1
Presence of kids younger than 12	0.3539	0.4782	0	1
Observation			33,348	

The estimation results of Equation (1) are reported in Table 2.

Table 2: Estimation results of employment equation

	Coefficient		Standard error[a]
Training indicator	0.1793	***	0.0055
Female	-0.2044	***	0.0044
Age	0.0644	***	0.0018
Age squared	-0.0009	***	0.0000
Education - Reference: Primary education			
Secondary education	-0.0868	***	0.0086
Tertiary education	-0.0587	***	0.0075
Presence of kids younger than 12	-0.1207	***	0.0051
Constant	-0.0923	**	0.0399
Observations		33,348	
R^2		0.2118	
$F(7, 33340)$		1,772.26	

Notes: * Significant at 10%; ** significant at 5%; *** significant at 1%.
[a] Standard errors are robust to heteroskedasticity.

After reading Table 2 answer the following questions:

(a) What is the marginal effect of attending (at least) a training course on the probability of being employed?

(b) Compute the t-statistic for the test of significance of the estimated parameter of the training indicator.

(c) Why does the author of this paper computed standard errors robust to heteroskedasticity? Would you have done it as well? Motivate your answer.

(d) What is the marginal effect of one more year of age on the probability of being employed?

(e) Why is the the indicator for primary education omitted from the set of regressors included in the model specification?

The paper also presents a second estimated employment equation. The equation is similar to the previous one, with the only difference that there is one more regressor, which is the interaction between the training indicator and the female indicator. The estimation results are presented in Table 3.

Table 3: Estimation results of employment equation with gender heteroegenous effect of training

	Coefficient		Standard error
Training indicator	0.0935	***	0.0063
Training indicator*Female	0.1841	***	0.0103
Female	-0.2167	***	0.0047
Age	0.0641	***	0.0018
Age squared	-0.0009	***	0.0000
Education - Reference: Primary education			
Secondary education	-0.0851	***	0.0086
Tertiary education	-0.0570	***	0.0075
Presence of kids younger than 12	-0.1190	***	0.0051
Constant	-0.0807	**	0.0400
Observations		33,348	
R^2		0.2142	
$F(8, 33339)$		1,596.84	

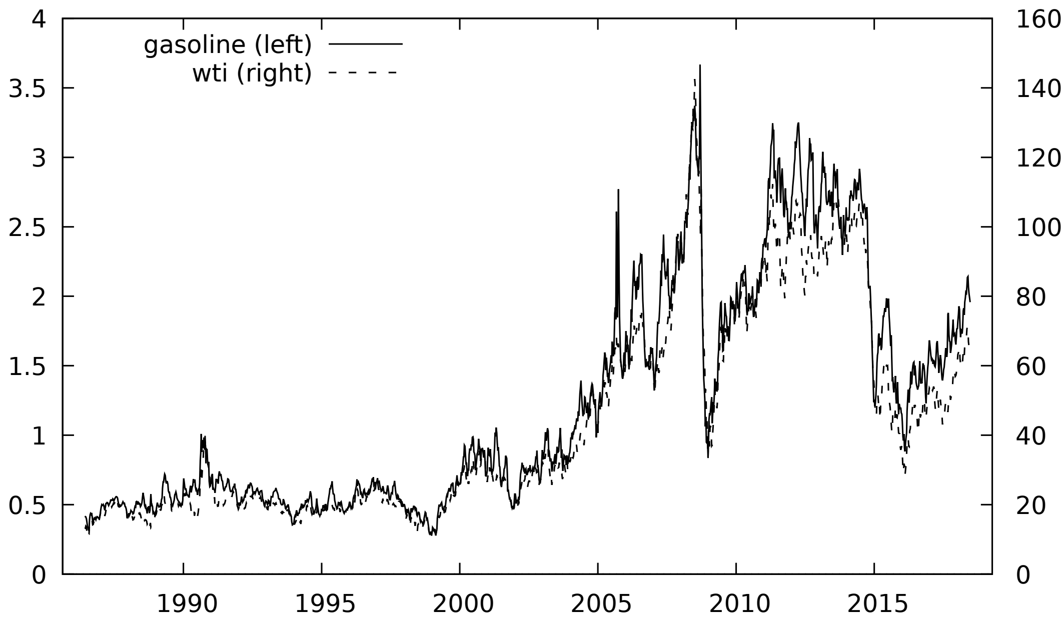
Notes: * Significant at 10%; ** significant at 5%; *** significant at 1%.
[a] Standard errors are robust to heteroskedasticity.

