vrije Universiteit amsterdam

Solution Preparation Exam
Financial Econometrics

Exam:	Financial Econometrics (3.5)
Code:	-
Coordinator:	dr. P. Gorgi
Co-Reader:	-
Date:	-
Time:	-
Duration:	2 hours and 45 minutes
Calculator:	Allowed
Graphical calculator:	Not allowed
Number of questions:	4
Type of questions:	Open
Answer in:	English
Credit score:	100 credits counts for a 10
Grades:	Made public within 10 working days
Inspection:	-
Number of pages:	-

- Read the entire exam carefully before you start answering the questions.
- Be clear and concise in your statements, but justify every step in your derivations.
- The questions should be handed back at the end of the exam. Do not take it home.

Good luck!

Question 1 [35 points]   Observation-Driven Models: Stochastic Properties

Consider the following ARCH(2) model:

$$y_t = \sigma_t \varepsilon_t \quad , \quad \{\varepsilon_t\}_{t\in\mathbb{Z}} \sim \text{NID}(0,1)$$

$$\text{where} \quad \sigma_t^2 = \omega + \alpha_1 y_{t-1}^2 + \alpha_2 y_{t-2}^2 \quad \text{for} \quad t \in \mathbb{Z},$$

where $\omega > 0$, $\alpha_1 \geq 0$, $\alpha_2 \geq 0$ and $\alpha_1 + \alpha_2 < 1$.

(a) Show that y_t has unconditional mean zero; i.e. show that $\mathbb{E}(y_t) = 0$.

(b) Derive the unconditional variance of y_t; i.e. derive an expression for $\mathbb{V}\text{ar}(y_t)$ in terms of the parameters ω, α_1 and α_2.

(c) Suppose that the following *for loop* is used in MATLAB to simulate data from the ARCH(2) model:

```
sig(1)  =  omega/(1-alpha1-alpha2);

for t=2:T

y(t) = sqrt(sig(t)) * epsilon(t);
sig(t+1) = omega + alpha1*y(t)^2 + alpha2*y(t-1)^2;

end
```

Consider the following statement: *"the for loop should start at* `t=1` *because we have specified the initial value* `sig(1)`*"*. Is the statement true or false? Justify your answer.

Consider the following updating equations for the conditional variances and conditional covariance between a bivariate vector $\boldsymbol{y}_t = (y_{1,t}, y_{2,t})^\top$ of stock returns:

$$\begin{bmatrix} \sigma_{1,t}^2 & \sigma_{21,t} \\ \sigma_{21,t} & \sigma_{2,t}^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0.5 \\ 0.5 & 2 \end{bmatrix} \odot \begin{bmatrix} y_{1,t-1}^2 & y_{1,t-1}y_{2,t-1} \\ y_{1,t-1}y_{2,t-1} & y_{2,t-1}^2 \end{bmatrix}$$

Suppose that the last observed returns for stocks 1 and 2 were given by $y_{1t-1} = 1$ and $y_{2t-1} = 0$. Additionally, consider three different portfolios which assign different weights k_1 and k_2 to stocks 1 and 2 respectively:

<u>Portfolio A</u>: consists only of stock 1 ($k_1 = 1$, $k_2 = 0$)

<u>Portfolio B</u>: consists only of stock 2 ($k_1 = 0$, $k_2 = 1$)

<u>Portfolio C</u>: gives the same weight to each stock ($k_1 = 0.5$, $k_2 = 0.5$)

(d) Which portfolio has lower risk? i.e. which portfolio has lower conditional variance?

(e) Suppose further that $\mu_{1,t} = \mathbb{E}(y_{1,t}|Y^{t-1}) = 1$ and $\mu_{2,t} = \mathbb{E}(y_{2,t}|Y^{t-1}) = 2$. Calculate the Sharpe ratio of each portfolio.

(f) Consider the following statement: *"Portfolio A is the best because it has the lowest Sharpe ratio"*. Is the statement true or false? Justify your answer.

