Introductory Econometrics
Minitests

Instructions Notation and terminology used below is in line with the lectures (see lecture presentations from the course website).

Questions

1. The following regression output was obtained in Gretl using a cross-sectional data set containing the variables *wage* (a respondent's hourly wage in USD) and *educ* (years of completed education):

```
Model 1: OLS, using observations 1-526
Dependent variable: wage

             coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------
  const      -0.904852     0.684968     -1.321    0.1871
  educ        0.541359     0.0532480    10.17     2.78e-022 ***

Mean dependent var   5.896103   S.D. dependent var   3.693086
Sum squared resid    5980.682   S.E. of regression   3.378390
R-squared            0.164758   Adjusted R-squared   0.163164
F(1, 524)            103.3627   P-value(F)           2.78e-22
Log-likelihood      -1385.712   Akaike criterion     2775.423
Schwarz criterion    2783.954   Hannan-Quinn         2778.764
```

   Write down the estimated regression function and interpret the estimated coefficients; for the coefficient on *educ*, provide both a descriptive and a causal interpretation.

2. Our aim is to quantify a causal link leading from x to y. Explain why a randomized experiment, where x is assigned to subjects at random, enables us to interpret the β_1 coefficient in the regression equation $y = \beta_0 + \beta_1 x + u$ in a causal fashion.

3. Using a cross-sectional dataset with data on 60 cities, we obtained the sample correlation between car thefts and policemen per capita. The value is -0.26, a result significantly different from zero at the conventional 5% significance level. Describe different causal relationships that might have caused (or contributed to) this correlation. (Try to justify each of the three causation schemes from the lectures.)

4. Give examples of a descriptive, causal and forecasting research question in empiric research.

5. Formulate a simple regression model both in a descriptive setting (using the population regression function) and in a causal setting (using a structural equation).

6. Give an example of a cross-sectional and a time-series dataset (sketch a part of the data table showing first three observations). Time series data are typically considered more difficult to analyze statistically; why is it so?

7. Explain the difference between a repeated (or pooled) cross section and a panel data set.

8. The table below shows the average wage in a population of interest, broken down by education groups. What is the average wage in the entire population? Formally put, the table lists the values of $\mathrm{E}(\mathit{wage} \mid \mathit{education})$ and the probability mass function of *education*. In this formalized setting, your calculation should follow the *law of iterated expectations*; write down expression that

formalizes the calculation you used to obtain the population average (or: the unconditioned expectation).

Education	Low	Medium	High
Average wage	2,500	2,900	4,100
% of population	20	60	20

9. Using the ordinary least squares method (OLS), we estimate the β_1 coefficient in the regression model $y = \beta_0 + \beta_1 x + u$. What is the relationship between the coefficient estimate and the sample covariance of x and y?

10. Using the ordinary least squares method (OLS), we estimate the β_1 coefficient in the regression model $y = \beta_0 + \beta_1 x + u$. Which of the following can possibly happen? Give a yes/no response for each item, explain all negative responses.
a) $\hat{\beta}_1 = 4.25$, sample correlation of x and y is 2.11.
b) $\hat{\beta}_1 = -2.27$, sample covariance of x and y is 0.11.
c) $\hat{\beta}_1 = -0.02$, sample correlation of x and y is 0.

11. What squares exactly are being minimized if OLS is applied to the model $y = \beta_0 + \beta_1 x + u$?

12. Using data on a sample of $n = 147$ respondents, we estimate using OLS the linear regression model $y = \beta_0 + \beta_1 x + u$. Write down an expression for the sum of squared residuals.

13. Using data on a sample of $n = 147$ respondents, we estimate using OLS the linear regression model $y = \beta_0 + \beta_1 x + u$. What is the range of possible values of the sum of residuals? Explain.

14. We are going to estimate a structural equation $y = \beta_0 + \beta_1 x + u$, with an intention to interpret the regression coefficients in causal manner. What is the crucial assumption about the random error that we need to make? Does the assumption imply anything about the correlation of x and u?

15. In a simple regression fitted by OLS, show that the sample mean of y (actual values) equals the sample mean of $\hat{y}$ (fitted values). *Hint*: From our derivations of the OLS estimator, we know a very useful fact about the sum of residuals ($\hat{u}$) that can be put to good effect.

