**Name: Luis Raúl Enriquez Tenorio** *Do you use STATA or GRETL?*

**TASK 1:** Use data from sheet "gdp". Select GDP time series of country: Latvia

Use 2 x 4- MA moving average (centered moving average, CMA) to smooth the series. After removing trend component estimate seasonality using:

a) model of additive seasonality

b) model of multiplicative seasonality

Then adjust the estimated seasonal indices.

For each model report:

* Estimated seasonal indices
* Adjusted seasonal indices

**TASK 2:** Use GDP of the same country for time series decomposition. Use time period 1995q1 - 2008q4 in this task.

1. Use following three trend models specified below and compare them:

a. Linear trend

b. Polynomial trend of 2nd order

c. Log trend

For each model report:

* Estimated coefficients (including intercept)
* R2 and adjusted R2
* Graph of estimated trend and original series (include it as PDF)

2. Use the best trend model to remove trend component and then estimate seasonality using 3 models specified below:

a. Model of additive seasonality

b. Model of multiplicative seasonality

c. Model of additive dynamic seasonality

For each model report:

* Estimated seasonal indices
* Graph of estimated seasonality and detrended time series (include as PDF)

3. Compute 3 versions of in-sample prediction adding together trend (using the best trend model) and seasonality (using each of the three seasonality models).

For each model report:

* RMSE (root mean squared error) of in-sample prediction
* Graph of in-sample prediction and original time series (include it as PDF)

**TASK 3: Testing stationarity and estimating ARIMA models**

Use data from sheet "arima". Data are generated time series (dataset structure is generic time series).

1. Test stationarity of variable x9. Use augmented DF test and test it down from 10 lags. (Criterion: Trim the longest lag one by one until the longest remaining lag is statistically significant). Determine if the variable is stationary or if it is integrated.

Report:

* Is the variable stationary? If not, is it integrated and of what order?
* Results of the ADF test of the variable based on which you decided whether it is stationary or non-stationary.
* Further results of ADF test based on which you determined the order of integration.

2. Use the same variable and estimate ARIMA models, using time period until (and including) time = 980. Fit ARIMA models up to AR process of order 4 and up to MA process of order 4. Use AIC and BIC and select the best two models. Test normality and independence of residuals of the two models. Make out-of-sample dynamic forecast for time period 981-1000, using each of the two models.

Report:

* What are the best two ARIMA models?
* Estimated coefficients of each of the two ARIMA models and their AIC and BIC.
* Results of the normality test of residuals and independence test of residuals of the two ARIMA models. What are your conclusions based on results?
* Out-of-sample predictions of each model
* RMSE (root mean squared error) and MAE (mean absolute error) of out-of-sample predictions of each model
* Graph of out-sample prediction and original series only for the time period 981-1000 (include it as PDF)

**TASK 4:** Use data from sheet "consumption". Use time period 1998q1-2010q4 in this task. Estimate the following specification of consumption function:

***consumption =*** ***a0 + a1.income + a2.interest\_rate*** ***+ u***

Interpret impact of explanatory variables. Evaluate basic statics (R2, F-test, t-test). Evaluate autocorrelation of residuals (up to the fourth order), heteroscedasticity of residuals, and normality of residuals. Use 5% significance level for each statistical test.

Report:

* Estimated coefficients and interpret impact of explanatory variables
* R2 and adjusted R2 and explain
* F-statistics of the model and critical value (use 5% significance level) and explain
* T-statistics and critical values (use 5% significance level) and explain
* Results of autocorrelation test of residuals (up to the fourth order) and explain
* Results of heteroscedasticity test of residuals and explain
* Results of normality test of residuals and explain

Compare the impact of the explanatory variables on consumption using beta coefficients, (i.e. estimate the model using the standardized variables). Report:

* Estimated beta coefficients and explain

**TASK 5:** Use data from sheet "audit". Use observations 51-450 in this task.

Estimate model 1:

***FEE = a0 + a1.Turnover + a2.FORSAL + a3.TA + a4.LOND + a5.BIG4 + a6.SUB + u***

Evaluate basic statics (R2, F-test, t-test). Analyse multicollinearity using two measures (two methods) of your choice. Evaluate heteroscedasticity of residuals, and normality of residuals. Use 5% significance level for each statistical test.

Report:

* Estimated coefficients
* R2 and adjusted R2 and explain
* F-statistics of the model and critical value (use 5% significance level) and explain
* T-statistics and critical values (use 5% significance level) and explain
* Diagnostics of multicollinearity using two measures (two methods) and explain
* Results of heteroscedasticity test of residuals and explain
* Results of normality test of residuals and explain

Estimate model 2:

***Ln(FEE) = a0 + a1.Ln(Turnover) + a2.FORSAL + a3.Ln(TA) + a4.LOND + a5.BIG4 + a6.SUB + u***

Evaluate basic statics (R2, F-test, t-test). Analyse multicollinearity using two measures (two methods). Evaluate heteroscedasticity of residuals, and normality of residuals. Use 5% significance level for each statistical test.

Report:

* Estimated coefficients
* R2 and adjusted R2 and explain
* F-statistics of the model and critical value (use 5% significance level) and explain
* T-statistics and critical values (use 5% significance level) and explain
* Diagnostics of multicollinearity using two measures (two methods) and explain
* Results of heteroscedasticity test of residuals and explain
* Results of normality test of residuals and explain
* What is the main difference in the specification of the two models related to interpretation of results?

**TASK 6:** Use data from sheet "equity". Construct a probability model to profile equity

funded companies using logistic regression (model1) and linear probability model (model2). Use the following set of explanatory variables:

charges trade\_debtors audited net\_worth

Interpret impact of explanatory variables from model1 using odds and probabilities. Calculate the accuracy matrix for two thresholds, t1 = 0.09 and t2 = 0.21 for each model. Plot ROC curve for each model and indicate points corresponding to the two thresholds.

Report:

* Estimated coefficients of model1 and model2
* Interpret impact of each of explanatory variables from model1 using odds and probabilities
* Accuracy matrix (confusion matrix) for each of the two thresholds for each model
* ROC curve for each model and indicate points corresponding to the two thresholds (include it as PDF)
* Which model do you consider better and why?

Total points of the assessment = 40 points

Task 1 = max 7 points

Task 2 = max 8 points

Task 3 = max 8 points

Task 4 = max 5 points

Task 5 = max 7 points

Task 6 = max 5 points

Evaluation criteria:

* Ability to do tasks correctly
* Clarity in presentation of your results (your report)

**Send your assignment back by email to** [**martin.alexy@euba.sk**](mailto:martin.alexy@euba.sk) **at the latest on Monday**

**4 May by 15:00 CET.**