Financial Econometrics Rauli Susmel
FINA 4397/7354-A

Second Midterm

Note: Please, follow the instructions received in the email you received this file.

1. (**30 points. Modeling Strategies**).

a. Starting from a General Unrestricted Model (GUM), using all the variables you can think of that make sense to include, select an appropriate (reduced) model for SF, (SF_c).
a1. Report GUM regression.

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  7.441e-04  1.989e-05  37.408  < 2e-16 ***
x_usa       -1.185e-03  4.869e-04  -2.434 0.015400 *  
u_sd         1.909e-05  5.248e-06   3.637 0.000315 ***
x_tech       1.948e-04  1.060e-04   1.838 0.066898 .  
Cind         2.371e-03  2.219e-03   1.069 0.285989    
Mkt_RF      -2.230e-05  9.017e-05  -0.247 0.804850    
SMB         -1.120e-04  8.160e-05  -1.372 0.170845    
HML          3.076e-05  9.993e-05   0.308 0.758360    
RMW         -4.258e-05  1.066e-04  -0.400 0.689695    
CMA          2.864e-06  1.475e-04   0.019 0.984519    
gold         4.776e-06  5.032e-05   0.095 0.924430    
oil          7.147e-07  2.220e-05   0.032 0.974337    
x_usa2      -5.951e-02  2.970e-02  -2.004 0.045830 *  
u_sd2       -1.499e-09  4.328e-07  -0.003 0.997238    
Cind2        1.223e-03  1.152e-02   0.106 0.915531    
x_tech2      3.314e-04  2.472e-04   1.341 0.180914    
Spring      -1.634e-05  1.759e-05  -0.929 0.353480    
Summ         1.750e-05  1.732e-05   1.010 0.312979    
Fall        -3.716e-06  1.754e-05  -0.212 0.832322    
Fin_c       -7.613e-05  1.773e-05  -4.293 2.26e-05 ***
u_sd_Cind    1.764e-06  2.182e-04   0.008 0.993554    
u_sd_tech   -1.539e-05  1.697e-05  -0.907 0.364891    
u_sd_Spring -4.419e-07  2.915e-06  -0.152 0.879602    
u_sd_Summ   -3.576e-06  2.849e-06  -1.255 0.210182    
u_sd_Fall   -3.010e-08  2.891e-06  -0.010 0.991698    
Cind_Spring  3.368e-03  2.092e-03   1.610 0.108271    
Cind_Summ    9.101e-04  2.261e-03   0.402 0.687589    
Cind_Fall    1.362e-03  1.873e-03   0.727 0.467511    
tech_Spring -2.501e-05  8.405e-05  -0.298 0.766172    
tech_Summ   -1.205e-05  8.477e-05  -0.142 0.887022    
tech_Fall   -1.059e-04  7.862e-05  -1.347 0.178690    
u_sd_Finc   -8.755e-06  2.849e-06  -3.073 0.002280 ** 
Cind_Finc   -5.161e-03  1.735e-03  -2.975 0.003123 ** 
tech_Finc   -1.363e-04  7.569e-05  -1.801 0.072597 .  
Finc_Spring  1.176e-05  1.315e-05   0.894 0.371997    
Finc_Summ    8.743e-06  1.316e-05   0.664 0.506908    
Finc_Fall   -1.932e-06  1.302e-05  -0.148 0.882141    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.152e-05 on 366 degrees of freedom
Multiple R-squared:  0.7511,	Adjusted R-squared:  0.7266 
F-statistic: 30.67 on 36 and 366 DF,  p-value: < 2.2e-16
```

a2. Report reduced regression.

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  7.635e-04  1.133e-05  67.392  < 2e-16 ***
x_usa       -1.551e-03  3.285e-04  -4.723 3.24e-06 ***
u_sd         1.643e-05  2.095e-06   7.843 4.12e-14 ***
x_usa2      -2.251e-02  2.496e-02  -0.902    0.368    
Fin_c       -8.643e-05  1.359e-05  -6.359 5.59e-10 ***
u_sd_Finc   -9.296e-06  2.354e-06  -3.948 9.31e-05 ***
Cind_Finc    1.238e-04  2.010e-04   0.616    0.538    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.236e-05 on 396 degrees of freedom
Multiple R-squared:  0.7197,	Adjusted R-squared:  0.7154 
F-statistic: 169.4 on 6 and 396 DF,  p-value: < 2.2e-16
```

c. Did the 2008 Financial Crisis affect SF home prices?