16. The sample mean of x and y is 4.5 and 6, respectively. Fill in the value of the slope parameter of a regression line fitted by OLS: $\hat{y} = -3 + \_\_ x$. Explain. *Hint*: See lecture 2, slide 13.

17. Using OLS, we estimated the sample regression function $\widehat{wage} = 1620 + 2.0\, height$, where *wage* is a respondent's monthly wage in EUR and *height* is measured in centimetres. Write down the estimated function that we would have obtained if we expressed *height* in inches instead of centimetres (1 inch = 2.54 cm) and wage in 000s EUR (thousands of euros) instead of EUR.

18. The following regression output was obtained in Gretl using a cross-sectional data set on 168 used cars (*Škoda Felicia*s) containing the variables *price* (price of a used car in CZK) and *age* (age of the used car in years); *l_price* is the natural log of *price*.

```
Model 1: OLS, using observations 1-168
Dependent variable: l_price

             coefficient   std. error   t-ratio    p-value
  --------------------------------------------------------
  const      12.3608       0.0506844    243.9     4.90e-214 ***
  age        -0.0994875    0.00703794   -14.14    2.74e-030 ***

Mean dependent var    11.66259   S.D. dependent var   0.218237
Sum squared resid     3.609194   S.E. of regression   0.147452
```

```
R-squared            0.546229   Adjusted R-squared   0.543496
F(1, 166)            199.8235   P-value(F)           2.74e-30
Log-likelihood       84.21860   Akaike criterion    −164.4372
Schwarz criterion   −158.1893   Hannan-Quinn        −161.9015
```

Write down the estimated regression function and interpret the estimated slope coefficient (use either causal or descriptive interpretation). Next, interpret the R-squared.

19. The following regression output was obtained in Gretl using a cross-sectional data set on 168 used cars (*Škoda Felicia*s) containing the variables *price* (price in CZK) and *km* (kilometres travelled); *l_price* is the natural log of *price* and *l_km* is the natural log of *km*.

```
Model 1: OLS, using observations 1-168
Dependent variable: l_price

             coefficient   std. error   t-ratio    p-value
  --------------------------------------------------------
  const       14.1950      0.385853     36.79     1.01e-081 ***
  l_km       -0.220007     0.0334957    -6.568    6.27e-010 ***

Mean dependent var   11.66259   S.D. dependent var   0.218237
Sum squared resid    6.313084   S.E. of regression   0.195014
R-squared            0.206279   Adjusted R-squared   0.201498
F(1, 166)            43.14152   P-value(F)           6.27e-10
Log-likelihood       37.25085   Akaike criterion    −70.50171
Schwarz criterion   −64.25378   Hannan-Quinn        −67.96600
```

Write down the estimated regression function and interpret the estimated slope coefficient (use either causal or descriptive interpretation). Next, interpret the R-squared.

20. Using a dataset on 450 manufacturing companies, you want to estimate the elasticity of *sales* (in 000,000s EUR) with respect to *labour* (number of employees). Write down the simple regression equation you are about to estimate.

21. Explain the term *sampling distribution of an estimator*. (Consider using as an example the OLS estimator of β_1 in simple regression.)

22. Give an example of variables x and y such that the random error in equation $y = \beta_0 + \beta_1 x + u$ is heteroskedastic. Explain why you expect the presence of heteroskedasticity.

23. Consider the usual simple regression equation $y = \beta_0 + \beta_1 x + u$. Give an example of variables x and y such that $\mathrm{var}(u\,|\,x)$ is a decreasing function of x.

24. Consider a linear regression model that satisfies assumptions SLR.1–SLR.5. How would you estimate the variance of random errors? Is your estimator unbiased?

25. Explain the term *standard error of* $\hat{\beta}_1$.

26. What exactly is the *standardized* (or *Studentized*) *estimator* $\hat{\beta}_1$? What sampling distribution does it have under the assumptions SLR.1–SLR.6?

27. Using data on 100 000 respondents, we estimated a simple regression by OLS. The value of the standard error for the slope coefficient is 4.2. What value (approximately) will you expect for the standard error if we re-estimate the same equation using a random subsample of 1000 respondents only? *Hint*: We have discussed this in the lectures, even though it is not explicitly contained in the slides. In the formula for the standard error of the slope coefficient, think about how the denominator changes if we change the number of observations.

