DEGREE OF MASTER OF SCIENCE IN FINANCIAL ECONOMICS

FINANCIAL ECONOMETRICS

HILARY TERM 2020 COMPUTATIONAL ASSIGNMENT 1
PRACTICAL WORK 3

February 2020.

This assignment must be submitted before
noon (12:00) Friday 5th Week (21 February 2020)
by uploading to SAMS**.**

This is group work. Groups of 3 or 4 are permitted.
Groups with less than 3 or more than 4 are not permitted without explicit permission.
All solutions must be submitted by the due date and time.
Do not write the names of members of your group on your submission.

*Candidates should answer **all** questions.*
Suggested Length: 7 pages; limit 10 pages.
Material on pages 11+ will not be assessed.
The limit does not include the cover sheet, academic honesty declaration or submitted code files.
All material, including figures, equations, and explanatory text must fit within 10 pages.

Assessment

This assignment is assessed in 3 parts:

- 67% - Report. This report should focus on the analysis of the results and not code or the numerical values of the problems.
- 33% - Autograder. 5 functions must be submitted to compute the required outputs using the inputs. The signature of each function is provided as part of the problem. You must ***exactly*** match the function name. Submissions in MATLAB or Python are accepted. MATLAB submissions must have 1 file per function. Python submissions should contain a single Python file containing all of the functions. MLX (MATLAB) or IPython notebooks (Python) are not accepted. If using Python, pay close attention to the input and output dimensions. All arrays should be NumPy arrays. **Note:** Please run your code in the function you submit to make sure that it does not produce an error. A function that produces an error is given a mark of 0. You must use either MATLAB or Python. Mixed submissions are not accepted and will be marked as MATLAB assignments.

Problem 1

The ID of your group members determines which data series you should analyze. The formula to determine your data set ID is:

$$\text{Data Set ID} = \sum_{i=1}^{n} ID_i \mod 4 + 1$$

where n is the number of members of your group. Your Data Set ID should be in 1, 2, 3 or 4. The data sets are:

Data Set ID	**FRED Code**
1	UNRATENSA
2	HOUSTNSA
3	UMCSENT
4	TOTALNSA

1. Report your Data Set ID and describe the data (quantitative, graphical, qualitative as your group finds useful).
2. Download data for your assigned data set from FRED.
3. Ignoring seasonality, is the data stationary? If not, transform it to be stationary.
4. Using the transformed data (if any was needed), build ARMA models for each series using the first 50% of the data. Assess the accuracy of this model using the second half.
5. Compare your models at horizons 1, 3 and 12 against random walk forecasts.

Code Problems

Mincer-Zarnowitz

Implement Mincer-Zarnowitz regression.

```
[alpha, beta, alpha_tstat, beta_tstat, joint_stat] = mincer_zarnowitz(realization, forecas
```

Outputs

- `alpha` - scalar, $\hat{\alpha}$ from MZ regression
- `beta` - scalar, $\hat{\beta}$ from MZ regression
- `alpha_tstat` - scalar, the t-stat for the null $H_0 : \alpha = 0$ in a MZ regression
- `beta_tstat` - scalar, the t-stat for the null $H_0 : \beta = 1$ in a MZ regression
- `joint_stat` - scalar, Wald statistic testing the joint null of correct specification

Inputs

- `realization` - R by 1 vector of realizations of variable being forecast
- `forecast` - R by 1 vector of 1-step ahead forecasts variable being forecast. The forecast in position i is the time $i-1$ forecast of the value in i so that the forecast errors are `realization - forecast`.

Diebold-Mariano

Implement a Diebold-Mariano test.

```
[avg_diff, std_err, dm_stat, concl] = diebold_mariano(loss_a, loss_b, nw_bandwidth)
```

Outputs

- `avg_diff` - Mean loss difference where $\delta_t = L_t^A - L_t^B$ (scalar).
- `std_err` - Estimated standard error of the difference (scalar).
- `dm_stat` - The Diebold-Mariano test statistic (scalar).
- `concl` - Conclusion. Should be -1 if null is rejected in favor of model A, 0 if null is not rejected and 1 if null is rejected in favor of model B. You should use a 5% size for the test (2.5% in each tail).

Inputs

- `loss_a` - R by 1 vector of losses from model A.
- `loss_b` - R by 1 vector of losses from model B.
- `nw_bandwidth` - Bandwidth (number of lags) to use in the Newey-West estimator (scalar).

Problem 2

Forecasting persistent time series is hard. Design and implement a Monte Carlo question to determine for which values of $\hat{\rho}$ it is better to assume that the true value of $\rho = 1$ instead of using the estimated value when forecasting the time series. You should simulate data from the model

$$y_t = \rho y_{t-1} + \epsilon_t \tag{1}$$

where $\epsilon_t \overset{iid}{\sim} N(0,1)$ for values of $\rho \in \{0.90, 0.91, \ldots, 0.99\}$ and $y_0 \sim N(0, 1/1-\rho^2)$. You should allow the model to have a constant, so that you estimate

$$y_t = \phi_0 + \phi_1 y_{t-1} + \epsilon_t.$$

Each time you simulate a time-series and estimate a model you get the estimated parameter values. You can then compare the forecast produced using the estimated parameters against a random walk forecast (which has $\phi_0 = 0$ and $\phi_1 = 1$) on using out-of-sample values. Use sample sizes of 100, 200 and 500 observations where you use 50% of the simulated data to estimate parameters and the other half to evaluate the model. For each of the sample sizes, describe the probability that the random walk forecast outperforms the estimated parameter forecast *as a function of* $\hat{\phi}_1$. Use your analysis to determine the point at which it is better, on average, to ignore your parameter estimate and use the simple RW model. Your analysis should indicate how sample size affects your conclusion.

Next, use recursive estimation starting with 50% of the sample and continuing until the end where you use observations 1, ..., t to estimate parameters when forecasting $t + 1$. Comment on how this affects your initial conclusion.

AR(1) Simulation

Implement an AR(1) simulator.

```
y = ar1_simulate(rho, errors, y0)
```

Outputs

- `y` - The simulated time series using eq. (1) (T by 1).

Inputs

- `rho` - Scalar autoregressive coefficient.
- `errors` - T by 1 vector of errors to use in the model.
- `y0` - Scalar initial value.