Università Commerciale
Luigi Bocconi

MSc. Finance/CLEFIN
2017/2018 Edition

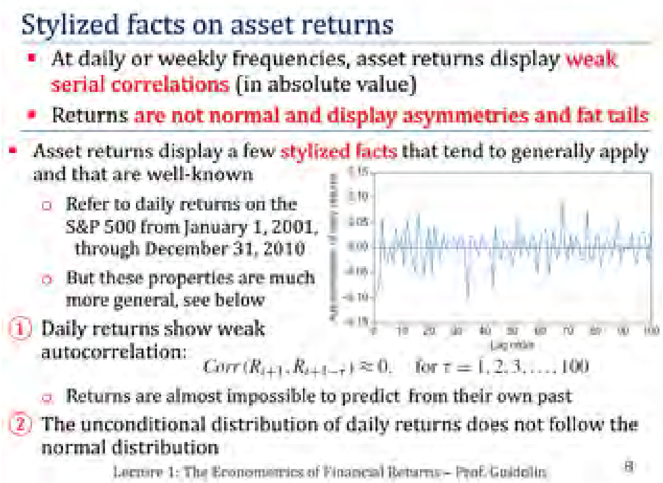
FINANCIAL ECONOMETRICS AND EMPIRICAL FINANCE - MODULE 2

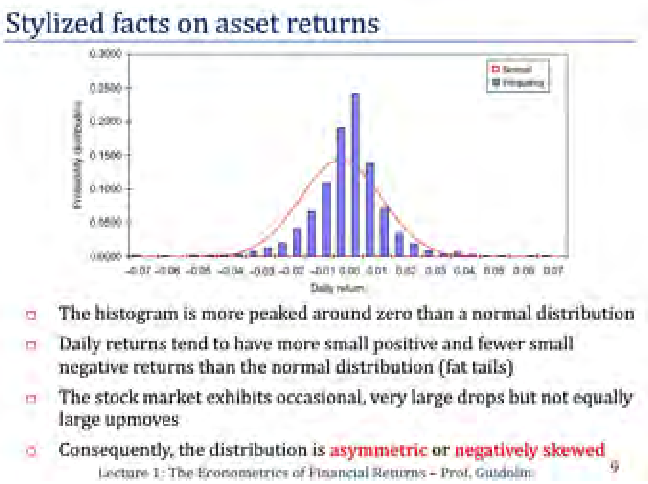
General Exam – October 2018
Time Allowed: 2 hours and 20 minutes

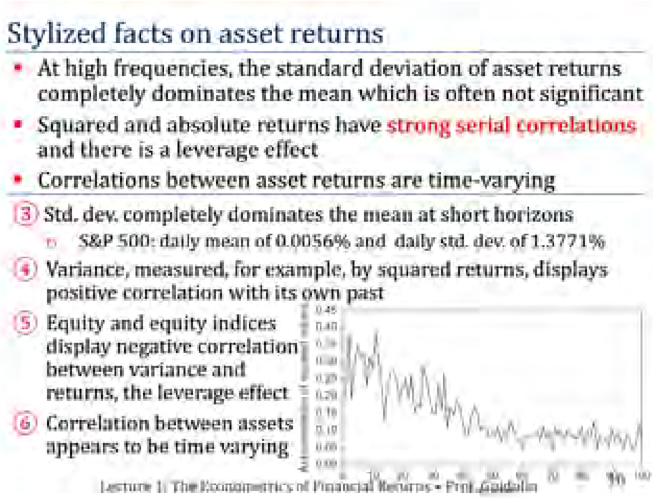
Please answer all the questions by writing your answers <u>in the spaces provided</u>. There are two optional questions (7 and 8). No additional papers will be collected and therefore they will not be marked. You always need to carefully justify your answers and show your work. The exam is closed book, closed notes. No calculators are useful or permitted. You can withdraw until 10 minutes before the due time.

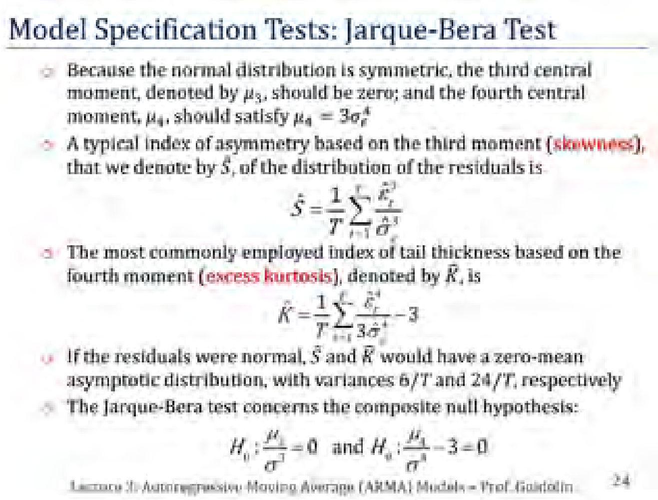
Question 1.A (10 points)
Describe in detail, also with reference to the examples that have been provided in the lectures, the six stylized facts that are typically displayed by asset returns. Discuss how you would proceed to test whether the unconditional distribution of asset returns is Gaussian. Make sure to also mention the main reasons why the returns may deviate from a Gaussian distribution.

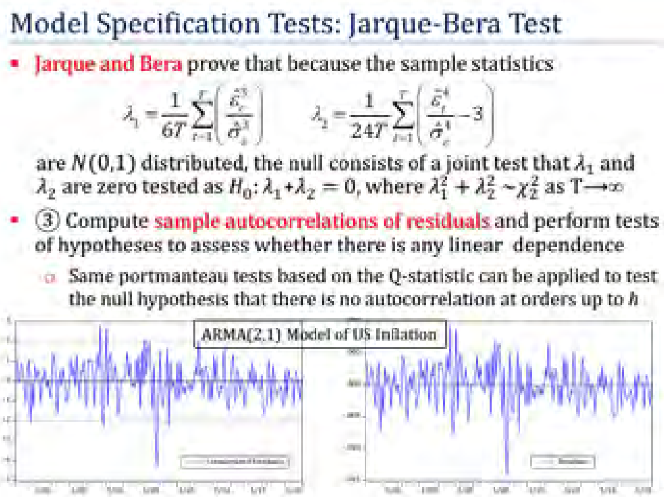
Debriefing







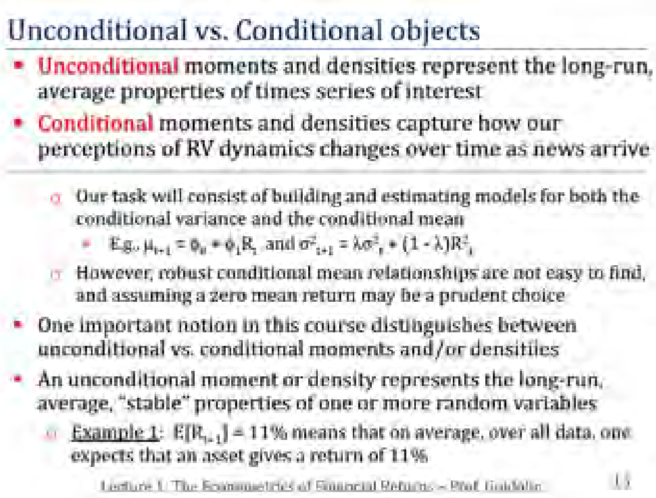


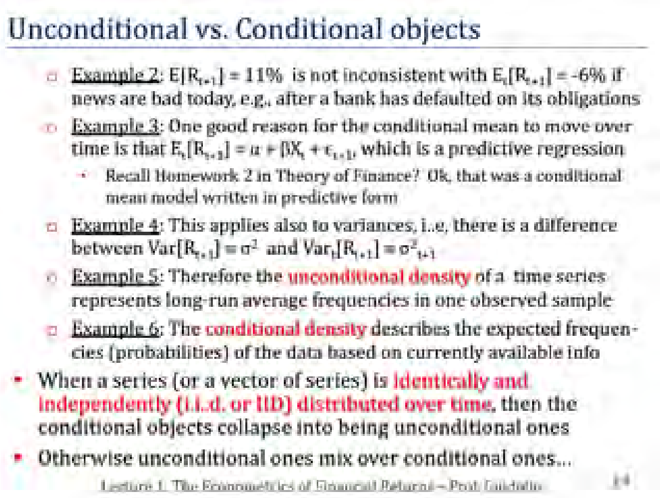


Question 1.B (4 points)

Ms. Granger, a junior analyst at Badcredit Bank, is trying to persuade her boss, Nic Dwarf, that $E[R_{t+1}] = 11\%$ is not incompatible with $E_t[R_{t+1}] = -11\%$. Do you agree with her claim? Support your answer by referring to the concept of conditional vs. unconditional distribution and moments.

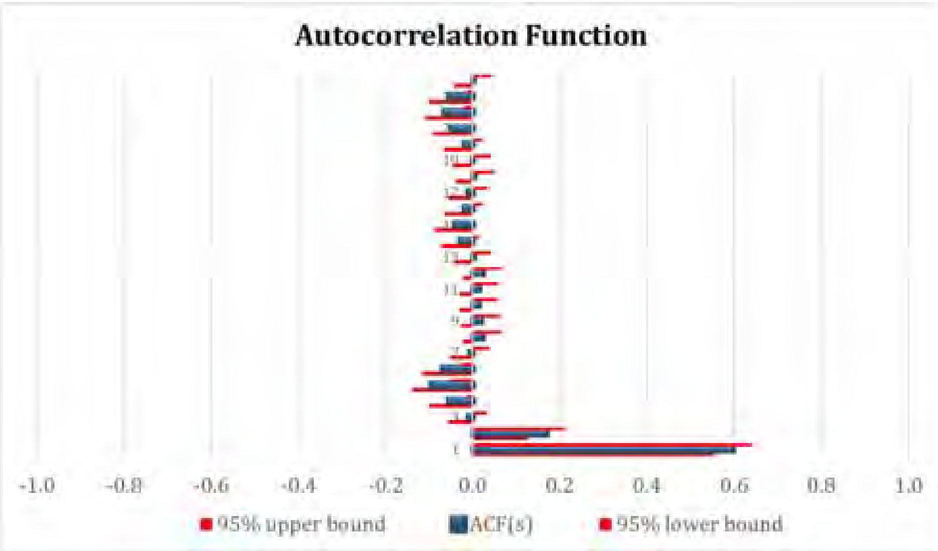
Debriefing





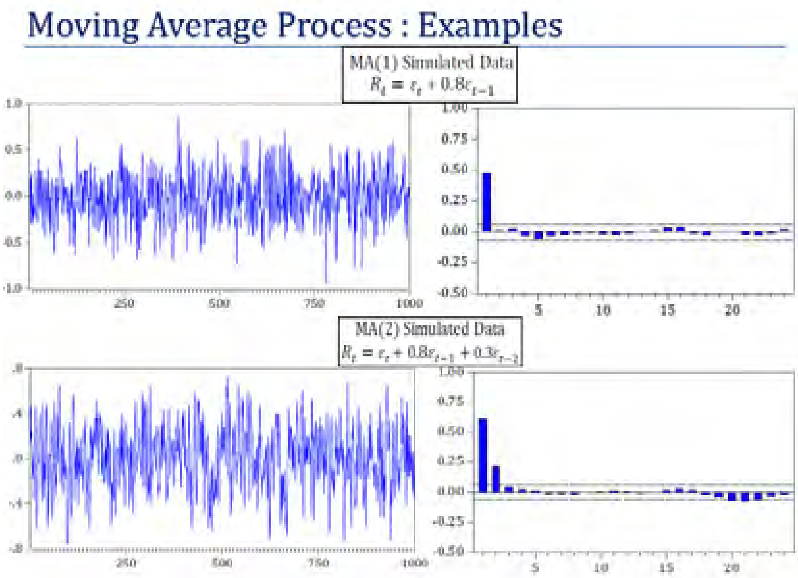
Question 1.C (3 points)

Mr. Dwarf has now assigned to Ms. Granger the task to analyze the features of the returns of an emerging market stock index. In particular, he is convinced that the returns follow the process $R_{t+1} = \mu + \sigma_{t+1} z_{t+1}$, with $z_{t+1} \sim IID\ D(0,1)$. As a first step, Ms. Granger has decided to estimate the correlogram of the series, which is reported below. On the basis of this evidence, do you believe that Mr. Dwarf's intuition about the process followed by the data is correct? Make sure to clearly justify your answer.



Debriefing.

The model assumed by Mr. Dwarf implies zero autocorrelation between the returns. However, this does not seem to be the case when we look at the correlogram. Indeed, the correlogram is compatible with an ARMA model, but not with a white noise process.