Solution to question 1:

(a) The unconditional mean $\mathbb{E}(y_t)$ can be obtained as follows

$$\mathbb{E}(y_t) = \mathbb{E}(\sigma_t \varepsilon_t) = \mathbb{E}(\sigma_t)\mathbb{E}(\varepsilon_t) = \mathbb{E}(\sigma_t) \times 0 = 0,$$

where the first equality is satisfied since y_t is generated by an ARCH model, the second equality is valid since ε_t is independent of σ_t and the third equality follows because $\mathbb{E}(\varepsilon_t) = 0$, which is implied by the assumption that $\varepsilon_t \sim N(0, 1)$.

(b) The unconditional variance of y_t is obtained by noting that

$$\mathbb{V}\text{ar}(y_t) = \mathbb{E}(y_t^2) = \mathbb{E}(\sigma_t^2 \varepsilon_t^2) = \mathbb{E}(\sigma_t^2)\mathbb{E}(\varepsilon_t^2) = \mathbb{E}(\sigma_t^2)$$

where the first equality follows from y_t having mean zero, the second equality is obtained from the observation equation of the ARCH model that generates y_t, the third equality follows from the independence of σ_t and ε_t and the fourth equality follows from the fact that the variance of ε_t is equal to 1 by assumption. Now using the updating equation, we have that

$$\mathbb{E}(\sigma_t^2) = \omega + \alpha_1 \mathbb{E}(y_{t-1}^2) + \alpha_2 \mathbb{E}(y_{t-2}^2).$$

Furthermore, using the fact that $\mathbb{E}(y_t^2) = \mathbb{E}(\sigma_t^2)$, shown above, we conclude that

$$\mathbb{E}(y_t^2) = \omega + \alpha_1 \mathbb{E}(y_{t-1}^2) + \alpha_2 \mathbb{E}(y_{t-2}^2).$$

The final step is to recognize that the condition $\alpha_1 + \alpha_2 < 1$ ensures that $\{y_t\}$ is weakly stationary, and hence we can set $\mathbb{E}(y_t^2) = \mathbb{E}(y_{t-1}^2) = \mathbb{E}(y_{t-2}^2)$, which implies that

$$\mathbb{E}(y_t^2) = \omega + \alpha_1 \mathbb{E}(y_t^2) + \alpha_2 \mathbb{E}(y_t^2).$$

Solving for $E(y_t^2)$ yields the desired result

$$\mathbb{E}(y_t^2) = \omega/(1 - \alpha_1 - \alpha_2).$$

(c) The *for loop* cannot start at $t = 1$ because the vector the updating equation in the loop contains an element `y(t-1)`. If the *for loop* is set to start at $t = 1$, then the *for loop* will call for `y(0)` which is invalid since the first element of a vector in MATLAB is indexed by 1, i.e. it is called using `y(1)`. Calling `y(0)` results in an error.

(d) First we note that the portfolio returns are obtained as a linear combination of the returns of the two stocks $y_{1,t}$ and $y_{2,t}$,

$$y_{p,t} = k_1 y_{1,t} + k_2 y_{2,t}.$$

Next, we note that conditional variance of the portfolio $\mathbb{V}\text{ar}(y_{p,t}|Y^{t-1})$ is given by

$$\mathbb{V}\text{ar}(y_{p,t}|Y^{t-1}) = k_1^2 \mathbb{V}\text{ar}(y_{1,t}|Y^{t-1}) + k_2^2 \mathbb{V}\text{ar}(y_{2,t}|Y^{t-1}) + 2k_1 k_2 \mathbb{C}\text{ov}(y_{1,t}, y_{2,t}|Y^{t-1}).$$

Using the notation of the updating equation, we have,

$$\sigma_{p,t}^2 = k_1^2 \sigma_{1,t}^2 + k_2^2 \sigma_{2,t}^2 + 2k_1 k_2 \sigma_{12,t}.$$

Since $y_{1,t-1} = 1$ and $y_{2,t-1} = 0$, we obtain from the updating equation that

$$\sigma^2_{1,t} = 1 + 1 \times y_{1,t-1} = 2 \; ,$$

$$\sigma^2_{2,t} = 1 + 2 \times y_{2,t-1} = 1 \; ,$$

$$\text{and} \quad \sigma_{12,t} = 0.$$

We thus conclude that portfolios A,B and C, have the following conditional variance

Portfolio A: $\sigma^2_{p,t} = \sigma^2_{1,t} = 2.$

Portfolio B: $\sigma^2_{p,t} = \sigma^2_{2,t} = 1.$

Portfolio C: $\sigma^2_{p,t} = 0.25\sigma^2_{1,t} + 0.25\sigma^2_{2,t} + 0.5\sigma_{12,t} = 0.75.$

Therefore the portfolio with lower risk is Portfolio C.