28. Formulate the MLR.4 assumption.

29. Formulate the MLR.5 assumption.

30. Formulate the MLR.6 assumption.

31. Explain the term *classical linear regression model*.

32. In a regression model that satisfies assumptions MLR.1–MLR.4, the random errors are heavily heteroskedastic. Can this make the OLS estimator biased?

33. After the estimation of $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + u$ using data on $n = 21$ respondents, we obtained $R^2 = 0.12$. Calculate the adjusted R-squared.

34. Which of the following can possibly happen after an OLS estimation of a linear regression? For each list item, give a yes/no answer (can /cannot happen); for each negative answer, provide a brief explanation.
    a) $R^2 = 0.50$, and the sample correlation of y and $\hat{y}$ equals 0.68.
    b) $R^2 = 1.25$, $\bar{R}^2 = 1.20$.
    c) $R^2 = 0.01$, $\bar{R}^2 = -0.02$.

35. The variables x_1 and x_2 are heavily correlated, their population correlation being 0.95. Can this cause a bias in the OLS-estimator of the regression coefficients in the equation $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$? Explain.

36. Using OLS, we estimated the equation $\hat{y} = 3.4 - 5.2x$. Next, we use the same data file to run another regression that contains an additional explanatory variable z (in addition to x, which has been retained from the previous model). The sample correlation between x and z is zero. Fill in the missing value in the estimated equation: $\hat{y} = -1.5 + \_\_\, x + 3.6z$. Explain.

37. What is the *bias-variance tradeoff* in multiple regression (in the context of estimating the partial effect of x on y)?

38. The following regression output was obtained in Gretl using a cross-sectional data set containing the variables *wage* (a respondent's average hourly earnings in USD) and *educ* (years of education). What is the approximate width of the 95% confidence interval for the intercept? Show and comment on your calculations.

```
Model 1: OLS, using observations 1-526
Dependent variable: wage

             coefficient   std. error   t-ratio    p-value
  --------------------------------------------------------
  const       -0.904        0.684        -1.321    0.1871
  educ         0.541        0.053        10.17     2.78e-022 ***
```

39. What is the *level of significance* of a hypothesis test?

40. Show that our formula for the endpoints of the 95% confidence interval works as intended, i.e. prove that it follows from the assumptions of the classical linear regression model that if c is the 97.5th percentile of Student's t distribution with $n - k - 1$ degrees of freedom, the following holds:

$$\Pr\left\{\hat{\beta}_j - c \cdot \mathrm{se}(\hat{\beta}_j) \le \beta_j \le \hat{\beta}_j - c \cdot \mathrm{se}(\hat{\beta}_j)\right\} = 0.95.$$

In your proof, you can use freely use our results about the distribution of the standardized estimator of the regression coefficient β_j.

41. What exactly is the p-value of a t-test about the regression parameter β_j?

42. The following regression output was obtained in Gretl using a cross-sectional data set containing the variables *wage* (a respondent's average hourly earnings in USD), *exper* (years of work experience) and *educ* (years of education).

```
^wage = 65.081 + 1.074*exper + 3.136*educ
        (5.120) (0.0067)      (1.600)

n = 8165, R-squared = 0.108
(standard errors in parentheses)
```

What is the *p*-value of a two-tailed test with the null H$_0$: $\beta_{educ} = 0$? Show your work.

43. The following regression output was obtained in Gretl using a cross-sectional data set containing the variables *wage* (a respondent's average hourly earnings in USD) and *educ* (years of education). What is the approximate width of the 95% confidence interval for the intercept? Show and comment on your calculations.

```
Model 1: OLS, using observations 1-526
Dependent variable: wage

             coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------
  const       -0.904        0.684        -1.321    0.1871
  educ          0.541       0.053        10.17     2.78e-022 ***
```

(a) Find the 95% confidence interval for β_{educ}; use the *approximate* calculation.

(b) Assuming that the random error is normally distributed and statistically independent of *educ*, how you would obtain the *exact* 95% confidence interval?