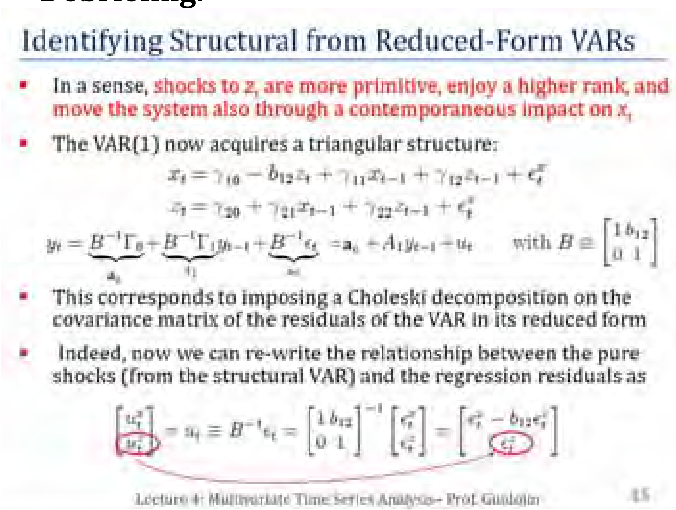
Question 2.A (10 points)

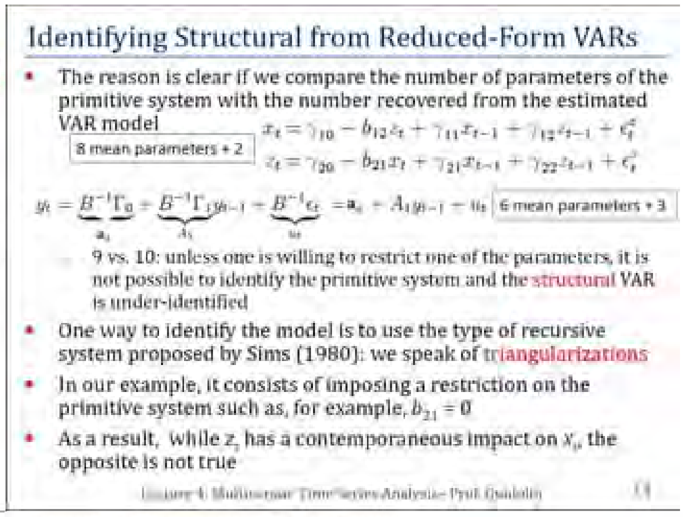
Consider the following VMA(∞) representation of a VAR(1) model

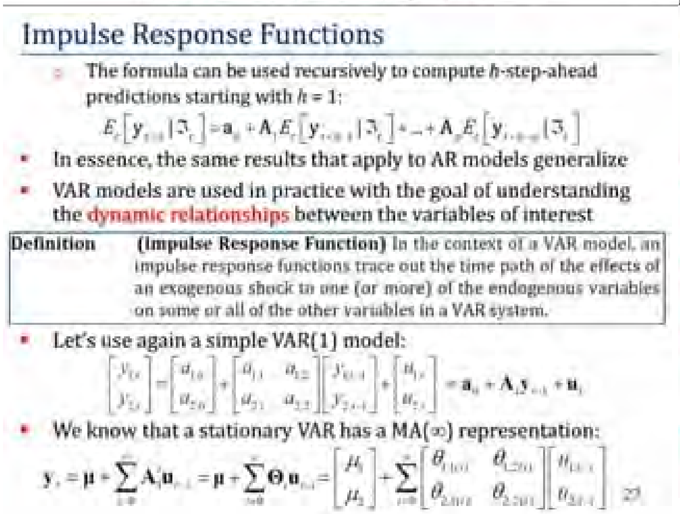
$$\boldsymbol{y}_t = \boldsymbol{\mu} + \sum_{i=1}^{\infty} \boldsymbol{\Theta}_i \boldsymbol{u}_{t-1} + \boldsymbol{u}_t,$$

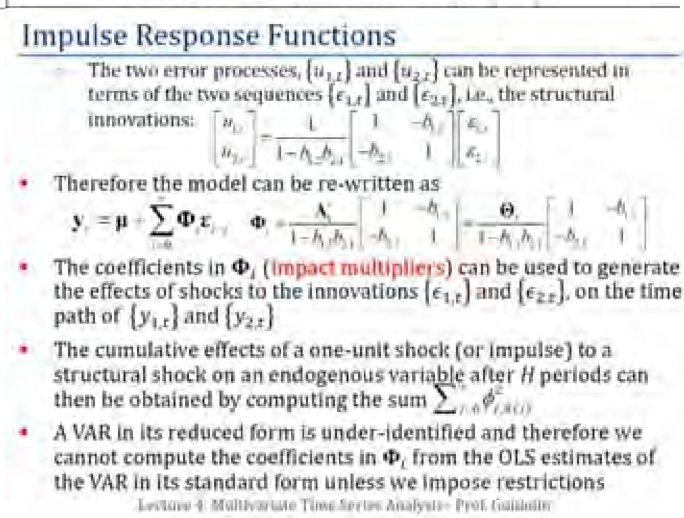
where $\boldsymbol{\Theta}_i = \boldsymbol{A}_1^i$ and $\boldsymbol{A}_1$ is the matrix of the coefficients of the *reduced form* VAR(1). Can we interpret the coefficients $\boldsymbol{\Theta}_i$ as impact multipliers of the true, *structural* innovations? If not, carefully explain why and discuss whether and under what conditions it is possible to retrieve the impact multipliers to structural innovations from the OLS estimates of a VAR in its reduced form. Finally, discuss which kind of information is entailed in a variance decomposition of forecast errors and specify whether some identification scheme must be imposed in order to retrieve such information.

Debriefing.









Impulse Response Functions

- One method to place these restrictions consists of the application of a Choleski decomposition: $\mathbf{y}_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \boldsymbol{\Theta}_i \mathbf{W}\mathbf{W}^{-1}\mathbf{u}_{t-i}$

  $(\boldsymbol{\Sigma}_u = \mathbf{W}\boldsymbol{\Sigma}\mathbf{W}',\ \boldsymbol{\varepsilon}_t = \mathbf{W}^{-1}\mathbf{u}_{t-i}$, and $\boldsymbol{\Phi}_i = \boldsymbol{\Theta}_i\mathbf{W})$
  - Because of the triangular structure of $\mathbf{W} = \mathbf{B}^{-1}$, a Choleski decomposition allows only the shock to the first variable to contemporaneously affect all the other variables in the system
  - A shock to the second variable will produce a contemporaneous effect on all the variables in the system, but the first one
  - This may of course be impacted in the subsequent period, through the transmission effects mediated by the autoregressive coefficients
  - A shock to the third variable will affect all the variables in the system, but the first two, and so on
- Therefore, a Choleski identification scheme forces a potentially important identification asymmetry on the system
- A different ordering of the variables in the system would have been possible, implying a reverse ordering of the shocks

Lecture 4: Multivariate Time Series Analysis– Prof. Guidolin

3.1 Impulse Response Functions

- VAR models can be used to understand the dynamic relationships between the variables of interest.

Impulse Response Function: In the context of a VAR model, an impulse response function traces out the time path of the effects of an exogenous shock to one (or more) of the endogenous variables on some or all of the other variables in a VAR system.

Impact multipliers: coefficients of the matrix Φ_i.

Starting from the moving average representation of a VAR(1)

$$y_t = a_0 + A_1 y_{t-1} + u_t = \mu + \sum_{i=0}^{\infty} A_1^i u_{t-i} = \mu + \sum_{i=0}^{\infty} \Theta_i u_{t-i}$$

17

Variance Decompositions

- Understanding the properties of forecast errors from VARs is helpful in order to assess the interrelationships among variables
- Using the VMA representation of the errors, the h-step-ahead forecast error is $\mathbf{u}_t(h) = \mathbf{y}_{t+h} - E_t[\mathbf{y}_{t+h}] = \sum_{i=0}^{h-1} \boldsymbol{\Phi}_i \boldsymbol{\varepsilon}_{t+h-i}$
  - See lecture notes for algebra of such representation
  - Because all white noise shocks the same variance, if we denote by $\sigma^2_{y1}(h)$ the h-step-ahead variance of the forecast of (say) y_1, we have:

$$\sigma^2_{y1}(h) = \sigma^2_{\varepsilon 1}\left[\phi^2_{11}(0) + \phi^2_{11}(1) + \dots + \phi^2_{11}(h-1)\right] + \sigma^2_{\varepsilon 2}\left[\phi^2_{12}(0) + \phi^2_{12}(1) + \dots + \phi^2_{12}(h-1)\right]$$

  - Because all the coefficients in $\boldsymbol{\Phi}_i$ are non-negative, the variance of the forecast error increases as the forecast horizon h increases
  - We decompose the h-step-ahead forecast error variance into the proportion due to each of the (structural) shocks

$$\frac{\sigma^2_{\varepsilon 1}\left[\phi^2_{11}(0) + \phi^2_{11}(1) + \dots + \phi^2_{11}(h-1)\right]}{\sigma^2_{y1}(h)} \qquad \frac{\sigma^2_{\varepsilon 2}\left[\phi^2_{12}(0) + \phi^2_{12}(1) + \dots + \phi^2_{12}(h-1)\right]}{\sigma^2_{y1}(h)}$$

- Such proportions due to each shock is a variance decomposition

Lecture 4: Multivariate Time Series Analysis– Prof. Guidolin

Variance Decompositions

- Like in IRF analysis, variance decompositions of reduced-form VARs require identification (because otherwise we would be unable to go from the coefficients in $\boldsymbol{\Theta}_i$ to their counterparts in $\boldsymbol{\Phi}_i$)
  - Choleski decompositions are typically imposed
  - Forecast error variance decomposition and IRF analyses both entail similar information from the time series
  - Example on weekly US Treasury yields, 1990-2016 sample:

Choleski ordering:
_ 1M Yield
_ 1Y Yield
_ 5Y Yield
_ 10Y Yield

Variance Decomposition of 10Y Yield:

Period	S.E.	1M Yield	1Y Yield	5Y Yield	10Y Yield
1	0.101	0.260	37.242	51.653	10.845
2	0.158	0.306	36.977	52.897	9.820
3	0.202	0.285	37.053	53.316	9.345
4	0.237	0.253	37.265	53.389	9.093
5	0.268	0.221	37.527	53.311	8.941
6	0.296	0.194	37.805	53.163	8.839
7	0.321	0.170	38.086	52.981	8.763
8	0.344	0.150	38.363	52.784	8.703
9	0.365	0.134	38.633	52.580	8.653
10	0.385	0.120	38.896	52.375	8.608
11	0.404	0.109	39.151	52.171	8.568
12	0.422	0.100	39.399	51.971	8.530

Lecture 4: Multivariate Time Series Analysis– Prof. Guidolin

or, for example in the case of a VAR(1)

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \sum_{i=0}^{\infty} \begin{bmatrix} \theta_{1,1(i)} & \theta_{1,2(i)} \\ \theta_{2,1(i)} & \theta_{2,2(i)} \end{bmatrix} \begin{bmatrix} u_{1,t-i} \\ u_{2,t-i} \end{bmatrix}$$

where

$$\begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix} = \frac{1}{1 - b_{1,2}b_{2,1}} \begin{bmatrix} 1 & -b_{1,2} \\ -b_{2,1} & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$

Then

$$\Phi_i = \frac{A_1^i}{1 - b_{1,2}b_{2,1}} \begin{bmatrix} 1 & -b_{1,2} \\ -b_{2,1} & 1 \end{bmatrix} = \frac{\Theta_i}{1 - b_{1,2}b_{2,1}} \begin{bmatrix} 1 & -b_{1,2} \\ -b_{2,1} & 1 \end{bmatrix}$$

and

$$y_t = \mu + \sum_{i=0}^{\infty} \Phi_i \epsilon_{t-i}$$

For example, $\phi_{1,2(0)}$ is the instantaneous impact on $y_{1,t}$ of a one-unit change in $\epsilon_{2,t}$.

Cumulative response of the variable j to a shock to the variable k:

$$\sum_{i=0}^{H} \phi_{j,k,(i)}$$

For example, $\sum_{i=0}^{H} \phi_{1,2(i)}$ is the cumulative effects of a one-unit shock (or impulse) to $\epsilon_{2,t}$ on the variable $y_{1,t}$ after H periods.

Long-run impact multipliers: impact multipliers when $H \to \infty$.

- The set of elements $\phi_{j,k(i)}$, with $i = 1, \dots, H$ is the impulse response function of the jth variable of the system, up to the period H.
- VAR in its reduced form is under-identified by construction and therefore $\phi_{j,k(i)}$ cannot be computed from the OLS estimates of the VAR in its standard form without imposing adequate restrictions.
- Choleski decompositions provide a minimal set of restrictions concerning the simultaneous relationships among variables that can be used to identify the structural model, but this method forces potentially important identification asymmetry on the system.
- IRFs are constructed using estimated coefficient, thus will contain sampling error. Therefore, it is advisable to construct confidence intervals around them to account for the uncertainty that derives from parameter estimation.

Question 2.B (3 points)

Max Earlgrey, a senior economist at BundBank Inc., is selecting the best VAR(p) model for a vector of time series that includes 1-month, 1-, 5-, and 10-year US Treasury bond rates. On the basis of a sample of 1,395 observations, he reports that a likelihood ratio test (LRT) of a VAR(1) vs. a VAR(2) gives a test statistic of 203.72. In addition, Max has determined that the LRT of a VAR(2) vs. a VAR(3) gives a test statistic of 38.56. Would Max be able to compute the (S)BIC for a VAR(2) using the information reported? In the affirmative case, please show how, otherwise clearly discuss why not and which additional information would he require.

Debriefing.

The formula for (S)BIC is the following

$$ln\left|\widetilde{\boldsymbol{\Sigma}}_u(p)\right| + \frac{2}{T} lnT(N^2p + N).$$

The number of observations (T) has been given to you in the text of the exercise and also N, the number of variables in the system is known (they are 4). However, you are not able to extract the value of $ln\left|\widetilde{\boldsymbol{\Sigma}}_u(2)\right|$ from the information that you were given. Therefore, in order to be able to perform the computation, one of the following information is sufficient:

_ $ln\left|\widetilde{\boldsymbol{\Sigma}}_u(1)\right|$
_ $ln\left|\widetilde{\boldsymbol{\Sigma}}_u(2)\right|$ (obviously)
_ $ln\left|\widetilde{\boldsymbol{\Sigma}}_u(3)\right|$ (obviously)

Question 2.C (4 points)

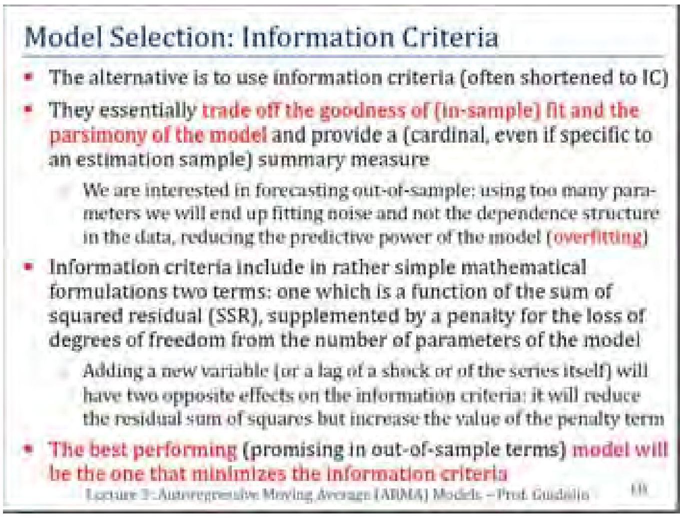
A younger colleague of Dr. Earlgrey, Miss Granger, pointedly suggests that they shall conduct a full specification search, and produces the table below. Which is the model selected by each of the three information criteria? Do they all lead to the selection of the same model and, if not, is this plausible?