(e) In general, the Sharpe Ratio $S_{p,t}$ is given by the ratio of the conditional expectation and the square root of the conditional variance deviation of portfolio returns. The conditional variance of the portfolios were derived in the previous question. Since $\mathbb{E}(y_{1,t}|Y^{t-1}) = 1$ and $\mathbb{E}(y_{2,t}|Y^{t-1}) = 2$, the conditional expectation the portfolio is naturally given by

$$\mathbb{E}(y_{t,p}) = k_1\mathbb{E}(y_{1,t}) + k_2\mathbb{E}(y_{2,t}) = k_1 + 2k_2.$$

The expected returns of portfolios A, B and C, are thus given by

Portfolio A: $\mu_{p,t} = 1 \times 1 + 2 \times 0 = 1.$

Portfolio B: $\mu_{p,t} = 0 \times 1 + 2 \times 1 = 2.$

Portfolio C: $\mu_{p,t} = 0.5 \times 1 + 0.5 \times 2 = 1.5.$

We thus conclude that the Sharpe ratio of each portfolio is given by

Portfolio A: $S_{p,t} = \frac{\mu_{p,t}}{\sigma_{p,t}} = \frac{\mu_{1,t}}{\sigma_{1,t}} = \frac{1}{\sqrt{2}} \approx 0.71.$

Portfolio B: $S_{p,t} = \frac{\mu_{p,t}}{\sigma_{p,t}} = \frac{\mu_{2,t}}{\sigma_{2,t}} = \frac{2}{\sqrt{1}} = 2.$

Portfolio C: $S_{p,t} = \frac{\mu_{p,t}}{\sigma_{p,t}} = \frac{0.5\mu_{1,t}+0.5\mu_{2,t}}{\sigma_{p,t}} = \frac{1.5}{\sqrt{0.75}} \approx 1.73.$

(f) The statement is incorrect because we prefer portfolios with large returns and low risk (variance). Therefore the best portfolio is the one with largest Sharpe Ratio.

Question 2 [25 points] Observation-Driven Models: Parameter Estimation

Consider the following GARCH(1,1) model:

$$y_t = \sigma_t \varepsilon_t \quad , \quad \{\varepsilon_t\}_{t\in\mathbb{Z}} \sim \text{NID}(0,1)$$

$$\text{where} \quad \sigma_t^2 = \omega + \alpha_1 y_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad \text{for} \quad t \in \mathbb{Z}.$$

(a) Consider the following MATLAB code that sets the initial value of the parameter vector $\theta = (\omega, \alpha_1, \beta_1)$ which is used with the `FMINCON` optimization package for obtaining estimates of parameters of the GARCH(1,1) model:

```
theta_ini = [0, 0, 0]
```

Consider the following statement: *"Setting the initial value of the parameter* $\theta = (0, 0, 0)$ *is problematic"*. Is the statement true or false? Justify your answer.

(b) You have estimated several competing GARCH(p,q) model specifications and obtained the following results for the log likelihood, the Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC):

Model	Log Likelihood	AIC	BIC
GARCH(1,1)	-4781.3	9558.6	9573.4
GARCH(1,2)	-4787.5	9583.0	9589.4
GARCH(2,1)	-4699.1	9406.2	9412.6
GARCH(2,2)	-4698.7	9407.4	9415.4

Are the following statements true or false? Please justify your answer.