After reading Table 3 answer the following questions:

(f) Is the effect of training significantly different between males and females?

(g) What is the marginal effect of training on the probability of employment for men? And for women?

Figure 1: Oil and gasoline prices



3. Figure 1 shows three weekly time series, from 06/06/1985 to 22/06/2018, defined as follows:

Name	Definition
`gas`	Natural logarithm of the spot price for U.S. Gulf Coast, regular gasoline (Source: eia.gov)
`oil`	Natural logarithm of the spot price for West Texas Intermediate crude oil price (Source: eia.gov)
`markup`	`gas - oil`

Economic theory would suggest that, since crude oil is by far the main component of the cost of producing gasoline, the long-run elasticity of the price of gasoline with respect to the price of oil should be 1.

Annex 1 contains a gretl script performing various tests and estimating procedures on these data, together with the unedited output from the program. Write **in no more than one page** the main conclusions you are able to draw from the output (and please please please, write legibly); if you don't comply with the space limitation, you will get 0 points, even if what you write is worth the Nobel Prize.

Annex 1

Gretl script file

```
# static regression
ols gas const oil

# take differences
d_oil = diff(oil)
d_gas = diff(gas)
# ecm model
scalar p = 6
ols d_gas const d_gas(-1 to -p) d_oil(0 to -p) gas(-1) oil(-1)
# Godfrey test
modtest 13 -a --quiet

# unit root tests
adf 13 oil --c --test-down=BIC
adf 13 gas --c --test-down=BIC
adf 13 markup --c --test-down=BIC

# lag selection
var 13 oil gas --lagselect
p = iminc($test[,2]) # use BIC

# Johansen test
coint2 p gas oil
# VECM
vecm p 1 gas oil

restrict
    b1 + b2 = 0
end restrict
```