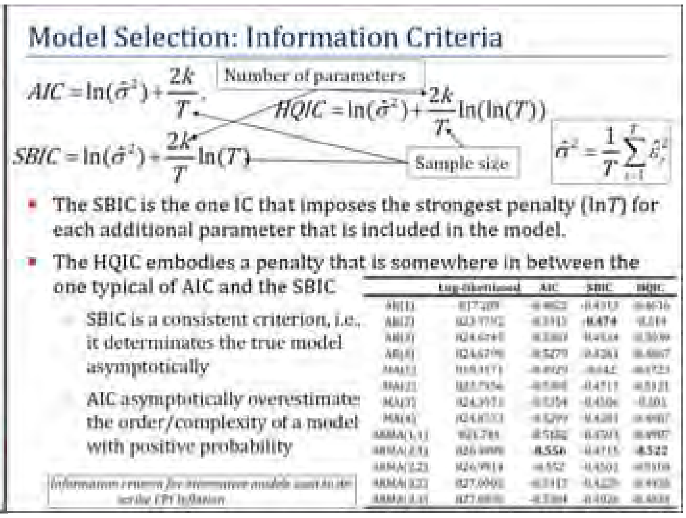
VAR Lag Order Selection Criteria
Endogenous variables: ONEMONTH ONEYEAR FIVEYEARS TENYEARS
Exogenous variables: C
Date: 10/11/18 Time: 17:43
Sample: 1/05/1990 12/30/2016
Included observations: 1395

Lag	LogL	LR	AIC	SC	HQ
0	-4984.879	NA	7.152514	7.167541	7.158133
1	6162.080	22214.01	-8.805849	-8.730715	-8.777757
2	6264.603	203.7234	-8.929897	-8.794654	-8.879331
3	6284.063	38.55692	-8.934857	-8.739507	-8.861818
4	6299.304	30.11002	-8.933769	-8.678310	-8.838256
5	6354.346	108.4274	-8.989743	-8.674177	-8.871756
6	6375.059	40.68298	-8.996500	-8.620826	-8.856039
7	6390.868	30.96155	-8.996226	-8.560445	-8.833292
8	6406.572	30.66431	-8.995802	-8.499912	-8.810394
9	6419.565	25.29749	-8.991491	-8.435494	-8.783610
10	6443.217	45.91304	-9.002461	-8.386356	-8.772106
11	6460.887	34.20005	-9.004855	-8.328643	-8.752027
12	6473.561	24.45873	-9.000088	-8.263767	-8.724785
13	6490.912	33.38322	-9.002024	-8.205596	-8.704248
14	6504.644	26.34175	-8.998773	-8.142236	-8.678523

Debriefing.

The exercise requires you to find the model that **minimizes** each of the three information criteria. Namely, the AIC selects a VAR(11) model while both the SC and the HQ criterion select a more parsimonious VAR(2) model. It is perfectly plausible that the three criteria lead to the selection of different models, see slides below.



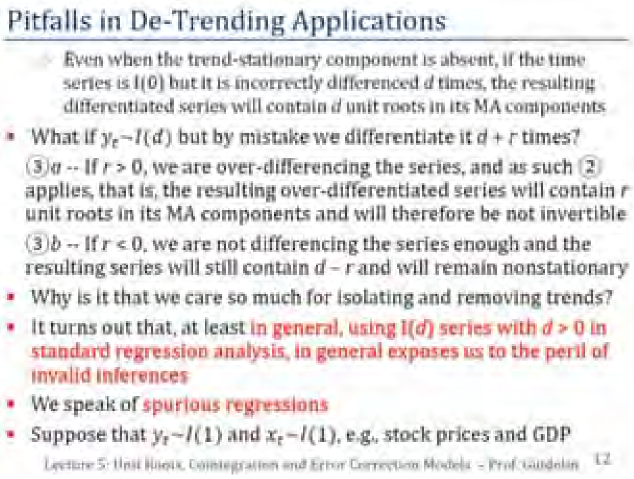


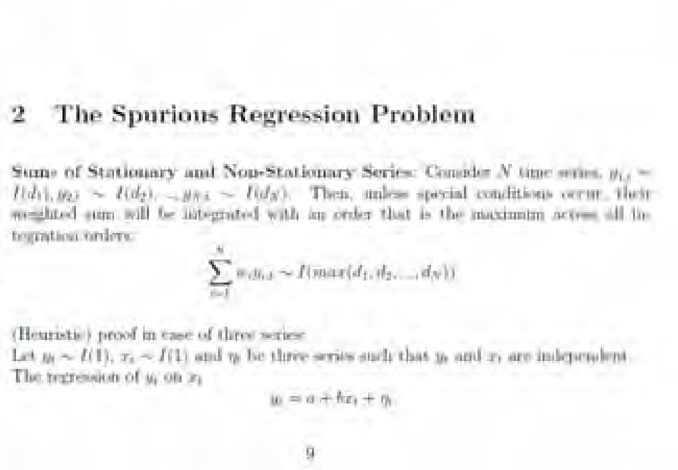
	Log-likelihood	AIC	SBIC	HQIC
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

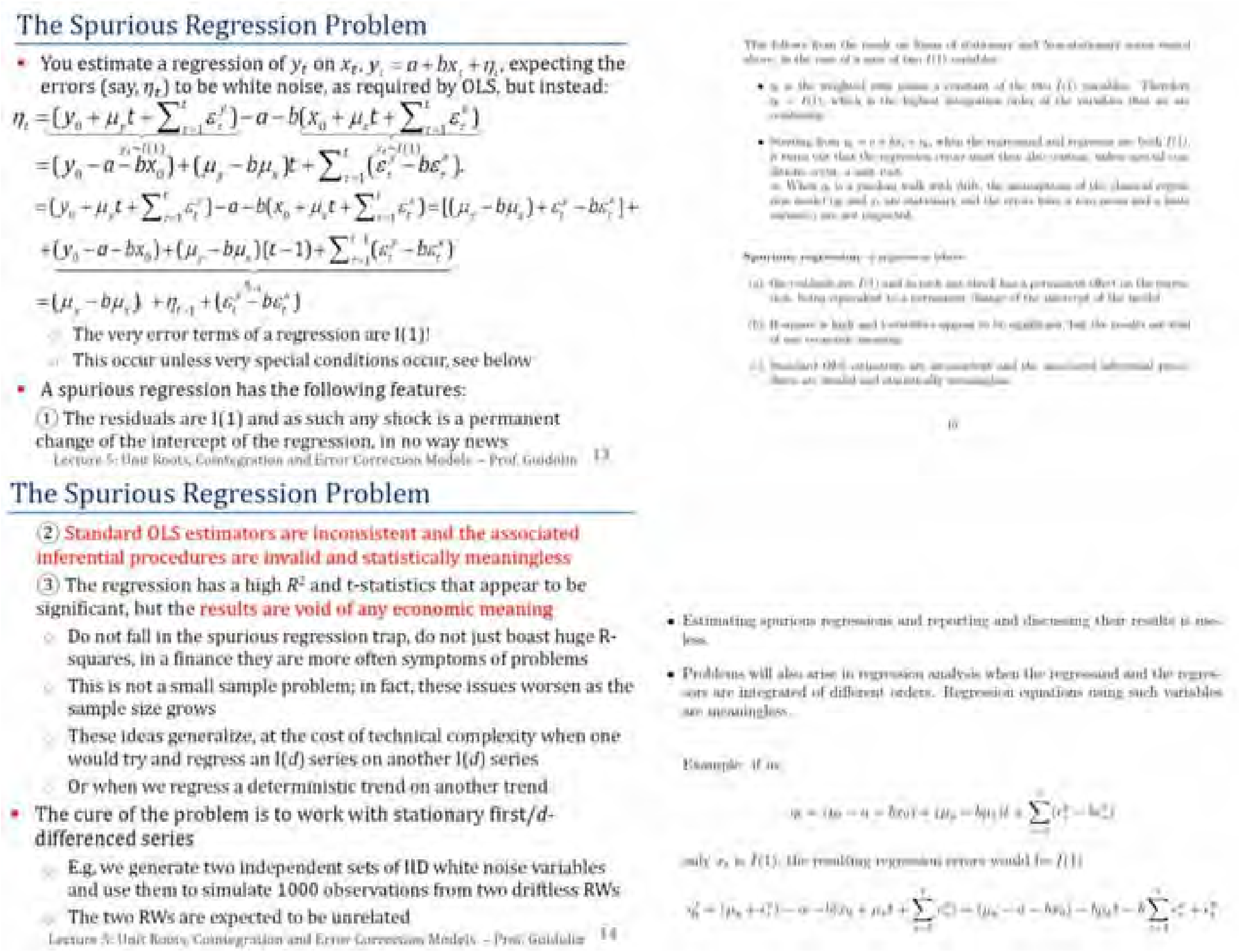
Question 3.A (9 points)

What is a spurious regression? Carefully define its causes and potential consequences for the validity of results from standard econometric procedures. How would you go about detecting the spurious nature of a regression? Suppose that one regresses a I(2) time series on a I(1) time series, would that cause a spurious regression problem? What are the remedies to avoid spurious regressions?

Debriefing.







Question 3.B (3.5 points)

Deron Cleanington is quant analyst that monitors Italian rates. He knows for a fact that 3-month T-bills (BoT) are I(1); however, he does not know much about 10-year note (BTp) rates. As a way to familiarize with the data, Deron regresses 10-year rates on 3-month rates, finding (p-values are in parenthesis):

$$r_t^{10Y} = \underset{(0.000)}{0.562} + \underset{(0.000)}{0.844} r_t^{3m} + \hat{\eta}_t,$$

Additional checks based on a Philipps-Perron unit root test reveal that the null hypothesis cannot be rejected both at a 5% and 1% test size. Deron concludes that *both* series contain a unit root but they are not cointegrated and that as such, being spurious, the estimated coefficients are invalid (biased and inconsistent). Do you agree with Deron's conclusions? Make sure to clearly justify your answer.

Debriefing.

Deron's may be correct but we have no evidence to back both his claims. In fact, the evidence provides is compatible with both:

_ $r_t^{10Y} \sim I(1)$, so that both r_t^{3m} and r_t^{10Y} contain a unit root but the regression (which is also a Engle-Granger's univariate cointegration test) indicates the absence of cointegration in the fact that the null of I(1) residuals cannot be rejected (here you needed to recall that a Phillips-Perron's test has a unit root null).

_ $r_t^{10Y} \sim I(0)$, so that the regression of r_t^{10Y} on r_t^{3m} is simply an "unbalanced" regression (another case of spurious regression) in which—by definition, because the sum of a I(0) and a I(1) series is I(1)—the residuals are I(1), as indeed established by the failure to reject by the PP test.

Yet, and in both cases, Deron is right when he claims that as result of the regression being either spurious or unbalanced, the estimated coefficients are invalid (biased and inconsistent).

Question 3.C (3.5 points)

Frank Tuvicci, a senior quant strategist at HappyHouse Hedge Fund, is having a heated discussion with a new junior colleague of his, John Marrone, about the nature of the time-series of US 1-month Treasury rates. Based on the evidence provided in Table 1, coming from a standard ADF test including both a constant and a trend, Frank has concluded that the series is I(0); however, John claims that relying on a KPSS test (for which results are reported in Table 2) he has failed to reject the null hypothesis and therefore that the series must be I(1). In general, is it possible that different tests may lead to different conclusions about the integration order of a series? In this specific case, based on the evidence displayed, do you think that both the claims of Frank and John were reasonable? Make sure to clearly justify your answer.

TABLE 1

Null Hypothesis: ONEMONTH has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 5 (Automatic - based on SIC, maxlag=23)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.975860	0.6135
Test critical values:	1% level	-3.964599	
	5% level	-3.413017	
	10% level	-3.128509	

*MacKinnon (1996) one-sided p-values.

TABLE 2

Null Hypothesis: ONEMONTH is stationary
Exogenous: Constant, Linear Trend
Bandwidth: 30 (Newey-West automatic) using Bartlett kernel

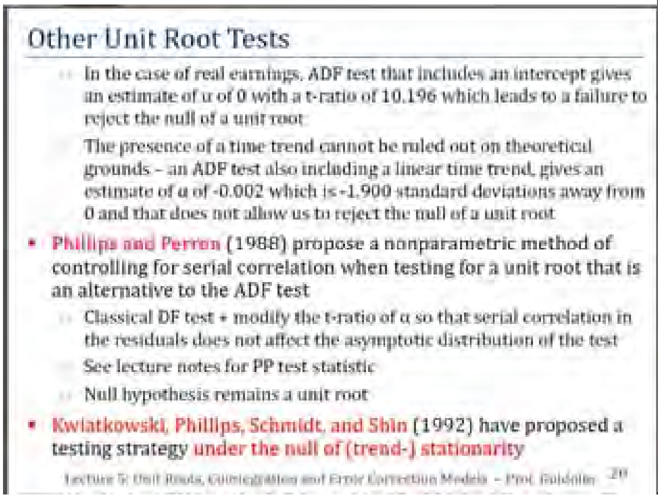
		LM-Stat.
Kwiatkowski-Phillips-Schmidt-Shin test statistic		0.138773
Asymptotic critical values*:	1% level	0.216000
	5% level	0.146000
	10% level	0.119000

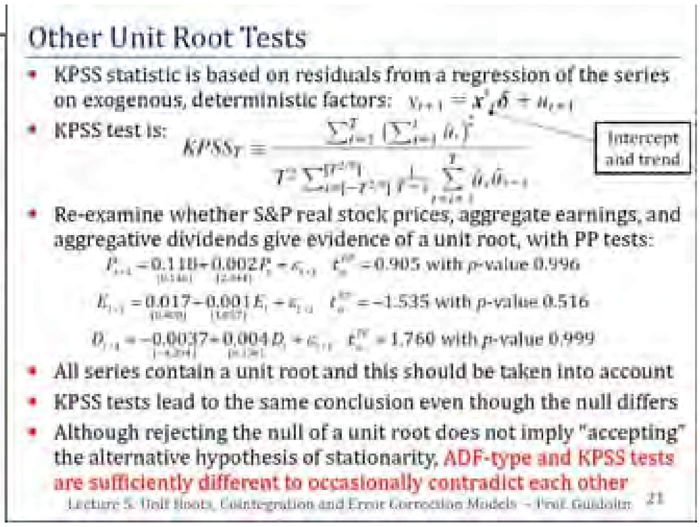
*Kwiatkowski-Phillips-Schmidt-Shin (1992, Table 1)

Residual variance (no correction)	2.292096
HAC corrected variance (Bartlett kernel)	65.75177

Debriefing.

Although in general ADF-type and KPSS tests are sufficiently different to occasionally contradict each other, this is not the case. John is clearly making a mistake in interpreting the results of the KPSS test that he conducted. Indeed, he is failing to recognize that the KPSS is conducted under the null of (trend) stationarity. Therefore, John is rejecting the null of stationarity, which is perfectly compatible with the conclusion by Frank, who is rejecting the null of the presence of a unit root in the series.