Yes! Fin_c is very significant in the reduced model. Interaction between Fin_c and Unemployment in SD is very significant (& also negative).

d. Do you have evidence of seasonality –i.e., are the dummy variables for Spring, Summer or Fall significant?

No. None of the seasonal dummies shows up as significant, directly or interacting with other variables.

e. Check if the errors of your reduced model are normal (use a Jarque-Bera test).

```
	Jarque Bera Test

data:  e_sd
X-squared = 3.3644, df = 2, p-value = 0.186
```

We cannot reject H_0. There is no evidence of non-normality!

f. Check that the model's errors do not show autocorrelation.

```
	Durbin-Watson test

data:  fit_sd_red
DW = 1.4276, p-value = 9.99e-10
alternative hypothesis: true autocorrelation is greater than 0

> bgtest(fit_sd_red, order=4)

	Breusch-Godfrey test for serial correlation of order up to 4

data:  fit_sd_red
LM test = 142.34, df = 4, p-value < 2.2e-16
```

Both tests reject H_0. There is evidence of autocorrelation

g. Check that the model's errors do not show heteroscedasticity.

```
> gqtest(fit_sd_red)

	Goldfeld-Quandt test

data:  fit_sd_red
GQ = 0.69346, df1 = 195, df2 = 194, p-value = 0.9945
```

```
alternative hypothesis: variance increases from segment 1 to 2
```

```
> bptest(fit_sd_red)

        studentized Breusch-Pagan test

data:  fit_sd_red
BP = 46.169, df = 6, p-value = 2.74e-08
```

BP test shows strong evidence of heteroscedasticity.

h. If they do show autocorrelation and/or heteroscedasticity, use proper SE to conduct tests of significance for the coefficients for the driver variables in the reduced model.

```
> t_b_NW
(Intercept)       x_usa        u_sd      x_usa2       Fin_c   u_sd_Finc 
Cind_Finc 
 27.1990601  -2.3140464   3.6458490  -0.6066121  -3.0638519  -1.9672615 
0.9483826
```

Relative to OLS, the t-stats get a bit lower, but the same variables are still significant at the 5% level. That is, once, we take into account autocorrelation and heteroscedasticity x_usa, u_sd, Finc_c and u_sd interating with Finc are significant at 5%.

i. Use an LM test to check if there is monthly seasonality in the residuals of your reduced model.

```
> lm_seas <- lm(e_sd ~ Feb + Mar + Apr + May + Jun + Jul + Aug + Sep + 
Oct + Nov + Dec)
> R2_r <- summary(lm_seas)$r.squared  # extracting  R^2 from fit_lm
> R2_r
[1] 0.2007784
> LM_test <- R2_r * length(e_sd)
> LM_test
[1] 80.91368
> qchisq(.95, df = 11)                # chi-squared (df=2) value at 5% 
level
[1] 19.67514
> p_val <- 1 - pchisq(LM_test, df = 11)                      # p-value 
of LM_test
> p_val
[1] 9.822143e-13
```

The p-value is very small. There is strong evidence of monthly seasonality in the residuals of the reduced model. We should go back to reformulate the model to include monthly seasonal variables. If you check summary(lm_seas), you will see that March, April, May and July are significant.

2. (**25 points. Forecasting**). You want to forecast log changes in home prices, p_t, using an AR(1) model.

a. Estimate an AR(1) model for p_t. using data from Jan 1990 to Dec 2020 (estimation period). That is, you estimate the following AR(1) model:

$$p_t = \mu + \phi_1 p_{t-1} + \varepsilon_t. \qquad \varepsilon_t.\sim \text{WN}.$$

Report the regression.

```
y <- x_sd
T1 <- 372
y_1 <- y[1:(T1-1)]                                   # Estimation period data
y_0 <- y[2:T1]                                       # Estimation period data
fit_y <- lm(y_0 ~ y_1)                               # Fit AR(1) model for e_f,t
b_y <- fit_y$coefficients
summary(fit_y)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.704e-04  2.572e-05   6.627 1.22e-10 ***
y_1         7.848e-01  3.222e-02  24.360  < 2e-16 ***
```