44. The following regression output was obtained in Gretl using cross-sectional data on young employed men containing the variables *lwage* (log of respondents' average hourly earnings in USD), *col2year* (years completed at a 2-year college), *col4year* (years completed at a 4-year college), *age* (in years) and *married* (an indicator of marital status, = 1 if the respondent is married):

```
^lwage = 5.08 + 0.0740*col4year + 0.0135*col2year + 0.0110*age + 0.203*married
        (0.159)(0.00674)         (0.00390)         (0.00489)    (0.0413)

n = 935, R-squared = 0.159
(standard errors in parentheses)
```

Your aim is to test whether the return to education is the same for years spent at 2-year and at 4-year colleges. Formulate (formally) the null hypothesis and briefly describe the statistical test you would use.

45. In the output shown below, the logarithmic price of a house (*l_price*) is being explained by several observed characteristics of the house. Test the null hypothesis H$_0$: $\beta_{colonial} = 0$ against an alternative implying that houses built in the colonial style (*colonial* = 1) are pricier than others (*colonial* = 0).

```
Model 1: OLS, using observations 1-88
Dependent variable: l_price

             coefficient   std. error    t-ratio    p-value
  ----------------------------------------------------------
  const      4.74538       0.0926914     51.20     1.45e-064 ***
  bdrms      0.00832141    0.0297933      0.2793   0.7807
  lotsize    5.65005e-06   2.01221e-06    2.808    0.0062    ***
  sqrft      0.000372785   4.17634e-05    8.926    9.18e-014 ***
  colonial   0.0814651     0.0458308      1.778    0.0791    *

Mean dependent var   5.633180   S.D. dependent var   0.303573
Sum squared resid    2.917374   S.E. of regression   0.187481
R-squared            0.636129   Adjusted R-squared   0.618593
```

46. The following regression output was obtained in Gretl using cross-sectional data on young employed men containing the variables *lwage* (log of respondents' average hourly earnings in USD), *col2year* (years completed at a 2-year college), *col4year* (years completed at a 4-year college), *age* (in years) and *married* (an indicator of marital status, = 1 if the respondent is married). Is the partial effect of age significant at the conventional 5% significance level? Report the formal statement of the null hypothesis and the conclusion of the test, including underlying calculations.

```
^lwage = 5.08 + 0.0740*col4year + 0.0135*col2year + 0.0110*age + 0.203*married
        (0.159)(0.00674)         (0.00390)         (0.00489)    (0.0413)

n = 935, R-squared = 0.159
(standard errors in parentheses)
```

47. What is the impact of heteroskedasticity on the OLS estimator in linear regression? What are the implications for statistical inference?

48. Briefly describe the Breusch-Pagan and the White heteroskedasticity tests. What are the auxiliary regressions run in these tests? What is the null hypothesis of the test?

49. The following regression output was obtained in Gretl using a cross-sectional data set on young employed men containing the variables *lwage* (log of respondents' average hourly earnings in USD), *col2year* (years completed at a 2-year college), *col4year* (years completed at a 4-year college), *age* (in years) and *married* (an indicator of marital status, = 1 if the respondent is married):

```
^lwage = 5.08 + 0.0740*col4year + 0.0135*col2year + 0.0110*age + 0.203*married
        (0.159)(0.00674)         (0.00390)         (0.00489)    (0.0413)

n = 935, R-squared = 0.159
(standard errors in parentheses)
```

According to the estimated equation, what is the *exact* partial effect of marital status on wage?

50. The following regression output was obtained in Gretl using a cross-sectional data set containing (among others) the variables *l_wage* (log of respondents' wage), *age* (in years), and *sq_age* = *age*^{2}. Calculate the turning point of the estimated relationship between age and wage. Does the relationship have the "u" shape or the "inverted u" shape?

```
Model 1: OLS, using observations 1-935
Dependent variable: l_wage

             coefficient    std. error   t-ratio    p-value
  ---------------------------------------------------------
  const        4.74683      1.70368       2.786    0.0054    ***
  educ         0.0618020    0.00583647   10.59     8.12e-025 ***
  married      0.208306     0.0415330     5.015    6.34e-07  ***
  age          0.0411124    0.103378      0.3977   0.6910
  sq_age      -0.000313632  0.00155276   -0.2020   0.8400
```