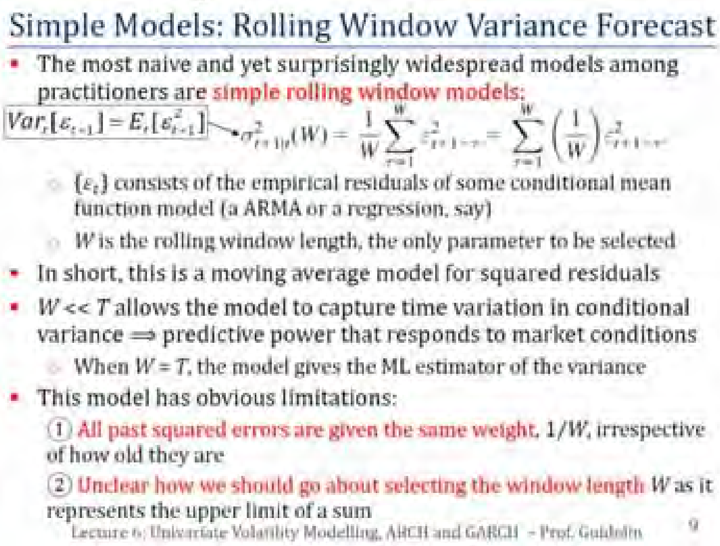


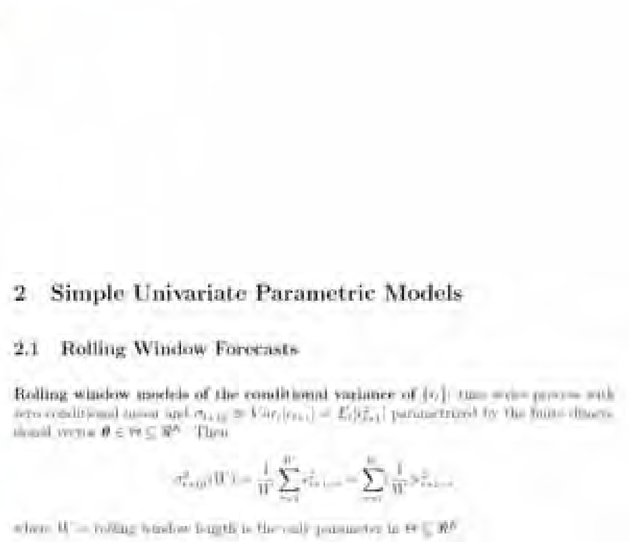


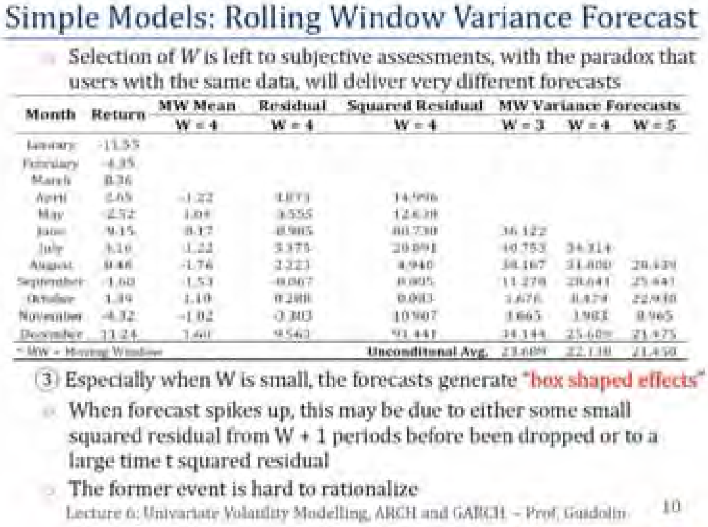
Question 4.A (10 points)

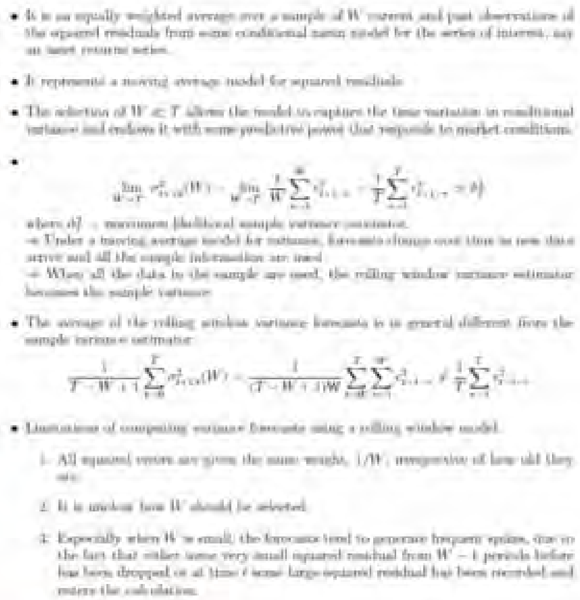
What is a rolling variance forecast model? What type of ARMA model does it represent and what random variables does it concern? Make sure to discuss the main pros and cons of this model.

Debriefing.









Question 4.B (3 points)

Mlado Vizov an analyst at Peeled & Head Ass. has just made a simple mathematical observation concerning a W-period rolling window variance estimator, $\sigma^2_{t+1|t}(W)$, namely that

$$\sigma^2_{t+1|t}(W) = \frac{1}{W}\sum_{\tau=1}^{W} \epsilon^2_{t+1-\tau} = \epsilon^2_t + \frac{1}{W-1}\sum_{\tau=1}^{W-1} \epsilon^2_{t-\tau} = \epsilon^2_t + \sigma^2_{t|t-1}(W-1).$$

Therefore, he claims that a rolling window variance estimator is just a special case of a RiskMetrics model, under the restriction that both the terms on the right-hand side are multiplied by a unit coefficient. Do you agree with his claim? Carefully explain your reasoning.

Debriefing.

As you know, a RiskMetrics model is simply written as

$$\sigma^2_{t+1|t} = (1-\lambda)\epsilon^2_t + \lambda\sigma^2_{t|t-1}.$$

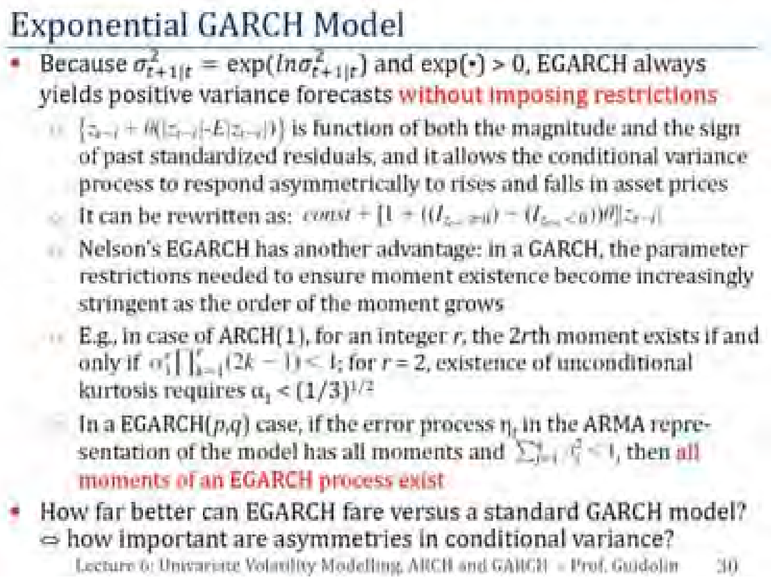
However, note that the variance process on the left- and right-hand sides of the RiskMetrics are the same: on the right we just have one lag of the process on the left. In the case pointed out by Mlado, we have instead that the process on the left, a W-observation rolling window variance estimator $\sigma^2_{t+1|t}(W)$ is structurally different from the process on the right, $\sigma^2_{t|t-1}(W-1)$, a (W -1)-observation rolling window variance estimator, which just uses less data. Therefore, we can say that there is no restriction on RiskMetrics that can take us to a rolling window variance process and as a result Mlado is wrong.

Question 4.C (4 points)

Mr. Manly Beverly is due to give a presentation on the process followed by the conditional variance of the log-price returns on the 3-month futures on Wheat, traded on the Chicago Mercantile Exchange. The audience is composed of homogeneous type of customers: risk managers. Manly has obtained by ML methods consistent and asymptotically efficient estimates for the parameters of an EGARCH(1,1) process such that $\hat{\alpha} + \hat{\beta} = 1.1$. Moreover, also the point estimate of the parameter θ is positive. Therefore, he decides to advise risk managers to purchase options to hedge long-run risks originated by the price of wheat as quickly as possible, as an explosive variance process may justify a progressive increase in the price of long-term options, which may make hedging progressively more expensive, for a given size of risks. Is Manly's advise a sensible one in the light of the parameter estimates that he has obtained?

Debriefing.

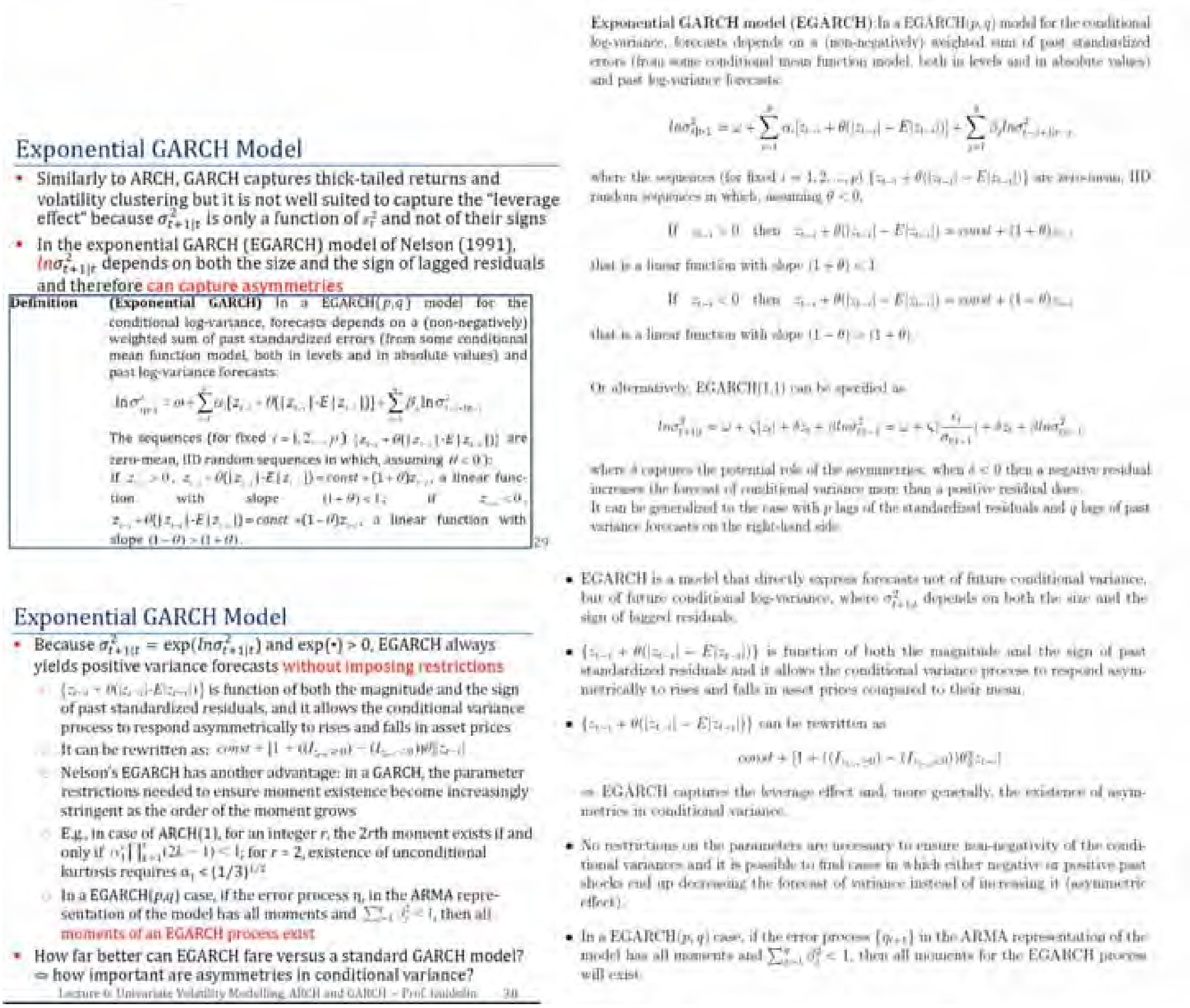
Manly's concern is legitimate in the sense that an explosive conditional variance process would imply a long-run, ergodic variance that diverges to infinity and therefore that option positions set up to hedge increasingly distant maturity positions on wheat will become very expensive (here recall that the price of both long puts and calls is monotone increasing in their variance). However, his concern suffers from a flaw: the condition for stationarity of a EGARCH(1,1) is NOT $\hat{\alpha} + \hat{\beta} < 1$! Such conditions are generally more complex than in the GARCH case and—based on what we have said in the lecture—in the (1,1) case, they appear to only require that $\hat{\beta} < 1$, which may happily be satisfied in this case (i.e., we do not have enough information). Therefore, Manly's conclusion may be premature and ill-advised: he should study his econometrics better, before making claims in public.



Question 5.A (10 points)

Define the statistical and economic nature of a EGARCH(p, q) model for conditional variance under a constraint of weak stationarity. How can you keep the forecasts of the conditional variance positive? What are in particular the reasons for the success of simple EGARCH(1,1) models in empirical applications? How would an EGARCH model capture the presence of (unconditional) skewness in typical time series of financial returns?

Debriefing.



Exponential GARCH Model

- Similarly to ARCH, GARCH captures thick-tailed returns and volatility clustering but it is not well suited to capture the "leverage effect" because $\sigma^2_{t+1|t}$ is only a function of ε_t^2 and not of their signs
- In the exponential GARCH (EGARCH) model of Nelson (1991), $ln\sigma^2_{t+1|t}$ depends on both the size and the sign of lagged residuals and therefore can capture asymmetries

Definition **(Exponential GARCH)** In a EGARCH(p,q) model for the conditional log-variance, forecasts depends on a (non-negatively) weighted sum of past standardized errors (from some conditional mean function model, both in levels and in absolute values) and past log-variance forecasts:

$$\ln\sigma^2_{t+1|t} = \omega + \sum_{i=1}^{p} \alpha_i [z_{t-i+1} + \theta(|z_{t-i+1}| - E|z_{t-i+1}|)] + \sum_{j=1}^{q} \beta_j \ln\sigma^2_{t-j+1|t-j}$$

The sequences (for fixed $i = 1, 2, \ldots, p$) $\{z_{t-i} + \theta(|z_{t-i}| - E|z_{t-i}|)\}$ are zero-mean, IID random sequences in which, assuming $\theta < 0$: if $z_{t-i} > 0$, $z_{t-i} + \theta(|z_{t-i}| - E|z_{t-i}|) = const + (1+\theta)z_{t-i}$, a linear function with slope $(1+\theta) < 1$; if $z_{t-i} < 0$, $z_{t-i} + \theta(|z_{t-i}| - E|z_{t-i}|) = const + (1-\theta)z_{t-i}$, a linear function with slope $(1-\theta) > (1+\theta)$.