Output

```
Model 1: OLS, using observations 1986-06-06:2018-06-22 (T = 1673)
Dependent variable: gas

             coefficient   std. error   t-ratio   p-value
  -------------------------------------------------------
  const       -3.53383     0.0130809    -270.2    0.0000  ***
  oil          0.989228    0.00361230    273.8    0.0000  ***

Mean dependent var  -0.011093   S.D. dependent var   0.657400
Sum squared resid    15.74986   S.E. of regression   0.097085
R-squared            0.978204   Adjusted R-squared   0.978191
F(1, 1671)           74993.80   P-value(F)           0.000000
Log-likelihood       1528.842   Akaike criterion    -3053.684
Schwarz criterion   -3042.839   Hannan-Quinn        -3049.666
rho                  0.911448   Durbin-Watson        0.176628


Model 2: OLS, using observations 1986-07-11:2018-06-22 (T = 1668)
Dependent variable: d_gas

             coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------
  const      -0.335603     0.0377083    -8.900    1.43e-18  ***
  d_gas_1     0.146425     0.0245458     5.965    2.98e-09  ***
  d_gas_2    -0.0733443    0.0248388    -2.953    0.0032    ***
  d_gas_3     0.103174     0.0245671     4.200    2.82e-05  ***
  d_gas_4     0.0514291    0.0246190     2.089    0.0369    **
  d_oil       0.754124     0.0226904    33.24     5.28e-186 ***
  d_oil_1    -0.0205386    0.0292229    -0.7028   0.4823
  d_oil_2     0.0688646    0.0293081     2.350    0.0189    **
  d_oil_3    -0.0378244    0.0288432    -1.311    0.1899
  d_oil_4    -0.0990733    0.0288769    -3.431    0.0006    ***
  gas_1      -0.0954441    0.0105602    -9.038    4.33e-19  ***
  oil_1       0.0939659    0.0105511     8.906    1.36e-18  ***

Mean dependent var   0.000989   S.D. dependent var   0.050424
Sum squared resid    2.375611   S.E. of regression   0.037875
R-squared            0.439511   Adjusted R-squared   0.435788
F(11, 1656)          118.0511   P-value(F)           4.3e-199
Log-likelihood       3099.351   Akaike criterion    -6174.703
Schwarz criterion   -6109.670   Hannan-Quinn        -6150.605
rho                  0.001530   Durbin-Watson        1.996874

Excluding the constant, p-value was highest for variable 12 (d_oil_1)


Breusch-Godfrey test for autocorrelation up to order 13

Test statistic: LMF = 1.361811,
with p-value = P(F(13,1643) > 1.36181) = 0.171

Alternative statistic: TR^2 = 17.781324,
with p-value = P(Chi-square(13) > 17.7813) = 0.166

Ljung-Box Q' = 12.9229,
with p-value = P(Chi-square(13) > 12.9229) = 0.454


Augmented Dickey-Fuller test for oil
testing down from 13 lags, criterion BIC
sample size 1669
```

```
unit-root null hypothesis: a = 1

  test with constant
  including 3 lags of (1-L)oil
  model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
  estimated value of (a - 1): -0.00267861
  test statistic: tau_c(1) = -1.7302
  asymptotic p-value 0.416
  1st-order autocorrelation coeff. for e: -0.003
  lagged differences: F(3, 1664) = 17.063 [0.0000]


Augmented Dickey-Fuller test for gas
testing down from 13 lags, criterion BIC
sample size 1669
unit-root null hypothesis: a = 1

  test with constant
  including 3 lags of (1-L)gas
  model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
  estimated value of (a - 1): -0.00393731
  test statistic: tau_c(1) = -2.13623
  asymptotic p-value 0.2305
  1st-order autocorrelation coeff. for e: 0.005
  lagged differences: F(3, 1664) = 25.501 [0.0000]


Augmented Dickey-Fuller test for markup
testing down from 13 lags, criterion BIC
sample size 1668
unit-root null hypothesis: a = 1

  test with constant
  including 4 lags of (1-L)markup
  model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
  estimated value of (a - 1): -0.103058
  test statistic: tau_c(1) = -9.49483
  asymptotic p-value 1.349e-17
  1st-order autocorrelation coeff. for e: 0.003
  lagged differences: F(4, 1662) = 15.387 [0.0000]

VAR system, maximum lag order 13

The asterisks below indicate the best (that is, minimized) values
of the respective information criteria, AIC = Akaike criterion,
BIC = Schwarz Bayesian criterion and HQC = Hannan-Quinn criterion.

lags        loglik    p(LR)       AIC          BIC          HQC

   1    5978.33476             -7.195584    -7.176013    -7.188330
   2    6014.30278  0.00000    -7.234100    -7.201482    -7.222010
   3    6030.09295  0.00000    -7.248305    -7.202640    -7.231379