(i) *"The GARCH(2,2) is the best model because it has the largest log likelihood value."*

(ii) *"The GARCH(1,2) is the best model because it has the largest AIC and BIC values."*

(iii) *"If the GARCH(2,1) model is better than the GARCH(1,1) model in terms of both AIC and BIC, then this means that the GARCH(2,1) model is well specified."*

Solution to question 2:

(a) The statement is correct. Setting the initial value to (0,0,0) is problematic because, according to the GARCH updating equation, in this way the conditional variance σ_t^2 that enters the log-likelihood is equal to zero

$$\sigma_t^2 = 0 + 0 \times y_{t-1}^2 + 0 \times \sigma_{t-1}^2 = 0 \quad \text{for every } t.$$

Therefore the log-likelihood will not exist at the initial value since the terms log(0) and $y_t^2/0$ are not defined.

(b) (i) The statement is false. The log-likelihood should not be used as a criterion to compare the performance of different models. This is because a larger model that nests another smaller model has always a larger log-likelihood, even if the smaller model gives a correct description of the data. For example, a GARCH(2,2) will always have larger log-likelihood than an GARCH(1,1) model, even if the GARCH(1,1) model provides a correct description of the data. Therefore we would always choose models with more parameters and overfit the data. We should instead use information criteria like the AIC and the BIC that penalize models with more parameters.

(ii) The statement is false. The AIC and BIC information criteria are inversely related to the model's fit (log likelihood) and positively related to the model's number of parameters. In general we prefer parsimonious models (i.e. models with few parameters) with good fit (i.e. large log likelihood) Hence, we should select models with low values of AIC and BIC. The best model is thus the GARCH(2,1).

(iii) The statement is false. Both the GARCH(2,1) and the GARCH(1,1) may be misspecified. When a model is better than another in AIC and BIC, this means simply that it fits the data better, even after penalizing for the difference in the number of parameters. The AIC and BIC do not tell us however if any of the models under comparison are well specified or misspecified. Questions about GARCH model specification should be answered using specification tests like testing for autocorrelation in residuals or testing for normality of the residuals.

Question 3 [20 points] Parameter-Driven Models: Stochastic Properties

(a) Explain and discuss the main differences in the estimation of parameters between *Generalized Autoregressive Conditional Heteroeskedasticity* (GARCH) models and *Stochastic Volatility* (SV) models?

(b) Let $\{y_t\}_{t \in \mathbb{Z}}$ be generated by the following SV-MA(2) model

$$y_t = \sigma_t \epsilon_t, \;\; \sigma_t^2 = \exp(f_t),$$
$$f_t = \eta_t + \phi \eta_{t-2},$$

where $\{\epsilon_t\}_{t \in \mathbb{Z}}$ is $NID(0, 1)$ and $\{\eta_t\}_{t \in \mathbb{Z}}$ is $NID(0, \sigma_\eta^2)$ with $\sigma_\eta^2 > 0$.

Recall that if z_t is a *normal* random variable $z_t \sim N(\mu, \sigma^2)$, then $\exp(z_t)$ has a *log-normal* distribution, which is denoted $\exp(z_t) \sim \text{log-}N(\mu, \sigma^2)$, and furthermore, has mean given by

$$\mathbb{E}(z_t) = \exp(\mu + \sigma^2/2).$$

(i) Show that the first-order autocorrelation function of y_t^2 is equal to zero, i.e. show that $\mathbb{C}\text{orr}(y_t^2, y_{t-1}^2) = 0$.

(ii) Show that the skewness of y_t is zero, i.e. show that $\mathbb{E}(y_t^3) = 0$.

(iii) Is the unconditional distribution of y_t Normal if $\phi = 0$? Justify your answer.

Solution to question 3:

(a) GARCH and SV models are have different updating equations. GARCH models are observation-driven models, and hence, conditional on the past Y^{t-1}, the conditional volatility σ_t^2 is a given constant. This means that the conditional distribution of y_t given the past is easily tractable. In particular, we have that

$$y_t|Y^{t-1} \sim N(0, \sigma_t^2)$$

and hence, we can factorize the joint density of the data as follows

$$p(y_1, ..., y_T; \theta) = \prod_{t=2}^{T} p(y_t|Y^{t-1})$$

where

$$p(y_t|Y^{t-1}) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{y_t^2}{2\sigma_t^2}\right).$$

The log-likelihood function of a GARCH model is thus analytically tractable and we can optimize it numerically. We can thus easily obtain maximum likelihood estimates for the parameters of a GARCH model.