29

Exponential GARCH model (EGARCH): In a EGARCH(p,q) model for the conditional log-variance, forecasts depends on a (non-negatively) weighted sum of past standardized errors (from some conditional mean function model, both in levels and in absolute values) and past log-variance forecasts:

$$ln\sigma^2_{t+1|t} = \omega + \sum_{i=1}^{p} \alpha_i [z_{t-i} + \theta(|z_{t-i}| - E|z_{t-i}|)] + \sum_{j=1}^{q} \beta_j ln\sigma^2_{t-j+1|t-j}$$

where the sequences (for fixed $i = 1, 2, \ldots, p$) $\{z_{t-i} + \theta(|z_{t-i}| - E|z_{t-i}|)\}$ are zero-mean, IID random sequences in which, assuming $\theta < 0$,

If $z_{t-i} > 0$ then $z_{t-i} + \theta(|z_{t-i}| - E|z_{t-i}|) = const + (1+\theta)z_{t-i}$

that is a linear function with slope $(1+\theta) < 1$

If $z_{t-i} < 0$ then $z_{t-i} + \theta(|z_{t-i}| - E|z_{t-i}|) = const + (1-\theta)z_{t-i}$

that is a linear function with slope $(1-\theta) > (1+\theta)$

Or alternatively, EGARCH(1,1) can be specified as

$$ln\sigma^2_{t+1|t} = \omega + \varsigma|z_t| + \delta z_t + \beta ln\sigma^2_{t|t-1} = \omega + \varsigma\left|\frac{\varepsilon_t}{\sigma_{t|t-1}}\right| + \delta z_t + \beta ln\sigma^2_{t|t-1}$$

where δ captures the potential role of the asymmetries: when $\delta < 0$ then a negative residual increases the forecast of conditional variance more than a positive residual does.
It can be generalized to the case with p lags of the standardized residuals and q lags of past variance forecasts on the right-hand side.

Exponential GARCH Model

- Because $\sigma^2_{t+1|t} = \exp(ln\sigma^2_{t+1|t})$ and $\exp(\cdot) > 0$, EGARCH always yields positive variance forecasts without imposing restrictions
  - $\{z_{t-i} + \theta(|z_{t-i}| - E|z_{t-i}|)\}$ is function of both the magnitude and the sign of past standardized residuals, and it allows the conditional variance process to respond asymmetrically to rises and falls in asset prices
  - It can be rewritten as: $const + [1 + (I_{z_{t-i} \geq 0} - I_{z_{t-i} < 0})\theta]z_{t-i}$
  - Nelson's EGARCH has another advantage: in a GARCH, the parameter restrictions needed to ensure moment existence become increasingly stringent as the order of the moment grows
  - E.g., in case of ARCH(1), for an integer r, the 2rth moment exists if and only if $\alpha_1^r \prod_{k=1}^{r}(2k-1) < 1$; for $r = 2$, existence of unconditional kurtosis requires $\alpha_1 < (1/3)^{1/2}$
  - In a EGARCH(p,q) case, if the error process η_t in the ARMA representation of the model has all moments and $\sum_{j=1}^{q} \beta_j^2 < 1$, then all moments of an EGARCH process exist
- How far better can EGARCH fare versus a standard GARCH model? ⇒ how important are asymmetries in conditional variance?

Lecture 6: Univariate Volatility Modelling, ARCH and GARCH – Prof. Guidolin 30

- EGARCH is a model that directly express forecasts not of future conditional variance, but of future conditional log-variance, where $\sigma^2_{t+1|t}$ depends on both the size and the sign of lagged residuals.
- $\{z_{t-i} + \theta(|z_{t-i}| - E|z_{t-i}|)\}$ is function of both the magnitude and the sign of past standardized residuals and it allows the conditional variance process to respond asymmetrically to rises and falls in asset prices compared to their mean.
- $\{z_{t-i} + \theta(|z_{t-i}| - E|z_{t-i}|)\}$ can be rewritten as

$$const + [1 + ((I_{z_{t-i} \geq 0}) - (I_{z_{t-i} < 0}))\theta]z_{t-i}$$

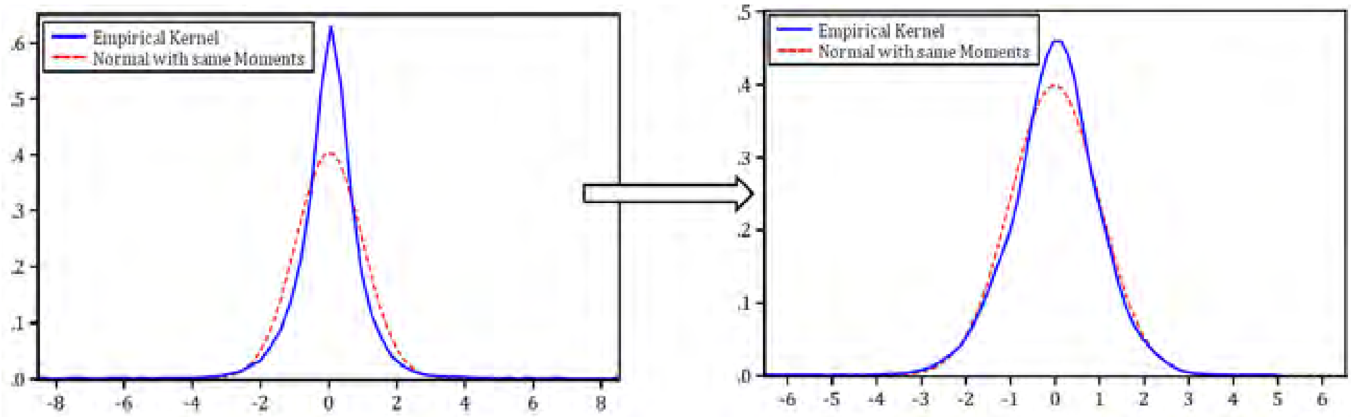
⇒ EGARCH captures the leverage effect and, more generally, the existence of asymmetries in conditional variance.

- No restrictions on the parameters are necessary to ensure non-negativity of the conditional variances and it is possible to find cases in which either negative or positive past shocks end up decreasing the forecast of variance instead of increasing it (asymmetric effect).
- In a EGARCH(p,q) case, if the error process $\{\eta_{t+1}\}$ in the ARMA representation of the model has all moments and $\sum_{j=1}^{q} \beta_j^2 < 1$, then all moments for the EGARCH process will exist.

As for the reference to skewness, as discussed in the lectures, when a conditional variance model implies asymmetric, leverage effects, this means that large negative returns imply an increase in conditional variance that exceeds the increase induced by large positive returns; therefore negative returns may induce even larger negative returns (because variance is high) and this will end up inflating the left tail of the unconditional distribution vs. the right tail, which will translate in the presence of unconditional skewness.

Question 5.B (3 points)

Ms. Martina Kalvin is analyzing the series of US daily excess stock returns for a long 1963-2016 sample. On the left, you can see the kernel density estimator of the standardized residuals from a homoskedastic ARMA(2,1) model with Gaussian shocks; as a benchmark, the kernel density estimator is compared to a N(0,1). On the right, you can see the kernel density estimator of the standardized residuals from a ARMA(2,1)/ EGARCH(1,1) model with Gaussian shocks; as a benchmark, the kernel density estimator is compared to a N(0,1). Martina does *not* specify whether she has estimated her model by MLE or QMLE.



Ms. Kelvin claims that moving from left to the right, the validity of the model records a considerable improvement: do you agree and why? However, Martina reckons that the ARMA(2,1)/ EGARCH(1,1) model with Gaussian shocks should be rejected: why would she claim that? Assuming you agree with her, what do you think may cause the difficulties that characterized the Gaussian ARMA(2,1)/EGARCH(1,1) model? Carefully motivate your replies.

Debriefing.

Probably it is trivial, but when moving from the left to the right panel, the improvements in the quality of the fit provided by the model to the data are evident: on the left, the blue empirical kernel density strongly departs from the Gaussian N(0,1) benchmark under which the conditional mean model was estimated; on the right, the empirical kernel density of the standardized residuals approaches, even though it remains visibly different to the N(0,1) benchmark under which the ARMA(2,1)/EGARCH(1,1) model has been estimated. However, whether or not the model can be rejected depends entirely on whether the model had been estimated by either MLE or QMLE:

__ if the model were estimated by MLE, then a rejection would be justified, because also in the right panel there is a significant deviation of the empirical kernel from the assumed N(0,1) distribution;

__ if the model were estimated by QMLE, then a rejection would *not* be justified, because the significant deviation of the empirical kernel from the assumed N(0,1) distribution in the right panel is not only admissible, but even expected, given that the Gaussian distribution for the shocks has been just assumed as an approximation.

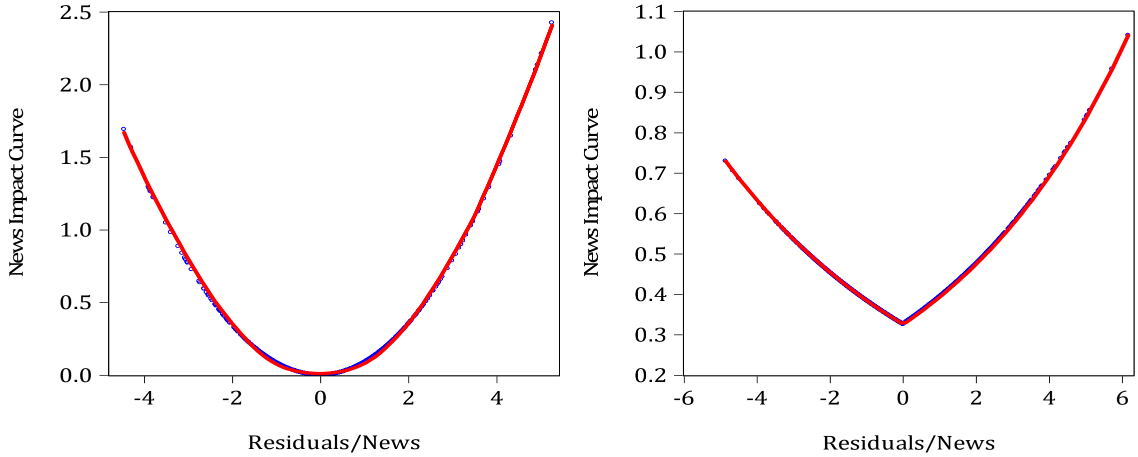
Question 5.C (4 points)

Eventually, Martina has estimated a ARMA(2,1)/EGARCH(1,1) model with GED errors that turns out to be as follows (standard errors are in parentheses):

$$R_{t+1} = \underset{(0.043)}{0.074} + \underset{(0.017)}{0.042}R_t + \underset{(0.019)}{0.053}R_t - \underset{(0.013)}{0.033}\epsilon_t + \underset{(0.008)}{0.022}\epsilon_t + \epsilon_{t+1}$$

$$ln\sigma^2_{t+1|t} = -\underset{(0.008)}{0.157} + \underset{(0.006)}{0.177}\left|\frac{\epsilon_t}{\sigma^2_{t|t-1}}\right| + \underset{(0.004)}{0.011}\frac{\epsilon_t}{\sigma^2_{t|t-1}}$$
$$+ \underset{(0.120)}{0.990}ln\sigma^2_{t|t-1} \qquad \epsilon_t\ IID\ GED(0, \sigma^2_{t+1|t}; \underset{(0.352)}{1.530})$$

However, Martina has forgotten to label the following two pictures concerning the news impact curve (NIC) derived from the estimated ARMA(2,1)/EGARCH(1,1) model and she no longer remembers whether it is the left or the right plots that represents the NIC of the model above.



Can you help her finding the appropriate NIC that refers to the model above? Carefully explain what has guided your selection/answer and why the remaining plot is not plausibly derived from a ARMA(2,1)/EGARCH(1,1) model. Martina is also confused as to whether her estimates are either ML or QML. Can you help her? Make sure to justify your answer.

Debriefing.

The ARMA(2,1)/EGARCH(1,1) NIC is the rightmost one. We can detect that from the existence of a kink induced by the appearance of the absolute value of the standardized error on the right-hand side of the conditional variance model. As you will recall from your basic math courses, the kink point occurs in correspondence to the change of sign of the absolute value, when the function fails to be differentiable. The leftmost plot is everywhere differentiable and as such it cannot represent the NIC from a EGARCH model. Finally, because the model has been estimated assuming GED and not Gaussian errors, clearly the estimation methods must be full-information maximum likelihood (MLE) and cannot be QML, that would incorrectly assume a pseudo normal density for the errors.

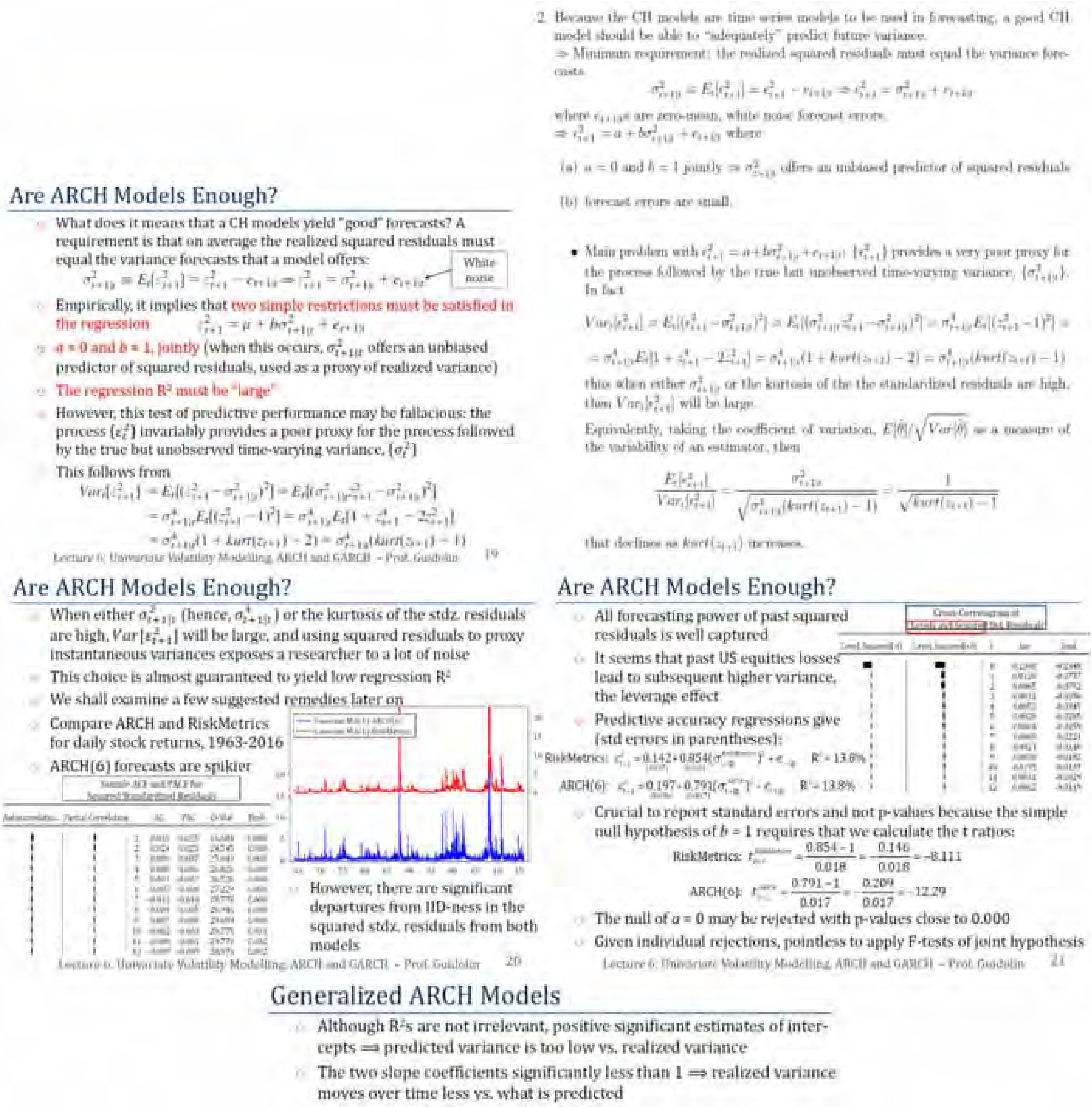
Question 6.A (9 points)

Describe the theoretical justifications as well as the practical implementations of tests of the forecasting validity of a conditional heteroscedasticity model based on the linear regression

$$\epsilon_{t+1}^2 = a + b\hat{\sigma}_{t+1|t}^2 + e_{t+1},$$

where e_{t+1} is a white noise shock and $\hat{\sigma}_{t+1|t}^2$ are the one-step ahead conditional variance forecasts derived from a given model. How would you estimate this linear model? Under what circumstances the null that the model yields unbiased and efficient forecasts will be rejected? Discuss whether you would also use the regression R-square to assess the validity of the variance model. Make sure to clearly justify your answers.