   4    6046.36271  0.00000    -7.263088    -7.204375*   -7.241326
   5    6056.61223  0.00040    -7.270617    -7.198858    -7.244020*
   6    6058.40761  0.46421    -7.267961    -7.183154    -7.236528
   7    6060.97515  0.27372    -7.266235    -7.168381    -7.229966
   8    6064.49229  0.13409    -7.265653    -7.154752    -7.224548
   9    6073.26383  0.00152    -7.271402    -7.147454    -7.225461
  10    6077.79432  0.05959    -7.272041*   -7.135046    -7.221264
  11    6078.28398  0.91292    -7.267812    -7.117770    -7.212199
  12    6083.75084  0.02732    -7.269579    -7.106490    -7.209131
  13    6084.44908  0.84481    -7.265601    -7.089465    -7.200317


Johansen test:
```

```
Number of equations = 2
Lag order = 4
Estimation period: 1986-07-04 - 2018-06-22 (T = 1669)
Case 3: Unrestricted constant

Log-likelihood = 10793.3 (including constant term: 6056.9)

Rank Eigenvalue Trace test  p-value  Lmax test  p-value
   0   0.047752     84.795 [0.0000]     81.664 [0.0000]
   1  0.0018739     3.1305 [0.0768]     3.1305 [0.0768]

Corrected for sample size (df = 1660)
Rank Trace test p-value
   0     84.795 [0.0000]
   1     3.1305 [0.0767]

eigenvalue    0.047752    0.0018739

beta (cointegrating vectors)
gas            -11.082    -0.010940
oil             10.967       1.5362

alpha (adjustment vectors)
gas          0.0064122   -0.0017113
oil         -0.0023742   -0.0017179

renormalized beta
gas             1.0000   -0.0071209
oil           -0.98959       1.0000

renormalized alpha
gas          -0.071061   -0.0026290
oil           0.026311   -0.0026391

long-run matrix (alpha * beta')
                   gas          oil
gas          -0.071043     0.067693
oil           0.026330    -0.028677


VECM system, lag order 4
Maximum likelihood estimates, observations 1986-07-04-2018-06-22 (T = 1669)
Cointegration rank = 1
Case 3: Unrestricted constant

beta (cointegrating vectors, standard errors in parentheses)

gas       1.0000
         (0.0000)
oil     -0.98959
       (0.015082)

alpha (adjustment vectors)

gas    -0.071061
oil     0.026311

Log-likelihood = 6055.337
Determinant of covariance matrix = 2.4193787e-06
AIC = -7.2347
BIC = -7.1762
HQC = -7.2130

Equation 1: d_gas
```

```
             coefficient   std. error   t-ratio   p-value
  -------------------------------------------------------
  const      -0.250569     0.0469438    -5.338    1.07e-07 ***
  d_gas_1     0.176205     0.0316676     5.564    3.06e-08 ***
  d_gas_2    -0.0199701    0.0315120    -0.6337   0.5263
  d_gas_3     0.122651     0.0315490     3.888    0.0001   ***
  d_oil_1     0.0531366    0.0375437     1.415    0.1572
  d_oil_2    -0.0247674    0.0368505    -0.6721   0.5016
  d_oil_3     0.00690781   0.0371647     0.1859   0.8526
  EC1        -0.0710615    0.0132732    -5.354    9.82e-08 ***

Mean dependent var   0.000953   S.D. dependent var   0.050430
Sum squared resid    3.976672   S.E. of regression   0.048930
R-squared            0.062571   Adjusted R-squared   0.058620
rho                  0.004186   Durbin-Watson        1.991006

Equation 2: d_oil

             coefficient   std. error   t-ratio   p-value
  -------------------------------------------------------
  const       0.0938062    0.0395033     2.375    0.0177   **
  d_gas_1     0.0419184    0.0266483     1.573    0.1159
  d_gas_2     0.0853092    0.0265174     3.217    0.0013   ***
  d_gas_3     0.0310184    0.0265485     1.168    0.2428
  d_oil_1     0.100191     0.0315931     3.171    0.0015   ***
  d_oil_2    -0.148400     0.0310097    -4.786    1.86e-06 ***
  d_oil_3     0.0633167    0.0312741     2.025    0.0431   **
  EC1         0.0263114    0.0111694     2.356    0.0186   **

Mean dependent var   0.000951   S.D. dependent var   0.042077
Sum squared resid    2.815980   S.E. of regression   0.041175
R-squared            0.046449   Adjusted R-squared   0.042430
rho                  0.000711   Durbin-Watson        1.995420

Cross-equation covariance matrix:

              gas          oil
gas     0.0023827    0.0012652
oil     0.0012652    0.0016872

determinant = 2.41938e-06

Restriction:
 b[1] + b[2] = 0

Test of restrictions on cointegrating relations

eigenvalue 1 = 0.0474912

Unrestricted loglikelihood (lu) = 6055.337
Restricted loglikelihood (lr) = 6055.1083
2 * (lu - lr) = 0.457361
P(Chi-square(1) > 0.457361) = 0.49886
```