In contrast, SV models are parameter-driven models, and this means that the conditional volatility σ_t^2 is not a constant, even conditional on the past Y^{t-1}. This fact renders the conditional distribution $y_t|Y^{t-1}$ analytically intractable. As a result, we cannot simply write down the log likelihood function and attempt to maximize it numerically. Instead, the parameters of the SV model can be estimated using simulation based methods like indirect inference. Rather than maximizing the log likelihood function, the indirect inference estimator attempts to find the parameter values that make data simulated from the SV model as similar as possible to observed data, as judged by a vector of auxiliary statistics that are used to describe both observed and simulated data.

(b) (i) The correlation between y_t^2 and y_{t-1}^2 is zero because y_t^2 and y_{t-1}^2 are independent and thus uncorrelated. This can be noted as

$$y_t^2 = \exp(\eta_t + \phi\eta_{t-2})\epsilon_t^2 \quad \text{and}$$

$$y_{t-1}^2 = \exp(\eta_{t-1} + \phi\eta_{t-3})\epsilon_{t-1}^2.$$

By the model's assumptions η_t, η_{t-2} and ϵ_t are all independent of η_{t-1}, η_{t-3} and ϵ_{t-1} and therefore y_t^2 is independent of y_{t-1}^2.

(ii) The desired result is obtained by noting that

$$\mathbb{E}(y_t^3) = \mathbb{E}(\sigma_t^3\epsilon_t^3) = \mathbb{E}(\sigma_t^3)\mathbb{E}(\epsilon_t^3) = \mathbb{E}(\sigma_t^3) \times 0 = 0,$$

where the first equality follows from the observation equation of the SV model that generates y_t, the second equality follows from the independence between ϵ_t and σ_t and the third equality follows from the fact that $\epsilon_t \sim N(0,1)$ and the skewness of the normal distribution is zero.

(iii) No, the unconditional distribution of the SV model is not normal. This can be shown by noting that the kurtosis of the unconditional distribution is not equal to 3, as it should be for a normal random variable. In particular, when $\phi = 0$, the model becomes

$$y_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = \exp(f_t),$$
$$f_t = \eta_t.$$

Therefore the kurtosis $k_u = \mathbb{E}(y_t^4)/\mathbb{E}(y_t^2)^2$ of y_t can be obtained as follows. First, we obtain that

$$\mathbb{E}(y_t^4) = \mathbb{E}(\sigma_t^4 \epsilon_t^4) = \mathbb{E}(\sigma_t^4)\mathbb{E}(\epsilon_t^4) = 3\mathbb{E}[\exp(2f_t)] = 3\exp(2\sigma_\eta^2),$$

where the first equality follows from the observation equation opf the SV model that generates y_t. The second equality follows by independence of σ_t^4 and ϵ_t^4. The third equality follows from the fact that ϵ_t is normally distributed and hence $\mathbb{E}(\epsilon_t^4) = 3$ and because $\sigma_t^2 = \exp(f_t)$ which implies that $\sigma^4 = (\sigma^2)^2 = (\exp(f_t))^2 = \exp(2f_t)$. The last equality follows from the fact that $2f_t = 2\eta_t \sim N(0, 4\sigma_\eta^2)$ and therefore $\sigma_t^4 = \exp(2f_t) \sim \log\text{-}N(0, 4\sigma_\eta^2)$. Second, we obtain that

$$\mathbb{E}(y_t^2) = \mathbb{E}(\sigma_t^2 \epsilon_t^2) = \mathbb{E}(\sigma_t^2)\mathbb{E}(\epsilon_t^2) = \mathbb{E}[\exp(f_t)] = \exp(\sigma_\eta^2/2),$$

where the first equality is implied by the observation equation of the SV model, the second equality follows from the independence of σ_t^2 and ϵ_t^2, the third equality follows from the fact that $\epsilon_t \sim N(0,1)$ and hence $\mathbb{E}(\epsilon_t^2) = \mathbb{V}\text{ar}(\epsilon_t) = 1$, and the last equality follows from the fact that $f_t = \eta_t \sim N(0, \sigma_\eta^2)$ and therefore $\exp(f_t) \sim \log\text{-}N(0, \sigma_\eta^2)$. Finally, we get that

$$k_u = \frac{\mathbb{E}(y_t^4)}{\mathbb{E}(y_t^2)^2} = \frac{3\exp(2\sigma_\eta^2)}{\exp(\sigma_\eta^2)} = 3\exp(\sigma_\eta^2).$$

Therefore, the kurtosis is bigger than 3 as long as $\sigma_\eta^2 > 0$. We conclude that y_t is not normally distributed.