Debriefing.



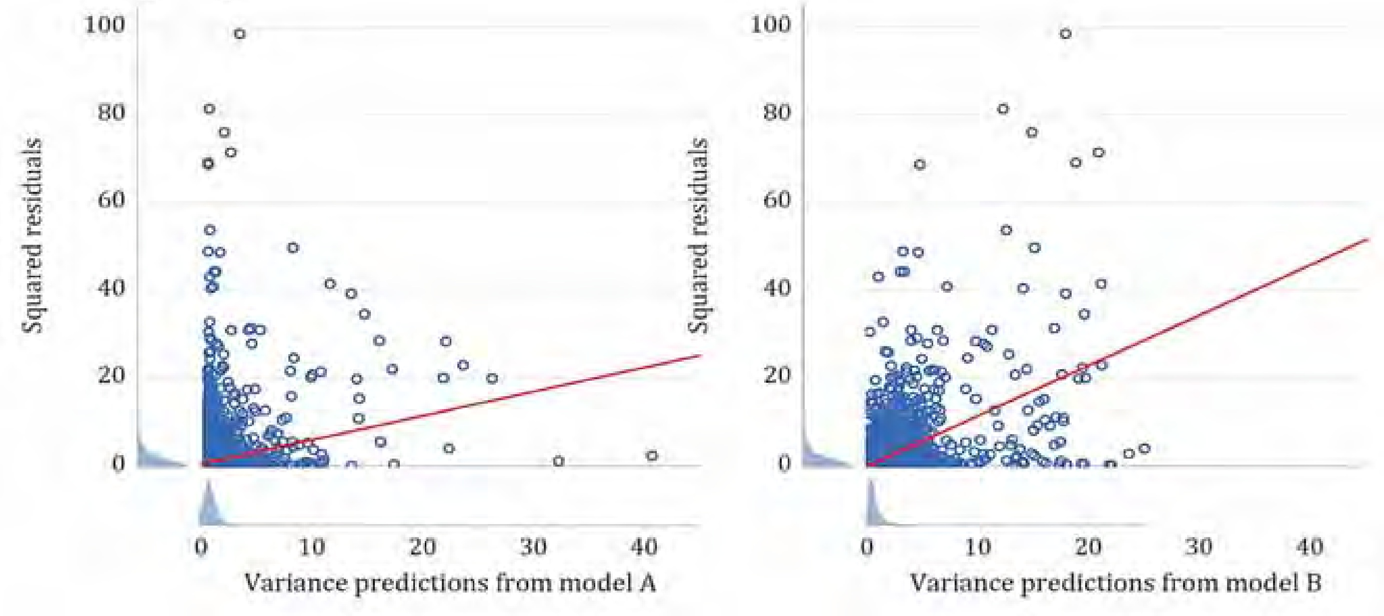
Question 6.B (4.5 points)

Mikki Paranoich, an independent researcher, has estimated two models to predict the one-day ahead variance of US aggregate excess stock returns. For two models, call them A and B, Mikki has obtained the following results from a regression of squared residuals (from a MA(1) model that has been pre-specified using Box-Jenkins analysis) on variance predictions, $\hat{\sigma}^2_{A,t+1|t}$ and $\hat{\sigma}^2_{B,t+1|t}$ (estimated standard errors are reported in parantheses):

$$\epsilon^2_{t+1} = \underset{(0.281)}{0.434} + \underset{(0.377)}{0.549}\hat{\sigma}^2_{A,t+1|t} + \hat{e}^A_{t+1} \qquad R^2 = 0.029,$$

$$\epsilon^2_{t+1} = -\underset{(0.038)}{0.106} + \underset{(0.023)}{1.146}\hat{\sigma}^2_{B,t+1|t} + \hat{e}^B_{t+1} \qquad R^2 = 0.159.$$

Moreover, in the case of model A, a test of the joint null of a = 0 and b = 1 using an F-test leads to a rejection. Scatter plots of the squared residuals vs. variance predictions with a regression line superimposed are as follows:



Which of the two models, if any, can be considered to be a valid prediction tool? Make sure to clearly justify your answers.

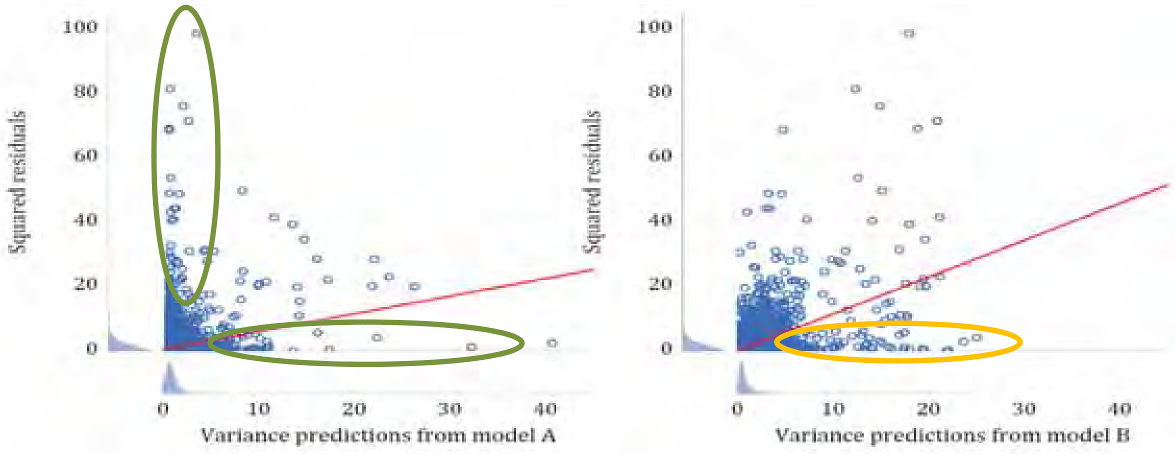
Debriefing.

In the case of model A, we have:

$$t^A_{a=0} = \frac{0.434}{0.281} = 1.545 < 2 \Rightarrow fail\ to\ reject\ null\ of\ a = 0$$

$$t^A_{b=1} = \frac{0.549 - 1}{0.377} = -1.196 \Rightarrow |1.196| < 2 \Rightarrow fail\ to\ reject\ null\ of\ b = 1$$

Therefore, in this case the model fails to be rejected. Yet, despite the parametric structure of the model is not rejected, the R-square in this case is largely disappointing. The left panel of the picture shows an interesting phenomenon: in a non-negligible fraction of the sample, recorded variance is large and exceeds 20 (careful, this is not a percentage!) but the model predicts a variance of almost zero, which is a reason for concern; in a few cases, we also record the opposite pattern: the recorded squared error is small and a below 1-2, but model A returns predictions that exceed 5 or even 10, see the green circles in the copy of the figures below. These regularities contribute to a rather small regression R-square.



In the case of model B, we have:

$$t^{B}_{a=0} = \frac{-0.106}{0.038} - 2.762 \Rightarrow |2.762| > 2 \Rightarrow reject\ null\ of\ a = 0$$
$$t^{B}_{b=1} = \frac{1.146 - 1}{0.027} = 6.448 \gg 2 \Rightarrow reject\ null\ of\ b = 1$$

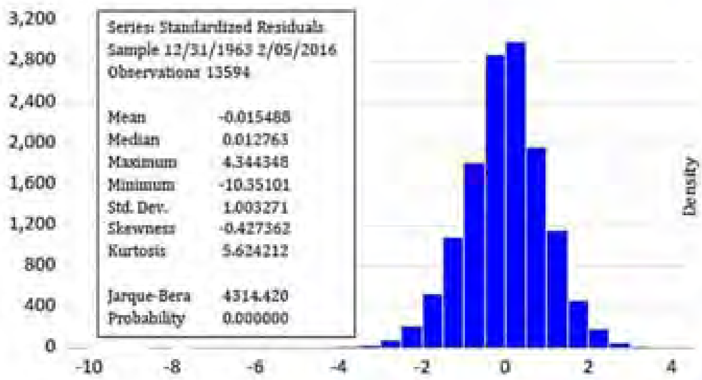
Clearly, in this case to test the joint null of a = 0 and b = 1 using an F-test will lead to a rejection. However, the R-square of this regression is not as low and disappointing as the one we have gotten for model A, and corresponds to almost the maximum one may hope to get with this type of data. This is qualitatively confirmed by the rightmost plot of the figure, in which high squared residuals are always matched by non-zero a substantial variance predictions: when variance will be high, the model will forecast that. However, remains visible and actually gets even stronger (see orange circle) the second type of bias: in a considerable fraction of the sample, the recorded squared error is small and a below 1-2, but model B returns predictions that exceed 5 or even 10, which implies that a fraction of the time, mode B predicts a high variance that fails to materialize in the data.

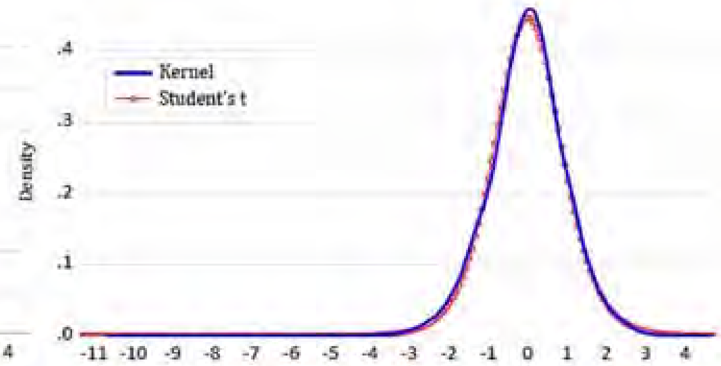
Question 6.C (2.5 points)

In the case of model B, Mikky proceeds then to look for ways to improve the model and its predictive performance. He obtains the following evidence:

Correlogram of Standardized Residuals Squared

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.012	0.012	2.0284	0.154
		2	0.013	0.013	4.4512	0.108
		3	0.003	0.003	4.5777	0.205
		4	0.006	0.006	5.0391	0.283
		5	0.005	0.005	5.3715	0.372
		6	-0.009	-0.009	6.4624	0.373
		7	-0.005	-0.005	6.7768	0.452
		8	0.007	0.007	7.3683	0.497
		9	0.008	0.008	8.1469	0.519
		10	0.003	0.003	8.2723	0.602
		11	0.001	0.001	8.3001	0.686
		12	-0.009	-0.009	9.3877	0.670
		13	-0.003	-0.003	9.5413	0.731
		14	0.006	0.006	9.9991	0.762
		15	-0.007	-0.007	10.644	0.777
		16	-0.006	-0.005	11.066	0.805
		17	-0.006	-0.005	11.488	0.830
		18	-0.006	-0.006	11.927	0.851
		19	0.003	0.003	12.058	0.883
		20	-0.012	-0.012	14.030	0.829





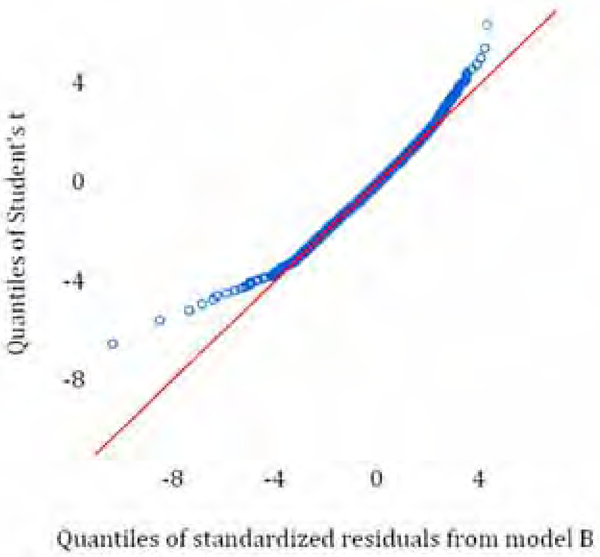
Heteroskedasticity Test: ARCH

F-statistic	0.708403	Prob. F(15,13563)	0.7786
Obs*R-squared	10.63025	Prob. Chi-Square(15)	0.7783

Test Equation:
Dependent Variable: WGT_RESID^2
Method: Least Squares
Date: 06/10/18 Time: 12:28
Sample (adjusted): 1/21/1964 2/05/2016
Included observations: 13579 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.977614	0.037462	26.09627	0.0000
WGT_RESID^2(-1)	0.011953	0.008586	1.392130	0.1639
WGT_RESID^2(-2)	0.013269	0.008587	1.545238	0.1223
WGT_RESID^2(-3)	0.002626	0.008588	0.305755	0.7598
WGT_RESID^2(-4)	0.005673	0.008588	0.660620	0.5089
WGT_RESID^2(-5)	0.004777	0.008588	0.556286	0.5780
WGT_RESID^2(-6)	-0.009433	0.008588	-1.098376	0.2721
WGT_RESID^2(-7)	-0.004899	0.008588	-0.570402	0.5684
WGT_RESID^2(-8)	0.006865	0.008588	0.799391	0.4241
WGT_RESID^2(-9)	0.007395	0.008588	0.861054	0.3892
WGT_RESID^2(-10)	0.002875	0.008588	0.334799	0.7378
WGT_RESID^2(-11)	0.001457	0.008588	0.169673	0.8653
WGT_RESID^2(-12)	-0.009217	0.008588	-1.073269	0.2832
WGT_RESID^2(-13)	-0.003398	0.008588	-0.395663	0.6924
WGT_RESID^2(-14)	0.006185	0.008587	0.720219	0.4714
WGT_RESID^2(-15)	-0.006734	0.008587	-0.784227	0.4329

R-squared	0.000783	Mean dependent var	1.007221
Adjusted R-squared	-0.000322	S.D. dependent var	2.171752



Keep in mind that model B has been estimated assuming that the standardized shocks are drawn from a t-Student distribution, which justifies the selection of benchmark in the kernel density plot (third plot going clockwise). What is your advice to Mikki as to ways to improve the predictive power of model B? Make sure to clearly justify your answer.