51. The following regression output was obtained in Gretl using a cross-sectional data set containing (among others) the variables *l_wage* (log of respondents' wage), *age* (in years), and *sq_age* = *age*^{2}. Calculate the relative change in the wage between the 35th and the 36th birthday, as implied by the

estimated equation. (Feel free to use the approximative interpretation of the regression coefficients in a log-level model.)

```
Model 1: OLS, using observations 1-935
Dependent variable: l_wage

            coefficient    std. error   t-ratio    p-value
  ---------------------------------------------------------
  const       4.74683      1.70368       2.786    0.0054    ***
  educ        0.0618020    0.00583647   10.59     8.12e-025 ***
  age         0.0411124    0.103378      0.3977   0.6910
  sq_age     -0.000313632  0.00155276   -0.2020   0.8400
  married     0.208306     0.0415330     5.015    6.34e-07  ***
```

52. The following regression output was obtained in Gretl using a cross-sectional data set on 320 used cars containing the variables *price* (price in CZK), *km* (kilometres travelled), age (in years), *diesel* (= 1 for cars running on diesel, = 0 for cars running on petrol; no other fuel types are present in the dataset). Additional variables were created via simple transforms of the existing variables: *l_price* = log(*price*), *l_km* = log(*km*), *ageXdiesel* = *age* × *diesel*. Your goal is to investigate whether the depreciation rate (loss of value with age) is lower for diesel cars than for petrol cars. What test will you use to find out? What is the conclusion?

```
Model 1: OLS, using observations 1-320
Dependent variable: l_price

                coefficient    std. error    t-ratio     p-value
  -------------------------------------------------------------
  const         13.3615        0.0548511     243.6       0.0000    ***
  km            -1.08453e-06   4.55554e-07    -2.381     0.0179    **
  age           -0.207967      0.0102957     -20.20      8.78e-059 ***
  diesel         0.298895      0.0862927       3.464     0.0006    ***
  ageXdiesel    -0.0122980     0.0174593      -0.7044    0.4817

Mean dependent var   12.18764   S.D. dependent var   0.656101
Sum squared resid    26.36943   S.E. of regression   0.289331
R-squared            0.807970   Adjusted R-squared   0.805532
F(4, 315)            331.3432   P-value(F)           1.7e-111
Log-likelihood      -54.68180   Akaike criterion     119.3636
Schwarz criterion    138.2052   Hannan-Quinn         126.8874
```

53. The following regression output was obtained in Gretl using a cross-sectional data set on 320 used cars containing the variables *price* (price in CZK), *km* (kilometres travelled), *age* (in years), *diesel* (= 1 for cars running on diesel, = 0 for cars running on petrol; no other fuel types are present in the dataset). Additional variables were created via simple transforms of the existing variables: *l_price* = log(*price*), *l_km* = log(*km*), *ageXdiesel* = *age* × *diesel*. What is the predicted difference between the price of a diesel car and a petrol car, both 5 years old and with the same *km* on the clock?

```
Model 1: OLS, using observations 1-320
Dependent variable: l_price

                coefficient    std. error    t-ratio     p-value
  -------------------------------------------------------------
  const         13.3615        0.0548511     243.6       0.0000    ***
  km            -1.08453e-06   4.55554e-07    -2.381     0.0179    **
  age           -0.207967      0.0102957     -20.20      8.78e-059 ***
  diesel         0.298895      0.0862927       3.464     0.0006    ***
  ageXdiesel    -0.0122980     0.0174593      -0.7044    0.4817
```

```
Mean dependent var   12.18764   S.D. dependent var   0.656101
Sum squared resid    26.36943   S.E. of regression   0.289331
R-squared            0.807970   Adjusted R-squared   0.805532
F(4, 315)            331.3432   P-value(F)           1.7e-111
Log-likelihood      -54.68180   Akaike criterion     119.3636
Schwarz criterion    138.2052   Hannan-Quinn         126.8874
```

54. Describe the Ramsey RESET test.

55. How would you decide whether to include an explanatory variable in the level or the log form? (Briefly discuss the main criteria you would base your decision on.)