Question 4 [20 points] Parameter-Driven Models: Parameter Estimation

(a) Consider the indirect inference estimator $\hat{\theta}_{HT}$ given by

$$\hat{\theta}_{HT} = \arg\min_{\theta \in \Theta} d(\hat{B}_T, \tilde{B}_H(\theta)),$$

where $\hat{B}_T$ is the auxiliary statistic obtained from the observed sample of data (which is of length T) and $\tilde{B}_H(\theta)$ is the auxiliary statistic obtained from the simulated sample of data (which is of length H).

Consider the following statement: *"the accuracy of the indirect inference estimator $\hat{\theta}_{HT}$ increases as the length H of the simulations increases".* Is the statement true or false? Justify your answer.

(b) Consider the following SV-MA(1) model

$$y_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \exp(f_t),$$
$$f_t = \eta_t + \phi \eta_{t-1},$$

where $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ is $NID(0,1)$ and $\{\eta_t\}_{t \in \mathbb{Z}}$ is $NID(0,1)$. The parameter vector is given by $\theta = \phi$ and the parameter set is $\Theta = (-1, 1)$.

We want to estimate the "true" parameter $\theta_0 = \phi_0 \in (-1, 1)$ by indirect inference. Consider the following auxiliary statistics

$$\hat{B}_T = \frac{1}{T} \sum_{t=2}^{T} y_t y_{t-1}, \quad \text{and} \quad \tilde{B}_H(\theta) = \frac{1}{H} \sum_{t=2}^{H} \tilde{y}_t(\theta) \tilde{y}_{t-1}(\theta)$$

Show whether or not the indirect inference estimator $\hat{\theta}_{HT}$ based on the above auxiliary statistics is consistent.

Solution to question 4:

(a) The statement is correct. The variance of the indirect inference estimator decreases as H increases. This is due to the fact that when H is large there is less uncertainty in the auxiliary statistics $\tilde{B}_H(\theta)$ obtained from the simulated data. Therefore, the indirect inference estimator is less exposed to the sample variability of $\tilde{B}_H(\theta)$ and thus it is more accurate.

(b) To show the consistency of the indirect inference estimator we first need to find the *binding functions* $B(\theta)$ and $B(\theta_0)$. Then, we have to show that the true parameter $\theta = \theta_0$ is the unique minimizer of $d(B(\theta), B(\theta_0))$ in the parameter space Θ. In this case $\theta = \phi$ and $\Theta = (-1, 1)$.

By the Law of Large Numbers we obtain that

$$\hat{B}_T = T^{-1} \sum_{t=2}^{T} y_t y_{t-1} \xrightarrow{p} B(\theta_0) = \mathbb{E}(y_t y_{t-1}) = 0,$$

where

$$\mathbb{E}(y_t y_{t-1}) = \mathbb{E}(\sigma_t \sigma_{t-1} \epsilon_t \epsilon_{t-1}) = \mathbb{E}(\sigma_t \sigma_{t-1}) \mathbb{E}(\epsilon_t) \mathbb{E}(\epsilon_{t-1}) = \mathbb{E}(\sigma_t \sigma_{t-1}) \times 0 \times 0 = 0.$$

In a similar way, for the statistics from the simulated data $\{\tilde{y}_h(\theta)\}_{h=1}^{H}$, we obtain that

$$\tilde{B}_H(\theta) = H^{-1} \sum_{h=2}^{H} \tilde{y}_h(\theta) \tilde{y}_{h-1}(\theta) \xrightarrow{p} B(\theta) = \mathbb{E}(y_h(\theta) \tilde{y}_{h-1}(\theta)) = 0.$$

Therefore we conclude that the indirect inference estimator is inconsistent because $d(B(\theta), B(\theta_0)) = d(0, 0) = 0$ for any $\theta \in \Theta$.