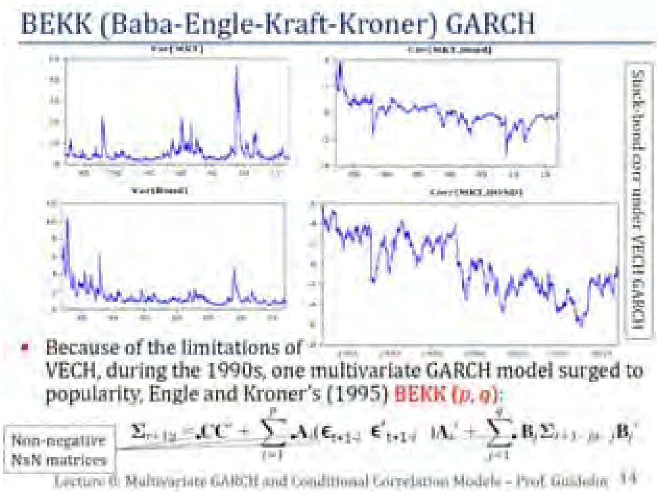
Debriefing.

In fact, it all looks rather good apart from one piece of evidence: the kernel density comparison and especially the quantile-quantile plot reveal that the t-student inflates the tails of the predicted density *excessively* given the tail thickness expressed by the data (also because the estimated number of degrees of freedom, less than 8, appears to be really small). On the contrary there is no evidence that any residual ARCH structure is left in the data or of asymmetries that are not captured (see the kernel plot), even though further tests for asymmetries using the LM principles or news impact curves might be explored. Finally, note that one piece of evidence is rather redundant and unhelpful—the histogram provides the background to test for normality, but there is no presumption here that the data may come from a Gaussian distribution.

OPTIONAL Question 7 (4 points)

Describe the structure of a BEKK (Baba-Engle-Kraft-Kroner) multivariate GARCH (p, q) model. Make sure to illustrate its key advantages and disadvantages. How many parameters would you need to estimate in the BEKK(1,1) case?

Debriefing.



2. Estimate constant correlations using a simple sample estimator based on the standardized returns, $\hat{z}_{t+1} \equiv R_{t+1}/\hat{\sigma}_{t+1}$, derived from the first step using GARCH-type models. Here, we exploit the fact that the conditional covariance of the $\hat{z}_{t+1}$ variables equals the conditional correlation of raw returns:

$$\hat{\rho}_{ij} = \frac{1}{T}\sum_{t=1}^{T} \hat{z}_{i,t}\hat{z}_{j,t}$$

Such constant correlations are then inserted inside Γ to estimate the constant correlation matrix.

4.3 BEKK GARCH

Given the picture provided above and the fact that DCC is a model popularized around the turn of the millenium, one may ask what was the state of multivariate GARCH modelling in practice before DCC became as popular as it is today. Apart from the uncomfortable case of CCC models that assume constant correlations over time, during the 1990s one of the most popular multivariate GARCH models had been Engle and Kroner's (1995) BEKK GARCH(p, q):[23]

$$\Sigma_{t+1} = \mathbf{C}\mathbf{C}' + \sum_{i=1}^{p} \mathbf{A}_i(\mathbf{R}_{t+1-i}\mathbf{R}'_{t+1-i})\mathbf{A}'_i + \sum_{j=1}^{q} \mathbf{B}_j\Sigma_{t+1-j}\mathbf{B}'_j$$

where the matrices $\{\mathbf{A}_i\}_{i=1}^{p}$ and $\{\mathbf{B}_j\}_{j=1}^{q}$ are non-negative and symmetric. This special product-sandwich form that is used to write the BEKK ensures the PSD property without imposing further restrictions, which represent the key reason for the success of BEKK models. In fact, this full matrix BEKK is easier to estimate than vech-GARCH models, even though it remains rather complex to handle. In practice, the popular form of BEKK that many empirical analysts have come to appreciate is a simpler (1,1) diagonal BEKK that restricts the matrices $\mathbf{A}$ and $\mathbf{B}$ to be diagonal matrices. BEKK models possess three attractive properties:

1. A BEKK is a truncated, low-dimensional application of a theorem by which all non-negative, symmetric $N \times N$ matrices (say, $\mathbf{M}$) can be decomposed (for instance) as

$$\mathbf{M} = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} = \sum_{i=1}^{\infty} \begin{bmatrix} m'_{i,1}m_{i,1} & m'_{i,1}m_{i,2} \\ m'_{i,2}m_{i,1} & m'_{i,2}m_{i,2} \end{bmatrix}$$

for appropriately selected vectors $\mathbf{m}_{i,j}$. In a sense, mathematically it is no surprise that BEKK models often offer a good fit to the dynamics of variance.

2. As already mentioned, it easily ensures PSDness of the covariance matrix.

3. BEKK is invariant to linear combinations: e.g., if $\mathbf{R}_{t+1}$ follows a BEKK GARCH(p, q), then any portfolio formed from the N securities or assets in $\mathbf{R}_{t+1}$ will also follow a BEKK.

[23] In case you wonder, BEKK means "Baba-Engle-Kraft-Kroner" and the acronym simply compacts the name of the four econometricians who contributed to its development.

31

BEKK (Baba-Engle-Kraft-Kroner) GARCH

- The special "sandwich" structure of the coefficient matrices guarantees that $\mathbf{\Sigma}_{t+1|t}$ is (semi-)PD without imposing other restrictions
  - The popular BEKK that many empiricists have come to appreciate is a simpler (1,1) diagonal BEKK that restricts the matrices **A** and **B**
- BEKK models possess three attractive properties:
  - ① When symmetry of **A** and **B** is imposed, a BEKK is a truncated, low-dimensional application of a theorem by which all nonnegative, symmetric *NxN* matrices (say, **M**) can be decomposed as:

$$\mathbf{M} = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} = \sum_{i=1}^{2N} \begin{bmatrix} \mathbf{m}_{i,1}\mathbf{m}_{i,1} & \mathbf{m}_{i,1}\mathbf{m}_{i,2} \\ \mathbf{m}_{i,2}\mathbf{m}_{i,1} & \mathbf{m}_{i,2}\mathbf{m}_{i,2} \end{bmatrix}$$

  for appropriately selected vectors $\mathbf{m}_{i}$
  - ② BEKK ensures (S)PD-ness of $\mathbf{\Sigma}_{t+1|t}$, because by construction, the sandwich form and outer vector products have this property
  - ③ BEKK is invariant to linear combinations, i.e., if $\mathbf{R}_{t+1}$ follows a BEKK GARCH(p, q), then any ptf. of the *N* assets in $\mathbf{R}_{t+1}$ will also follow a BEKK, see lecture notes for examples and counterexamples under VECH ARCH
  - However, the number of parameters in BEKK remains rather large

However, the number of parameters in BEKK remains rather large:

$$0.5N(N+1) + 0.5pN(N+1) + 0.5qN(N+1) = 0.5N(N+1)[1+p+q] = O(N^2)$$

Often, this has still made DCC models preferable in practice. However, the number of parameters in BEKK is substantially inferior to those appearing in a full VEC specification. This happens because the parameters governing the dynamics of the covariance equation in BEKK models are the products of the corresponding parameters of the two corresponding variance equations in the same model.

The second and third properties of BEKK models can only be appreciated contrasting the features of BEKK under linear aggregation with the properties of alternative multivariate GARCH models, for instance even a simple diagonal vech ARCH. Not all multivariate GARCH models are invariant with respect to linear transformations.[23] For instance, for the case of two asset return series ($N=2$), consider a simple diagonal multivariate ARCH(1) model obtained from a simplification of the diagonal GARCH(p, q) introduced early on:

$$vech(\mathbf{\Omega}_t) = (\mathbf{I}_3 - \mathbf{A})vech\left(T^{-1}\sum_{t=1}^{T}\mathbf{R}_t\mathbf{R}_t'\right) + \mathbf{A}vech\left(\mathbf{R}_{t-1}\mathbf{R}_{t-1}'\right) \quad (17)$$

where the helpful variance targeting restriction has already been imposed and **A** is a diagonal matrix. Because we have set $N=2$, $\mathbf{\Omega}_t$ will be a 2×2 matrix, **A** is a 3×3 diagonal matrix, $\mathbf{R}_t$ is 2×1 vector of asset returns, $vech(\mathbf{\Omega}_t)$ is a 3×1 vector of unique elements from $\mathbf{\Omega}_t$, $vech(T^{-1}\sum_{t=1}^{T}\mathbf{R}_t\mathbf{R}_t')$ is a 3×1 vector of unique elements from the sum of cross-product matrices $T^{-1}\sum_{t=1}^{T}\mathbf{R}_t\mathbf{R}_t'$, $vech(\mathbf{R}_{t-1}\mathbf{R}_{t-1}')$ is a 3×1 vector of unique elements from the lagged cross-product matrix $\mathbf{R}_{t-1}\mathbf{R}_{t-1}'$. The number of coefficients to be estimated is of course 3: a^{11}, a^{12}, and a^{22} in the representation

$$\begin{bmatrix} \sigma_{11,t} \\ \sigma_{12,t} \\ \sigma_{22,t} \end{bmatrix} = \left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} a^{11} & 0 & 0 \\ 0 & a^{12} & 0 \\ 0 & 0 & a^{22} \end{bmatrix}\right)\begin{bmatrix} T^{-1}\sum_{t=1}^{T}R_{1t}^2 \\ T^{-1}\sum_{t=1}^{T}R_{1t}R_{2t} \\ T^{-1}\sum_{t=1}^{T}R_{2t}^2 \end{bmatrix}$$

$$+\begin{bmatrix} a^{11} & 0 & 0 \\ 0 & a^{12} & 0 \\ 0 & 0 & a^{22} \end{bmatrix}\begin{bmatrix} R_{1t-1}^2 \\ R_{1t-1}R_{2t-1} \\ R_{2t-1}^2 \end{bmatrix} = \begin{bmatrix} (1-a^{11})T^{-1}\sum_{t=1}^{T}R_{1t}^2 + a^{11}R_{1t-1}^2 \\ (1-a^{12})T^{-1}\sum_{t=1}^{T}R_{1t}R_{2t} + a^{12}R_{1t-1}R_{2t-1} \\ (1-a^{22})T^{-1}\sum_{t=1}^{T}R_{2t}^2 + a^{22}R_{2t-1}^2 \end{bmatrix}$$

As for the conditions that guarantee that $\sigma_{11,t} > 0$ and $\sigma_{22,t} > 0$ at all times, i.e., that ensure PSDness of the model, clearly

$$(1-a^{11})T^{-1}\sum_{t=1}^{T}R_{1t}^2 + a^{11}R_{1t-1}^2 > 0 \text{ if and only if } a^{11} \in (0,1)$$

$$(1-a^{22})T^{-1}\sum_{t=1}^{T}R_{2t}^2 + a^{22}R_{2t-1}^2 > 0 \text{ if and only if } a^{22} \in (0,1)$$

[23] By invariance of a model, we mean that it stays in the same class if a linear transformation is applied to $\tilde{\mathbf{R}}_{t+1} = \mathbf{F}\mathbf{R}_{t+1}$, where **F** is a square matrix of constants and $\tilde{\mathbf{R}}_{t+1}$ corresponds to new assets (portfolios combining the original assets). It seems sensible that a model should be invariant, otherwise the question arises which basic assets should be modeled.

At this point the filtered (predicted) correlation coefficient has expression

$$\rho_{12,t} = \frac{c^{12} + a^{12}R_{1t-1}R_{2t-1}}{\sqrt{c^{11} + a^{11}R_{1t-1}^2}\sqrt{c^{22} + a^{22}R_{2t-1}^2}}$$

and, as it is obvious, $\rho_{12,t}$ should belong to $[-1,1]$ $\forall t \geq 1$. Here we have shortened the notation setting $c^{11} \equiv (1-a^{11})T^{-1}\sum_{t=1}^{T}R_{1t}^2$, $c^{22} \equiv (1-a^{22})T^{-1}\sum_{t=1}^{T}R_{2t}^2$, and $c^{12} \equiv (1-a^{12})T^{-1}\sum_{t=1}^{T}R_{1t}R_{2t}$. Focussing on the upper bound of the interval this means that

$$(c^{12} + a^{12}R_{1t-1}R_{2t-1})^2 \leq (c^{11} + a^{11}R_{1t-1}^2)(c^{22} + a^{22}R_{2t-1}^2)$$

or

$$(c^{12})^2 + (a^{12})^2R_{1t-1}^2R_{2t-1}^2 + 2c^{12}a^{12}R_{1t-1}R_{2t-1} \leq c^{11}c^{22} + c^{22}a^{11}R_{1t-1}^2 + c^{11}a^{22}R_{2t-1}^2 + a^{11}a^{22}R_{1t-1}^2R_{2t-1}^2$$

which is equivalent to

$$[a^{11}a^{22} - (a^{12})^2]R_{1t-1}^2R_{2t-1}^2 + [c^{11}c^{22} - (c^{12})^2] + c^{22}a^{11}R_{1t-1}^2 + c^{11}a^{22}R_{2t-1}^2 - 2c^{12}a^{12}R_{1t-1}R_{2t-1} \geq 0$$

which cannot hold for a continuous distribution for the asset return series as: even constraining $[a^{11}a^{22} - (a^{12})^2] \geq 0$ and $[c^{11}c^{22} - (c^{12})^2] \geq 0$,[24]

$$c^{22}a^{11}R_{1t-1}^2 + c^{11}a^{22}R_{2t-1}^2 - 2c^{12}a^{12}R_{1t-1}R_{2t-1} \geq 0$$

in general does not hold for $a^{12} \neq 0$. However, notice that if one sets $a^{12} = 0$, then the previous inequality simplifies to

$$a^{11}a^{22}R_{1t-1}^2R_{2t-1}^2 + \left\{\left[(1-a^{11})T^{-1}\sum_{t=1}^{T}R_{1t}^2\right]\left[(1-a^{22})T^{-1}\sum_{t=1}^{T}R_{2t}^2\right] - \left[T^{-1}\sum_{t=1}^{T}R_{1t}R_{2t}\right]^2\right\} +$$
$$+ c^{22}a^{11}R_{1t-1}^2 + c^{11}a^{22}R_{2t-1}^2 \geq 0$$

which has a chance to hold if a^{11} and a^{22} are such that

$$\left[(1-a^{11})T^{-1}\sum_{t=1}^{T}R_{1t}^2\right]\left[(1-a^{22})T^{-1}\sum_{t=1}^{T}R_{2t}^2\right] \geq \left[T^{-1}\sum_{t=1}^{T}R_{1t}R_{2t}\right]^2,$$

which also means that

$$\bar{\rho}_{12} = \frac{\bar{\sigma}_{12}}{\bar{\sigma}_{11}\bar{\sigma}_{22}} = \frac{T^{-1}\sum_{t=1}^{T}R_{1t}R_{2t}}{\sqrt{(1-a^{11})T^{-1}\sum_{t=1}^{T}R_{1t}^2}\sqrt{(1-a^{22})T^{-1}\sum_{t=1}^{T}R_{2t}^2}} \leq 1,$$

the unconditional correlation implied by the data and the diagonal bivariate ARCH(1) process is well-behaved. Therefore, if $a^{11} \in (0,1)$ and $a^{22} \in (0,1)$, then $a^{12} = 0$ (and some other restriction on a^{11} and a^{22}) must be imposed. This means that it is impossible to model the dynamics of volatilities