56. After estimating a wage equation, we obtained the VIF factors, presented in the Gretl output below. The following variables have been used in the regression: *exper* is the respondents' work experience (in years); *educ* their education (in years); *female*, *nonwhite* and *smsa* are indicators (dummies) of gender, race and urban residence; the remaining variables were obtained as $sq\_educ = educ^2$, $femaleXeduc = female \times educ$. Is there a reason to worry about multicollinearity issues? Discuss.

```
Variance Inflation Factors
Minimum possible value = 1.0
Values > 10.0 may indicate a collinearity problem

      exper   13.216
   sq_exper   13.493
       educ    1.867
     female   22.899
femaleXeduc   22.869
   nonwhite    1.013
       smsa    1.059
```

57. The following regression output was obtained in Gretl using a cross-sectional data set on 327 used cars (*Škoda*s) containing the variables *price* (price of a used car in CZK), *km* (kilometres travelled), *age* (age in years), and a categorical variable *model* with three levels: *felicia*, *octavia* and *superb*. The variable *l_price* is the the natural log of *price*. Predict the price of a used Škoda Felicia that is 10 years old and has travelled 100,000 km.

```
Model 1: OLS, using observations 1-327
Dependent variable: l_price

             coefficient    std. error    t-ratio    p-value
  ---------------------------------------------------------
  const      12.6247        0.0444711     283.9     0.0000    ***
  km         -1.22462e-06   2.78073e-07    -4.404   1.45e-05  ***
  age        -0.121677      0.00720144    -16.90    2.67e-046 ***
  octavia     0.584548      0.0264572      22.09    1.75e-066 ***
  superb      1.11115       0.0541394      20.52    1.92e-060 ***

Mean dependent var   12.18048   S.D. dependent var   0.650827
Sum squared resid    10.42042   S.E. of regression   0.179893
R-squared            0.924767   Adjusted R-squared   0.923599
F(5, 322)            791.6110   P-value(F)           1.7e-178
Log-likelihood       100.2646   Akaike criterion    -188.5291
Schwarz criterion   -165.7710   Hannan-Quinn        -179.4493
```

58. The following regression output was obtained in Gretl using a cross-sectional data set on 327 used cars (*Škoda*s) containing the variables *price* (price of a used car in CZK), *km* (kilometres travelled),

age (age in years), and a categorical variable *model* with three levels: *felicia*, *octavia* and *superb*. Additional variables have been created using the formulas *km_100000* = *km* – 100000 and *age_10* = *age* – 10. Find the 95% prediction interval for the price of a used Škoda Felicia that is 10 years old and has travelled 100,000 km.

```
Model 1: OLS, using observations 1-327
Dependent variable: price

               coefficient     std. error    t-ratio    p-value
  --------------------------------------------------------------
  const        33879.9        11201.4         3.025     0.0027    ***
  km_100000       -0.521768       0.116296   -4.487     1.01e-05  ***
  age_10      -29973.3         3002.94       -9.981     1.26e-020 ***
  octavia     112915          11058.8        10.21      2.14e-021 ***
  superb      429509          22378.8        19.19      2.55e-055 ***

Mean dependent var   246395.1   S.D. dependent var   193028.5
Sum squared resid    1.83e+12   S.E. of regression   75235.32
R-squared            0.849943   Adjusted R-squared   0.848085
F(4, 323)            457.3796   P-value(F)           1.3e-131
Log-likelihood      -4145.800   Akaike criterion     8301.600
Schwarz criterion    8320.565   Hannan-Quinn         8309.166
```

59. In a time-series regression, we often need remove a long-term (linear) trend from the dependent variable. Describe the procedure you would use to obtain the detrended version of a time series $\{y_t : t = 1, \ldots, n\}$.

60. The following regression output contains the variables x and y (time series for the dependent and the explanatory variable), and their detrended versions *x_detrended* and *y_detrended* (obtained as the residuals from the regression of the original series on time, like the *salmon* and *gdp* variables from the lecture).

    a) Fill in the blanks in the output below.

    b) Which of the two R^2s presented below would you use to quantify the strength of the relationship between x and y? Justify your choice.

```
Model 1:
^y_detrended = 1.73*x_detrended
              (0.218)

T = 55, R-squared = 0.538

Model 2
^y = 1.46 + ________*x + 0.000550*time
    (1.30) (________)  (0.00101)

T = 55, R-squared = 0.756
```