b. Using your AR(1) estimates from 2.a, forecast p_t from Jan 2021 to July 2023 (validation period). Compute the MSE of your AR(1) forecast.

```
T_for <- T-T1
xx_cons <- rep(1,T_for)
T_val <- T1 + 1
y_f0 <-  cbind(xx_cons,y[T1:(T-1)])%*% b_y          # b_est coef from estimation period reg
S_ar1_f0 <- S[T1:(T-1)]*(1+y_f0)                    # Forecast for S_t, using validation data
e_ar1_f0 <- S[T_val:T] - S_ar1_f0                   # Forecasat error
mse_e_ar1_f0 <- sum(e_ar1_f0^2)/k_for               # MSE
> mse_e_ar1_f0                                               # MSE(2)
[1] 0.2188236
```

c. Using a the random walk model (RW) for p_t, forecast p_t from Jan 2021 to July 2023. Compute the MSE of your RW forecast.

```
e_rw_f0 <- S[T_val:T] - S[T1:(T-1)]                 # Error for RW model =>  et (1)
mse_e_rw_f0 <- sum(e_rw_f0^2)/k_for
> mse_e_rw_f0
[1] 0.2264603
```

d. Test the equality of MSEs using an MGN/HLN test. Interpret the test results.

```
> ## 2.d Testing Equality of MSE: Mod vs RW
> z_mgn <- e_rw_f0 + e_ar1_f0
> x_mgn <- e_rw_f0 - e_ar1_f0
> fit_mgn <- lm(z_mgn ~ x_mgn)
> summary(fit_mgn)

Call:
lm(formula = z_mgn ~ x_mgn)

Residuals:
    Min      1Q  Median      3Q     Max
-1.5166 -0.3839 -0.0428  0.3279  1.1995

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    2.665      3.928   0.679    0.503
x_mgn        -40.732     61.574   0.662    0.514

Residual standard error: 0.7107 on 29 degrees of freedom
Multiple R-squared:  0.01487,	Adjusted R-squared:  -0.0191
```

```
F-statistic: 0.4376 on 1 and 29 DF,  p-value: 0.5135
```

The p-value of the t-stat of x_mgn is **0.514** (not significant): We cannot reject that both MSEs are equal.

3. (**15 points. ARMA identification**). For this question you will use the log changes in oil prices (oil). Below, we show the first 17 autocorrelations (you can do this by using the R command: acf(oil)):

```
Autocorrelations of series ‘oil’, by lag

 0       1      2      3      4      5      6      7      8      9
1.000  0.249 -0.038 -0.106 -0.136 -0.053 -0.034  0.003  0.009 -0.004

   10     11     12     13     14     15     16     17
0.042  0.035 -0.007 -0.032 -0.050 -0.049  0.015  0.009
```

a. You have T=**403** observations. How many autocorrelation are statistically different from 0?.
C.I.: $\pm$ .0976 $\Rightarrow$ Only 3 autocorrelations are different from 0 (1, 3 & 4)

b. Compute the Ljung-Box test with 3 lags. Interpret the test result.
LB <- T*(T+2)*((-0.249)^2/(T-1) + -0.038^2/(T-2) + (-0.106)^2/(T-3))

```
> LB
[1] 29.16984    # very significant at 5%, chi-squared[df=3] = 7.31
```

Interpretation: The first 3 autocorrelations are jointly different from 0. .

c. Using the above ACF, you consider an AR(1) model for changes in oil prices. Is the AR(1) process stationary? (Hint: No need to compute moments here, just get ϕ_1.)
Yes! ϕ_1 = **0.249** $\Rightarrow$ Since $|\phi_1| < 1$, the AR(1) process is stationary.

d. Using your ϕ_1 parameter, compute the accumulated impact of a shock after 3 months. That is, the IRF at J=3. Interpret the result

```
> IRF_3 <-    (phi_1 + phi_1^2 + phi_1^3)
[1] 0.3264392
```

Interpretation: After 3 months the accumulated shock is 0.326.

4. True of False (**20 points**). Briefly justify all your answers.
a. If errors are heteroscedastic, OLS is biased, but consistent.
False. Under usual assumptions, if errors are heteroscedastic, OLS is unbiased, and consistent.

b. On average, FGLS estimates should be similar to the OLS estimates.

True. OLS and FGLS are both consistent.

c. NW SE are not useful when we have no autocorrelation in the residuals.
True. White SE, if there is heteroscedasticiy, or just OLS SE, if there is no heteroscedasticiy will be fine.

d. A Random Walk model without a drift is stationary, since it has a constant mean.
False. The variance is time dependent; it will be explosive as T grows

e. If the coefficients of an MA(2) process are all greater than 1, the MA(2) process is not stationary.
False. MA processes are always stationary.