[24] "At all times" here really means "for all possible realizations of the continuous bivariate vector $\mathbf{R}_t$, which as domain $(-1, +\infty) \times (-1, +\infty)^2$", which alludes to the fact that even under limited responsibility, in finance asset returns may in principle take very large values.

and covariances simultaneously while satisfying the positivity requirement for the volatilities and keeping $\mathbf{\Omega}_t$ semi-positive definite at all times. Equivalently, if one wants to impose that the diagonal vech ARCH(1) model delivers a filtered covariance matrix $\mathbf{\Omega}_t$ that is semi-positive definite at all times, the diagonal model itself must be turned into a constant covariance multivariate ARCH model, as you understand that $a^{12} = 0$ implies $\sigma_{12,t} = T^{-1}\sum_{t=1}^{T}R_{1t}R_{2t} = \bar{\sigma}_{12}$ so that

$$\rho_{12,t} = \frac{\bar{\sigma}_{12}}{\sqrt{c^{11} + a^{11}R_{1t-1}^2}\sqrt{c^{22} + a^{22}R_{2t-1}^2}}$$

and dynamics in conditional correlations will exclusively come from dynamics in volatilities.[25]

Let's now examine the issue concerning the fact that while BEKK is "closed" under linear aggregation, a simpler diagonal vech-GARCH model is not. Consider a portfolio of the two assets with weights w and $(1-w)$. We show that in spite of the fact that $\mathbf{R}_{t-1}$ is characterized by a diagonal bivariate ARCH(1), the portfolio returns $R_t^p = wR_{1t} + (1-w)R_{2t}$ has a variance process $\sigma_{pp,t} \equiv Var_{t-1}[R_t^p]$ that fails to display the typical "diagonal form", i.e., $(1-a^{pp})T^{-1}\sum_{t=1}^{T}R_t^p + a^{pp}(R_{t-1}^p)^2$. Note first that

$$\begin{aligned} \sigma_{pp,t} &\equiv Var_{t-1}[R_t^p] = Var_{t-1}[wR_{1t} + (1-w)R_{2t}] \\ &= w^2\sigma_{11,t} + (1-w)^2\sigma_{22,t} + 2w(1-w)\sigma_{12,t} \end{aligned}$$

$$= w^2(1-a^{11})T^{-1}\sum_{t=1}^{T}R_{1t}^2 + w^2a^{11}R_{1t-1}^2 + (1-w)^2(1-a^{22})T^{-1}\sum_{t=1}^{T}R_{2t}^2 +$$
$$+(1-w)^2a^{22}R_{2t-1}^2 + 2w(1-w)(1-a^{12})T^{-1}\sum_{t=1}^{T}R_{1t}R_{2t} + 2w(1-w)a^{12}R_{1t-1}R_{2t-1}$$

which cannot be written in diagonal form, $(1-a^{pp})T^{-1}\sum_{t=1}^{T}[wR_{1t} + (1-w)R_{2t}]^2 + a^{pp}[wR_{1t} + (1-w)R_{2t}]^2$ because for no definition of a^{pp} it is possible to show that

$$w^2(1-a^{pp})T^{-1}\sum_{t=1}^{T}R_{1t}^2 + (1-w)^2(1-a^{pp})T^{-1}\sum_{t=1}^{T}R_{2t}^2 + 2w(1-w)(1-a^{pp})T^{-1}\sum_{t=1}^{T}R_{1t}R_{2t} =$$
$$= w^2(1-a^{11})T^{-1}\sum_{t=1}^{T}R_{1t}^2 + (1-w)^2(1-a^{22})T^{-1}\sum_{t=1}^{T}R_{2t}^2 + 2w(1-w)(1-a^{12})T^{-1}\sum_{t=1}^{T}R_{1t}R_{2t}$$

[25] In case you are curious, notice that the heuristic proof above is in itself sufficient to derive that $a^{12} = 0$ from $a^{11} \in (0,1)$ and $a^{22} \in (0,1)$ and that you do not need to deal with the lower bound of the filtered correlation coefficient derived from a^{11}, a^{12}, and a^{22}. Just for completeness, let us also consider that case of imposing that $0 \geq \rho_{12,t} \geq -1$ $\forall t \geq 1$. This lower bound means that

$$-(c^{12} + a^{12}R_{1t-1}R_{2t-1}) \leq \sqrt{(c^{11} + a^{11}R_{1t-1}^2)(c^{22} + a^{22}R_{2t-1}^2)}$$

or

$$(c^{12} + a^{12}R_{1t-1}R_{2t-1})^2 \leq (c^{11} + a^{11}R_{1t-1}^2)(c^{22} + a^{22}R_{2t-1}^2)$$

$$(c^{12})^2 + (a^{12})^2R_{1t-1}^2R_{2t-1}^2 + 2c^{12}a^{12}R_{1t-1}R_{2t-1} \leq c^{11}c^{22} + c^{22}a^{11}R_{1t-1}^2 + c^{11}a^{22}R_{2t-1}^2 + a^{11}a^{22}R_{1t-1}^2R_{2t-1}^2$$

which is equivalent to

$$[a^{11}a^{22} - (a^{12})^2]R_{1t-1}^2R_{2t-1}^2 + [c^{11}c^{22} - (c^{12})^2] + c^{22}a^{11}R_{1t-1}^2 + c^{11}a^{22}R_{2t-1}^2 - 2c^{12}a^{12}R_{1t-1}R_{2t-1} \geq 0$$

which is the same condition used above.

and especially that

$$w^2a^{11}R^2_{1t-1} + (1-w)^2a^{22}R^2_{2t-1} + 2w(1-w)a^{22}R_{1t-1}R_{2t-1} = w^2a^{pp}R^2_{1t} + (1-w)^2a^{pp}R^2_{2t} + 2w(1-w)a^{pp}R_{1t-1}R_{2t-1}$$

This means that the Diagonal multivariate ARCH model fails to be invariant to linear combinations: if you start with N assets that follow a Diagonal multivariate ARCH model, the resulting portfolio of assets will fail to follow a similar Diagonal model, which is of course problematic if not confusing. As you should be reading in the paper by Bauwens et al. (2006), the problem of (12) that causes it to fail the invariance property is very simple to visualize: while in

$$vech(\mathbf{\Omega}_t) = (\mathbf{I}_2 - \mathbf{A})vech\left(T^{-1}\sum_{t=1}^{T}\mathbf{R}_t\mathbf{R}_t'\right) + \mathbf{A}vech(\mathbf{R}_{t-1}\mathbf{R}_{t-1}')$$

$\mathbf{A}$ is diagonal, R^p_t can be written as $[w\ 1-w]\mathbf{R}_t = \mathbf{w}'\mathbf{R}_t$ and $Var_{t-1}[R^p_t] = \mathbf{w}'\mathbf{\Omega}_t\mathbf{w}$ implies the need to use a vector of coefficients $\mathbf{w}'\mathbf{A}$ which is no longer a diagonal matrix (of course, it is not even a matrix).

It is also easy to see what you need to do in order for the invariance property to obtain: if you set $a^{11} = a^{22} = a^{22}$, then when $a^{pp} = a^{11}$

$$w^2(1-a^{pp})T^{-1}\sum_{t=1}^{T}R^2_{1t} + (1-w)^2(1-a^{pp})T^{-1}\sum_{t=1}^{T}R^2_{2t} + 2w(1-w)(1-a^{pp})T^{-1}\sum_{t=1}^{T}R_{1t}R_{2t} =$$
$$= w^2(1-a^{11})T^{-1}\sum_{t=1}^{T}R^2_{1t} + (1-w)^2(1-a^{22})T^{-1}\sum_{t=1}^{T}R^2_{2t} + 2w(1-w)(1-a^{12})T^{-1}\sum_{t=1}^{T}R_{1t}R_{2t}$$
$$= w^2a^{11}R^2_{1t-1} + (1-w)^2a^{22}R^2_{2t-1} + 2w(1-w)a^{22}R_{1t-1}R_{2t-1}$$
$$= w^2a^{pp}R^2_{1t} + (1-w)^2a^{pp}R^2_{2t} + 2w(1-w)a^{pp}R_{1t-1}R_{2t-1}$$

will trivially hold. But this means that the only way for a Diagonal multivariate ARCH to possess the invariance property is for it to actually be a Scalar multivariate ARCH, in which the same ARCH coefficient applies to all conditional equations.

Question 8 (4 points)
One analyst working on your desk has been given the task to identify and estimate alternative regimes in the dynamic relationship among monthly US excess equity returns, Japanese excess equity returns, and the rate of change in the implied volatility of SPX options. The analyst has reported the following estimation output (p-values are in parentheses).

$$\begin{bmatrix} x^{US}_{t+1} \\ x^{Japan}_{t+1} \\ rvxo_{t+1} \end{bmatrix} = \begin{cases} [\underset{(0.000)}{-2.088}\ \ \underset{(0.000)}{-6.157}\ \ \underset{(0.000)}{-6.785}]' & \text{if } S_{t+1}=1 \\ [\underset{(0.000)}{0.958}\ \ \underset{(0.424)}{0.296}\ \ \underset{(0.019)}{-3.780}]' & \text{if } S_{t+1}=2 \\ [\underset{(0.000)}{2.256}\ \ \underset{(0.000)}{4.227}\ \ \underset{(0.000)}{20.226}]' & \text{if } S_{t+1}=3 \end{cases} + \begin{cases} \begin{bmatrix} \underset{(0.000)}{0.237} & \underset{(0.592)}{0.033} & \underset{(0.000)}{-14.914} \\ \underset{(0.875)}{0.018} & \underset{(0.788)}{0.028} & \underset{(0.271)}{-1.891} \\ \underset{(0.000)}{0.002} & \underset{(0.389)}{0.002} & \underset{(0.304)}{-0.050} \end{bmatrix} & \text{if } S_{t+1}=1 \\ \begin{bmatrix} \underset{(0.007)}{0.167} & \underset{(0.767)}{0.011} & \underset{(0.000)}{-12.254} \\ \underset{(0.064)}{0.204} & \underset{(0.704)}{0.027} & \underset{(0.000)}{-7.586} \\ \underset{(0.213)}{0.004} & \underset{(0.457)}{0.002} & \underset{(0.000)}{-0.390} \end{bmatrix} & \text{if } S_{t+1}=2 \\ \begin{bmatrix} \underset{(0.265)}{-0.135} & \underset{(0.864)}{0.011} & \underset{(0.143)}{-8.715} \\ \underset{(0.064)}{0.235} & \underset{(0.071)}{0.214} & \underset{(0.136)}{-6.385} \end{bmatrix} & \text{if } S_{t+1}=3 \end{cases} \cdot \begin{bmatrix} x^{US}_{t} \\ x^{Japan}_{t} \\ rvxo_{t} \end{bmatrix} +$$

$$+ \begin{cases} \begin{bmatrix} 3.104 & -0.154 & -0.151 \\ -2.269 & 4.746 & -0.321 \\ -5.118 & -16.64 & 10.92 \end{bmatrix} & \text{if } S_{t+1}=1 \\ \begin{bmatrix} 2.332 & 0.312 & 0.097 \\ 3.067 & 4.215 & -0.009 \\ 2.914 & -0.489 & 12.88 \end{bmatrix} & \text{if } S_{t+1}=2 \\ \begin{bmatrix} 3.661 & 0.118 & -0.335 \\ 2.924 & 6.768 & -0.468 \\ -36.45 & -94.14 & 29.72 \end{bmatrix} & \text{if } S_{t+1}=3 \end{cases} \cdot \begin{bmatrix} z^{US}_{t+1} \\ z^{Japan}_{t+1} \\ rvxo_{t+1} \end{bmatrix} \qquad \hat{P} = \begin{bmatrix} 0.397 & 0.055 & 0.548 \\ 0.000 & 0.841 & 0.159 \\ 0.392 & 0.340 & 0.268 \end{bmatrix}$$

How many regimes did he specify for this model? Are the regimes persistent? What is the duration of each of the regimes? Does the past rate of growth of implied volatility forecast positive or negative excess US returns, and when? After taking such vector autoregressive structure into account, is the correlation between shocks to US excess equity returns and the rate of growth of implied volatility positive or negative, and when? Please justify your answers with reference to the estimation outputs provided.

Debriefing.
The analyst has specified three Markov regimes (states), as one can see from the fact that the estimated transition matrix is a 3x3 one. Only one regime, the second, is persistent, in the sense that $\Pr(S_{t+1} = 2|S_t = 2) = 0.814 > 0.5$; the other two regimes are non persistent, in the sense that $\Pr(S_{t+1} = j|S_t = j) < 0.5$ for j = 1 and 3. The durations of the three regimes are $(1 - 0.397)^{-1}$ = 1.658, $(1 - 0.841)^{-1}$ = 6.289, and $(1 - 0.268)^{-1}$ = 1.366 months. Based on the p-values reported, the past rate of growth of implied volatility forecasts negative excess US returns (coefficients of -14.91 and -12.25, with zero p-values) in regimes 1 and 2, while the predictive power is more doubtful in regime 3 (coefficient of -8.72, but with a p-value exceeding 0.10). Net of the VAR effects, the correlation between shocks to US excess equity returns and the rate of growth of implied volatility is negative and rather sizable in regimes 1 and 3 (correlations are -0.15 and -0.34), and positive but smaller (0.10) in regime 2.