61. In order to analyze the long-term trend and seasonality of the series $\{y_t : t = 1, \ldots, n\}$, we ran a regression in Gretl, the output of which is shown below. The *time* variable is the sequence 1 through n and *dqj* is an indicator (dummy) variable for quarter j. What statistical test would you use to find out if y_t exhibits a significant seasonal pattern? Formulate the null hypothesis of the test and give the distribution of the test statistic under H_0 (no formulas needed, just report the name of he distribution, along with its parameters).

```
^y = 1.48e+03 + 247*dq2 + 347*dq3 - 67.3*dq4 - 21.5*time
    (76.2)     (60.6)    (60.7)    (61.0)     (4.57)
```

```
T = 36, R-squared = 0.721
```

62. In the regression output from Gretl below, *l_GNP* is the log of GNP in Sweden (in billions of current dollars), *time* is the sequence 1 through 58. Interpret the coefficient on *time* and explain why this regression is said to describe an *exponential trend*.

```
Model 1: OLS, using observations 1960-2017 (T = 58)
Dependent variable: l_GNP

             coefficient   std. error   t-ratio    p-value
  --------------------------------------------------------
  const       9.49542      0.0698728    135.9     3.02e-072 ***
  time        0.0698100    0.00205999    33.89    5.29e-039 ***
```

63. In the regression output from Gretl below, *qdgp* is the quarterly GDP in the U.S. (in billion USD); *l_qgdp* = log(*qgdp*); is the sequence 1 through 258 and *dqj* is an indicator (dummy) variable for quarter *j*. What is the average annual percentage growth of GDP, as implied by the equation? Explain your answer.

```
^l_qgdp = 5.41  - 0.00173*dq2 + 0.00232*dq3 - 0.000136*dq4 + 0.0176*time
         (0.0217)(0.0231)      (0.0232)      (0.0232)       (0.000110)

T = 258, R-squared = 0.990
(standard errors in parentheses)
```

64. In a regression with a *finite distributed lag* (FDL), e.g. $y_t = \beta_0 + \delta_0 x_t + \delta_1 x_{t-1} + \delta_2 x_{t-2} + \delta_2 x_{t-3} + u_t$, we often observe very wide confidence intervals for the marginal effects of a unit impulse in x_t (temporary unit change) on y_t, y_{t+1}, y_{t+2}, and y_{t+3}. However, the long-run propensity is typically estimated much more accurately (its confidence interval is much narrower). Why is it so?

65. Consider a regression with a *finite distributed lag* (FDL), $y_t = \beta_0 + \delta_0 x_t + \delta_1 x_{t-1} + \delta_2 x_{t-2} + \delta_3 x_{t-3} + u_t$.
    a) What is the impact propensity and the long-run propensity (LRP) in this model?
    b) If we need to obtain the 95% confidence interval for the LRP, we can use a suitable transform of the explanatory variables that makes the LRP emerge as one of the model coefficients. Describe this transform and demonstrate that it works.

66. Sketch a plot of the *lag distribution* for the equation $\hat{y}_t = 1.5 + 0.7x_t + 1.3x_{t\text{-}1} + 0.5x_{t\text{-}2} + 1.5x_{t\text{-}3}$.

67. Calculate the *long-run propensity* in the FDL equation $\hat{y}_t = 1.5 + 0.7x_t + 1.3x_{t\text{-}1} + 0.5x_{t\text{-}2} + 1.5x_{t\text{-}3}$ and interpret its value. (Assume that x_t and y_t are annual time series.)

68. Formulate assumptions TS.1–TS.3 for regression with time series. Explain why TS.3 is likely violated in a model that explains the number of monthly car thefts per capita in a particular city (y_t) with the average monthly number of policemen per capita (x_t).

69. Describe the random processes denoted as AR(1) and MA(1); for each, give a full name and a mathematical definition.

70. Give an example of a *weakly dependent* and a *strongly dependent* random process.

71. Explain the term *impulse response curve*. As an example, draw a plot of an impulse response curve for the AR(1) process.

72. Formulate the assumptions about homoskedasticity and no serial correlation in a regression with time series. Use the version of assumptions that relies on strict exogeneity of regressors.

73. Formulate assumptions TS.1’–TS.3’ for regression with time series. Why do we need assumptions about stationarity and weak dependence? (A short and non-technical answer is expected for the latter question.)